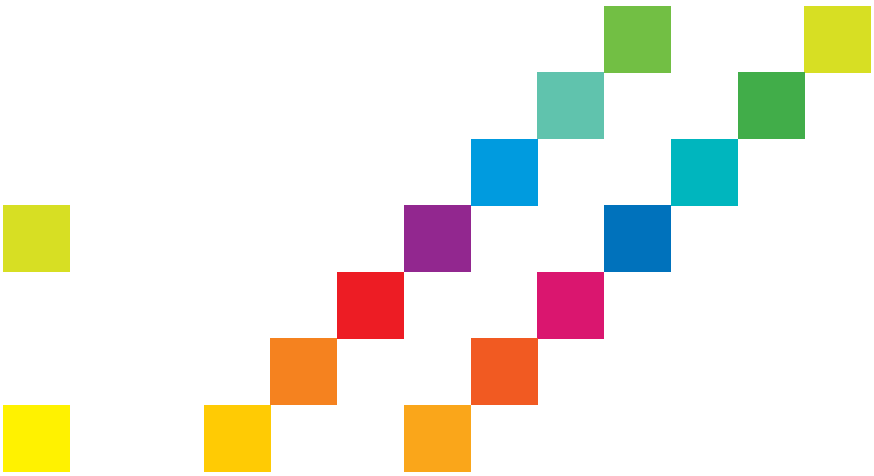


MATTHIAS C. KETTEMANN, JOSEFA FRANCKE, CHRISTINA DINAR AND LENA HINRICHS (EDS.)

Platform://Democracy

Research Report Europe

Platform://Democracy

Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Europe

edited by Matthias C. Kettemann, Josefa Francke, Christina Dinar and Lena Hinrichs
LEIBNIZ INSTITUTE FOR MEDIA RESEARCH | HANS-BREDOW-INSTITUT, HAMBURG, GERMANY
ALEXANDER VON HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY, BERLIN, GERMANY

Cite as: Kettemann, Matthias C.; Francke, Josefa; Dinar, Christina; Hinrichs, Lena (eds.)(2023), Platform://Democracy–Perspectives on Platform Power, Public Values and the Potential of Social Media Councils: Research Report Europe. Hamburg: Verlag Hans-Bredow-Institut. https://doi.org/10.21241/ssoar.86528

This publication is part of the project *Platform://Democracy: Platform Councils as Tools to Democratize Hybrid Online Orders*. The project was carried out by the Leibniz Institute for Media | Hans-Bredow-Institut, Hamburg, the Alexander von Humboldt Institute for Internet and Society, Berlin, and the Department of Theory and Future of Law of the University of Innsbruck und funded by Stiftung Mercator.

Publisher: Leibniz Institut für Medienforschung | Hans-Bredow-Institut (HBI)
Rothenbaumchaussee 36, 20148 Hamburg
Tel. (+49 40) 45 02 17-0, info@leibniz-hbi.de, www.leibniz-hbi.de

LEIBNIZ INSTITUTE FOR MEDIA RESEARCH HANS-BREDOW-INSTITUT

ALEXANDER VON HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY

STIFTUNG MERCATOR

Contributors

Name(s)	Affiliation
Pierre François Docquir	Independent Researcher
Rachel Griffin	SciencesPo
Bettina Hesse	Verdi
Niklas Eder	Meta Oversight Board
Dominik Piétron	Humboldt-Universität zu Berlin
Clara Iglesias Keller & Theresa Züger	Leibniz Institute for Media Research | Hans-Bredow-Institut & Alexander von Humboldt Institute for Internet and Society
Aleksandra Kuczerawy	KU Leuven
Pierre François Docquir	Independent Researcher
Rachel Griffin	SciencesPo

Table of Contents

Towards More Legitimacy in Rule-Making

Matthias C. Kettemann and Josefa Francke
LEIBNIZ INSTITUTE FOR MEDIA RESEARCH | HANS-BREDOW-INSTITUT, HAMBURG, GERMANY
ALEXANDER VON ALEXANDER VON HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY, BERLIN, GERMANY

The idea of creating platform councils to support the legitimacy of rules and recommender algorithms on social media platforms has gained traction in recent years. The concept involves establishing independent bodies, comprised of experts and stakeholders, to provide oversight and guidance to social media platforms on the development and implementation of their policies.

These councils can be responsible for developing, and ruling on, content moderation rules, data privacy guidelinces, and algorithmic transparency. They could also be tasked with conducting audits and reviews of platform policies and practices, and making recommendations for improvements.

One of the main goals of platform councils is to enhance the legitimacy and transparency of social media platforms' decision-making processes, and to increase public trust in their policies and practices. By involving independent experts and stakeholders in the development and implementation of platform policies, the hope is that social media platforms can create fairer rules and improve their algorithmic governance practices.

The following studies offer perspectives on the potentials of social media councils from European researchers.

Pierre François Docquir has utilized his research experience with ARTICLE 19 to emphasize the significant gap between local contexts and global company policies. He suggests that Social Media Councils (SMCs) fuelled by independent multi-stakeholder coalitions at the national level, may be able to bring local voices into content moderation, thereby bridging this gap.

Rachel Griffin takes a feminist perspective in her examination of SMCs, questioning how they can address existing online communication structures that perpetuate the marginalization of vulnerable groups. She emphasizes the importance of ensuring that SMCs' normative setup does not ignore structural problems, maximally includes diverse voices, and exercises sufficient power to demand substantive changes in business practices.

Bettina Hesse explores broadcasting councils as a potential role model for SMCs, identifying areas where they fall short and what SMCs can learn from them. Hesse highlights unchecked power relationships within the councils, with political parties and state representatives overrepresented at the expense of civil society. Additionally, the composition of most broadcasting councils lacks diversity and shows itself to be random in its choice of members. Hesse argues that broadcasting councils are quite removed from general society and should be more independent from overarching media organizations.

Niklas Eder's focus is on DSA regulation with regard to systemic risks. He suggests that platforms are required to do risk assessment and develop mitigation measures to mitigate the harm of platforms. However, given the conflict of interests, the platforms may not be the best-suited actors to do these assessments, and public institutions' involvement could conflict with their obligation to human rights such as freedom of expression. Eder argues that involving stakeholders and civil society via a Social Media Council could be the best way to evaluate risk and design effective mitigation measures.

Dominik Piétron argues for the need for collective data governance in the context of large-scale data extraction. He suggests that individual rights of informational self-determination are not sufficient for adequate protection. Piétron uses the example of a German city platform to demonstrate that SMCs

could be a suitable institution for such collective governance. Piétron proposes inclusive discussions and feedback opportunities, democratic accountability, algorithmic transparency, and effective implementation as design criteria.

Clara Iglesias Keller and Theresa Züger debate whether SMCs are suitable to combat democratic deficits in platform governance or merely validate existing power structures. They argue that citizens' roles and ways of contributing to democracy have changed under online conditions. Drawing on the concept of "digital citizenship," Keller and Züger suggest that SMCs should focus on improving input legitimacy, i.e., public participation.

Aleksandra Kuszerawy examines the potential role of SMCs within the Digital Services Act framework. She notes that internal complaint-handling mechanisms lack a crucial requirement for SMCs, namely independence. However, external out-of-court dispute settlement bodies could be congruent with the idea of independent SMCs, even if limited in their role to reviewing individual content moderation decisions. Kuszerawy maintains that independence and signaling of independence are necessary for legitimacy. Assessing the potential of the Meta Oversight Board to become an out-of-court dispute settlement body under the DSA, Kuszerawy concludes that the current OB structure does not allow for the necessary scaling up to meet the demands of the DSA.

Riku Neuvonen sheds light on the nature of SMCs as self-regulatory bodies and their relevance in the present debate around regulation of online communication. He draws on past instances where self-regulation has been deemed necessary and/or useful, particularly in the context of media councils.

In his analysis of SMCs Giovanni Di Gregorio approaches the issue through the lens of the European constitutional framework. He argues that, as long as SMCs are regarded as private entities, the application of constitutional principles is limited within their decision-making processes. Although SMCs have the potential to enhance oversight in the context of social media platforms, this is not a given, especially if the public has no involvement in the development of their procedures. Di Gregorio further notes that the global scale of SMCs may lead to a conflict between the adoption of international human rights law as a standard and local values and regional variations in its application.

Bringing local voices into conversations about content moderation

Pierre François Docquir
INDEPENDENT RESEARCHER, FORMER HEAD OF MEDIA FREEDOM PROGRAMME OF ARTICLE 19

In most countries of the world, looking at social media companies from the point of view of local actors is tantamount to contemplating distant planets that in any conceivable way appear to be out of reach. The same national observers, however, are acutely aware of the impact that online content has on the society they live in. In this contribution, I argue that, while content moderation processes often fail to integrate a robust understanding of the local circumstances in which online content deploys its effect, national stakeholders (such as for instance civil society organisations, media and academics) are well positioned to bring this specific knowledge and expertise to global companies, which would contribute to a more effective alignment of content moderation outcomes with the realities lived at the local or national level. I then look at two different possible approaches to enabling local voices to partake in conversations about content moderation and argue that an independent, multi-stakeholder and diverse Social Media Council (as defined in the model developed by free speech organisation ARTICLE 19) presents distinctive advantages over a consultative forum operated by a single company. Throughout the text, I'm using the terms local and national as almost interchangeable. Whether based on practical considerations (such as safety or the state of civil society), a 'local Social Media Council' could end up being set up as a national, sub-national or even regional mechanism is not the point of focus here: the goal is to shed light on the question of how to bridge the flagrant gap between global companies and many of the societies impacted by the massive circulation of online content.

The role of local voices in content moderation

While the content rules of social media platforms are elaborated as global norms, the resolution of any content moderation case inevitably requires an informed understanding of the circumstances in which a particular instance of speech has been uttered: in order to decide whether what is being said is in violation of, for instance, rules that prohibit speech that incites violence or discrimination, it is necessary to understand the language in which a particular message is expressed, as well as the social, cultural, historical and political dimensions of the societal context in which this message is disseminated. As research shows, a message that may seem innocuous when seen from Silicon Valley might contribute to reinforcing polarisation and a climate of violence in a complex and diverse country whose history has been defined by frequent episodes of violence. Content moderation systems that ignore local circumstances may cause severe societal harms (such as real-world violence, increased polarisation between social groups, or the undermining of trust in democratic institutions and electoral processes) or aggravate the risk of such societal harms (ARTICLE 19, 2022). In a normative perspective, the Santa Clara Principles on Transparency and Accountability in Content moderation (https://santaclaraprinciples.org/), a global civil society initiative aimed at encouraging global companies to comply with their responsibilities to respect human rights, set forth a principle of 'Cultural Competence' that requires that 'those making moderation and appeal decisions understand the language, culture, and political and social context of the posts they are moderating.' The principle entails that content moderation decisions need to be informed by sufficient understanding of the linguistic, cultural, social, economic and political dimensions of the relevant local or regional context.

However, it appears that content moderation systems – either human or algorithmic, or a combination of both – are often not designed to take local circumstances into consideration. More exactly, social media platforms seem to focus their attention and resources on a very limited number of countries (Newton, 2021). Other countries end up being entirely ignored: this means for instance that no locally-relevant classifier is built into the automated content moderation systems to detect problematic content (such as misinformation or hate speech) and that no partnership is established with local fact-checkers (TBIJ, 2022).

Local stakeholders could naturally play a role in helping social media companies integrate a solid understanding of the local context in their content moderation systems (Access Now, 2020). Partnerships, including regular meetings, with in-country organisations that have a deep knowledge of conflict dynamics could help identify and resolve cultural and social barriers to content reporting, and develop mitigating interventions in response to problematic content (SFCG, 2021). Multi-stakeholder coalitions at the national level could empower local communities to monitor and detect speech that incites violence or discrimination, and a regular cooperation between such coalitions and social media companies could contribute to ensuring that content moderation is aligned with the local context (UNESCO 2021). Regular and transparent consultations with a local coalition that would form a critical mass of local actors could provide social media companies with an effective approach to ensuring that their content moderation processes are better informed by a robust understanding of the local context. Both sides would benefit from such regular dialogue: in exchange for providing a valuable service to social media companies, national stakeholders would find themselves in a position where they can better achieve their own goals in relation to content moderation (ARTICLE 19, 2022a)

As has been documented by research, individuals and organisations who would be particularly well positioned to reflect on the meaning and potential impact of problematic speech disseminated in their society, often find themselves unable to engage with giant global entities that operate on a global scale. They are confronted with various obstacles that range from the absence of translation of the content rules, the lack of effectiveness of internal remedies provided by platforms, or the sheer inability to contact a representative of the company. Even civil society organisations that partake in partnerships with social media platforms (such as trusted flaggers' programmes) often report that they find it uneasy to make their viewpoints heard by their giant partner (ARTICLE 19, 2022a). And in this way, there remains a wide gap between global social media companies and local communities, while the moderation processes often remain out of touch with the specific societal context in which online content deploys its impact.

The participation of local voices in conversations on content moderation

As has been suggested, a form of dialogue between social media companies and local stakeholders could serve to better align content moderation with an informed understanding of the local or national context. To shed light on how this interaction could be designed, I will compare two different approaches: a consultative community forum recently announced by Meta and the model of Social Media Council that was developed by global free speech organisation ARTICLE 19 (2021). And I will argue that an independent Social Media Council (SMC) that represents the diversity of society would be in a better position to ensure an effective participation of local stakeholders than a consultative forum set up – and controlled – by social media companies.

Meta's Community Forum is described as a 'new method for making decisions' in the development of apps and technologies (Facebook, 2022). Drawing on the model of deliberative citizen assemblies, the Community Forum is meant to gather 'nearly 6,000 persons from 32 countries'. Organised by a global

social purpose company and a research centre of Stanford University, this global forum took place in December 2022 on a virtual closed platform that enables participants to discuss amongst themselves and to have access to experts. The non-binding recommendations of the forum will be made public at the end of the process.

The fact that the initiative is ongoing at the time of writing should not forbid initial thoughts on how it could contribute to solving the equation of integrating local context into content moderation processes. In this perspective, the forum certainly appears to be open to viewpoints from a broad diversity of countries. The partnership with expert organisations contributes to the legitimacy and credibility of the consultative process. However, for the specific matter discussed in this contribution, the community forum's main weakness lies in that it aims at making decisions globally, in a 'one size fits all' perspective. In addition, and even without looking at how participants are selected, its inclusivity inevitably remains limited: many countries, many languages, many cultures will simply not be represented in the deliberations. This being said, we could imagine a similar mechanism set at the national or regional level: assuming that participants would be selected in a manner that ensures a fair and transparent representation of the broad diversity of the living forces of society, such a deliberative process that combines learning (through access to expertise) with discussions could potentially deliver relevant recommendations on the alignment of content moderation with local circumstances. But there are two elements that could undermine the credibility and legitimacy of the process. First, the recommendations are not binding for the social media company, which ultimately remains free to base its decisions on whichever considerations it deems relevant. And the forum is commissioned, designed and controlled by a social media company: the participants do not have the opportunity to discuss the process itself. It also means, as illustrated by the recent disbandment of Twitter's Trust and Safety Council, that the company can at will bring an end to the forum, even very abruptly so (Former Members, 2022).

While a Social Media Council (SMC) could offer a comparable cocktail of deliberation and expertise (either through its own membership or external consultation), it presents significant differences. As defined in ARTICLE 19's approach (2021) an SMC is a multi-stakeholder mechanism that would oversee content moderation decisions on the basis of international human rights law both through the review of individual cases and the elaboration of general recommendations. While there have been debates on whether the SMC should be instituted as a global forum or as a national mechanism (GDPi, 2019). ARTICLE 19 argued in favour of a national SMC, except in situations where safety concerns would prevent the feasibility of such initiative.

For the purposes of this contribution, a first distinctive feature is that the SMC is meant to be created and operated by all the participating stakeholders, which means that its rules, processes and even its mere existence are not left to the discretion of a social media company alone – or, if it is thought of as a multi-platform council, of the participating social media companies. The funding of the SMC, which would come either from contributions of social companies, public funds or donations from international foundations, would need to be organised in a manner that prevents threats to its independence. A second point of contrast with a consultative Community Forum comes from the fact that the outcomes of the SMC's deliberations would be binding for the participating social media companies, which would have committed to executing its decisions in good faith. This ensures that stakeholders' voices could effectively carry some weight in terms of influencing content moderation practices. And finally, the ARTICLE 19 model proposed that the SMC should represent the whole diversity of society, including marginalised groups, which would confer additional democratic legitimacy to its deliberations.

While they present strong features, SMCs are not easy policy proposals. That is in part because legislators and stakeholders may be reluctant to the idea that the industry would take part in the definition of the rules that govern its behaviour, and that this very participation alongside other actors and under a regime of publicity, would ensure that compliance slowly percolates through the sector's practices. The complexity also comes from the need to resolve issues that are merely touched upon here, such as ensuring that funding does not undermine its sustainability and its independence, or the need to achieve an effective balance of powers between the various groups of stakeholders. The difficulties in setting it up notwithstanding, a national SMC could enable local stakeholders to bring their knowledge of the local circumstances to the analysis of content moderation issues affecting their society. In this model, the complexity of the local context would be analysed and discussed in the deliberations of a representative and diverse group of stakeholders, which ensures a higher degree of credibility and legitimacy than a consultative forum with limited representativeness. The same could indeed be said of the application of international human rights law to content moderation. In a multi-stakeholder Social Media Council, a social company would not be in a position to discreetly determine what the requirements of freedom of expression, privacy or due process are: instead, the definition of the rules that will govern content moderation would result from a collective deliberation of all the stakeholders that are impacted by the circulation of online content. The independence of the SMC and the (voluntarily) binding nature of its decisions reinforce its capacity to serve as an effective enabler for local stakeholders to contribute to the alignment of content moderation processes with the needs and realities of the society they live in.

At the time where a new European legislative framework for the digital world is entering into force, it should be noted that, as discussed elsewhere in this series, a national SMC could correspond to the definition of the out-of-court settlement mechanism provided for in Article 18 of the DSA. In any case, an effective and credible participation of local stakeholders in content moderation would provide a response to the disconnect between global companies and the societies that they impact: in the light of whistleblowers' revelations and of chaotic change of ownership, a local SMC could be key to restoring users' trust in social media platforms.

Sources

Access Now, https://www.accessnow.org/cms/assets/uploads/2020/03/Recommendations-On-Content-Governance-digital.pdf

ARTICLE 19, Content moderation and local stakeholders in Indonesia, June 2022, https://www.article19.org/wp-content/uploads/2022/06/Indonesia-country-report.pdf

ARTICLE 19, Bridging the Gap, June 2022a, https://www.article19.org/wp-content/uploads/2022/06/Summary-report-social-media-for-peace.pdf

ARTICLE 19, Social Media Councils: one piece in the puzzle of content moderation, Oct. 2021, https://www.article19.org/wp-content/uploads/2021/10/A19-SMC.pdf

Facebook, Improving People's Experiences Through Community Forums, Nov. 2022, https://about.fb.com/news/2022/11/improving-peoples-experiences-through-community-forums

Former members of the Twitter Trust and Safety Council, Statement, 14 Dec. 2022, https://www.article19.org/resources/twitter-disbanding-of-trust-and-safety-council

GDPi, From Concept to Reality, June 2019, https://cyber.fsi.stanford.edu/gdpi/content/social-media-councils-concept-reality-conference-report

Newton, C., The tier list: how Facebook decides which countries need protection, 25 Oct. 2021, https://www.theverge.com/22743753/facebook-tier-list-countries-leaked-documents-content-moderation

SFCG, Search for Common Ground, Handling harmful content online: Cross-national perspectives of users affected by conflict, April 2021, https://www.sfcg.org/wp-content/uploads/2021/07/SearchForCommonGround_Handling-harmful-content-online-report_April-2021.pdf

TBIJ, Facebook accused by survivors of letting activists incite ethnic massacres with hate and misinformation in Ethiopia, Feb. 2022, https://www.thebureauinvestigates.com/stories/2022-02-20/facebook-accused-of-letting-activists-incite-ethnic-massacres-with-hate-and-misinformation-by-survivors-in-ethiopia

UNESCO, Addressing hate speech on social media: Contemporary challenges, 2021, https://unesdoc.unesco.org/ark:/48223/pf0000379

Feminist perspectives on Social Media Councils

Rachel Griffin
SCIENCESPO, PARIS, FRANCE

'Better never means better for everyone…It always means worse, for some.'

— Margaret Atwood, The Handmaid's Tale, 1985

'When they enter, we all enter.'

— Kimberlé Crenshaw, 'Demarginalizing the Intersection of Race and Sex', 1989

Introduction

The overarching question we were asked to consider in these essays is whether Social Media Councils can reconcile private ordering of online spaces with public values. This seems obviously desirable, especially compared to the values that currently dominate online media: surveillance, consumerism and private profit. But 'public values' is a slippery concept, as the Handmaid's Tale quote suggests. Who is 'the public'? What do they value, and what should they? These questions are ideological and contestable. Values that enjoy consensus among elite groups, majority populations or the media don't necessarily serve everyone. Social Media Councils offer one way to constrain corporate power and make platform companies more accountable. But this raises more questions: accountable to whom, and for what? Who will they make social media better for?

As I'm exploring in my PhD research, discrimination and inequality are pervasive in contemporary social media – from content moderation systems which disproportionately censor marginalised groups, to recommendation algorithms which decide whose voices will be heard, and design features that systematically expose marginalised users to hate speech and harassment. If Social Media Councils are to address such issues, we should be asking how they can be designed to redistribute power towards those who are most marginalised within existing institutions, and how their own governance structures could reproduce existing inequalities.

As Kimberlé Crenshaw's work on discrimination suggests, this involves taking politically contentious positions, not just upholding universally-shared values – and it would make life more difficult for some people, like state security agencies or company shareholders. Yet prioritising marginalised communities can also create more inclusive, safe and diverse online spaces that in some sense benefit society as a whole: when they enter, we all enter.

In this essay, I argue that feminist legal theory can guide such an analysis. While it's not possible to do justice here to the diversity of feminist theory, I am influenced by Adam Romero's framing of feminism as methodology: one which is context-sensitive, empirically-informed, and attentive to how power relations are reproduced within legal institutions. In this sense, feminist legal theory aligns with an intersectional approach – not only asking the 'woman question', but highlighting multiple, distinct yet interrelated ways that legal institutions exclude certain perspectives and interests. On this basis, I highlight three analytical moves that I think are characteristic of feminist legal scholarship and suggest how they could be relevant to Social Media Councils.

Elements of a feminist analysis

What are the law's underlying assumptions?

Feminist methods question the perspectives and assumptions built into legal frameworks that are presented as neutral, consensual or objective. How does the law frame problems and potential solutions? What kinds of harms does it recognise? And where is the law silent? As Mary Anne Franks suggests in a feminist analysis of US privacy law, often the harms and concerns recognised in legal and political debate are those that trouble elite and/or majority groups, while harms affecting marginalised communities are not deemed matters of public interest. In the context of Social Media Councils, this suggests several questions that deserve more investigation.

First, many harms and injustices associated with social media are primarily experienced at the systemic or collective level, rather than by identifiable individuals – for example, because ad targeting or moderation software is probabilistically biased against certain groups. Will Social Media Councils be able to recognise, represent and redress such harms? Existing press councils, a major inspiration for Article 19's proposals, generally investigate issues through individual complaints, and focus primarily on protecting individual interests like reputation. This approach would prevent Social Media Councils from investigating or redressing the most consequential decisions platform companies make, such as the design, operation and resources of moderation systems. It also excludes entirely some bigger questions that feminist analyses might highlight, such as whether largely white, male, Western company executives pursuing shareholder profits should be making those decisions in the first place.

Second, what are the assumptions built into the normative framework Social Media Councils apply? This is typically assumed to be international human rights law, which is often presented as an objective or universally-shared set of values to guide platform governance. Feminist theorists, especially those writing from postcolonial perspectives, have questioned the supposed objectivity or neutrality of international human rights norms and institutions. Might this framework bias Social Media Councils towards a liberal, individualistic approach to social media governance, overlooking more structural, material and distributional issues?

It is certainly possible to discuss such broader questions, like the ownership and control of media resources, in human rights terms – but these more collective, structural interpretations of international human rights law are far from universally accepted. If Social Media Councils pursue consensual interpretations of human rights principles that all stakeholders agree on, this will in practice create a 'levelling down' effect where only the least disruptive, most corporate-friendly interpretations of human rights norms win – an outlook which is anything but politically neutral.

Whose voices are excluded?

As well as asking which values are prioritised, feminist legal theorists question whose voices are heard in practice in legal institutions. This is a crucial question for Social Media Councils, since inviting more multistakeholder participation in overseeing platforms is often seen as a way to make social media more inclusive and egalitarian – yet multistakeholder governance structures often fall short of this ideal in practice.

As Brenda Dvoskin shows, while civil society participation has many advantages, it cannot be assumed to offer an objective, unbiased or representative picture of the 'public interest'. Different social groups have different abilities to make their voices heard in multistakeholder fora, depending on factors like

economic and informational resources, popular and media support, and geographic position. This also reflects the preferences of platform companies – who, through funding and partnership programmes, exercise significant influence over civil society and research in this field. Often, these inequalities will track gender, race, class and other entrenched social divisions.

As well as composition and participation, a feminist approach would question what kinds of expertise and authority Social Media Councils value. Current proposals often seem to envisage them as quite technocratic, with authority deriving from legal expertise, in particular in human rights law. Human rights lawyers and academic experts are, to state the obvious, not representative sections of society. If Social Media Councils prioritise legal and academic expertise, they will primarily include people from privileged racial and class backgrounds – replicating the unrepresentative industry they are meant to oversee, which has been linked to biases and blindness towards the particular experiences and risks faced by marginalised users.

Moving forward, we should ask how Social Media Councils can be set up to include a maximally diverse range of voices, and to value lived experiences and grassroots activism as well as legal expertise and prestigious credentials. We should also ask whether stigmatised or marginalised social 'groups, like sex workers and migrants, will have meaningful opportunities to participate. As things stand, they are largely excluded – both discursively, because the interests of unpopular minorities are not seen as part of the general 'public interest', and practically, because they are less likely to have the necessary resources and access. As well as valuing and respecting their perspectives, Social Media Councils should materially support their participation, for example through grant funding. This of course then raises further questions about how to ensure the councils themselves are adequately funded, while remaining independent from platforms.

What are the gaps between theory and practice?

Finally, feminist research often incorporates more empirical or sociolegal methodologies, highlighting gaps between formalist interpretations of the law and how legal institutions actually operate in practice. One aspect of this is the much-discussed tension between formal and substantive equality. I have previously argued that formal protection of human rights in the DSA and other EU social media laws is likely to offer very unequal protection in practice. Similarly, formally equal rights to petition to Social Media Councils are unlikely to benefit everyone equally: people with more economic resources, free time, education and digital literacy will be much more able to press their case. This is another point in favour of a less individualistic approach. Social Media Councils should not only hear individual complaints, but should be empowered to proactively investigate structural and systemic governance issues.

Another related move is to emphasise the socioeconomic context in which legal institutions operate. Here, a particularly relevant question is the relationship between Social Media Councils and platform companies. Feminist sociolegal scholarship has highlighted that where the law delegates responsibility for enforcing the law or policy goals to companies – even seemingly progressive goals, like preventing gender discrimination – they are in practice often enforced in a way that suits corporate interests and demands little change to business practices. This is particularly the case where the law is abstract or focused on procedure, leaving companies to fill in the details.

In this context, given the sprawling complexity of contemporary social media, there is an obvious risk that Social Media Councils will mostly provide rather high-level, abstract guidance on how companies should approach social media governance – which the companies will then be able to interpret in selective and self-interested ways. If Social Media Councils are to operate within the industry as it is

currently configured, dominated by a few large private corporations, we need to ask how they can exercise sufficient power to demand substantive changes in business practices. However, we should also question whether meaningfully inclusive and equal social media governance is possible within this privatised corporate landscape – and whether Social Media Councils could help guide a transition to less commercial alternatives.

Three paths forward for Social Media Councils

On this note, I want to briefly offer some suggestions as to how these approaches could guide us in analysing and potentially constructing Social Media Councils in different contexts. Feminist legal theory can help us analyse the strengths and shortcomings of existing institutions and proposals. It can also help critique and inspire new institutions in platform governance – in both reformist and radical forms.

An existing example: the Meta Oversight Board

Feminist approaches could help us critique – and perhaps improve upon – the most-cited existing example: the Meta Oversight Board. Some aspects of the Board's approach are quite aligned with feminist perspectives, such as its attempts to use selected individual cases and advisory opinions to highlight systemic failings and unequal treatment in content moderation. However, feminist analyses might highlight the Board's unrepresentative composition, which clearly overrepresents Western perspectives and legal expertise. This can be linked both to its rather liberal, individualistic approach to human rights law, tending to prioritise individual freedom of expression over other values, and to its lack of independence from Meta, which continues to oversee the selection of Board members and has vetoed candidates with expertise in targeting, surveillance and other aspects of its business model.

Building on this, feminist analyses might highlight that the Board ultimately functions only within the parameters set by Meta. Its policy recommendations depend on Meta's willingness to implement them. The Board's responsibility is also defined by Meta, as asking whether the company is living up to its own stated values and policies. This means it can't address more fundamental questions, such as whether Meta's business model, surveillance practices, and global influence are conducive to creating an inclusive and egalitarian online environment. Feminist analyses might investigate whether alternative approaches to corporate self-regulation could give more weight to marginalised voices and exercise more leverage over companies.

A reformist project: the Digital Services Act

Second, Social Media Councils could be part of current reformist agendas, like the DSA, which aim to improve accountability within the current ecosystem of corporate social media. Importantly, many DSA provisions – especially very large platforms' obligations to address systemic risks, which Niklas Eder's contribution highlights as a key mechanism enabling regulators to address structural harms and discrimination – remain quite vague and abstract. How they will be enforced in practice is thus still up for debate. As Niklas suggests, civil society and academics can play a crucial role in shaping more progressive and radical interpretations. Through independent research, oversight and recommendations, Social Media Councils could push platforms and regulators to use these provisions to address inequality and discrimination in social media governance, highlighting systemic issues and underrepresented or minority concerns.

However, it is essential to consider how Social Media Councils themselves could become unequal and unrepresentative. How can they be made as diverse and inclusive as possible, which will require

extensive effort and resources, while also remaining independent? Instead of relying on abstract 'public values' which generally reflect majority or elite consensus, how can they enable inclusive political debate and contestation? As the DSA starts being implemented and enforced, feminist legal theory can provide guidance on how to incorporate Social Media Councils into its framework, but will also offer important methodological and analytical tools to investigate how such institutions function in practice and which perspectives and interests they prioritise.

A radical alternative: independent non-commercial governance

Finally, Social Media Councils could be part of more radical structural reform. Arguably, having the online public sphere ruled by profit-making companies – whose priorities are promoting advertisers' brands and selling people things they don't need, not creating safe and inclusive public spaces – is simply not compatible with feminist ideals. Socialist writers like James Muldoon and Ben Tarnoff have argued that governing internet infrastructure in line with public values requires 'demarketisation', with a bigger role for non-commercial and public services. More recently, introducing an explicitly feminist perspective, Rachel Coldicutt has argued for the provision of public resources to replace centralised corporate platforms with diverse online 'counterpublics'. Given the power of today's big tech companies, radical reform might seem distant – but the rapid recent ascent of decentralised non-profit platform Mastodon shows that alternative governance structures are possible.

Could Social Media Councils play a role in governing public and non-commercial platforms? In this context, a key issue is how to ensure that state funding and publicly-owned infrastructure do not translate into direct government control of platforms, which raises obvious issues around democracy and freedom of expression. Delegating governance or oversight functions to councils of civil society representatives could thus provide a path forward for inclusive and independent governance. Such models have successfully been established for older media, such as Germany's Rundfunkräte which oversee public broadcasters. We can learn from such examples, as Bettina Hesse's contribution (link) shows.

Yet this example also reminds us that ensuring that such bodies are inclusive and egalitarian is not easy. The Rundfunkräte remain highly skewed towards white Germans and dominant social segments like Christian organisations, while almost entirely excluding minorities like Muslims and migrants. Considering whose voices would be excluded or included, the international scope of social media also poses tricky problems. If public-service digital infrastructure is provided at the national or European level, is it possible to implement governance mechanisms which do not centre Western perspectives and interests?

Conclusion

If Social Media Councils are to make platform governance better – more rule-bound, more accountable, more democratic – we must ask: better for whom? Feminist legal theory suggests that behind the universalising rhetoric of public values, the default answer is often better for existing elites and privileged groups.

In a worst-case scenario, Social Media Councils might just be groups of mostly white, male, upper- and middle-class lawyers, offering non-binding recommendations on how currently-dominant companies could slightly tweak their procedures and policies to pay lip service to human rights law. On the other hand, if they are given the resources and authority they need, and if their design and operations consistently prioritise the inclusion and elevation of marginalised and minority groups, they could also

push platforms and regulators to address systemic inequalities, make platforms safer and more inclusive spaces, and establish more egalitarian forms of social media governance.

Achieving this will not be easy and will require critical analysis and concerted efforts to excavate both how women and minorities are excluded within contemporary social media, and how they might continue to be excluded within new governance institutions. In this essay, I have tried to identify some first questions that are suggested by feminist approaches to legal research, and which deserve further investigation as proposals for Social Media Councils move forward.

Learning from Broadcasting Councils

Bettina Hesse
VER.DI – UNITED SERVICES UNION, BERLIN, GERMANY

When considering the design of platform councils, it makes sense to look at possible role models. Independent institutions whose task it is to control and align with society communication spaces free of state interventions have existed for a long time: the twelve public broadcasters in Germany are supervised by broadcasting councils. For 70 years these boards have been meeting to discuss programmes, elect management and decide whether to release funding for new projects. Even though according to the Federal Constitutional Court, they are a very crucial component of the independent, non-governmental public service system, they operate mostly independent of the public and have barely been known in society. This changed abruptly when an alleged corruption scandal was uncovered in the summer of 2022, in which the chairman of the board and the director of the Berlin-Brandenburg Broadcasting Corporation were suspected of being involved. The following article explores the question of which conclusions can be drawn from the current crisis in German broadcasting supervision in order to set up platform councils more efficiently than the broadcasting councils have proven to be.

Social supervision as condition for independence

One of the constituent features of the publicly funded broadcasting system founded after WW II in Germany has been social control – with the aim to prevent political instrumentalisation and to foster democracy instead. Broadcasting is supposed to be free of any state interference and has to provide relevant contributions to everyone in society. In order to align it with public values, the responsibility for the journalistic-editorial mission of the public broadcasters has to be carried by the society. In 1961 the Federal Constitutional Court stipulated that society should be represented in the broadcasting councils by members of all significant political, ideological and social groups (appointed by the respective Federal State). According to the Federal Constitutional Court, they would serve as the "trustees of the general public interest" (1961) and constituted the "highest organ of the institution" (1971). This ruling was prompted by the observation that political representatives were increasingly dominating the oversight bodies. Despite the verdict, this tendency persisted and led to another court decision in 2014 that asked the councils to limit the amount of governmental and administration representatives to one third at maximum.

The composition of each broadcasting council is determined in different ways, depending on the legal basis in the respective Federal State. The State parliaments can send a certain number of members to the councils. In addition, organisations and associations are appointed, which may also send representatives. These are usually religious groups, employers' associations, trade unions, and in some States social associations, LGBTIQ*-groups, environmentalist groups etc. In addition, there are sometimes "citizens' mandates" for which citizens can apply to the State parliament, and in some cases, the broadcasting council can elect additional representatives.

Broadcasting Councils

The main task of the broadcasting councils is to review the fulfilment of the public service mandate in the form of ex-post control (decision-reviewing) – without intervening in the program, because the freedom to shape programming lies with the broadcaster's director. The council has to monitor legal

requirements in the programming, such as standards for the protection of children, respect for human dignity, freedom of expression and religion, freedom from discrimination, gender equality, the representation of existing political opinions etc.

Furthermore, the broadcasting council elects an administrative council, which has the task to supervise the finances of the broadcaster and to permit larger expenses (starting from 30.000 to 5.000.000 Euros).

In addition, the broadcasting council elects directors, suggested by the administrative council. Except for the election of the director-general and other directors, the broadcasting council has no right of initiative, yet almost all far-reaching measures taken by the director-general depend on the approval of the broadcasting council or the administrative board.

Since 2009 the councils also decides upon the introduction of new digital services by assigning "three-step-tests" – an assessment of the new service's contribution to the fulfilment of the given constitutional task.

The members of the broadcasting and administrative councils are volunteers. However, they receive expense allowances, meeting allowances and travel allowances – in very different amounts. In addition, depending on the broadcaster, a council member receives an amount starting from a fixed rate of 264 Euros per year (Radio Bremen) to 1.000 and up to 2.800 Euros per month (WDR). The councils receive (more or less) support from an office (with 0.5 to 7 FTE), which is sometimes equipped by the broadcasters, other councils can decide on the financial and personnel resources of their offices themselves.

Criticism and design issues of Broadcasting Councils

Systematic comparative studies on the staffing, working methods, decisions or the intra-organisational division of labour between broadcasting council and administrative council are rare in German research. Also in the public debate on public broadcasting and in the legislative process of structural development, the supervisory bodies play at best a secondary role. Therefore, the following remarks refer to a debate of a usually small professional public.

Various points of criticism can be extracted from the academic as well as the media debate. They indicate doubts on whether the councils actually fulfil their function or whether there are issues with their design:

Influence of political parties

Despite the 2014 verdict, 41% of the members in the council of "Deutsche Welle" belong to government or administration. With at least 18%, state representatives form the largest group in every council. Members with political affiliation tend to have strong influence in the councils due to the resources provided by their sending organisations, their respective expertise and opinion leadership and their degree of organisation within the councils ("Freundeskreise"). As a result, there is an imbalance of power and representation between the political and civil society councillors.

Representation in the councils

According to a study by "Neue deutsche Medienmacher*innen" the power relations prevailing in society are perpetuated in broadcasting councils: The socially marginalized (poor or disabled people, religious minorities, PoC, LGBTIQ*) are underrepresented, the average age of the members is relatively high (57.8 years) and more men than women or persons of other gender are members in the councils. This

gives rise to concrete problems, since the broadcasting councils monitor compliance with the programme mandate. Those programme standards typically include respect for human dignity, freedom from discrimination or gender equality – requirements whose non-observance particularly affects disadvantaged social groups. In case of contentious issues in broadcasting, it would be helpful to hear experts and stakeholders of these groups in the board, especially if they are not represented in the board themselves. The above-mentioned study points out that in most broadcasting councils such hearings either do not take place at all or only very rarely.

Associations without a seat on the broadcasting council often complain that the allocation process for seats on the council is not transparent and that there is no broad public consultation in the legislation process. With the last amendment to the WDR Act, achieving diversity in the broadcasting council was made even more difficult by the deletion of citizens' mandates. In contrast, the number of seats for state representatives remained untouched.

It is often argued that the broadcasting councils represent outdated social relations. The majority of councillors are chosen by political parties and associations – whose role in society has changed considerably over the last 75 years. Today, professional communities of users hardly organise themselves as classic interest groups. Therefore, young and digitally skilled people are structurally underrepresented in all councils.

Alignment between broadcasters and the public

Criticism is often raised that broadcasting has distanced itself from society more than a publicly funded system should (although representative surveys show a high level of general trust in public broadcasting). This could be interpreted as criticism in the broadcasting councils, since discussing the broad lines of the programme with the broadcaster is their key task (with the broadcaster holding the program design authority). In fact, many people contact the broadcaster with submissions, complaints and suggestions about the programme. However, submissions are answered by the directorate. Only if complainants are not satisfied with a reply, they can demand that their complaint is dealt with by the broadcasting council. Only then does the council hear about the audience's criticism. There is no institutionalized dialogue between the council and the public.

High engagement indicates a sense of ownership and voice, which certainly originates in the fact that all citizens in Germany (have to) finance broadcasting. Regarding the councils, the results of the survey #meinfernsehen2021 express the audience's strong desire for reform and change – even though it was not always clear to the participants how the supervisory bodies work, what exactly the thematic focus of their work is and what their relationship to the broadcasters is.

Transparency of the council's work

Since the beginning of the Covid-19 pandemic there have been livestreams for every council's full sessions, yet not from the meetings of the committees (or of the administrative boards), so the process of decision-making is not accessible to the public. The agendas are public, but difficult to find, and the minutes of meetings are not sufficiently informative. When important decisions are made or personnel changes, the councils (often: the broadcasters) publish press releases (sometimes only one release per year). Usually there is no media coverage of the meetings and only a few councils use communication instruments such as newsletters. All in all, the councils mostly act with distance to the public.

Independence from the media organisation

For many decisions the councils depend on information provided by the house's director, the house's legal department or the council's office (usually equipped by the house). No clear rules are communicated to most councillors on whether they are even allowed to commission external expert reports themselves, which is therefore a rarely used instrument.

Expense allowances are paid by the broadcaster, which certainly does not contribute to a sense of independence from this institution.

Self-concept of councillors

Some members of broadcasting councils identify as inspectors, whereas others are accused to act as co-managers of the broadcaster. The size of some councils (from 26 to 74 members) may contribute to diffusion of responsibility and low commitment of individuals.

Qualification and volunteering

There has been a long debate whether the broadcasting or administrative councils should consist of (more) media/law/accounting experts instead of representatives of the broad public. It can hardly be questioned that expertise is required for effective control. However, defenders of the pluralism model argue that expertise can be acquired on the job – as long as the term of office is long enough and the members find the necessary time for further training in addition to their main job. The concept of volunteering is in question as well, since - including all committee sessions - there can be up to 34 meetings per year – which is difficult to combine with a full-time job (and explains why many of the members join the councils when retired).

Continuity and identification

Many councillors report that it takes several years to be able to participate effectively in the council. When new members come into office every 5-6 years – or even earlier, in the case of shared mandates – the council's monitoring work lacks continuity. On the other hand, there are demands for term limits to prevent too much identification with the respective broadcaster.

Lack of flexibility

Media use and production have changed considerably since the councils were first introduced seven decades ago. The shortened production cycles of cross-media content require forms of handling and control which the old structures cannot react to adequately. Some, but not all councils have established telemedia committees, but no bigger changes have been introduced.

Efficiency

The efficiency of the control of broadcasting councils has been contested for a long time, as some councils had the reputation to just pass uncritically what the director suggested to them. A recent supposed case of nepotism including the director of one of the public broadcasting stations (RBB) and the chairman of its administrative council exposed the system's shortcomings quite obviously.

A clear lack of accountability was revealed: the members of the administrative council each worked alone, their work was not sufficiently transparent to the other members. This resulted in a lack of mutual

checks and balances, which apparently was exploited at least by the chairman and led to serious mismanagement.

Both broadcasting and administrative council realised that the decision papers they had received from the broadcaster's directorship and on which their decisions depended, were not as neutral and complete as assumed. The council's office apparently was not able to generate independent information. There was no legal base for the council or its office to commission external appraisals. And it turned out that the broadcaster's employers had received explicit instructions to withhold information from the administrative board.

Investigations around the incidents are still in progress, and it is well possible that personal misconduct played a bigger role. Nevertheless, these findings indicate shortcomings in the design of broadcasting councils. The fact that the irregularities at RBB have been revealed by journalists from a commercial news outlet and not by members of the internal supervision bodies speaks for itself.

Current legislation: more responsibility for Councils

In reaction to this scandal, the federal state of Hessen just passed a law, which is supposed to guarantee the independence of the council's office from the broadcaster.

In the next months, a new amendment to the Interstate Treaty on Media will be debated in the state parliaments. The amendment addresses some of the previously mentioned, long known concerns regarding the broadcasting councils. The new amendment to the Interstate Treaty on Media will grant them more responsibilities. Councils will no longer be limited to discussing the programme in the aftermath, but they will also set up quality standards for the programme in advance – a shift to rule-making. Furthermore, their responsibility in financial control will grow. They will review compliance with the principles of economy, efficiency and financial rigour, and monitor resources efficiency, which was previously the task of the internal revision as well as that of the commission for determining financial requirements in broadcasting (KEF). Given the recently disclosed control deficit in the RBB and the already known weaknesses in the board-design, these new tasks will be an enormous challenge.

The Interstate Treaty also provides for more dialogue between broadcasting stations and audience – a notable decision, given that the already existing broadcasting councils have been designed as a forum to align the programme with the audience. The legislator reacts to an increased need for participation in public broadcasting. A large number of dialogue formats between broadcasters and audience have been instituted following this widespread demand in the past years (online-panels, townhall-meetings, public debates, open-door days, focus groups for the further development of broadcast formats, WDR "Userlab" for testing digital applications) – without a legal mandate. In addition, accompanying research projects (like #meinfernsehen2021, similar to ARD-Zukunftsdialog 2021) gathered feedback on the quality of and ideas for public service broadcasting.

Leveraging experiences for supervision of new platforms

If the broadcasting councils are to contribute to a sense of ownership of the public, then public awareness of this institution and its work is a necessary prerequisite. It is therefore obvious that the main problem of the broadcasting councils is that the public knows far too little about them, their work and its impact. The very existence of broadcasting councils seems not to bring the intended legitimacy (anymore), there seems to be a higher demand for explanation and participation and an actual need for stronger societal control. If broadcasting councils are to serve as inspiration for Social Media Councils, the following

conclusions from experiences of broadcasting councils should be considered to improve the design of new supervisory bodies:

Legitimacy and alignment through visibility and transparency

- Transparency of the council's work, of consideration processes and of the impact of the council's decisions is necessary. The council's attitude towards the public must be one of openness. It has to communicate actively in order to raise public awareness for its existence and its work.
- A dialogue between council and audience should be established – as an occasion to explain the structures of the council and its work and to learn about the audience's concerns. The audience/users must be able to contact the council directly.

Adequate representation and accessibility

- A continuous process to guarantee the representation of an evolving society should be enabled. The general public should have the chance to participate in the composition process of a council.
- In order to prevent political dominance, it should be considered to include more social groups on a parity basis instead of proportional political representation that leaves room for fewer groups.
- Sending organisations should be encouraged to consider members of different ages, backgrounds, religions, sexual orientation and gender identities and disabilities when selecting their representatives. Previous experiences concering gender equality show that only binding regulations work.
- Representation deficits of people without affiliation could be reduced by partly random selection of council members, by publicly elected members or by institutionalised dialogue formats that are promoted and open to the public. Other possibilities would be to oblige political parties (or equivalents with fixed seats) to select representatives from a pool of defined organisations or to publicly advertise their positions.

Quality and independence

- When sending delegates, organisations should prioritise relevant expertise over prominence of individuals.
- Councils should be provided with support in form of independent and reliable information/preparation of decisions. The council must have access to independent media research. Ensure that the council has the legal competence to commission expert opinions itself.
- Members should be provided with resources to fulfil their task with the necessary diligence. Civil society members may need paid leave from their main job in order to adequately fulfil the tasks of the council and for trainings. Financial compensation must be based on their needs.
- Allowances should be paid by an independent entity, not by the supervised organisation.
- Committees/discussion forums should consist of a small number of people so that everyone has their say and feels commitment.

Platforms considerably differ from public service broadcasting in many aspects. However, since there are decades of experience – and criticism – with broadcasting councils, it might be worth taking note of these supervisory bodies when designing social oversight structures for new communication spaces. Undoubtedly, learning from broadcasting councils, the design must be carefully crafted to provide for

efficiency, transparency, independence and representation, so that councils can really improve alignment with the public interest and contribute to an infrastructure's legitimacy.

Assessing the systemic risks of curation

Niklas Eder
OVERSIGHT BOARD, LONDON, GREAT BRITAIN[1]

The future of content moderation (self-)regulation is systemic

Systemic risk assessments are the next big thing in content moderation. They carry a gigantic promise, which is to address content moderation where it really matters. If things go well, they will force platforms to engage in the "proactive and continuous form of governance" that we have been waiting for. Systemic risk assessments, so the hope, will constitute the second core pillar of content moderation (self-)regulation, and overcome the limitations of individual remedy mechanisms, which constitute the first pillar. The amplification of content, the design of recommender systems, the newsfeed — here referred to as curation practices — are out of reach for individual remedy mechanisms, yet are believed to be responsible for the deterioration of civic discourse, electoral processes, public health, security and fundamental rights. Articles 34 and 35 of the Digital Services Act (DSA) come to rescue, and finally establish obligations for platforms to assess and mitigate these risks, ideally with participation of stakeholders. These risk assessments are then sent to auditors, the Commission and the Board for Digital Services, which will review the risk assessments, develop best practices and guidelines, and will potentially require platforms to take alternative mitigations measures or even fine them. This is huge leap forwards for content moderation. It's also a jump in the unknown.

The compelling idea behind systemic risk assessments is that, to mitigate the harms of platforms, we need to go beyond individual remedy. The great challenge, which this approach entails, is that we have no idea how to measure and mitigate systemic risks of content moderation. The DSA vaguely defines the sources of risk (Article 34 § 2), the kinds of risks (Article 34 § 1) and the kinds of mitigation measures platforms must consider (Article 35). Beyond that, systemic risk assessment in content moderation are unknown territory for industry, academics, and regulators alike. We need to make systemic risk assessments work to address the societal harms of content moderation, but we don't know yet how. The next months offer a unique opportunity for academia to help transform systemic risk assessments into the effective mechanism they could be, rather than the harmless paper tiger they might become.

A virtuous loop to assess systemic risks

Beyond defining sources of risk (Article 34 § 2), kinds of risks (Article 34 § 1) and the mitigation measures platforms must consider (Article 35), the DSA outlined responsibilities and processes around systemic risk assessments. It basically creates a loop in which, once a year, platforms efforts to assess and mitigate risks are evaluated and recalibrated. Platforms are obliged to work with stakeholders and assess risks and develop mitigation measures, and to submit a yearly report on their efforts to the EU Commission, the Board for Digital Services and auditors. They will then assess whether platforms comply with their obligations under the DSA, or whether they should alter their risk assessment efforts and take alternative mitigation measures. After the Commission and the Board for Digital Services

[1] All views expressed in this article are strictly Niklas Eder's own.

received the first round of risk assessments from platforms, they will begin developing best practices and guidelines which platforms, in their future risk assessments, will be expected to apply. While the setup of this process is reasonable, it raises as many questions as it answers. The success of systemic risk assessments ultimately depends on who takes what role in that loop, and how those involved execute their particular functions.

Neither platforms nor EU institutions are particularly well equipped to decide over what systemic risk are and how they should be mitigated. Platforms are ill equipped, because self-assessments by corporations whose ultimate goal is to maximise profits, lack credibility. The Commission and the Board for Digital Services are ill equipped, because public institutions should not dictate the rules governing content moderation and with that public discourse. It also appears rather unlikely that auditors will play a significant role in shaping systemic risk assessment. Conventionally, and this matches the description in the DSA, the role of auditors is to assess whether a company has complied with legal requirements. Auditors are usually large companies, likely with no significant expertise in content moderation and certainly with no particular legitimacy with regard to defending the public interest. The cannot set the standards governing systemic risks.

This leaves one more role open, on which we may rest our hopes, which is the role of stakeholders. This role is only established in the recitals of the DSA, indicating minor relevance. However, taking into account the practical and theoretical concerns raised here, stakeholders might be best equipped to shape the standards based on which systemic risks should be measured and mitigated. The Commission could easily empower stakeholders, if it decided to give their contributions significant weight in its evaluations of the submissions of platforms, and if it reflects them in their best practices and guidelines. Perhaps this is the true ingenuity of the DSA's systemic risk assessment provisions: They create a loop in which the Commission can use its enforcement powers to empower stakeholders. This virtuous loop might solve the conundrum of public actors regulating speech: It allows them to assure that platforms will account for public interest, without public actors having to define what that concretely means themselves.

The hard question, which will be addressed in the remainder of this essay, is what shape stakeholder engagement should take, if it is them who are supposed to develop the standards based on which to assess and mitigate risks. Simple stakeholder consultations, conducted by platforms, might have little effect. Much speaks for a more substantial cooperation between platforms and stakeholders. We can think of a variety of constellations, many of which already exist to a certain degree. For example, platforms can — breathe in, breathe out — work with Social Media Councils to establish the standards which govern systemic risk assessments.

Social Media Councils

The cooperation between platforms and Social Media Councils can take different shapes and different degrees of intensity. Fuelled by the DSA, different models could emerge.

On one end of the spectrum could be cooperations which leave control largely in the hands platform, allowing them to decide what questions the Social Media Council should answer, what access they get to information and leaving them discretion if and how to implement solutions suggested by the council. On the other end of the spectrum, the cooperation could consist in platforms conveying certain competencies to Social Media Councils, providing them with access to relevant data, allowing them to chose the questions they wish to engage with, to develop their own frameworks and empowering them to make recommendations on some of the most important questions pertaining to curation and systemic

risks. The platforms would not necessarily have to follow all recommendations that Social Media Council make, but at least would commit to respond to them — and justify before the Commission and the Board for Digital Services if they did not follow the recommendations the Social Media Councils made. The Social Media Councils would in turn work with relevant stakeholders to reach their recommendations, ultimately strengthening the impact all stakeholders have on the focus and outcome of systemic risk assessments.

Relying on Social Media Councils has significant advantages for all parties involved, including platforms and the Commission: Platforms would benefit from involving Social Media Councils in their process of assessing and mitigating risks, as Social Media Councils would ultimately help decide the hardest questions platforms face. If platforms chose to implement the recommendations by the Social Media Council, and incorporated them into their strategy to assess and mitigate risks, this conveyed credibility to the systemic risk assessments they submit to the EU. In response, platforms could expect positive evaluations, non-interference from the side of the EU institutions and would minimise the risks for fines. The involvement of Social Media Councils allows the Commission to restrain from defining the concrete standards against which to measure content moderation and curation practices, as any public institution should, while catalysing a productive, a virtuous loop which empowers civil society. Thorough the platforms' work with external experts and Social Media Councils, the Commission and the Board for Digital Services receive high quality risks assessments, based on which they can develop practical and effective best practices and guidelines without itself unduly interfering with the working of platforms.

Empowering positive data rights with platform councils

Dominik Piétron
HUMBOLDT UNIVERSITY BERLIN, BERLIN, GERMANY

Every social interaction on the Internet, whether in social media, e-commerce, or digital public services, presupposes a certain form of data collection, data storage, data processing and data transfer. However, data governance practices mostly remain opaque. Especially the most influential data intermediaries, digital gatekeeper platforms such as Google, Facebook, or Amazon, are autocratically ruled by a non-transparent managerial elite. Numerous court cases have highlighted how the platform's data governance can be significantly harmful to individuals and groups, workers, and businesses, as well as the public discourse. Nevertheless, users are increasingly disclosing very intimate personal information about their behavior, social relationships, needs, and desires, thus boosting the data power of platform companies. This central contradiction between informational self-determination and algorithmic heteronomy illustrates the blatant lack of fundamental democratic principles in the platform economy.

In this article, I argue that rights-preserving data governance that adequately addresses the individual and social harms caused by digital platforms requires a new institutional form for collective decision-making. Given that every platform operator is confronted with the need to govern personal data, democratic platform councils are a promising institutional framework for many areas of the digital society – not only for social media platforms, but also for e-commerce and public service platforms. Just as English entrepreneurs in the 16th century demanded a representative parliament to decide how to spend their tax revenues, and workers in the 19th century fought for the right to establish works councils, today's platform users require some sort of platform council to debate on the data governance mechanisms and enforce their data rights.

I will proceed in four steps: First, I will elaborate on the need for democratic data governance on a collective level that originates from the individual right to informational self-determination. Second, I argue that platform councils are a well-suited institutional framework for democratic data-governance. Third, I develop design principles for platform councils that form the minimum conditions for empowering data rights of individuals and collectives. Fourth, I empirically illustrate my argument using the example of "Berlin.de," a German city platform that acts as a central digital access point to Berlin's public infrastructure.

The need for democratic data governance

The challenge of data governance arises, inter alia, from the paradox that the right to informational self-determination is legally enshrined in the EU with the GDPR but can hardly be implemented in today's platform economy. There are two reasons for this, which can be explained along the vertical-horizontal-axes of data relations developed by Simone Viljoen (2021: 607): First, there is such a strong vertical asymmetry of power between users and platform operators that user's consent to processing of personal data is often coerced rather than voluntary (ibid.). Due to the strong concentration tendencies in the platform economy, some platforms like Alphabet's and Apple's smartphone OS and app stores, Amazon's e-commerce empire, Microsoft 365 or Meta's social media group have become essential infrastructures of our everyday life and important preconditions for social participation. But sector-

specific platforms such as AirBnB or Uber are also increasingly displacing analog services and are already without alternative in some regions of the world. The problem is that users must agree to far-reaching data usage declarations when they sign up for platforms, thus allowing platforms to gain "quasi-ownership" through the enclosure of personal data. As the digital socialization of the individual extends to more and more areas of life, the individual is severely overburdened with a rapidly increasing number of decisions to be made about the processing of personal data.

In addition to vertical power asymmetry, a second reason for the inadequacy of individual contract-based data rights, according to Viljoen, is horizontal data relationships between different users (ibid: 609). Horizontal data relationships occur automatically when many people are given a digital identity that tracks their behavior, records it in a digital form, and aggregates it in a database with other data subjects. As people are assigned different data values, relative differences become visible and social patterns respectively horizontal data relations emerge. Individuals can be correlated and categorized into groups, enabling population-level insights and predictions that form the basis of most data-driven value creation – as well as many data-driven social harms such as algorithmic discrimination, disinformation, or behavioral control. The crucial point is that the sharing of data by one person in a group also allows conclusions to be drawn about all the other people in the group. As social entities, humans are always affected by the data decisions of their friends, peers, neighbors, coworkers, etc. This is most obvious in communication between users, where the message becomes public as soon as only one user shares the data. Thus, individual data reflect social relationships with others and thus represent an inherently collective resource that cannot be reduced to individual choices.

Consequently, the individualistic approach of the GDPR is increasingly criticized as insufficient. There are calls from various sides to treat aggregated personal data as "data commons" (Bria 2018, Shkabatur 2019, Zygmuntowski et al. 2021), that belong to a community and therefore require a collective governance framework – just as Nobel laureate Elinor Ostrom has emphasized the commons nature of natural resources such as forests, waters, or traditional knowledge (Ostrom 1998). Singh and Gurumurthy (2021) even propose an additional legal codification of collective data rights. They hold that the rights to determine the use of a social group's data, be it the users of a digital platform or the residents of a city, should belong to the respective community as a collective subject. But even without the demand for collective data rights, a superindividual data governance is necessary, since the "bundle of data rights" (Kerber 2021) must be assigned among various actors, that include rights on non-personal data as in the case of IoT products. Either way, a supra-individual decision-making process is required, in which the rights of the individual, such as the right to data portability or the right to be forgotten, fully exist and are supported. Starting from such a rights-based data governance approach requires a collective decision-making process in which all individuals decide collectively on the use of their aggregated personal data. Since that can only be legitimized by the aggregated will of those affected by data processing, Viljoen concludes that a "democratic data governance" (Viljoen 2021: 648) is needed to enforce informational self-determination.

Platform councils as an institutional form for democratic data governance

Since platforms are technically algorithmic infrastructures for controlling data flows, data governance is at the heart of every "platform governance" (Gorwa 2019). Each digital platform has established its own data governance structure, that can be defined as the sum of decision rules, rights and obligations for the collection, storage, processing, and sharing of data within an organization and between an organization and external actors (Abraham et al. 2019, Khatri/Brown 2010). As information brokers it

is their very function to reduce transaction costs by deciding on the adequate social embedding of data flows. By specifying a data governance framework, platform operators have to give answers to a multitude of difficult data questions: Which product should be listed first in the search results? At what point is a social media post no longer acceptable? How should algorithms standardize social interaction in the digital space? How much personal information is necessary for a service to deliver the best user experience?

Numerous scholars have highlighted the legitimacy deficit of platform operators, especially regarding oligopolistic decision-making at big digital platforms such as Amazon, Google, Facebook, or Twitter (Griffin 2022). While on the one hand European data laws such as the Digital Services Act or the Digital Markets Act ensure that platforms follow rule of law principles, on the other hand the idea of "multistakeholderism" is emerging to increase civil society's influence on platform governance through transparency, consultation, and participation (ibid). In this context, platform councils have recently become a normative approach through which platforms renegotiate their relationship with their users and demonstrate their commitment to public interests. Especially in the field of social media platforms, the first platform advisory boards were established, such as Meta's Oversight Board or the advisory boards of Twitter, Tiktok, Twitch and Spotify (Kaye 2019, Kettemann/Fertmann 2021).

There are good reasons why platform councils could be a useful addition to social media platforms as well as e-commerce and public service platforms. Numerous court cases have shown how data management on e-commerce platforms like Amazon can harm individuals and groups, workers and businesses through its power to select and amplify certain data objects. The same applies to public service platforms, which often hold a local monopoly and thus, in the course of digitization, act as a central information provider for a city and increasingly moderate content.

Beyond that, however, platform councils could be an appropriate institutional form for democratic data governance, which necessarily arises from the tension between data rights and algorithmic heteronomy. In the platform economy, it is the rule that users de facto give up their rights when they agree to the terms and conditions of a gatekeeper platform and have no further say in the processing of their data. This is particularly problematic because many users are tied to specific platforms for which there are no alternatives, and platform operators can easily reprogram the platforms' algorithmic infrastructure, leaving users with the choice of accepting the new terms or losing access to the platform. Bringing platform governance in line with European data rights would thus mean putting informational self-determination at the center and understanding them not just as a negative right to protection, but as a positive right to co-design data infrastructures. To this end, a collective decision-making process is needed that gives platform users a say in the design of the platform architecture that channels their data. Platform councils could strengthen this positive data right and overcome the two main shortcomings of the General Data Protection Regulation: They are collective rather than individualistic, so that horizontal data relationships are considered, and they are dynamic and process-oriented rather than static contractual relationships, so that data rights are also taken into account as platforms evolve.

Design principles for positive data rights-based platform councils

The crucial question then is, how does a platform council that supports a democratic data governance look like in practice? When it comes to designing governance mechanisms for collective goods, Elinor Ostrom's work on the institutional diversity of natural commons all over the world is a valuable point of reference. Especially Ostrom's approach to develop certain "design principles" (Ostrom 2010: 652) as opposed to a one-size-fits-all model can be of great help here. Ostrom starts with a resource to be governed collectively and a community of people who have a stake in the resource, since they are

producing or using it or are somehow affected by its usage. Applied to the challenge of data governance in the platform economy, the resource in question is aggregated personal data, and the community boundaries can be constituted by platform users as well as external individuals affected by decision making based on the aggregated personal data of the platform. Based on this assumption we can derive the following design principles for data rights-based platform councils in various forms:

Inclusive discussions and feedback opportunities

All platform users – but also non-platform users affected by data governance of a platform – should be able to participate in an inclusive discussion about how user-generated data is collected, aggregated, analyzed, shared, and fed back to users. In addition to the primary use of data on the platform, it should also be decided in particular which data can be made available for secondary use by third parties. The agenda-setting should be open and autonomous so that every governance mechanism can be debated. This requires an institutionalized discussion forum in which proposals for change and concerns can be raised and discussed. To organize this process, the platform council should select moderators to facilitate the discussion and develop individual proposals based on stakeholder comments.

Democratic accountability

To ensure democratic accountability, clear guidelines need to be established about who decides on proposals and how – from a consensus democracy approach, in which all stakeholders must agree, on the one hand, to a more representative approach, in which a board of elected representatives decides on new governance mechanisms with a relative majority, on the other hand. Given the complexity of data governance issues, platform councils should act as a coordination body that organizes this decision-making process. The more decisions the platform council makes, the more important it becomes to legitimize its composition and to hold it accountable through democratic elections. In free, fair and secret online elections, the various stakeholder groups of the platform (including platform users, platform workers, and external stakeholders) should select their representatives for the platform council. If appropriate, quota can ensure representation of individual groups or minorities. Elections should be held at regular intervals every one to five years so that citizens can hold council members accountable for their decisions.

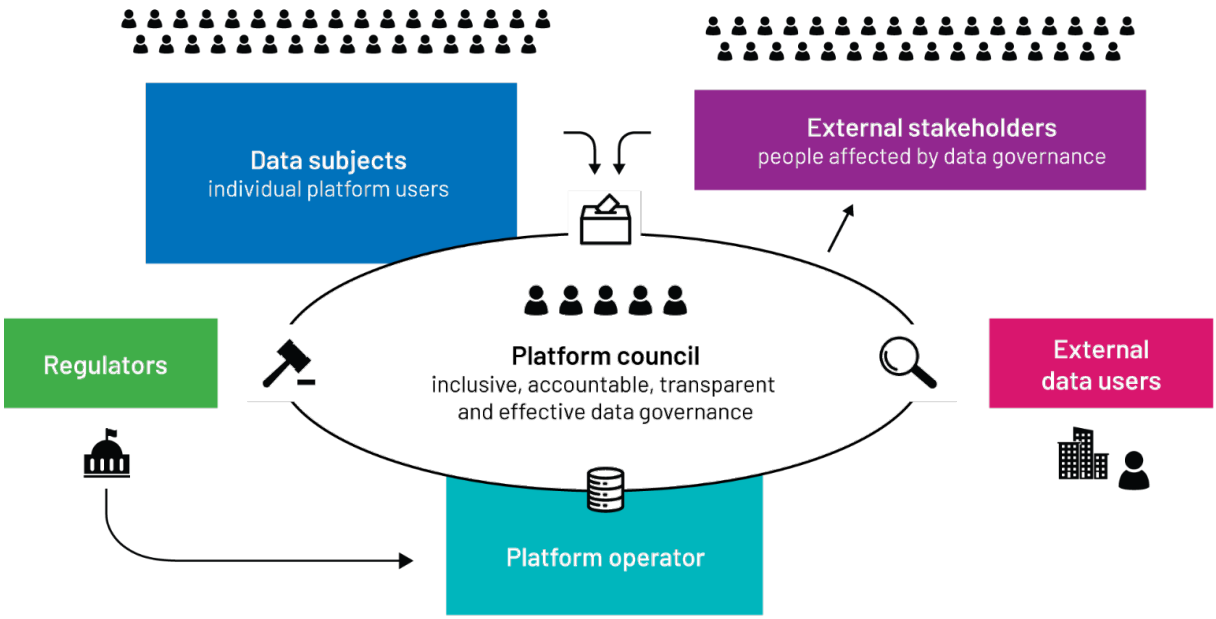
Algorithmic transparency

Since platform infrastructures often act as algorithmic black boxes, transparency of algorithms and data governance practices is key for a rights-based data governance. Only if stakeholders know, how their digital identities are constructed and how their online and offline behavior is affected by the algorithms of the platform, they can perform their right on informational self-determination. To verify the implementation of the resolutions, the platform council must have comprehensive rights of access to information vis-à-vis the platform operator's technical infrastructure. Platform councils shall contribute to informed decision-making of stakeholders by highlighting potential risks and clarifying trade-offs among competing interests of stakeholders.

Effective implementation

It must be ensured that the decisions of the Platform Council are effectively implemented. To this end, the councils must be technically and legally independent of the platform operators in order to enforce democratically made decisions against the will of the platform operator if necessary, as research on data

trusts has shown (Delacroix and Lawrence 2019). The independence of the platform council must also be ensured by means of a secure financial basis, whereby funding can be provided either by platform operators or by publicly funded media regulators. Furthermore, the voluntary commitment of the platform operators to implement the decisions of the platform council may not be sufficient. If a platform operator does not cooperate, platform councils must have the ability to seek assistance from data and consumer protection authorities to hold the platform operator accountable.



Graph 1: Visualization of a platform council based on the design principles described above.

The case of city platform 'Berlin.de'

While platform councils are currently being established primarily on social media platforms, similar institutions are just as relevant for e-commerce platforms or public service platforms. Public service platforms are also playing an increasingly important role in the digital accessibility of public infrastructure. The behavioral data generated in the process is an important feedback mechanism for administrative modernization and should at the same time be publicly accessible wherever possible, but it also harbors risks and must be protected in certain areas. Also, the question of what personal data should be collected and how it should be processed is not trivial and should be developed in the spirit of informational self-determination, at best with the participation of the users themselves. In the following, the proposed data governance structure based on platform councils is illustrated with the case of "Berlin.de", the central public city platform of Berlin. Recently, the public administration has initiated a process to introduce a platform council, and now develops its own framework to institutionalize collective decision-making in the context of an infrastructural city platform.

The platform Berlin.de was launched in 1995 and successively expanded to become the central digital access point to Berlin's e-government services, providing information on urban infrastructure such as mobility, culture and events, tourism and travel, shopping, and business. It also includes the Berlin Open Data portal and the participation portal mein.berlin.de, which makes the platform a complex digital ecosystem used by various user groups such as citizens, public servants, local businesses, and tourists.

In 2021, the platform was bought by the municipality and turned into public property. In this way, a process was started that continues to this day and, inter alia, led to the idea of a platform council. Already during the communalization, the Berlin House of Representatives decided to do a thorough relaunch of Berlin.de, including a central focus on data protection and inclusive platform development together with local civil society organizations (Abgeordnetenhaus Berlin 2021). Shortly before, several civil society organizations organized in the Bündnis Digitale Stadt Berlin had already published an open letter in which they proposed a participatory process for co-designing the city platform (Bündnis Digitale Stadt Berlin 2021). In the following month, the civil society alliance organized an expert workshop and a public discussion event together with public servants, where the demand for an own platform council was formulated for the first time. In summer 2022, the city government agreed to start a process to develop a platform council for the city platform Berlin.de. A first internal meeting with representatives from the administration and civil society took place in September where both sides agreed to the need of platform councils as a tool for participatory platform-co-design and collective data governance. In October 2022, the common goal of a platform council for Berlin.de was re-affirmed in the House of Representatives (Abgeordnetenhaus Berlin 2022).

However, the concrete institutional form of the Platform Council Berlin.de has not yet been determined. The House of Representatives in October 2022, and the Alliance for the Digital City of Berlin proposed three guidelines for the design of the city's administration, which are largely based on the design principles developed above in section 3:

- The Platform Council should act as an interface between the administration and the population. It should be a tripartite multi-stakeholder body composed of elected representatives of the various parties of civil society, of the citizens of Berlin and of the public administrations involved, thus combining elements from direct and representative democracy.
- The main objective of the Platform Council is to formulate recommendations for the administration on the development of planned or ongoing projects related to the city portal. To this end, its agenda-setting may not be restricted to certain topics but be open able to respond to users' needs.
- For the users to participate and hold their representatives accountable, the advisory board should meet in public, obtain information from the administration, and conduct its own surveys of platform users.

Sources

Abgeordnetenhaus Berlin (2021): Das Stadtportal berlin.de in öffentlicher Hand neu aufstellen, Nr. 2021/81/42 Drucksache 18/3834, https://pardok.parlament-berlin.de/starweb/adis/citat/VT/18/PlenarPr/p18-081bs3240.pdf.

Abgeordnetenhaus Berlin (2022): 13. Sitzung des Ausschuss für Digitalisierung und Datenschutz am 05.10.2022, Youtube, https://youtu.be/USRCs54JxJY?t=2381.

Bündnis Digitale Stadt Berlin (2021): Info-Sheet „Rekommunalisierung des Stadtportals Berlin.de" – Geschichte, Technik, Herausforderungen, https://digitalesberlin.info/info-sheet-stadtportal/.

Bria, F. (2018): New Deal on Data, London/New York: verso.

Sylvie Delacroix, Neil D Lawrence (2019): Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance, *International Data Privacy Law*, 9 (4): 236–252, https://doi.org/10.1093/idpl/ipz014

Gorwa, Robert (2019) What is platform governance?, Information, *Communication & Society*, 22(6): 854-871, https://doi.org/10.1080/1369118X.2019.1573914.

Griffin, Rachel (2022): Public and Private Power in Social Media Governance: Multistakeholderism, the Rule of Law and Democratic Accountability, http://dx.doi.org/10.2139/ssrn.4190500

Kaye, David (2019): Social Media Councils. From Concept to Reality, Conference Report, https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/gdpiart_19_smc_conference_report_wip_2019-05-12_final_1.pdf

Kettemann, M./ Fertmann, M. (2021): Platform-Proofing Democracy. Social Media Councils as Tools to Increase the Public Accountability of Online Platforms, Friedrich Naumann Stiftung (eds.), https://shop.freiheit.org/#!/Publikation/1084.

Kerber, W. (2022): Specifying and Assigning "Bundles of Rights" on Data, in: Hofmann/Raue/Zech (ed.): Eigentum in der digitalen Gesellschaft (06.06.2022), Mohr Siebeck, Tübingen, pp. 151-176.

Ostrom, E. (2010): Beyond Markets and States: Polycentric Governance of Complex Economic Systems. *American Economic Review* 100(3): 641–72.

Singh, P. J./ Gurumurthy, A. (2021): Economic Governance of Data Balancing individualist-property approaches with a community rights framework, IT for Change (ed.), http://dx.doi.org/10.2139/ssrn.3873141

Viljoen, S. (2021): A Relational Theory of Data Governance, *The Yale Law Journal*, 131(2): 573-654.

Zygmuntowski, J. J./ Zoboli, L./ Nemitz, P. F. (2021). Embedding European values in data governance: a case for public data commons, *Internet Policy Review*, 10(3). https://doi.org/10.14763/2021.3.1572.

Promises and perils of democratic legitimacy in Social Media Councils

Clara Iglesias Keller
WEIZENBAUM INSTITUE, BERLIN, GERMANY

Theresa Züger
ALEXANDER VON HUMBOLDT INSTITUTE FOR INTERNET AND SOCIETY, BERLIN, GERMANY

This piece approaches Social Media Councils (SMCs) as institutional responses to the democratic deficit of social media platforms. Our first goal is to analyze their promises and perils to advance democratic legitimacy as framed in democratic theory, and thus assess their potential to address the ends that inspired their creation. Against the concentration of power of private platforms (van Dijck, Nieborg, and Poell 2019; Cohen 2019) and depending on their format, SMCs represent an opportunity to improve compliance with procedural guarantees, or to include more stakeholders in platform-led decision-making processes. Meanwhile, their perils lie in the possibility that, instead of indeed promoting the idea of democratic legitimacy that founded them, SMCs end up contributing to validating private self-regulatory institutions that might improve internal procedures, but bring little remedy for existing power imbalances. Against this assessment, our second goal is to outline relevant possibilities for governance designs of SMC and discuss them regarding their potential to create a more legitimate governance of social media platforms. In particular, we will focus on enhancing input legitimacy of such institutional forms, all the while supported by the concept of digital citizenship. Increasing input legitimacy, as we see it, is the necessary first step to increase the legitimacy of platform power overall.

Digital platforms have developed wide reaching political power, as they affect the lives and rights of citizens worldwide. They "play a unique public role in society, shaping culture, economies and societal relationships" (Haggart and Iglesias Keller 2021) and their effective power can be seen as an aspiration to govern in ways which were reserved for states alone in the past. Social media companies are some of the private actors that "aspire to displace more government roles over time, replacing the logic of territorial sovereignty with functional sovereignty", subjecting users to ever more "corporate, rather than democratic, control" (Pasquale 2018). In social media, this control is notably exercised through content moderation, as the set of practices and policies through which platforms interfere in access to information and the right to freedom of expression (a task that encompasses the collection and treatment of personal data). As private companies, they do so with no bound to a democratic process, which has been broadly scrutinized by the literature (Klonick 2020, Cohen 2019, Gillespie 2018).

In recent years, SMCs were increasingly established under the premise to help rectify these deficits. As we argue in the following, they can play an important role in the direction of a stronger democratic legitimation of plattforms's power to address this democratic deficit. However, many of the existing examples still inherit structural shortcomings which leave wide areas of democratically illegitimate power untouched.

Contextual remarks - SMCs as a "piece of a puzzle"

In order to analyze SMCs potential, we start by contextualizing their development in relation to the greater platform governance ecosystem. Overall, platform governance encompasses several "layers of

governance relationships structuring interactions between key parties in today's platform society, including platform companies, users, advertisers, governments, and other political actors" (Gorwa 2019). Due to a "distinct redistribution of governance capacity away from states" (Wagner 2013), platform governance has long been materialized in informal and hybrid forms of rule making. The growing relevance of digital platforms has enhanced claims for embedding their operations with law enforcement and democratic legitimacy standards (Haggart and Iglesias Keller 2021), both through co-regulatory initiatives (Marsden 2011) and state regulation (exemplified in Europe by the recently approved Digital Services Act). Along these initiatives, we also witness the emergence of a hybrid governance landscape formed by institutions that were "developed in a consent based fashion independently, although occasionally steered by state action" (Gorwa, Iglesias Keller, Ganesh 2022). This "hybrid governance" landscape can be seen as a continuation of the important developments in platform governance that have unfolded outside national states' capacities, through voluntary, informal and collaborative governance arrangements (Gorwa 2019, Douek 2019). From expertise focused and industry initiated bodies, like the Global Internet Forum to Counter Terrorism (Ganesh 2021), to governmental or co-regulatory bodies, this landscape includes a variety of institutions marked by different motivations, initiators, members, competences, and levels of state participation (including none).

Against this background, SMCs arise as institutional initiatives distinguished by their aim to improve legitimacy in social media, by providing a solution to companies' "unchecked system" for users' speech governance (Klonick, 2020, p. 2476; Douek, 2019, p. 46). However, even within the very idea of SMCs we can find different institutional forms that operate on different premises and promote different ideas of democratic legitimacy. For instance, platform-initiated SMCs are usually implemented on single companies' own terms, to bolster their internal processes with specific procedural safeguards. We want to call these **internal SMCs**. Meta's Oversight Board, as an example of such, is meant to provide a second instance to Facebook content moderation decisions and to issue recommendations on its policies. Similarly implemented by an internal process, Twitter's Trust and Safety Council gathered a group of organizations that advised on the development of its products, programs and rules. The council was dissolved by Elon Musk after he took charge of Twitter, which exemplifies one of the core problems with internal SMC - their dependance on the companies' leadership decisions. Despite its resolution, we want to recognize the Trust and Safety Council as one of the probed models of an SMC. Twitter's approach differed from Meta's Oversight board, since the Council was constituted by civil society organizations and not staffed by handpicked individuals by the company. However, it was still entirely self-regulatory, which depends (as Elon Musk demonstrated) on the good will of the platform and yet does not effectively change input legitimacy.

In a different approach, which we want to call **external SMCs**, proposals for cross-platform multistakeholder SMCs - like the one from Article 19 - focus on voluntary-compliance mechanisms that provide "a transparent and independent forum to address content moderation issues on social media platforms on the basis of international human rights standards." Despite also being pinned on self-regulation, this version of SMCs presumes "independence from social media companies and participation of representatives of various stakeholders to regulate the practices of the sector" (Article 19).

These (mostly unaccounted) variations of ways through which SMCs can improve legitimacy reflects what is indeed an issue to the approach of democratic legitimacy in platform governance. As previously pointed out by Haggart and Iglesias Keller (2021), there is a need for a multidimensional approach to assessing governance in this field, since focusing on single specific aspects - like multistakeholderism, procedural principles of the rule of law (Suzor 2019) or even international human rights-based

frameworks - reduces the complexity of platform-governance legitimacy, particularly as it relates to democratic legitimacy. Moreover, it overlooks the importance of infiltrating these policy-making processes with safeguards regarding how these decisions are made, and by whom.

Breaking down democratic legitimacy in SMCs

Therefore, we apply the framework tailored by Haggart and Iglesias Keller (2021), according to which democratic legitimacy in the realm of platform governance is not a one-dimensional concept, but rather as a multifaceted phenomenon that must be assessed as such. The authors build on Vivian Schmidt's (2013) modification of Scharpf's (1999) taxonomy of democratic legitimacy as it relates to the European Union - i.e., in an exercise of drawing democratic legitimacy outside of the nation-state. According to this framework, the democratic legitimacy of a policy regime can be divided into three parts. As per Scharf's original contribution, input legitimacy refers to the "responsiveness to citizen concerns as a result of participation by the people," while output legitimacy refers to the "effectiveness of the policy outcomes for the people (Schmidt, 2013, p. 2). To this, Schmidt adds a third category, "throughput legitimacy," which highlights the quality of the governance process and "is judged in terms of the efficacy, accountability and transparency of the (...) governance processes along with their inclusiveness and openness to consultation with the people" (Schmidt, 2013, p. 2).

As a sphere of policy making beyond state's capacities, platform governance legitimacy assessments can profit a lot from Schmidt's framework, especially with regard to the dynamics between the three different strains of legitimacy (Haggart and Iglesias Keller 2021). Although conceptually separate, these different forms of legitimacy interact in ways that may reinforce or undermine each other, and this nuance is mostly overlooked in platform governance policy analysis. For instance, high input legitimacy may be able to compensate for low output legitimacy. Low input or low output legitimacy can mitigate eventual positive effects of high throughput legitimacy. "Importantly, high throughput legitimacy tends to have a minimal effect on input and/or output legitimacy, while low throughput legitimacy tends to have a negative effect on input and/or output legitimacy" (Haggart and Iglesias Keller 2021).

The latter outtake is of particular importance for SMCs, as they largely rely on improving procedures. However, as Schmidt's framework shows, low input and output can compromise the effectiveness of policies that hang on throughput, and this might just be the case for some key SMC conceptions. Adding a second adjudicatory instance to content moderation, increasing transparency or diversifying policy advice has little potential to alter the power status quo if the rules and policies, as well as these throughput legitimacy standards, are still determined by the same players. In this sense, SMCs follow the tendency of other limited platform governance initiatives that "rely excessively on throughput legitimacy for their overall legitimacy", when in fact, they should also "focus on whether they involve rule by and for the people, not just on processes" (Haggart and Iglesias Keller 2021).

Lack of people participation and of public awareness not only decrease the democratic legitimacy of SMCs, but can also compromise whatever developments they allow in the realm of throughput legitimacy. Ultimately, their essence of democratic legitimacy will be more rhetorical than concrete. To avoid this, we further elaborate on ways to improve input legitimacy in SMCs.

Who are "the people" represented by an SMC?

Beyond arguing along the lines of our claim to further input legitimacy in platform governance, we introduce the idea of digital citizenship as a helpful concept. In particular, it addresses the problem that any body of people governed requires an identity and self-understanding, as argued by Schmidt (2013).

For any process of democratic governance, one essential question to begin with is, who are "the people" - both governed and in power? And in the case of platform governance? The idea that platform users are the governed body and that user-engagement can be seen as a participation providing democratic legitimacy is ultimately insufficient. Many of the effects of platform power, like surveillance and a change of communication structures, affect people despite having chosen themselves to use a specific platform. User-engagement can "still not be equated to the expansion of the legitimate power of citizens" (Haggart and Iglesias Keller 2021).

Instead, the idea of the "digital citizen" might be helpful to the question of who "the people", whose input legitimacy should be improved, is in this scenario. The concept of digital citizenship is far from being a universally defined and agreed upon term, but as Jørring Valentim and Porten-Cheé (2018) phrase it: "The concept of digital citizenship has the potential to capture the shifting role of citizens under online conditions." Some understandings of digital citizenship reduce it to the literacy of using the internet ("ones' abilities to access, use, create, and evaluate information and to communicate with others online". A second approach focuses on "the ethics of internet user's behaving appropriately, safely, ethically, and responsibly" (Choi 2016). A third understanding introduces digital citizenship as "different types of online engagement, including political, socio-economic, and cultural participation" (Choi 2016) which most often still refers to state governed politics rather than the self-governance of digital citizens beyond the state.

Last but not least, digital citizenship can also be seen as "another front in citizens' struggle for justice" (Emejulu and McGregor 2019). Understanding citizenship in this context encompasses not only as a set of given rights and duties in relationship with the state but also a claim to new rights and means for participation, as Isin & Ruppert describe in their book "Being digital citizens" (2020). Even though they argue that digital citizenship is a citizenship yet to come, they allow for the possibility of rights claims being made by all humans as potential digital citizens towards old powers of state institutions, as well as new power structures such as tech companies and platforms. This understanding, which is often framed as radical digital citizenship, "is the insistence that citizenship is a process of becoming - that it is an active and reflective state for individual and collective thinking and practice for collective action for the common good. Radical digital citizenship is a fundamentally political practice of understanding the implications of the development and application of technology in our lives" (Emejulu & McGregor 2019). It "is to resist the idea that a neutral technology exists" (Emejulu & McGregor 2019).

As Schmidt envisioned, the governed people would need "a sense of collective identity and/or the formation of a collective political will" (Schmidt, 2013, p. 5). A collective identity of a digital citizenry certainly does not already exist entirely, but we think that it is also an undeniable fact that most humans living today are confronted with a digitized world and oftentimes with a rising awareness that the technology which is the digital infrastructure of this world is not neutral towards them and others. Digitization impacts people's lives very differently but most often widely. Most humans living today act in a digital world in some way. Despite the question, if people are making use of it, in the spirit of human rights they have a right to deliberation, participation and for collective self-governance regarding the digital world. One could say that also the identity of a digital citizenry is a body "yet to come", a dynamic public that is in a process of formation, and constant re-formation.

How can Social Media Councils be designed to start addressing the democratic deficit that platform's power creates?

To answer this question, we outline relevant possibilities for governance designs of SMCs and discuss their potential to create a more legitimate governance of social media platforms. While our focus here is on features that could potentially improve input legitimacy, we have also listed throughput and output related ones, considering how these different fronts interact.

Firstly, the existence and structure of SMCs could be mandated by law, which would give their existence and governance structure a due process, setting out the rules for participation in and communication to such SMC. Moreover, once bound by legislation, SMCs would be within nation-states' democratic design, supported by the flux of legitimacy incoming from popular vote. Such a normative form would be no novelty either for regulatory theory or communication and technology policies, where a number of public-private hybrid arrangements run under the label of co-regulation (Iglesias Keller 2022). Here, Ayres and Braithwaites' idea of "enforced self-regulation" stands out, as the cases where private actors are mandated by government to elaborate on a set of rules aiming at specific purposes (1992).

Secondly, to serve as a legitimate organ, such boards should be constituted in such a way that they ideally represent all stakeholders' interests. The concept of multistakeholderism, which is long practiced in internet governance is helpful in this regard, while the question remains, how to identify and include *all* relevant stakeholders. One way to approach this, could be an inclusive process prior to the actual constitution of the SMC arranging participatory events to identify and recognize local stakeholders.

To meet all digital citizen's needs, an SMC governance structure must be tailored to global and local challenges and interests. This requires a set-up on the national level. International governance structures might be relevant but we will leave them out of the scope of this paper for it exceeds the level of complexity we can address here. If the goal is to make SMC responsive to citizens' concerns as a result of participation, channels must exist that allow for citizens' concerns to be heard and for threats to human rights to be considered.

The national board could be a mix of

- political representatives, nominated by the executive power and approved by Parliament, as it is the case in several independent specialized regulatory bodies;
- a number of elected citizens (users and non-users) of the platform, meeting broad diversity to increase a sense engagement for citizens and allow for more direct representation;
- a number of local advocates of vulnerable groups; and
- potentially changing guests to the board processes, e.g. academics, journalists, activists or human-rights advocates to represent in specific issues.
- Additionally, the framework of the cross-platform SMC would need to be designed in a way to allow for equal terms of participation.
- For this, all board members must be equally paid for their efforts and reimbursed for missing their normal jobs, in the case of temporary participation; legal oversight needs to be organized to ensure due process of the SMC election, and
- Continuous work of the council membership in the board could be restricted to the given election period of the parliament to ensure the change of personnel.

Lastly, what rights and setup should the board have?

- The SMC should have far reaching rights, which are not necessarily restricted to internal functions of the platform such as content moderation but also strategic decisions of the platform policies that might affect citizens;
- the SMCs work could be organized as interventional acts, meaning they have the right to question or veto a decision in the process and intervene in different types of decisions of the SM platform on a national level;
- Regarding content moderation, their competences would ideally include not only oversight over second-instance to content moderation decisions, but also the possibility of reviewing the internal policies upon which these decisions are taken (including terms of use, community standards and their update);
- Moreover, SMCs could be equipped to receive and process complaints brought to their attention by users or non-users of the platform through different, easy to access channels. This could include channels for dialogue between the public and the SMC, to be established on a periodic basis or not;
- Decisions of the SMCs should (with possible exceptions for privacy reasons) be documented and made public, to increase public awareness of the SMC;
- Also, to increase public knowledge and transparency of the SMC, council meetings should be open for visitation by the public at certain hours and in appropriate settings.

As it shows, such a design of an SMC is meant to serve the specific purpose of promoting a more comprehensive idea of democratic legitimacy of social media governance. Depending on their specific competences, SMCs potential to fulfill this goal will vary, but should not be expected to rectify all democratic shortcomings of social media. Our approach is not to be taken as the single way for social media platforms to become spaces driven by the public interest, but a proposal to address more severe democratic deficits and improve institutional forms currently implemented in such spaces.

References

Article 19, Social Media Councils. Available at: https://www.article19.org/wp-content/uploads/2021/10/A19-SMC.pdf. Accessed on: 01 dec. 2022.

Choi M. (2016). A concept analysis of digital citizenship. Theory & Research in Social

Education, 00:1–43, DOI: 10.1080/00933104.2016.1210549

Cohen, J. E. (2019). *Between Truth and Power: The Legal Constructions of Informational Capitalism* (1st ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190246693.001.0001

Emejulu A. & McGregor C. (2019). Towards radical digital citizenship in digital education.

Critical Studies in Education, 60:1, 131-147, DOI: 10.1080/17508487.2016.1234494

Ganesh, B. (2021, March 16). *Platform Racism: How Minimizing Racism Privileges Far Right Extremism* [Social Sciences Research Council]. https://items.ssrc.org/extremism-online/platform-racism-how-minimizing-racism-privileges-far-right-extremism/

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

Gorwa, R. (2019). What is platform governance? *Information, Communication & Society*, *22*(6), 854–871. https://doi.org/10.1080/1369118X.2019.1573914

Iglesias Keller, C. (2022). The perks of co-regulation: An institutional arrangement for social media regulation? In E. Celeste, A. Heldt, & C. Iglesias Keller, *Constitutionalising Social Media*. Hart Publishers.

Klonick, K. (2018). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, *131*, 1598–1669.

Marsden, C. T. (2011). *Internet Co-regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace*. Cambridge University Press.

Oversight Board, Ensuring respect for free expression, through independent judgment. Available at: https://www.oversightboard.com/. Accessed on: 01 dec. 2022.

Pasquale F. (2018). Frank Pasquale on the Shift from Territorial to Functional Sovereignty, https://blog.p2pfoundation.net/frank-pasquale-on-the-shift-from-territorial-to-functional-sovereignty/2018/01/16

Scharpf, F. (1999). *Governing in Europe: Effective and Democratic?* Oxford University Press.

Schmidt, V. (2013). Democracy and legitimacy in the European Union revisited: Input, output and "throughput", *Political Studies, 61(1)(2013), pp. 2-22.*

Wagner, B. (2013). Governing Internet Expression: How Public and Private Regulation Shape Expression Governance. *Journal of Information Technology & Politics*, *10*(4), 389–403. https://doi.org/10.1080/19331681.2013.799051

Social Media Councils under the DSA: a path to individual error correction at scale?

Aleksandra Kuczerawy*
CENTER FOR IT AND IP, KU LEUVEN, BELGIUM

* *This chapter benefited from funding from FWO grant nr. 1214321N and a stipend from the Cyber Policy Center at Stanford University. I would also like to acknowledge helpful feedback and advice from Assistant Professor Evelyn Douek at Stanford University.*

Introduction

Content moderation by online social media is an arduous process. It consists of private entities making decisions that impact the effective exercise of the right to freedom of expression, oftentimes without meaningful accountability. It also tends to upset everyone who, even remotely, cares about the outcome. Social Media Councils (SMCs) have been presented as a solution to resolve this conundrum. The idea has not been widely picked up so far, although examples can be found. In Germany, a self-regulatory NGO "Freiwillige Selbstkontrolle Multimedia-Diensteanbieter" (FSM) has been set up by YouTube and Facebook, under the NetzDG, to decide difficult content removal cases. The most prominent example, of course, is the Meta Oversight Board. Both bodies resolve only a small selection of cases and do not fully reflect the original idea of SMCs. But is this about to change thanks to the Digital Services Act (DSA)? In an attempt to regulate content moderation practices of online platforms, the DSA mandates the use of internal complaint-handling mechanisms as well as external out-of-court dispute settlements for individual error correction. This essay looks at both paths to remedy to see if they align with the original idea of SMC. Could SMCs fit the framework of the DSA? This essay will focus on two aspects in particular, namely independence and scalability. Furthermore, this essay looks at the special case of the Meta Oversight Board (MOB) to answer whether, and under what circumstances, the MOB could serve as an out-of-court dispute settlement body within the meaning of the DSA.

Social Media Councils in a nutshell

In 2018, the UN Special Rapporteur for Freedom of Speech David Kaye recommended that "all segments of the ICT sector that moderate content or act as gatekeepers should make the development of industry-wide accountability mechanisms (such as a social media council) a top priority". It was a strong endorsement of Social Media Councils (SMCs), an idea that had been propagated by Article 19 to address the unaccountable power of social media over people's right to freedom of expression. According to the original concept, a Social Media Council is a multi-stakeholder voluntary-compliance mechanism for the oversight of content moderation practices on social media. An SMC would provide a transparent and independent forum to address content moderation issues on the basis of international human rights standards. An SMC would have several roles. For starters, it would serve as an appeal body for individual content moderation decisions. It would also provide general guidance on content moderation policies in line with international human rights standards and act as a forum where all stakeholders could discuss and adopt recommendations. Compliance with SMC decisions and recommendations would be on a voluntary basis. Its core characteristics include independence from government and from any particular social media company. It should include representatives from all

relevant stakeholders, such as media associations, media regulatory bodies, freedom of expression experts, academia, and civil society appointed through a democratic and transparent process. Moreover, SMCs should receive a stable and appropriate level of funding to ensure its independence and capacity to operate. Funding could come from social media companies (at least partially), as well as public subsidies, other stakeholders, or philanthropic organisations. Ideally, SMCs would operate on national (or regional) basis to ensure participation of decision-makers with the best knowledge of the local context and understanding of cultural, linguistic, historical, political, and social nuances.

So far, the idea has not been widely picked up, even though it received considerable praise from academics, international policy experts, and civil society organisations. Is it, perhaps, about to change with the arrival of the DSA? Can the provisions on redress mechanisms encourage the implementation of the SMC and facilitate its operationalization?

Remedies under the DSA

Interestingly, during the negotiation process of the DSA the idea of establishing Social Media Councils was contemplated. A group of MEPs in the IMCO committee proposed an amendment (1686) that would require the establishment of an independent advisory group, representing recipients of the service, or groups potentially impacted by services. The European Social Media Council (ESMC) would be independent of commercial interests, with expertise and competence in the area of international human rights law, content moderation, algorithmic systems, media, consumer protection, disinformation, hateful speech, etc. The ESMC would issue non-binding guiding principles and recommendations to improve content moderation processes; foster a participative and transparent public debate around content moderation; and issue policy and enforcement recommendations. The ESMC, therefore, would not carry out any review or correction of individual decisions. The proposed amendment, however, did not make it into the final position of the European Parliament.

The DSA, aiming to strengthen fundamental rights online, focuses strongly on the right to freedom of expression but also the right to effective remedy. This latter right, in the context of content moderation, is translated into a requirement that any erroneous decision regarding content should allow for rectification. The DSA offers three different redress mechanisms for content moderation decisions: internal complaint-handling mechanisms (Art. 20), out-of-court dispute settlement (Art. 21) and judicial redress (no specific provision but several mentions). They can be used in sequence or separately, as self-standing mechanisms.

According to Art. 20, providers of online platforms should create an internal complaint-handling system and make it available for at least six months from the time a measure against a piece of content was taken. The complaint-handling system should be easy to access and user-friendly. It should handle complaints in a timely, non-discriminatory, diligent and non-arbitrary manner. And most importantly, review by such a complaint-handling system should lead to a reversal of a previous content moderation decisions, if the complaint contained sufficient grounds to justify such reversal.

For disputes where a platform stands firmly behind its action (e.g. because it is in line with their internal policy), the DSA foresees an external redress mechanism. For that purpose, Art. 21 introduces the out-of-court dispute settlement. It is meant for complaints that could not be satisfactorily resolved via the internal mechanism, or would not be resolved via the internal mechanism because the complainant went straight for this second option.

A complainant should be able to select any out-of-court dispute settlement body certified by a competent Digital Services Coordinator. Such certification is then valid in all EU countries. Conditions for

certification include impartiality, and independence, including financial independence and independence at the level of the persons resolving disputes. Certification also requires the necessary expertise in one or more areas of illegal content or application of terms and conditions. The rules of procedure under which the settlement body would operate should be clear, fair and easily accessible. The bodies should reach a decision within a specified time frame (90 days with possible extension of another 90 days for complex disputes). The costs of the procedure are spread between the platform and a complainant, depending on whose claim succeeds. But the main financial burden is on the platforms, while the users should be able to access the settlement bodies free of charge or at a nominal fee.

Parties to the dispute should engage in good faith with the selected body in an attempt to find a solution. The dispute settlement bodies, however, do not have the power to impose a binding settlement of the dispute on the parties.

Match or no match?

So how do these two redress mechanisms correspond with the original idea of SMCs? And more specifically, can they constitute a pathway to operationalize the SMC idea in fulfilling its intended role in individual error correction?

Internal complaint handling systems are, arguably, a necessary first instance redress path that offers a possibility of a quick dispute resolution free of charge. It is also, usually, able to handle a large amount of complaints. The main focus of the internal review is a compliance assessment with the platform's existing policies, rather than questioning if the policies are in fact reasonable. Although important to efficiently resolve (at least some) content moderation disputes, this mechanism clearly does not amount to an SMC. Internal complaint mechanisms are, as is explicitly stated, internal. This means that they are administered by the platform, financed by the platform, accountable to the platform and dependent on the platform's choices regarding content policies. It is therefore, everything but independent from the platform.

The out-of-court dispute settlement, on the other hand, seems to be a better match. First of all, the out of court settlement bodies under the DSA would be external, outside of the platform's own structure. Like for any other ADR mechanism, this is a crucial element and it aligns the requirements for an SMC. The DSA does not make any demands regarding the composition of the settlement body. It might be difficult for any such body to achieve the ideal composition of an SMC that calls for broad participation of all relevant stakeholders (i.e. media associations, media regulatory bodies, freedom of expression experts, academia, and civil society appointed through a democratic and transparent process). But it could definitely strive for a broad representation of interests. In any case, as long as the decision-making members are independent and impartial, the body would still be able to fulfill its intended role.

Another similarity between SMCs and dispute settlement bodies under the DSA is that the decisions of the SMC and the settlement bodies shall not be binding. The idea of an SMC is based on the tradition of self-regulation and voluntary participation. No legal obligations should be necessary, therefore, to ensure voluntary compliance with the SMC decisions. Arguments for not making the decisions binding under the DSA were discussed during the negotiation process. It was argued that binding decisions would create extra-judicial bodies, would lead to must-carry orders and would prevent platforms from obtaining redress. Either way, the non-binding approach corresponds with the general requirements articulated in Directive 2013/11 on alternative dispute resolutions.

The DSA requires certification of the settlement body by (at least one country's) Digital Services Coordinator and specification of the languages in which the procedure is available. In theory, that could

mean only one dispute settlement body for the whole EU. Considering the amount of languages and cultures, as well as types of content to cover, a variety of specialized bodies (thematically, geographically and linguistically) would be more likely. This d the requirement that the decision-making in SMCs happens at a level that would ensure knowledge of local context and nuances.

According to the original concept, a complaint to the SMC would be possible only after all possibilities of remedying the issue with the social media company have been exhausted. In contrast, a complaint to an out of court settlement body can be initiated at any stage, even without using the platform's internal complaint mechanism first. The difference in approach here, however, is not a source of major incompatibility and mainly impacts the number of potential complaints. Charging users for using the out of court dispute settlement procedure, even if at nominal fee, could have a similar limiting result (especially since the internal mechanism can be used first, for free).

Effectively, the out of court settlement bodies whose role would be to review specific content moderation decisions, upon complaint by those affected by the decision, closely resemble the original SMC concept. This is also a view held by the proponent of the original SMC idea, Article 19, which indicated that a SMC could serve as the out-of-court dispute settlement mechanism required under the DSA.

It should be highlighted, however, that a dispute settlement body under the DSA would fulfil only one of the roles foreseen for SMCs, namely the rectification of errors in individual instances. It would not function as a body designed to provide guidance or give recommendations to social media platforms.

Independence is key

One element that keeps coming back in the description of both SMCs and dispute resolution mechanisms is the requirement of independence. But which facets of independence should be considered in the context of content moderation disputes? And since we're talking about private enforcement mechanisms, why does it even matter?

Access to an independent and impartial tribunal is at the core of the right to fair trial and the right to effective remedy. Coming from international human rights instruments, such as the ECHR and CFREU, these rights are primarily addressed to States, instructing them how to organize their judicial system. In the context of private enforcement, these rights serve mainly as guideposts. An ideal to look up to, without an obligation of strict compliance. Yet they should be given careful consideration, as the aim of the out of court settlement mechanisms is, in fact, to provide effective remedies.

To be clear, neither the out of court dispute settlement bodies under the DSA, nor the SMCs can be considered independent tribunals in the understanding of art. 6 ECHR. Apart from being private bodies, they do not have the power to issue binding decisions. Nevertheless, it is still worth looking at how the ECHR and ECtHR understand independence. It is also worth checking how this concept has been translated for the purpose of assessing the independence of regulators.

First and foremost, independence is understood as a guarantee that there is no undue influence on the decision-making process. It is a condition for parties to accept the outcome, trusting that the process was fair and there was no external interference. Independence refers to freedom from interference by both the executive power and by the parties that may have interests in the outcome. Independence could be impacted in different ways. It is possible, therefore, to distinguish components, such as the manner of appointment of the members of the decision-making body, the existence of sufficient safeguards against the risk of outside pressures, as well as financial freedom. There is no formula to calculate independence

that specifies the weight of individual components. Depending on the circumstances, deficiencies in one area (e.g. financing by a party subject to the decisions) may be compensated by other elements (e.g. existence of strong safeguards against undue influence).

One component of independence that can be hard to assess is the appearance of independence. Appearance of independence comes from the idea that "justice must not only be done, it must also be seen to be done". The main purpose is maintaining confidence and trust of the public in the independence of the decision making body. The appearance of independence may be compromised when one of the parties has reason to believe that independence of the decision-maker is not guaranteed and it may be triggered by various factors. This fear, however, must be objectively justified.

In the context of private enforcement, independence is crucial to legitimize operations of such out of court dispute settlement bodies. The source of financing as well as the manner of appointment of the members must therefore be clear and transparent. Strong involvement of one party in either of those elements does not necessarily exclude independent decision-making but it must be accompanied by sufficient safeguards against the outside pressure. It might impact, however, the appearance of independence. A complainant might question, for example, the independence of a body that happens to be fully financed by the opponent in the conflict. External positioning of the out of court dispute settlement bodies alone does not always guarantee independence but it strongly supports it. Adding additional degrees of separation between the decision-makers and those affected its decisions contributes to the appearance of independence. It also helps to increase confidence in objectivity and impartiality of decisions. Providing clear, fair and accessible rules of the procedure ensuring that the process of resolving conflicts is handled in a fair manner is another element that improves the trust and, as a result, perception of legitimacy.

Meta Oversight Board as a global cross-platform dispute settlement body?

The remaining question is whether the Meta Oversight Board (MOB) could become a body that satisfies the requirements of an SMC and a dispute settlement body under the DSA.

Dubbed as Meta's own Supreme Court, the MOB is an example of voluntary corporate self-regulation. Its role is to review a selection of Facebook's and Instagram's removal decisions, to oversee the enforcement process and to provide policy recommendations. The cases handled by the MOB are selected by a committee that focuses on cases that raise important policy questions across the platform, that have a major effect on public discourse, and have impact on large numbers of users. The number of cases that the MOB has actually analyzed remains limited. To illustrate, in Q2 of 2022, out of 347,304 cases submitted the MOB has resolved three.

In order to ensure the independence of the MOB's decision-making process, several structural and procedural measures have been introduced. While the first members of the MOB have been selected by Meta, the MOB will select its future members on its own. The members are not Meta employees and cannot be removed by Meta. The initial members of the MOB include a Nobel Prize-winning press freedom activist, a former prime minister, and several legal academics and former judges. Funding of the MOB comes from Meta, but it has been organised through a trust to protect the independence. Moreover, the MOB operates according to a charter and bylaws, describing its operational procedures. With all these precautions, and despite criticism of the whole concept, the independence of the decision-making process of the MOB is generally not questioned.

Despite all these properties, the MOB is not actually an SMC within the original understanding of the concept. The initial excitement over this endeavor has somewhat faded, and it is currently perceived more of a PR move created to legitimize Meta's content moderation practices. But could the MOB become something more and provide an answer to both civil society and EU policymakers calls for independent oversight? In particular, could the MOB serve as an out-of-court dispute settlement body within the meaning of the DSA, also for other online platforms?

In theory, the MOB could evolve to fulfill the role of a dispute settlement body within the meaning of the DSA. The number of changes it would have to undergo, however, are rather significant. First of all, Meta would need to make even more concessions to distance itself sufficiently from the new body to maintain the appearance of independence. Moreover, keeping the focus only on Meta's activities would be incompatible with the requirements listed in the DSA.

Moreover, the MOB would have to become certified and established in at least one EU country. Ideally, it should be operating in multiple EU countries to make sure that all the languages as well as specific social, cultural, historical and political nuances can be properly taken into account. In addition, the process of selecting the members and the financing structure would have to change, especially if this new version of the MOB was to review content moderation decisions not only by Facebook and Instagram but also other social media platforms. And here comes a major difficulty that the new MOB would encounter: the number of decisions to make. From a body that very selectively picks a handful of cases to review, it would have to become a body that is able to review each individual complaint. And it would need to be able to decide not only complaints about removals and non-removals but also complaints about any other action taken against content, such as restrictions on visibility, suspension or termination of a service or an account, and decisions restricting the ability to monetize content.

The amount of people required to cover the incoming complaints would be enormous. For this purpose, the current MOB structure cannot be replicated at scale. Workforce appropriate in size to process all the complaints cannot be filled with Noble-prize winners and law professors receiving 6-figure salaries. After all, it is not the idea of the DSA to end up with another layer of low-paid overworked content moderators processing complaints against the clock. Unless its operations remain local (e.g. focusing only on one country and one language), creating an out of court dispute settlement body of this size would basically require Meta to open a completely new and parallel business venture. While it is not impossible, of course, the necessary changes would be far from trivial. And the final result would differ significantly from the MOB in its current form.

Conclusion

The idea of SMCs is still waiting to be picked up and implemented in practice. With the help of the DSA, it could be operationalized to become a more realistic tool able to address the problem of unaccountable decision making with an impact on fundamental human rights. Even though there are some differences between the original SMC concept and the out of court dispute settlement mechanism under the DSA, they are not fundamentally incompatible. The main challenge would be to earn the trust and confidence of the public that such a body is able to review the content moderation decisions in a fair, independent, and efficient manner. Only then, the public would refer their grievances in a search of effective remedy for content moderation decisions. The MOB, at the moment, only vaguely resembles the original SMC idea. In order to become an SMC that would also fulfill the requirements of Art. 21 DSA, it would have to undergo a massive makeover. For starters, it would have to significantly grow in size to become a mechanism for individual error correction. It would also have to distance itself even

further from Meta to maintain the appearance of independence. The necessary changes would make it nothing like it is right now. But perhaps for the better?

Social Media Councils as Self-Regulatory Bodies

Riku Neuvonen
TAMPERE UNIVERSITY, UNIVERSITY OF HELSINKI, FINLAND

Introduction

In principle, Social Media Councils are self-regulatory bodies; in this context, self-regulation can be defined as regulation of the private sector by nongovernmental entities. Hence, it is the opposite of public regulation, that is, hard regulation (Black 1996). During the past few decades, these modes of regulation have been mixed, with the hybrid regulatory approaches being termed coregulation. Coregulation has been typical for audio-visual media, in which classification systems have been enacted by law or at least endorsed by authorities (Marsden 2011, 157–160). In the EU, innovative regulation has been part of the so-called 'new approach' towards legislation, in which industry-led standardisation and regulation are the key components. However, regulating public debate is very different from regulating a cucumber trade (Quintel and Ulrich 2020).

In addition to the privatisation of public regulation, private companies and organisations have developed different models of self-regulation. One incentive driving these ventures is the global nature of the internet, which means a lack of standards and overlapping regulations. In fields related to human rights, especially freedom of speech, demands for private bodies to respect and even enforce human rights have grown. The obligation to actively protect rights and the horizontal effect of human rights are disputable, but treaties such as the United Nations Guiding Principles on Business and Human Rights (UNGP) have increased the pressure on companies to do something. The Oversight Board has been Meta's answer to several scandals and regulatory pressures. On the other hand, NGOs such as Article 19 have developed models for Social Media Councils, and there have been several attempts to bring democratic values and processes to the internet.

In the present paper, my hypothesis is that, compared with existing models of self-regulation and coregulation, Social Media Councils can learn from previous experiments but still face challenges that are unique to the environment in which these councils operate. I first demonstrate why there are several self-regulatory and coregulatory bodies in the field of media. My question is why there is a need for self-regulation, here in the form of Social Media Councils. My second question is how other self-regulatory bodies are formed and what legal basis they operate within. The third question relates to what kind of needs the Social Media Councils could address.

Why Self-Regulation?

Self-regulation is often caused by the fear of public regulation. For example, in April 1916, the Swedish newspaper Ny Dagligt published a private letter, causing a scandal in that country. The scandal led to calls for greater regulation of the press. As a result, the first modern press council, Pressens Opinionsnämnd (PON), was established in Sweden in 1916. However, PON only covered the press and suffered from a lack of funding in its early years (Weibull and Börjesson 1995). In Great Britain in 1949, the first Royal Commission on the Press recommended that a General Council of the Press should be formed to govern the behaviour of print media publishers (Frost 2000). The commission was founded amid public concern that a concentration of ownership was inhibiting free expression, leading to factual inaccuracies and allowing advertisers to influence editorial content. In Finland, the first guidelines for

journalists were drawn up in 1958 by Finnish media associations; this effort did not prevent the press from publishing scandalous news. The public blamed a few scandalous publications for causing the death of famous writers. Hence, the Council of Massa Media was established out of fear of government regulation in 1968, but it did not prevent the criminalisation of the dissemination of information violating personal privacy in 1973 (Neuvonen 2005).

Similar developments have been seen with the game industry. Violence in video games caused moral panic in the USA in the early 1990s. As a result, members of the combined United States Senate Committees on Governmental Affairs and the Judiciary held congressional hearings on the subject. The game industry was given options to organise self-regulation or be the object of public regulation. During these hearings, the industry developed a model of self-regulation, the Entertainment Software Rating Board (ESBR). In Brown v. Entertainment Merchants Association, 564 U.S. 768 (2011) the ruling was that law prohibiting the sale or rental of violent video games to minors violated the First Amendment. However, the court stated that the ESBR provides parents with sufficient information about the content of the games, so there is no need for hard regulation (Wuller 2013). The system is similar to other private rating systems in the USA: MPAA/MPA (Motion Picture Association), which is controlled by CARA (Classification and Rating Administration); the RIAA (Recording Industry Association of America) record rating system; and television companies' TV Parental Guidelines.

In Europe, long-running negotiations led to the establishment of Pan European Game Information (PEGI) (Marsden 2011). Before this, games had been regulated and even banned based on film regulations and criminal law in different countries. The EU Commission has repeatedly considered PEGI's activities to promote the goals of human rights, especially a safe environment for children. The classification system has been strongly involved in the recommendations of projects promoting safer internet and media literacy. In its resolution, the European Parliament has stated that, in addition to their disadvantages, games also have clear benefits and are part of the current digital operating environment. Parliament urges member states to continue and intensify cooperation for the development of the PEGI system. According to European Parliament, modes of self-regulation should be sought in the gaming industry, thus avoiding the need for EU-level legislation. The Council of Europe, which is behind the European human rights system, in cooperation with the Interactive Software Federation of Europe, has created the principles for considering human rights in online games. Therefore, it can be said that self-regulation has not only been approved by the EU, but also by the CoE.

Social media has caused several uproars and scandals in the late 2010s and 2020s. Consequently, demands for the regulation of social media have grown. Meta's (Facebook and Instagram) answer has been the Oversight Board, while other platform companies have taken different initiatives to make moderation more transparent and even democratic. The incentives for these initiatives are regulatory demands on both sides of the Atlantic. However, in the EU, the Digital Services Act is now nationally implemented, and it has elements of coregulation such as trusted flaggers and dispute settlement mechanisms.

These developments seem to be similar. The abolition of censorship and the rise of the press as an industry caused the need for press ethics and self-regulation as regulation instead of hard regulation. The significance of game industry growth during the 1990s, but games did not fit into the framework of electronic communication regulation. As a result, games were censored (Germany) or self-regulated (Netherlands), or the regulation authorities did not recognise them for a long time (UK, Finland).

In the early 1990s, the internet was seen as an area of freedom and a good business opportunity. The lack of effective regional or global norms and exemptions of liability, that is, "Good Samaritan" principles, led both sides of the Atlantic to enable the free growth of social media. Today, the empires

are striking back, and now, companies and NGOs are looking for self-regulation as a way to fix the biggest problems and avoid harsher regulations. It should be remembered that self-regulation is not always the best answer. In the US, the Production Code (i.e., Hays Code) for movies and Comic Code were self-regulatory systems but implemented harsher censorship than similar activities by public authorities in different countries.

Who Are the Regulators and Who Are the Subjects of Regulation?

It is necessary to be aware of who the regulators are; especially in self-regulation, there is (usually) no standards or authorisation. Self-regulatory bodies claim that they represent the field in question, but this demands critical evaluation. Another essential aspect is the scope of regulation. The Alliance of Independent Press Councils of Europe (AIPCE) has described press (media) council functions as monitoring the codes of ethics/practice and defending freedom of the press. The AIPCE is the most important union for press councils. The World Association of Press Councils (WAPC) describes itself as a defender of free speech, but none of the members of the AIPCE members are members of the WAPC; instead, members include, for example, Turkey, Kenya, Zimbabwe and other councils from countries not known for free media. In this paper, I focus on the AIPCE.

According to the comparative data on media councils collected by the AIPCE, media councils differ from each other. There is no clear blueprint on how to organise media councils and on what basis. The data from the AIPCE cover 32 media councils. First, four councils have been established by a decree, and two are recognised in law. The Danish Press Council is even a public entity, though it is an independent tribunal. Second, membership in press councils varies. Some have accepted individual journalists as members, and others have accepted only organisations, media outlets or both. Six councils also accept other members than those mentioned before. Third, print newspapers and magazines, as well as websites (of media outlets), are within scope in all of them, but 75% also cover television and radio. However, the mandate of most councils has covered all media only in recent decades.

Management of the PEGI system was handed to PEGI s.a., an independent, not-for-profit company with a social purpose established under Belgian law. The Netherlands Institute for the Classification of Audio-Visual Media and the VSC Rating Board (British) administrate the system on behalf of PEGI. In the PEGI council, 35 media authorities from the member states are represented. In some countries, for example, in Finland, PEGI ratings are accepted in law as legal ratings. Therefore, even though the PEGI is an independent company, it exercises public power and is endorsed and acknowledged by the EU and member states.

Meta set up an irrevocable trust, which is a legal entity in Anglo-American law but not so much in continental Europe. The trust created a limited liability company that owns, facilities and hires staff for the Oversight Board (OB). As Lorenzo Gradoni demonstrated, the OB itself is not a legal entity, and the members of the OB are in a contractual relationship with the background organisation (Gradoni 2022). As an interesting detail, the Oversight Board Trust Agreement allows for the accession of other platforms with the consent of Meta, as long as they are US persons and contribute to the funding of OB. The nature of SMCs is important because SMCs need to acquire facilities and potentially hire staff and pay for expert services to at least cover the expenses of the members. This is important because the (legal) form affects how practical matters are organised, such as accounting, transparency, employee rights and liability.

The proposals for SMCs include reviewing individual content moderation decisions made by social media platforms based on international standards on freedom of expression and other fundamental rights

but without creating legal obligations. This is a similar feature in most media councils. The SMCs should be established via a fully inclusive and transparent process that includes broad and inclusive representation, and SMCs must be transparent. There are many demands that SMCs should be inclusive and democratic, which increases the demands for both the nature of background organisation and for day-to-day management.

When compared with media councils, I have earlier demonstrated that one of the key issues is trust between stakeholders (Neuvonen 2022). It is necessary to define and commit stakeholders and to assure that SMCs benefit all parties. This requires that the initiatives for SMCs define stakeholders with care. Article 19 suggests that the SMC should be made up of representatives from social media companies, media, journalists, media regulators, press councils, the advertising industry, civil society organisations and academics. These groups have been included in the Irish SMC. They might leave open questions regarding inclusivity, third parties and the public. For example, in the Finnish Mass Media Council, eight members represent the media and eight the general public (some of them are members of academia). The Finnish Council is respected and operated for more than 50 years. This means that the SMCs must balance between the requirements of adding all stakeholders, inclusivity and democracy in both background organisation and management without compromising efficiency. It is not possible to fulfil all wishes.

The role of the media councils is to supervise media ethics. These ethical guidelines are connected to the ethics of journalism and journalism as a profession. Journalists and media outlets identify with the ideals of journalism and free speech. An individual journalist has the desire to follow the ethical code, and here, journalists are important stakeholders of self-regulation. Five councils are responsible for the distribution processes for press cards for journalists. According to many studies, moderators are outsourced employees and artificial intelligence is used a lot. There is no clear profession to identify with or professional ethos for moderators. The subject of OB is moderation as a whole and single processes and practices, not for a single moderator as such. An individual journalist is committed to their work, but there is a big gap between the moderator who made the decision and the complaint process. In the proposals for SMCs, moderators are never mentioned as stakeholders, which is understandable because moderation is outsourced and AI handles most of the job. Nevertheless, moderators are often ignored when social media practices are discussed. This also makes a difference compared with media councils.

Another issue is the scope. Most of the media councils cover all media (print, radio, television) but how to define social media. The scope of OB covers Facebook and Instagram, and PEGI requires recognition from states. The concept of platforms also covers food delivery companies, Uber and streaming services. Streaming services could also establish chats, for example. What is the status of chats and forums in games? To be effective, the SMC should cover most social media, but the field is very large, from small discussion forums to social media giants. It should also be noted that there might be competing councils. In the UK, the Leveson Report led to a system based on the 2013 Royal Charter on self-regulation of the press (hereinafter the Charter). The Charter created the Press Recognition Panel (PRP) to ensure that regulators of the press and other news publishers would be independent, properly funded and able to protect the public. The first regulator recognised by the PRP was the Independent Monitor for the Press (IMPRESS). However, none of the large national publishers were members of IMPRESS. Instead, most national publications were members of the Independent Press Standards Organisation (IPSO), which the PRB did not recognise. In addition, several prestigious newspapers (e.g., The Guardian, Financial Times) have established their own independent complaint systems.

What Can Added Value SMCs Create?

Social media has benefitted from the ethos of freedom found in the early days of the internet. However, China and Russia have developed mechanisms to monitor and block internet routes. Similarly, the biggest democracy, India, among other countries, uses shutdowns to control network activities. In Europe, the EU's digital package is changing the rules, and many hope that in a similar way to the GDPR, these regulations can create the Brussels Effect, that is, making regional rules de facto global rules. In the USA, Texas and Florida have introduced social media laws. These laws are already contested in courts, and their nature is very political.

Communications rights, digital rights, epistemic rights and digital constitutionalism, among other buzzwords, are attempts from civil society and academia to create principles and normativity on the internet. SMCs have similar backgrounds, but as self-regulatory bodies, SMCs are more than just frameworks. For example, the OB is a sui generis attempt to create some kind of higher body to handle the fundamental issues of single company platforms. The OB has been compared with the US Supreme Court, especially regarding the Marbury v. Madison case. The fact it can be compared with the Supreme Court is noteworthy because it is also necessary to note the language used in discussions about SMC.

Using legal metaphors and adding the digital into rights and constitutionalism is tempting. However, in the case of SMCs, this could mean moving from self-regulation to the domain of law. Even though the Danish Press Council is a public entity, it maintains a wall between law and self-regulation. In none of the media councils has the government been involved in the appeal process. Media councils are committed to self-regulation, which is seen as separate from the legal system. Indeed, the self-regulation of media is about ethics.

In addition, because of the AIPCE, only four media councils can sanction the media or journalist for an upheld complaint, and only two councils can order financial consequences for the breach of a journalistic principle. It is also necessary to note what grounds complaints can be made upon; for example, in Finland, complaints are quite open, and anyone can complain. In comparison, in Sweden, the grounds for complaints are more regulated, including the eligibility as a complainer.

The references to Marbury v. Madison can be seen as a part of digital constitutionalism. Digital constitutionalism can be divided into two sects: the internet's own (techno-utopian) legal system or the growing importance of existing constitutional systems on the internet. The proposals for SMCs highlight the importance of human rights, especially UN human rights. This means that SMCs can be seen as human rights interpreters. The OB refers to the UNGP and the International Covenant on Civil and Political Rights (ICCPR). Here, the UNGP requires companies to follow and respect human rights, especially regarding the ICCPR. However, this does not constitute a horizontal effect of human rights. Therefore, using legal metaphors and the language of law, SMCs are considered more quasi-courts than self-regulatory bodies. There are many nonlegal and private human rights actors, so what can new SMCs bring in? Or should SMCs have guidelines of their own?

There are several trends that have been occurring. First, the internet can be seen as a system of its own, and since Barlow's declaration, many have thought that the internet community itself could achieve sovereignty. Second, states and supranational entities are controlling the internet more intensively. Third, traditional media councils are increasingly engaged with social media, and media regulation is increasingly affecting media councils. Where is there room for SMCs? Should SMCs remain in the private world of self-regulation, or is coregulation a more suitable form?

Finally, there is the question of global versus local. Here, I will make a distinction that is more specific to global, regional (i.e., Europe), local (state) and hyperlocal (community). The PEGI is a regional

European operator and requires recognition from the member states. Media councils are national, and some councils are hyperlocal; for example, in Belgium, there are the Council for Journalism in Flanders and the Council for Ethical Journalism in Wallonia or the Information Council of Catalonia. The Swiss Press Council members represent different language regions. Looking at these examples, we need global, at least regional, rules for local and hyperlocal problems (Azzi 2021). According to Facebook Files, Facebook and Instagram have focused their moderation resources mostly on the US and English-speaking world. This can lead to problems. For example, The Guardian revealed during the COVID-19 pandemic that there has been a lot of misinformation and fake news in Spanish, which has the second most native speakers in the world. How many resources are there for Finnish, Latvian, Basque or Sami speakers is a mystery. How can SMCs strike a balance between global and regional impact and regional and linguistic representativeness? Without representativeness, there is no commitment of users.

Conclusion

In Ireland, SMCs have been interesting experiments both intellectually and in practice. I have compared the proposals and ideals for SMCs with existing self-regulatory and coregulatory bodies. My conclusions are based on the idea that Social Media Councils would be established. SMCs can be different, and there can be several of them.

1) The need for SMCs is similar to the pressure to establish media councils and organise the classification of games other than through hard regulation. Both platform companies and civil society are looking for solutions for content-monitoring issues in social media. However, at the same time, the EU and other legislative bodies are tightening the regulation of social media.

2) It is important to note the legal nature of the SMC or its background organisation. The Oversight Board does not exist in normal legal standards, but the company and trust operate its day-to-day management and funding. The OB itself operates in a void called global law. As a legal person, the SMC must be established in some state. This will affect how to make contracts, requirements for audits and the status of employees.

3) The SMC's credibility requires representativeness. The most effective media councils have representatives from different groups. All stakeholders must be engaged, and the public must be represented. However, how do we account for moderators?

4) It is necessary to clearly think of who can make a complaint and on what grounds.

5) Should we think that SMCs are courts? The media councils have drawn a clear line between self-regulation and law. The PEGI is a coregulatory body. So which path should SMCs follow? It is very tempting to use legal concepts and metaphors, but then, we are in the domain of law. Instead of legal concepts and human rights, a separate set of norms and concepts could be created for SMCs.

Sources

Abderrahmane Azzi: "Avoiding Imperialism: Merging the Global and the Local" in Handbook of Global Media Ethics. Edited by Stephen J. A. Ward, Springer, 2021.

Julia Black: "Constitutionalising self-regulation" (1996) 59 Modern Law Review 24.

Chris Frost: Media Ethics and Self-Regulation. London, 2000.

Lorenzo Gradoni: “Twitter Complaint Hotline Operator: Will Twitter Join Meta’s Oversight Board?” VerfBlog, 2022/11/10, https://verfassungsblog.de/musk-ob/, DOI: 10.17176/20221110-095537-0.

Christopher T. Marsden: Internet Co-Regulation: European Law, Regulatory Governance and Legitimacy in Cyberspace. Cambridge: Cambridge University Press.

Riku Neuvonen: “Internet and Self-Regulation: Media Councils as Models for Social Media Councils?” in: Matthias C. Kettemann (ed.), How Platforms Respond to Human Rights Conflicts Online. Best Practices in Weighing Rights and Obligations in Hybrid Online Orders. Hamburg: Verlag Hans-Bredow-Institut, 2022.

Riku Neuvonen: Sananvapaus, joukkoviestintä ja sääntely. Helsinki, 2005.

Teresa Quintel and Carsten Ullrich: “Self-Regulation of Fundamental Rights? The EU Code of Conduct on Hate Speech, Related Initiatives and Beyond.” in: Bilyana Petkova and Tuomas Ojanen (eds), Fundamental Rights Protection Online the Future Regulation of Intermediaries. Edward Elgar, 2020.

Lennart Weibull and Brit Börjesson: Publicitiska seder. Falun, 1995.

Lindsay Wuller: “Losing the Game.” (Winter 2013) 57 Saint Louis University Law Journal, no. 2 457–492.

Social Media Councils in the European Constitutional Framework

Giovanni Di Gregorio
CATÓLICA GLOBAL SCHOOL OF LAW, LISBON, PORTUGAL

Introduction

The introduction of Social Media Councils (SMCs) has captured the attention on a global scale. The attempt of Meta to set up an independent oversight has been one example of the steps towards institutionalising internal systems of review in the field of content moderation. SMCs are increasingly called upon to make decisions on the governance of online content, thus promising to provide answers to the limits of content moderation as a global system.

Still, the introduction of SMCs is not neutral from a constitutional point of view. First, Social Media Councils are boards conceived within the private sector, and this setting primarily influences the scope of constitutional law that traditionally follows a vertical logic in the relationship between freedoms and powers. Second, SMCs adopt and mirror constitutional narratives and procedures, particularly focused on human rights, to make decisions and provide recommendations. This process influences how constitutional values such as freedom of expression or access to justice are interpreted in the digital age not only by law-makers and courts but also by private actors. Third, the scope of SMCs is usually global. Such a perspective leads to a tendency to not consider the constitutional nuances of the local dimension. Despite the universal dimension of human rights, their protection and enforcement are connected to regional and national nuances.

This contribution aims to trigger a conversation about the constitutional role of SMCs. By adopting a European constitutional perspective, this contribution examines the nature of these boards, the use of constitutional narratives and the scope of their activities. The role of SMCs raises peculiar questions in Europe, as underlined by the adoption of the Digital Services Act as the European constitutional reaction to the challenges raised by content moderation.

Private Actors...

SMCs are conceived as private actors. Even if independent from social media, these boards make decisions that are not bound by the rule of law but primarily influenced by the "rule of platforms", such as terms of services, community guidelines and bylaws. Even if SMCs would contribute to increasing awareness of the challenges raised by content moderation while also fixing some pitfalls, they still exercise discretion in making their decisions on human rights.

These actors raise constitutional questions, particularly about how to frame the performance of quasi-public functions by private actors and their private governance on a global scale. Indeed, SMCs contribute to different roles, from adjudication to advisory, but primarily solve conflicts between human rights and private standards. Besides, similarly to constitutional courts, they also make recommendations on community standards in order to align their private standards with human rights. This framework underlines how, in this space, the rule of law has played a limited role so far. The private nature of SMCs narrows the scope of constitutional principles that become only an informal reference to the rule of platforms.

The consolidation of SMCs constitutes another example of the path towards the consolidation of the "rule of platforms". This process is outside the traditional scope of constitutional law. Rights and procedural safeguards directly apply only to public actors. In the lack of any regulation, it is not possible to require private actors to respect constitutional rights such as freedom of expression or, broadly, the rule of law. Constitutions are a critical part of the social contract that, for good reasons, limit governmental powers and ensure individual freedoms from interference by public authorities. Nonetheless, this mission has traditionally focused on limiting the authority of public actors rather than private powers.

At the same time, European constitutionalism is not based on a rigid vertical model. Both the European Convention on Human Rights and the European Charter of Fundamental Rights provide a constitutional system that can react to the challenges raised by private powers. The abuse of rights clause is an example of European constitutionalism that does not tolerate that the protection of economic freedoms turns into justifications to exercise (private) powers. This approach has led to extending constitutional safeguards into horizontal relationships, thus underlining how the European constitutional framework tends to react against unaccountable exercises of powers. Besides, the challenges for fundamental rights driven by a private system of adjudication can also trigger the positive obligation of States to protect human rights, thus leading to a new regulatory landscape for SMCs.

...Watching Over Human Rights

SMCs can play a critical role in fostering oversight, particularly in those areas where the exercise of public authority is absent or excessive, or even when States do not have the power to negotiate safeguards to increase transparency and accountability in content moderation. Nonetheless, these self-regulatory bodies contribute to the institutionalisation of procedural rules that raise questions about access and due process in these private spaces.

In this case, the primary constitutional issue is that human rights work as parameters to review social media decision-making and assess terms of services and community guidelines. This process entails that the subject of (constitutional) scrutiny is not primarily the rule of law but private standards. In this case, SMCs lead to the hybridisation of human rights enforcement and judicial review. Unlike the traditional rule where fundamental rights play the role of parameters to review legislation, in this case, private standards are de facto equivalent to the law, thus confirming the institutionalisation of the "rule of platforms" as the primary system of governance. In other words, these private rulebooks constitute a quasi-legal basis according to which platforms exercise their powers, without relying on public authorities.

Besides, even if SMCs' oversight is based on human rights law, it is not possible to exclude the influence resulting from private standards defined in community guidelines. This is a critical point about participation and representation in the digital environment. If private standards are developed outside a system of transparency and accountability but are only legitimised by users' adherence, the role of oversight, and review, loses its power, and its primary role, i.e. to ensure that constitutional values agreed by a community limit discretionary decisions. If private standards are still defined from the top, any system of oversight would only be an important but empty exercise. This issue also suggests why it is challenging to frame the rise of SMCs as an expression of societal constitutionalism, considering SMCs more as a transnational attempt to institutionalise procedures and rules on a global scale.

The consolidation of SMCs oversight contributes to the marginalisation of public actors in ensuring the enforcement of rights and freedoms online. This framework leads to looking at SMCs as additional

system of adjudication to protect fundamental rights in the digital age. These boards also provide perspectives on the hybridisation of the global approaches to online speech. This additional and alternative system of oversight and adjudication constitutes another call for European constitutionalism that, as already stressed, tends to be intolerant to decision-making processes affecting constitutional values outside the scope of public safeguards. Likewise, the model advanced by SMCs could also be a "positive" trigger for public actors to improve the enforcement and oversight of fundamental rights in the digital age.

Clashing Territorial Dimensions

The predominance of a global approach to SMCs also leads to another constitutional question. SMCs usually tend to operate on a global scale, and this approach is also the reasons why human rights law is particularly appealing for their decision-making. The Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, and regional systems such as the European Convention on Human Rights, are only some of the instruments that protect human rights, thus recognising their universal dimension.

Nonetheless, the protection of rights and freedoms is far from being equal geographically. Already in the European context, human rights are subject to different institutional and political dynamics shaping their protection. Even more importantly, human rights do not always overlap with the protection of fundamental rights as recognised by supranational systems and national constitutions. Without addressing the relationship between human rights and fundamental rights, it is critical to stress how the use of a universal approach by SMCs lead to reducing the importance of local constitutional nuances. Taking as example the EU and the US as two consolidated constitutional democracies, the protection of fundamental rights and freedoms do not always overlap and, in some cases, leads to opposite constitutional paths as underlined by the protection of free speech.

These constitutional nuances are underlined by the European path toward mitigating social media powers by proceduralising content moderation. The Digital Services Act is just an example of a path towards the injection of due process safeguards in the activities of online platforms taking decisions on content on a global scale. While the Union is at the forefront of a new constitutional phase addressing the challenges raised by the exercise of private powers in the digital age, the US has not shown the same concern, following an opposite path. For instance, the Communication Decency Act immunises online intermediaries, including modern online platforms, from liability when moderating users' content.

Even if SMCs take into account global community standards and human rights, they address local challenges as underlined by the cases addressed by the Meta Oversight Board. In this case, in order to represent constitutional nuances, the proliferation of regional SMCs can take into account the protection of fundamental rights into the local dimension. This approach would also lead to overcome the limited composition of these bodies that usually are not adequate to represent the nuances of rights and freedoms on a global scale. Besides, this change would also lead to increasing impact. The Meta Oversight Board is still a small move in the field considering the very few cases that this board will likely consider and also that the board provides an ex-post remedy which is not able to address the large amounts of cases raised by users.

Constitutional Perspectives

SMCs will not probably solve the primary challenges of content moderation such as content monetisation and profiling, or the biases of artificial intelligence moderating content. Nonetheless, these

bodies can play a critical role in improving the process of content moderation, particularly making the governance of online speech more transparent and accountable.

The question is how and to what extent SMCs will be part of the new strategy of the Union to proceduralise content moderation. The introduction of the Digital Services Act has defined a landmark step by providing procedural safeguards and redress mechanisms, and it is likely to play an important role also in relation to SMCs. For instance, the obligation for online platforms to take into account the protection of fundamental rights in their terms of services would also impact how SMCs review the compatibility of private standards with human rights. This rule will not only contribute to making the rule of law part of the discussion but also to increase the nuances of SMCs decision-making.

The European approach contributes to the expansion of SMCs. Even if social media are not required to establish SMCs, still these bodies can show accountability while assessing and mitigating risks such as the spread of harmful content. In this case, in Europe, SMCs are likely to become critical parts of social media architecture, that, from a constitutional perspective, is also governed by the rule of law.