

## Moderating the Regulators/Regulating the Moderators: NetzDG and online content moderation in Germany

Fichtner, Laura

Erstveröffentlichung / Primary Publication

Konferenzbeitrag / conference paper

*Diese Arbeit wurde durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert (Förderkennzeichen: 16DII121, 16DII122, 16DII123, 16DII124, 16DII125, 16DII126, 16DII127, 16DII128 – "Deutsches Internet-Institut"). / This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) (grant no.: 16DII121, 16DII122, 16DII123, 16DII124, 16DII125, 16DII126, 16DII127, 16DII128 – "Deutsches Internet-Institut").*

### Empfohlene Zitierung / Suggested Citation:

Fichtner, L. (2021). Moderating the Regulators/Regulating the Moderators: NetzDG and online content moderation in Germany. In *Proceedings of the Weizenbaum Conference 2021* (pp. 1-4). Berlin: Weizenbaum Institute for the Networked Society - The German Internet Institute. <https://doi.org/10.34669/wi.cp/3.5>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see: <https://creativecommons.org/licenses/by/4.0>

**Proceedings of the Weizenbaum Conference 2021**

**Democracy in Flux**

Order, Dynamics and Voices in Digital Public Spheres

**Moderating the Regulators/Regulating the Moderators**

NetzDG and online content moderation in Germany

**Fichtner, Laura**

Universität Hamburg

Hamburg, Germany

fichtner@informatik.uni-hamburg.de

**KEYWORDS**

Content Moderation; Public Sphere; Digital Democracy

DOI: 10.34669/wi.cp/3.5

The proceedings of the Weizenbaum Conference 2021 "Democracy in Flux: Order, Dynamics and Voices in Digital Public Spheres" have been funded by the Federal Ministry of Education and Research of Germany (BMBF) (grant no.: 16DII121, 16DII122, 16DII123, 16DII124, 16DII125, 16DII126, 16DII127, 16DII128 – "Deutsches Internet-Institut").

# 1 THE PROBLEM OF CONTENT MODERATION

Content moderation describes the rules by which platforms decide to allow or take down content and the practices that enforce these decisions. While it has always been a part of social media (Gillespie, 2018), content moderation has received increasing public attention as hate speech and hate campaigns, right-wing extremism, ‘election-hacking’, ‘fake news’ and disinformation have become heavily debated issues (see for example Tobin, Varner & Angwin, 2017; Marwick & Lewis, 2017; Van Zuylen-Wood, 2019; Satariano, 2019; Klonick, 2021; Edelman, 2021). Content moderation practices draw sometimes difficult distinctions between hate speech, political criticism and satire and freedom of press and disinformation. They determine what we can see and say online and who can participate on social media, in what ways, with what topics and on whose expense (Pasquale, 2020). In this way, platforms can set norms of legitimate and acceptable speech and decide which content is politically and publicly relevant and which is undesired, dangerous, harmful or detrimental to public discourse (Klonick, 2017). It is this public discourse, meant to guarantee citizens political participation, freedom of expression and access to information, that is often considered essential for liberal democracies (Calhoun, 1992). Thus, content moderation touches upon fundamental values of democracy and calls for politically legitimate ways of governance and implementation.

## 2 REGULATING CONTENT MODERATION IN GERMANY

In 2018, Germany introduced the Network Enforcement Act or *NetzDG* as a response to the challenges of content moderation. This law obliges social media platforms to delete “manifestly unlawful” content within 24 hours and decide on uncertain cases in 7 days (Bundesministerium der Justiz und für Verbraucherschutz, 2017, p. 3353). Systematic failure to comply can lead to fines of up to 50 million euros. This was possible, because the German legal code already included speech regulations regarding for instance insults, defamation and incitement to hatred and violence. *NetzDG* refers to a total of 21 legal paragraphs it seeks to enforce (Ibid.). I present and discuss my analysis of media articles reporting on and discussing the law over a period from its initial proposition until implementation (March 1<sup>st</sup> 2017 – August 15<sup>th</sup> 2018). In order to identify different framings of *NetzDG*, I coded a sample of articles from sources across the political spectrum such as *Die Tageszeitung*, *Süddeutsche Zeitung*, *Zeit Online*, *die Welt*, *Frankfurter Allgemeine Zeitung* and *netzpolitik.org*. My analysis gives insight into the public reasoning on content moderation, major points of contestation and ethical and political challenges of regulating social media platforms. While the discussion was tied to the national and legislative setting in which it took place, it posed fundamental questions about the governance of social media platforms and the social order that reigns online.

## 3 FRAMING NETZDG

The threat *NetzDG* might pose to the fundamental right to freedom of expression, and to related rights like freedom of information and freedom of press, presented one of the biggest criticisms levelled against it. This was fueled by a fear of overblocking: while *NetzDG* imposed fines for the failure to delete unlawful content, it did not prohibit platforms from deleting lawful content. This was seen to create a lopsided incentive structure that would motivate companies to take down more than necessary, resulting in a restriction on freedom of expression and undermining liberal democracy. In addition, *NetzDG* was characterized as a violation of the rule of law under which juridical entities should judge the legality of speech and law enforcement should implement these judgements. *NetzDG* was interpreted as an impermissible privatization of law enforcement that signified a failure of the state

and gave too much power to private companies already characterized by a lack of transparency. Others criticized NetzDG for giving too much power to the state to interfere with public discourse. Such interference was understood as detrimental to innovation and technological progress and to the internet's spirit of freedom, openness and plurality. Regulations like NetzDG were taken as unacceptable infringements on users' right to informational self-determination and autonomy. Moreover, such concentration of power within the hands of few big actors seemed to undermine the decentralized character of the internet which was believed to enable free exchange in a market-like structure.

Proponents of the law on the other hand saw NetzDG as a strong move by the state to uphold the law and the rule of law on the internet. NetzDG appeared as a necessary step to implement legal certainty online and get internet corporations 'under control'. Such state engagement was understood to transform the internet's lawlessness, chaos and anarchy into a space of democratic exchange within the rules set by the law. By implementing instances of regulatory control, it was also found to present a means of protecting social media users as consumers against overbearing companies and opaque and discriminatory algorithms as well as against malicious actors who exploited algorithms for their own gain. In this view, NetzDG was an attempt to hold companies accountable and force them to take up the social responsibility that came with their power and societal function. By enforcing existing legal regulations tailored to this purpose, the law was held to contribute to the creation of a public discursive space online. This space was envisioned as inclusive, constructive and safe, enabling everyone to participate by protecting them from harassment, intimidation, discrimination and silencing.

#### 4 CREATING DIGITAL DEMOCRACY

The controversy over NetzDG can be read as a broader controversy over the future of digital democracy. The discussion asked difficult questions about the limits of freedom of speech, but also prominently featured questions about the institutional arrangements and practices best suited to draw these lines and decide on and enforce deletion practices. While all perspectives claimed to support open and plural discourse as a way to enable and sustain liberal democracy, they had very different visions of how this discourse should look like. They centered different values such as freedom of expression, the rule of law, legal certainty, law and order, consumer protection and corporate accountability, and different aspects of these values. They evoked a diverse set of potential governing mechanisms meant to bring this discourse forth. These included civil liberties defending against overbearing state interference and enabling a 'marketplace of idea', the institution of 'the law' and law enforcement online, state engagement and consumer protection or, following early internet utopias, decentralization and personal autonomy. These stood in as abstract yet potent *sources of democratic legitimacy* for governing online content and speech and promised a politically impartial and procedurally sound basis for designing a public sphere that was both located within the confines of *and* produced democracy.

But the technologies themselves and the particular socio-technical conditions reigning on social media also complicated these 'old' ways of responding and gave rise to new questions about the design and governance of public discourse online. How are different frameworks to be applied to contemporary social media platforms? What are the roles and responsibilities of governmental institutions and state agencies, of corporate actors and platforms and of citizens and users online? How can and should they act on the internet and relate to each other? What kind of power are they supposed to carry? The different framings of NetzDG here give insight into a particular set of ways in which social media platforms *could* be governed, their ways of socio-technical ordering as well as their limits.

## 5 ACKNOWLEDGMENTS

As this abstract is part of my dissertation work, I would first of all like to thank my supervisor Prof. Judith Simon for her guidance and support and her and my entire group at Universität Hamburg for their invaluable feedback on my ongoing research. Secondly, I would like to thank Prof. Sheila Jasanoff for her constructive feedback and the Harvard-STS fellows of Spring 2020 for their comments on my project. Finally, I would like to thank two anonymous reviewers for their feedback on an earlier draft of this abstract.

## 6 REFERENCES

1. Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape Social Media*. New Haven, CT: Yale University Press.
2. Tobin, A., Varner, M., & Angwin, J. (2017). Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up. *ProPublica*. <https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes>
3. Marwick, A., & Lewis, R. (2017). *Media manipulation and Disinformation Online*. Data & Society Report. New York, NY.
4. Van Zuylen-Wood, S. (2019). „Men Are Scum“: Inside Facebook's War on Hate Speech. *Vanity Fair*. <https://www.vanityfair.com/news/2019/02/men-are-scum-inside-facebook-war-on-hate-speech>
5. Satariano, A. (2019). Facebook Can Be Forced to Delete Content Worldwide, E.U.'s Top Court Rules. *The New York Times*. <https://www.nytimes.com/2019/10/03/technology/facebook-europe.html>
6. Klonick, K. (2021). Inside the Making of Facebook's Supreme Court. *The New Yorker*. <https://www.newyorker.com/tech/annals-of-technology/inside-the-making-of-facebooks-supreme-court>
7. Edelman, G. (2021). On Social Media, American-Style Free Speech is Dead. *Wired*. <https://www.wired.com/story/on-social-media-american-style-free-speech-is-dead/>
8. Pasquale, F. (2020). 4. The Alien Intelligence of Automated Media. In *New Laws of Robotics* (pp. 89-118). Cambridge, MA: Harvard University Press.
9. Klonick, K. (2017). The New Governors: The People, Rules, and Processes Governing Online Speech. *Harvard Law Review*, 131, pp. 1598-1670.
10. Calhoun, C. J. (Ed.). (1992). *Habermas and the Public Sphere*. Cambridge, MA: MIT Press.
11. Bundesministerium der Justiz und für Verbraucherschutz. (2017). Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (Netzwerkdurchsetzungsgesetz–NetzDG). *Bundesgesetzblatt Jahrgang 2017, Teil I, Nr. 61*, pp. 3352-3355. [https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/BGBl\\_NetzDG.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/BGBl_NetzDG.pdf?__blob=publicationFile&v=2)