

WEIZENBAUM JOURNAL OF THE DIGITAL SOCIETY  
Volume 2 \ Issue 2 \ w2.2.2 \ 12-31-2022  
ISSN 2748-5625 \ DOI 10.34669/WI.WJDS/2.2.2

Information on this journal and its funding can be found on its website:  
<https://wjds.weizenbaum-institut.de>

This work is available open access and is licensed under Creative Commons Attribution 4.0 (CC BY 4.0):  
<https://creativecommons.org/licenses/by/4.0/>

---

KEYWORDS

content moderation  
internet governance  
digital democracy  
platform regulation  
freedom of expression

FULL LENGTH ARTICLE

# Content Moderation and the Quest for Democratic Legitimacy

Laura Fichtner

University of Hamburg  
[laura.fichtner@uni-hamburg.de](mailto:laura.fichtner@uni-hamburg.de)

---

ABSTRACT

The paper analyzes the public controversy incited by the introduction of the Network Enforcement Act (NetzDG) in Germany. This law obliges social media platforms to delete unlawful content from their sites and has received international attention as a regulatory blueprint for governing corporate content moderation. The paper describes different ways in which NetzDG was framed in German media reporting, which offered distinct assessments of whether the new law endangered or supported democratic principles and values. Major differences in the public controversy over NetzDG revolved around, for instance, what freedom of expression and the rule of law meant for content moderation and how NetzDG's regulatory intervention would interact with platforms. The paper finds that a major point of contention thus concerned how to ground content moderation practices and policies in democratic legitimacy. Its analysis demonstrates that the governance of content moderation on social media platforms can open up a site for renegotiating democratic values and principles. As the NetzDG case shows, this can happen without substantively challenging existing laws but by raising the question of how to legitimately apply them to platforms. At stake in this controversy were the underlying logics by which to govern speech online. Different perspectives on this built on distinct understandings of democracy, attributing particular roles and responsibilities to platforms, state institutions, and users. Thus, the paper illuminates that the

public controversy over NetzDG, and over the right way to uphold speech laws on platforms, concerned more fundamental questions about the shape of democracy and the distribution of power, agency, and responsibility.

## 1 Introduction

In *Custodians of the Internet*, Tarleton Gillespie has suggested that studying content moderation provides a “prism for understanding what platforms *are*, and the ways they subtly torque public life” (Gillespie, 2018, p. 13). Similarly, one might say that studying governments’ regulatory interventions provides a prism for understanding what states are and how their institutions and actions shape public life. Studying the *regulation of content moderation* combines both perspectives. This paper analyzes the controversy around one such regulatory approach: the Network Enforcement Act (NetzDG) in Germany, which was first proposed in March 2017 and came into full effect in January 2018.

This analysis shows how the seemingly straightforward exercise of applying existing law to social media platforms prompted more deep-seated questions about what democracy means online and about the roles played by state institutions, corporate platforms, and internet users. It investigates how social media platforms (as technological infrastructures) and ideas about democracy are *co-produced* (Jasanoff, 2004), and how social media platforms (as actors on the political stage) and content moderation (as a practice of public discourse) interact with democratic ideals. This analysis emphasizes “dimensions of meaning, discourse and textuality” and looks at how scientific and technological developments are incorporated into – or transform – social structures, meanings, and contexts (Jasanoff, 2004, p. 4). In this paper, I investigate media reporting on NetzDG as a source of insight into how this new law’s regulatory interventions in the area of content moderation were variously interpreted in the ensuing public controversy. This forms the basis for a critical reflection on NetzDG as a site for (re)negotiating the shape of digital democracy and searching for the right ways to ground content moderation in principles of democratic legitimacy.

In the following, I first introduce NetzDG and explain the study’s methodology. I then describe different ways in which NetzDG’s intervention to content moderation was framed; these evaluated NetzDG’s impact on democratic values and principles and put different democratic ideals to work in the context of social media. Thus, these framings describe how different conclusions were reached as to whether NetzDG’s intervention protected or threatened democratic principles and values. I therefore argue that the controversy represented a public dispute over how to ground content moderation in principles of *democratic legitimacy*. In the last part of the paper, I explicate what is meant by this

observation and what we can learn from it about the governance of content moderation on social media platforms. The paper examines the assumptions and consequences of analyzing content moderation as a quest for democratic legitimacy and closes by outlining the questions and aspects this perspective both illuminates and obscures.

## 2 Content moderation and its regulation in Germany

In March 2017, SPD<sup>1</sup> politician and then-Minister of Justice and Consumer Protection Heiko Maas proposed a *Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*, a much-debated piece of internet legislation known as the Network Enforcement Act or *NetzDG* (short for *Netzwerkdurchsetzungsgesetz*). Despite much criticism, the act has been in effect since 2018 and applies to social media platforms with more than 2 million registered users in Germany. When introducing NetzDG, the government at the time described it as a compliance regulation for social media platforms that would establish German law as the standard for content deletion and address an increasingly hateful, aggressive, and hurtful culture of debate online (CDU/CSU; SPD, 2017, p. 1). NetzDG was presented as a necessary effort to fight online hate crimes that posed a danger to a “free, open and democratic society” (CDU/CSU; SPD, 2017, p. 1). It followed what the Ministry of Justice, led by Maas, deemed the failure of a task force it had set up in collaboration with social media platforms to encourage platforms to react more strongly to hate speech and other illegal content.

Since 2018, NetzDG has required social media platforms to delete “manifestly unlawful content” within 24 hours and to reach a decision on unclear cases within 7 days (*Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)*, 2017). Furthermore, the act requires that platforms establish a point of contact for users and law enforcement inside Germany; law enforcement inquiries via such points of contact must receive responses within 48 hours (*ibid.*, p. 5). Platforms must also issue biannual reports on NetzDG’s implementation, including information on the procedures in place and statistics on reported and deleted content (*ibid.*, p. 1ff.). In addition, NetzDG enables platforms to “disclose information about subscriber data” to claimants in civil lawsuits, provided a court order has been obtained (*ibid.*, p. 6). The version which was finally adopted in 2017 also includes options for self-regulation: Platforms can transfer decisions on ambiguous content to a governmentally recognized institution of self-regulation. Such institutions are

---

<sup>1</sup> The SPD (Social Democratic Party of Germany) is one of the major parties and is located on the center-left of the political spectrum. In the latest election (2021), the party garnered 25.7% of the votes and became part of the governing coalition (Der Bundeswahlleiter, 2022). It is also the party of current German Chancellor Olaf Scholz.

to be staffed with independent experts but financed by companies and approved by the Ministry of Justice (*ibid.*, p. 3).<sup>2</sup>

As a unique regulatory approach to tackling the problems of *online content moderation*, the introduction of NetzDG has received international attention as well as critiques from academia and beyond (Bernau, 2018; Douek, 2022, p. 48; Hülsen & Müller, 2018; Roberts, 2019, p. 213). *Content moderation* refers to how internet platforms decide what is and what is not allowed on their site, as well as how they take down content or regulate its visibility and appearance. This can sometimes raise difficult questions about how to distinguish between hateful comments and satire, between hate speech and political criticism, and between legitimate exercise of press freedom and illegitimate disinformation. Kate Klonick has described platforms' moderation systems as systems of governance that influence users' right to expression and political participation and shape democratic culture (Klonick, 2018). Historically, US legal norms, and especially §230 of the US Communications Decency Act, have influenced such systems (*ibid.*). They exempt platforms from liability if they do not engage in an editorial process, giving them considerable discretion in their moderation. And although platforms are free to delete content as private corporations, in the past, legal and cultural views on free speech prevalent in the United States have often shaped platforms' moderation practices. As per the First Amendment to the US Constitution, freedom of speech in this context is often viewed as an absolutely protected right that requires acceptance of nearly every kind of speech that does not cause direct harm. However, with increasing public attention being devoted to the circulation and effects of hate speech and misinformation, platforms have in recent years started to more explicitly and proactively moderate content.

NetzDG intervened in this space and sought to change the way platforms made decisions on reported content, forcing them to uphold German law. This was made possible by Germany's existing legal framework, which differs from US jurisprudence. §5 of Germany's constitution, the Basic Law (or *Grundgesetz*), also guarantees freedom of expression, including freedom of speech, information, arts, and sciences, and the absence of censorship. However, it subjects these rights to the boundaries of general laws as well as to the protection of minors and personal honor (Bundesministerium der Justiz und für Verbraucherschutz, 1949, Article 5(2)). There are several laws that may reasonably limit what a person can say or what content they can circulate. NetzDG specifically lists the laws it seeks to enforce: Articles 86, 86a, 89a, 91, 100a, 111, 126, 129 to 129b, 130, 131, 140, 166, 184b, 185 to 187, 201a,

---

<sup>2</sup> Since my analysis was conducted, NetzDG has been amended in several ways. As one of the most important changes, since 2020, platforms must report serious offenses to the Federal Criminal Police, including the IP addresses and port numbers of offenders ("Gesetz Zur Änderung Des Netzwerkdurchsetzungsgesetzes," 2021). Moreover, the disclosure of subscriber data for civil law claims under court order has been made mandatory, while previously being voluntary.

241 and 269 of the German Criminal Code<sup>3</sup> (*Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act)*, 2017, p. 1). Under NetzDG, platforms are obliged to delete reported posts that violate these laws.

### 3 Analyzing the public debate on NetzDG

My analysis of content moderation as a quest for democratic legitimacy is based on a study of the German media reporting that surrounded NetzDG.<sup>4</sup> This study builds on a framing studies approach, which enables me to investigate how the same policy issues can be made sense of in multiple ways through processes of selection, emphasis, and ordering. Entman describes four different elements used in such framing processes: “a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation” (Entman, 1993, p. 52). Although frames so constructed can also be analyzed as static entities strategically employed by political actors, I build my analytical approach on van Hulst and Yanow’s conceptualization of framing as a dynamic, interactional process of *sense-making* and *meaning-making* (van Hulst & Yanow, 2016). I understand the different frame elements as being constituted in such processes, which happen “in a conversation with the situation,” organize already-held knowledge and values, and give guidance for future actions (van Hulst & Yanow, 2016, p. 98). To highlight this dynamic process, van Hulst and Yanow talk about framing rather than about frames; in my paper I use the term *framing* (or ways of framing) as shorthand for such interpretive processes of sense-making. Therefore, the different framings presented in this paper describe different ways of making sense of NetzDG, its impact, and its desirability.

For my study, I analyzed a sample of articles from mainstream media sources across the political spectrum and from a popular technology politics blog.<sup>5</sup> These articles covered a time span from the initial proposal of NetzDG until approximately 8 months after its implementation (March 1, 2017–August 15, 2018). The sample was compiled by taking every third article reporting on the law; these were found through the Nexis and Factiva databases using different terms for NetzDG.<sup>6</sup> The articles were qualitatively analyzed and coded with a scheme adapted from the elements Entman described. The codes categorized

<sup>3</sup> This broad list includes laws against the use of unconstitutional symbols; laws criminalizing incitement to commit serious violent offenses against the state; laws criminalizing reward or approval of criminal offenses; laws against insults and defamation; laws criminalizing defamation of religious or ideological associations; and laws criminalizing incitement to hatred against individuals and groups based on their racial, religious, ethnic, or national identity.

<sup>4</sup> This study is part of an ongoing dissertation project on the politics of platform governance.

<sup>5</sup> Die Tageszeitung (17 articles), Süddeutsche Zeitung (39 articles), Zeit Online (22 articles), Welt Online (22 articles), Frankfurter Allgemeine Zeitung (faz.net) (28 articles), and netzpolitik.org (49 articles).

<sup>6</sup> “Netzwerkdurchsetzungsgesetz,” “NetzDG,” “Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken,” and “Facebook-Gesetz.”

different types of problems, the principles and values that were evoked, as well as contextual factors, cases, and examples used to characterize and illustrate them. Positive effects of NetzDG or suggestions of alternative solutions were also coded. I then analyzed how the different aspects were related in framing the problem at hand and how to best react to it.

I found a variety of often overlapping and intersecting issues that were framed in the NetzDG controversy. They included problems with content on social media platforms, with platforms' corporate practices, with categorizing these practices, as well as with internet governance and platform regulation in general. Not all of these aspects can be presented here. Instead, I focus specifically on different ways in which media reporting framed NetzDG's impact on democratic values and principles, which also expressed particular attitudes toward the desirability and acceptability of regulatory interventions in content moderation. These framings centered on different concerns, drew upon different interpretations of democratic values and principles, and prompted different conclusions on NetzDG. The observation of these differences leads to the insight that the framings described competing ideas of what the democratically legitimate governance of content moderation looks like and of how public, democratic discourse can be safeguarded on social media.

Accordingly, the framings presented in this paper are differentiated by three criteria: 1) the central values and principles they used for their evaluation of the problem at hand, 2) the concern they centered on in their evaluation, and 3) their attitude towards NetzDG and the regulation of content moderation. The framings could relate to one another in two different ways: Firstly, they could draw from the same democratic value or principle but also differ in how they interpreted its meaning and application to social media platforms or in the concern on which they centered. Such divergences illustrate the conflicts that played out in the NetzDG controversy over how to interpret democratic values and principles in the context of platform governance and the regulation of content moderation. Secondly, the framings could share similar assumptions about or attitudes toward regulatory interventions to content moderation but also differ in the values they used to justify this attitude. This illustrates different legitimation strategies that could be employed for moderation policies and regulations, which centered on different values and concerns but nevertheless supported one another's conclusions.

The table below schematizes different framings that emerged from the analysis, describing their central values and concerns. The framings are grouped by four more general democratic principles. In my description of the framings which follows in the next section, I also include examples used for illustration and explicate further concerns that were derived from central concerns and values as well as alternative solutions suggested.

Table 1: Classification of framings described in the paper

<b>Title of framing</b>	<b>Central value(s)</b>	<b>Central concern(s)</b>	<b>Attitude toward NetzDG</b>
<i>NetzDG's impact of freedom of speech and democratic discourse</i>			
A Threat to Freedom of Speech and Civil Liberties	Freedom of expression	NetzDG incentivizes overblocking and endangers free expression	Critical: NetzDG's interaction with platforms' financial incentives brings about threats to freedom of expression
State Engagement to Defend Free Expression	Equal right to freedom of speech	Undemocratic voices on platforms threaten others' freedom of speech and political participation	In support: NetzDG protects everyone's right to freedom of speech and counters hate
Creating a Public Discursive Space Online	Public discourse	Regulation must do justice to the public character of social media platforms	In support: NetzDG creates democratically accountable procedures for setting norms on social media
<i>NetzDG's impact on the rule of law</i>			
Enforcing Law and Order and the Rule of Law Online	Law and order + law enforcement	Lawlessness on the internet and arbitrariness on platforms in the absence of regulation	In support: NetzDG ensures that platforms adhere to laws and respond to illegal activities
A Threat to the Rule of Law	Rule of law + due judicial process	Privatization and outsourcing of legal decision-making	Critical: NetzDG transfers the state's judicial process to private corporations
<i>NetzDG's impact on user empowerment and openness on the web</i>			
Reinforcing Platforms' Opacity	Transparency	Platform opacity disempowers users and prevents accountability	Critical: NetzDG reinforces platforms' lack of transparency and accountability
Governmental Overinterference and Overregulation on the Internet	Openness + innovation	Governmental overregulation compromises the internet's openness, democratic nature, and innovative potential	Critical: NetzDG exemplifies a trend of overregulation and overinterference
Protecting Users from Economic Exploitation	User safety + protection	Platforms' business models can compromise user rights	In support: NetzDG counters negative effects of business models on users
<i>NetzDG's adherence to democratic standards of legislation</i>			
A "Bad" or Even Illegitimate Law	Due process + legitimacy of political intentions	NetzDG as a "bad" law, due to legislative process, legal text, or political intentions	Critical: NetzDG's legislative process did not adhere to democratic standards

The media reporting on NetzDG, which was investigated for this study, illuminates how platform regulation and content moderation were made sense of in the context of a particular national law and legal and cultural background. The insights from the material are limited because the articles may not encompass all viewpoints or equally consider all voices and ways of thinking about the issue. Nevertheless, the analysis draws on a broad sample of voices from across the political spectrum, thereby providing a picture of the prominent perspectives discussed. This paper concentrates on the *content* of different framings and not on their strategic use; further exploring how and with what effects different political actors employed and shaped these framings provides an avenue for future research. The framings that are presented in this paper emerged as overarching clusters from the analysis of all articles, which drew from them at many points. Speakers and articles are nevertheless sometimes cited for illustration. In general, a diverse range of critics were cited as having spoken out against NetzDG who focused on potential problems arising from its stipulations and implementation. It was particularly the minister and party that had proposed NetzDG who were cited in framing NetzDG most strongly in terms of its positive effects. Articles that reiterated such positive framings often described these effects as positive *intentions* of NetzDG. They highlighted a need for regulatory intervention to content moderation based on what they saw as pressing problems with content online and with social media platforms.

## 4 The controversy around NetzDG

### 4.1 A threat to freedom of speech and civil liberties

One of the most common ways in which NetzDG was negatively framed positioned the new law as a potential threat to *freedom of speech*. This was discussed by many articles in various sources and brought together the concerns of different actors, such as opposition parties, civil rights activists, legal and technological experts, social media companies, industry associations, journalists, and academics. Central to this framing was a problematization of NetzDG in terms of its negative impact on the right to free expression. This was attributed to NetzDG's incentive to *overblock*, which would incentivize platforms to delete more content than necessary and censor permissible speech. NetzDG's lopsided incentive structure was cited as the reason for this: Whereas fines of up to 50 million euros could be issued for systematic failures to delete, there appeared to be no fines for wrongful deletions and no obligation to retain permissible content.<sup>7</sup> Because judging the legality of speech was

---

<sup>7</sup> The concerns were also aggravated by the circumstance that NetzDG initially included no right to re-upload falsely deleted posts, which implied that appeals could only be filed in a court of law.



understood as a difficult and ambiguous endeavor, even for experts, the position which this framing described implied that freedom of speech called for leaving posts up when their permissibility was merely in doubt.

Such evaluations characterized NetzDG as an *unconstitutional* law threatening to infringe on people's ability to freely express themselves and exchange their opinions online. Its regulatory intervention in content moderation was problematized as an undue and dangerous form of interference with free expression on the internet, which could also compromise freedom of the press, information, and arts. Consequently, this framing articulated concerns over a potential threat to civil liberties and even over NetzDG's lack of conformity with the European Charter of Human Rights. In this view, the protection of such freedoms was important because they guaranteed liberal democracy; safeguarded democratic discourse; and enabled political participation, societal critique, and control of powerful (political) actors. One *Welt* article<sup>8</sup> therefore suggested that the infringements on them, which NetzDG potentially incentivized, represented harbingers of undesirable and undemocratic political developments.

Taken-down satirical pieces stood in as symbolic examples of ambiguous and politically relevant but hard-to-judge speech that could fall prey to overblocking. They demonstrated how overblocking might compromise the potentially important and critical function of such speech in public discourse. One popular example was the blocking of the Twitter account of the satire magazine *Titanic* after its satirical impersonation of AfD<sup>9</sup> politician Beatrix von Storch.<sup>10</sup> Other examples of speech that was at risk of overblocking included testimonies of hate crimes and critical presentations of hate speech. Such examples were invoked to show that NetzDG could not do justice to the contextuality of speech and could lead to self-censorship and impermissible restrictions on public discourse. Sharing such concerns, the so-called *Declaration on Freedom of Expression*,<sup>11</sup> which was cited in its entirety in *Netzpolitik*<sup>12</sup> and signed by academic and legal experts as well as associations from civil society, media, journalism, and the internet industry, feared that NetzDG could compromise free expression and a "*plurality of views*."

<sup>8</sup> "Tag der Pressefreiheit; Die Freiheit muss jeden Tag neu erschrieben werden," *Welt Online*, May 5<sup>th</sup>, 2018.

<sup>9</sup> The AfD (the Alternative for Germany) is the leading right-wing populist party in Germany, founded only in 2013. In the latest election (2021), the party garnered 10.3% of the votes (Der Bundeswahlleiter, 2022).

<sup>10</sup> After issuing a tweet that condemned groups of Muslim men, the politician's own account had previously been temporarily blocked for a potential violation of rules against hateful content; *Titanic*'s satirical impersonation had been a critical reaction to this tweet. Story cited for example in "Titanic bleibt gesperrt: Tweets der Wahrheit," *Frankfurter Allgemeine Zeitung (faz.net)*, January 4<sup>th</sup>, 2018; "Medien: Twitter löscht Satire-Tweet der "Titanic"," *Süddeutsche Zeitung*, January 3<sup>rd</sup>, 2018.

<sup>11</sup> <https://deklaration-fuer-meinungsfreiheit.de/en/>

<sup>12</sup> "Breites Bündnis stellt sich mit Deklaration für die Meinungsfreiheit gegen Hate-Speech-Gesetz," *Netzpolitik.org*, April 11<sup>th</sup>, 2017.

In relation to this, *counter-speech*, which opposes problematic speech with counter-arguments, critical reflection, or contrary evidence – and which requires a lot of civil initiative – was discussed as a potentially better approach to act against hate speech that would ensure important political speech stayed up and that neither threatened free expression nor motivated overblocking. The platform companies themselves embraced self-regulation, which would allow them to coordinate their own responses to issues such as hate speech and fake news, possibly under governmental oversight.

Taking a defensive approach to freedom of speech, this framing represented NetzDG’s regulatory intervention in content moderation on social media as bringing more risks than benefits. A central concern it focused on was that platforms’ attempts to comply with NetzDG, in combination with their business operations, could result in limits on free speech. Part of this issue was that NetzDG did not provide an obligation or incentive to protect legal content, which could also open up the question of whether platforms should have to actively uphold a right to free expression.

## 4.2 State engagement to defend free expression

In opposition to the view described above stood a framing in which NetzDG appeared as *defending* freedom of speech. This framing was characterized by a view of freedom of speech as a collective democratic right that required some limitations to ensure that everyone could exercise it equally and participate in public discourse. This sometimes meant that, to *defend democratic rights*, individual liberties had to be restricted where hate speech and illegal content infringed on others’ right to free expression or political participation. Maas, for instance, quoted in *Die Zeit*<sup>13</sup> and *FAZ*,<sup>14</sup> emphasized that freedom of speech could not be used as a “carte blanche” for committing crimes and that it was not NetzDG, but the hate online – which it sought to counter—that was the “true enemy of freedom of speech.”<sup>15</sup>

In contrast to the previous one, this framing presented NetzDG not as a threat to democratic expression and discourse, but as an important form of state engagement to *promote* free expression and democracy by countering undemocratic voices and forms of communication, as well as by providing the right conditions to facilitate safe and open public discourse. Furthermore, this framing viewed NetzDG not as infringing on constitutional rights, but as *defending* them by applying existing democratic laws. Centering on problems with hate speech, misinformation and undemocratic voices online, this framing

---

<sup>13</sup> “Netzwerkdurchsetzungsgesetz: Was Sie über das NetzDG wissen müssen,” *Zeit Online*, January 4<sup>th</sup>, 2018.

<sup>14</sup> “Titanic bleibt gesperrt: Tweets der Wahrheit,” *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), January 4<sup>th</sup>, 2018.

<sup>15</sup> “Justizminister Maas: ‘Im Netz wird viel zu wenig gelöscht,’” *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), May 19<sup>th</sup>, 2017.

described NetzDG's regulatory interventions to platforms' moderation practices in positive terms and as being necessary to protect democratic discourse as well as citizens' liberties and political participation. In defense of NetzDG, Maas also pushed back on concerns over overblocking, stating that platforms had financial incentives not to delete more than necessary.<sup>16</sup>

### 4.3 Creating a public discursive space online

This view that regulatory intervention in content online is necessary and important, which was just described in Section 4.2, was also shared by a framing that described NetzDG as doing justice to the *public* character of social media platforms and as working toward a public discursive space online. This framing did however not so much focus on requirements for freedom of speech, but instead centered on the argument that platforms' public character necessitated publicly and democratically accountable procedures to set norms and rules. This positioned social media as a *public sphere* that ought not be subject to the whims of corporations. In this framing, NetzDG was presented as an effort to take democratic control over this space. This again raised a question as to whether platforms have an obligation to *uphold* the right to freedom of speech, with one SPD politician even suggesting that platforms should be obligated to display a plurality of sources and content.<sup>17</sup>

According to this framing, regulatory intervention could cancel out negative influences from business models and create democratic legitimacy for content moderation. A *Zeit* article<sup>18</sup> further described the discussion around NetzDG as being valuable because it shed light on horrible content and invited a broader public debate on how to establish the norms for democratically permissible speech. Echoing this view, one *FAZ* article<sup>19</sup> referred to the belief that issues with problematic content on social media would resolve themselves without governmental interference as an "illusion." In this view, content moderation required clear institutional frameworks and politically neutral rules that protected minorities, safeguarded plurality, and recognized not only freedom of speech but also the inviolability of human dignity. The creation of new institutions and special oversight authorities that controlled platforms and enforced transparency could also provide new paths forward.

---

<sup>16</sup> "#Netzwerkdurchsetzungsgesetz: Was Sie über das Gesetz gegen Hass im Internet wissen müssen," *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), June 30<sup>th</sup>, 2017.

<sup>17</sup> "Freiheit im Internet: Facebook löscht Meinungen nach eigenen Regeln," *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), June 27<sup>th</sup>, 2018.

<sup>18</sup> "Der Storch-Effekt," *Zeit Online*, January 9<sup>th</sup>, 2018.

<sup>19</sup> "Netzwerkdurchsetzungsgesetz: Freiheit für Heiko Maas," *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), January 1<sup>st</sup>, 2018.

## 4.4 Enforcing law and order and the rule of law online

Existing law was identified as a source for such democratically legitimate and politically neutral rules as referred to above. This view was promoted by a framing of NetzDG *as enforcing law and order and the rule of law on the internet*. This framing described the fact that NetzDG made criminal law – and not the power of social media platforms – the source for deletion decisions in positive terms. The main concern addressed by this framing was that, in the absence of regulatory intervention on social media, criminal hate would run free and lawlessness would govern. This presented a push-back on critiques of NetzDG as it emphasized that the new law did not introduce new rules but simply applied existing rules – democratically legitimated and widely accepted – more consistently. This framing was often cited in support of NetzDG and explicitly promoted by Heiko Maas, the minister who had drafted NetzDG. Maas suggested that the new legislation would put an end to a “verbal club-law”<sup>20</sup> that prevailed online. It was here characterized as having the power to ensure that companies acted to counter hate speech in a timely and efficient manner. It appeared that by upholding existing laws on the internet, it provided tools for tackling problems on social media.

In this framing, the internet was depicted as a chaotic and lawless space where hate crimes and fake news flourished and where algorithms, echo chambers, and filter bubbles reigned. Its liberating potential appeared to have been poisoned by toxic communication, and platform companies’ business models were incentivizing them to push conspiracies, sensationalist stories, and provocative content; hence, governmental institutions were called upon to step up. Thus, this framing supported the view that regulatory action was needed to bring it into the territory of the state because it identified the internet as an anarchic place deteriorating under the influence of algorithmic exploitation and the whim of private corporations.

Therefore, it described NetzDG as countering lawlessness and “cleaning up” the internet, expressing the hope that this would enforce legal and political primacy over private corporations. NetzDG was consequently understood as a law that defended the pluralism, diversity, and protection of minorities that democratic societies require: It stood in the interest of citizens because it upheld democratic structures and lawfulness. By inviting the view that laws could be directly applied on platforms as much as anywhere else, this framing did not create much space for thinking about content moderation’s *specific* challenges for traditional law enforcement.

---

<sup>20</sup> “Soziale Netzwerke: Das Löschen beginnt,” *Süddeutsche Zeitung*, June 30<sup>th</sup>, 2017.

## 4.5 A threat to the rule of law

In contrast, however, NetzDG was also frequently framed as a threat to the rule of law. In this framing, NetzDG appeared as *endangering the rule of law* because it delegated legal decisions to private corporations, and thus as effectively outsourcing and *privatizing* law enforcement. Therefore, it seemed that NetzDG created what one journalistic association called a “private media police”<sup>21</sup> and one *Süddeutsche Zeitung* article characterized as a “Facebook self-court.”<sup>22</sup> Comparisons to the analog world were used to illustrate this framing, with one *TAZ* article<sup>23</sup> suggesting that we should not let companies decide on the limits of speech online, just as we would not allow private security firms to enforce law and order on the streets.

Central to this perspective were the modalities by which decisions over the legality of posts were made, which were seen as the task of the state’s judicial apparatus and its courts. Thus, according to this framing, the main problem with NetzDG was that it forced private corporations to make quasi-legal decisions, while seeking to generate as much revenue as possible, and without having to uphold fundamental elements of jurisprudence. This framing therefore did not (directly) address the question of how the benefits of freedom of speech on social media could be maintained or balanced with the negative effects of hate speech but instead reframed the issue as a question about what the rule of law meant online.

It articulated a worry that, under NetzDG, laypeople – as opposed to legal experts and judges – would decide on the legality of speech and that private platforms would determine the boundaries of constitutional rights, while perpetrators of criminal acts were not prosecuted. It found such privatization to be in violation of the rule of law and to signify the state’s failure to do its duty. The state was seen as renegeing on its responsibilities to uphold democratic forms of jurisprudence and as delegating the capacity to set norms and exercise coercion to private corporations. This problematized NetzDG as giving platforms too much power over the right to freedom of speech and the rules of public discourse, and as unduly endowing them with legal and governmental responsibilities. Facebook itself echoed this criticism when it emphasized that it did not want to be an “arbiter of truth.”<sup>24</sup> This framing suggested that decisions should be determined in legal processes and political procedures requiring public accountability, and that NetzDG could undermine democratic processes and the principle of the rule of law.

---

<sup>21</sup> “Titanic bleibt gesperrt: Tweets der Wahrheit,” *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), January 4<sup>th</sup>, 2018.

<sup>22</sup> “Gesetz gegen Hasskommentare: ‘Die Justiz muss entscheiden, nicht Facebook,’” *Süddeutsche Zeitung*, April 5<sup>th</sup>, 2017.

<sup>23</sup> “Hass bleibt privat,” *Die Tageszeitung*, April 5<sup>th</sup>, 2017.

<sup>24</sup> “Cheflobbyist im Gespräch: ‘Wir wollen kein Debattenwächter sein,’” *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), February 5<sup>th</sup>, 2018.

According to this assessment, NetzDG meant that the state surrendered its responsibilities to private corporations without the wherewithal or moral justification, while the rule of law in fact demanded judicial institutions such as law enforcement agencies and courts to judge and enforce the legality of speech. In line with this fell the proposition – made, for instance, by FDP<sup>25</sup> and Green<sup>26</sup> politicians, internet and publishers’ associations, and even Facebook itself – to instead adequately fund, staff, and train law enforcement to prosecute internet criminality and even create new institutions such as special courts. Thus, this framing put forward the idea that, although state institutions ought to make important (legal) decisions on speech, they needed to take these decisions away from platforms, which are viewed as being ruled by logics full of bias and exploitation, and put them under the auspices of an apparently impartial legal apparatus.

#### 4.6 Reinforcing platforms’ opacity

The view that NetzDG gave too much power and authority to platforms resonated with a framing of NetzDG as potentially reinforcing, and even increasing, the *opacity of powerful* platforms. This view was expressed by articles across the political spectrum, but especially in *Netzpolitik*, a source focused on technology politics. This framing was built upon the observation that platforms devised content moderation policies and community standards in a rather random and obscure fashion. The way platforms operated, including their content moderators’ working conditions, were cited as a cause of this and were described as opening the door to biases and injustices. To illustrate, one *Netzpolitik* article<sup>27</sup> criticized leaked Facebook moderation rules which showed that content moderation practices protected powerful people but put vulnerable groups at risk. That NetzDG would reinforce and even aggravate this trend was a main concern. One such scenario described that NetzDG might incentivize *automated* content moderation, which was seen as even harder to control than human moderation, as well as biased and prone to manipulation.

According to this framing, platform companies’ lack of transparency was part of a greater problem, namely the increasing concentration of power on the internet in the hands of a few companies, which rendered platforms’ opacity unaccountable to users, state institutions, and the public. Where NetzDG was framed in terms of platform opacity, the worry was that it could further disempower users, violate their rights, and prevent scientific inquiry into phenomena

<sup>25</sup> The FDP, (Free Democratic Party) is a major party in Germany supporting economic (and social) liberalism and free markets. In the latest election (2021), the party garnered 11.5% of the votes and became part of the governing coalition (Der Bundeswahlleiter, 2022).

<sup>26</sup> *Alliance 90/The Greens* is a major center-left party in Germany with a focus on environmental protection. In the latest election (2021), the party garnered 14.8% of the votes and became part of the governing coalition (Der Bundeswahlleiter, 2022).

<sup>27</sup> “Warum Facebooks Löschregeln weiße Männer schützen, aber nicht schwarze Kinder,” [Netzpolitik.org](https://netzpolitik.org), June 29<sup>th</sup>, 2017.

such as hate speech. This framing characterized the state as weak and as failing to counter platforms' business models, and NetzDG as not holding platforms accountable or ensuring responsibility and transparency. Instead, it suggested that NetzDG would transfer *more* power to platforms rather than giving it to the public and civil society, which could hold platforms accountable and engage in a public debate on the rules of content moderation (which also required the cultivation of media literacy). A well-balanced and diverse media system that was not centered on immediate exploitation, commercialization, and the pursuit of private interests was then contrasted with platforms' opaque and algorithmic content moderation; in this regard, Germany's public broadcasting provided a positive example.<sup>28</sup>

This framing drew attention to corporate power, accountability, and the structural conditions of speech. In this view, it appeared that NetzDG did not address the "real" problem, which was linked to platforms' opacity, resulting in user disempowerment and unfair treatment. Calls for a stronger state intervention to shape the *conditions* under which internet users can *interact* with platforms, rather than to the content of speech itself, were consistent with this. The framing then raised concerns that NetzDG provided platforms with democratically illegitimate power to avoid transparency and accountability, and, at the same time, implied that platforms could act responsibly when under pressure from users, civil society organizations, or governmental authorities. Transparency toward users, civil society, and oversight institutions was here cited as one of the most important factors for the implementation of effective and democratic forms of self-regulation. Despite its criticism of NetzDG, approvals of the introduction of an obligatory point of contact in Germany and the requirements for a transparent and efficient complaint management system can be included in this framing.

#### 4.7 Governmental overinterference and overregulation on the Internet

A framing of NetzDG as part of a broader trend of governmental *overregulation* and *overinterference* shared with the previous framing the concern that NetzDG could compromise user empowerment and openness online. In doing so, however, it did not focus on the concern that platforms would attain too much power, but rather on a perceived trend of states encroaching upon a free and empowering internet and the civil rights it guaranteed, turning it into a place of control and exerting undue influence over communications. Concerns cited in previous criticisms of NetzDG may exemplify potential outcomes of this. Accordingly, this framing suggested that incompetent (over)regulation on the internet was compromising its inherent openness, democratic nature, and innovative potential.

---

<sup>28</sup> "Demokratisch-mediale Öffentlichkeiten im Zeitalter digitaler Plattformen," Netzpolitik.org, May 7<sup>th</sup>, 2018.

This framing described a broader concern that internet regulations could pose a risk to openness on the web, expand state surveillance and hinder technological progress and innovation. This was illustrated by connections drawn between NetzDG and other internet legislation, such as the EU's copyright reform, which, it was feared, would – just like NetzDG – incentivize the use of upload filters. Such regulations fueled concerns that the over-interference that characterized them was a precursor for authoritarian governance, instituted surveillance, undermined anonymity and civil rights, and weakened freedom of speech and free information flows. A view of internet users as autonomous and *self-determined* individuals who ought to take responsibility for their technology use can be classified as part of this framing. One *Welt* article,<sup>29</sup> for instance, argued that the preservation of freedom and autonomy ought to take priority over the provision of security through state interventions. Such a view implied that users ought to be able to deal with terrible or problematic content online.

By this framing, the government was seen as *incompetent* in the realm of digitalization and as displaying an exaggerated focus on technology as the source of societal problems. This included a view of governmental interference as obstructing the free flow of information through open and accessible networks and as endangering freedom and democracy on the web, but also as *stifling technological innovation and progress*. This framing emphasized the need to keep networks *open* and found many regulatory interventions to be counter-productive in this regard. It included the stipulation that governments should focus their efforts on furthering digitalization, assuring openness, plurality, and competition; upholding net neutrality; providing good infrastructure; and strengthening data protection and IT security. Generally, the position this framing described, seemed to put more trust in the powers of platforms and technological developments and innovations to bring about desirable effects, whereas the state – judged incompetent in and unsuited for this realm – was called on to refrain from limiting access or steering developments. The liberal party FDP's draft for a *Law for Strengthening Civil Rights* could here provide a specific alternative to NetzDG.

## 4.8 Protecting users from economic exploitation

However, contrary to these framings that described NetzDG as a problematic regulation with undesirable results, NetzDG was elsewhere also framed as a *desirable regulatory* intervention that protected *users* from the negative – if not disastrous – consequences of economic exploitation and hence empowered them against platforms. This framing established a connection between NetzDG and consumer protection efforts like data protection and focused on platforms' unsavory business models. NetzDG was specifically cast in the role

---

<sup>29</sup> “Datenschutz; Der Facebook-Skandal, das sind wir selbst,” *Welt Online*, March 25<sup>th</sup>, 2018.



of protecting and empowering users by enforcing transparency and safety on social media. This was mirrored in Maas's call for platforms' complaint management to become more "user-friendly."<sup>30</sup> Regulations and regulatory control and oversight were described as necessary in this framing because companies were deemed untrustworthy. The focal concern was companies' business models, which included targeted advertising, attention capturing, and practices of opaque algorithmic filtering. To exemplify this, one *TAZ*<sup>31</sup> article described Russian-sponsored election manipulation efforts as "business as usual" on Facebook. Platforms' own efforts at combating fake news and hate speech were characterized as strategies to avoid regulation and not as measures with real teeth. NetzDG, by contrast, presented an attempt to hold companies accountable, to ensure they stuck to rules and took responsibility for the content their sites exposed users to.<sup>32</sup>

This framing shared with others a positive view on NetzDG's intervention but centered on the empowerment of *users as consumers* that needed protection from economic exploitation rather than the conservation of citizens' speech rights. It could thus form a bridge between two sides: Rather than describing NetzDG as state interference into the realm of expression and communication, which might risk infringing on freedom of speech, it positioned it as a mode of state engagement that held companies accountable, enforced the rights of users, and created the right conditions for free consumer choice. This thus shifted the focus from regulating citizens' speech to shaping the relationship between users as consumers of platform services. NetzDG was understood here as empowering users against corporations so that they could successfully participate in shaping their online communication spaces.

#### 4.9 A "bad" or even illegitimate law

Finally, articles from various sources and voices cited across the political spectrum also raised concerns about the quality of NetzDG's *legislative process*. They articulated a framing that stood apart from the others in that it did not center on the content of NetzDG and its consequences, but rather on the *legislative process* or even the intentions of its proponents. When it comes to the democratic legitimacy of regulating content moderation, this framing is particularly interesting because it shows that it is not only the content of a legislation, but also the process behind it, that determines its democratic acceptability.

---

<sup>30</sup> "Internet: Facebook richtet zweites deutsches Löschzentrum in Essen ein," *Süddeutsche Zeitung*, August 9<sup>th</sup>, 2017.

<sup>31</sup> "Debatte Hass im Netz: Die Sensationsschleuder," *Die Tageszeitung*, April 3<sup>rd</sup>, 2018.

<sup>32</sup> The articles, however, differed in how well they believed NetzDG succeeded in this, with some even finding it to not go *far enough* in holding internet corporations to account.

This framing was motivated by ideas of what *due democratic process of legislation* ought to look like and what signified a “good” law. This included concerns that, aside from its content, NetzDG was a technically “bad” or even illegitimate law, and criticisms of the process by which it had been instituted. Observations that NetzDG’s implementation had been rushed, lacked careful public deliberation, and was not based on expert opinions were part of these criticisms. They pointed to procedural and technical flaws that indicated NetzDG’s lack of legal precision, certainty, and compliance with other laws and its *poor quality as a law*. This framing then also cast doubts on the new law’s practical effectiveness and ability to counter hate and radicalization.

In line with this, NetzDG was frequently cited in relation to ongoing political developments. In the context of the upcoming elections, it was featured as a political football in election campaigning, political profiling, and coalition negotiations. One AfD politician<sup>33</sup> even went so far as to describe it as a censorship law that resembled the methods of the *Stasi*, former East Germany’s notorious secret police. This comparison fell in line with the party’s strong opposition to NetzDG and worked to question the democratic credibility of the political actors involved and of NetzDG’s intention.

## 5 Content moderation and the quest for democratic legitimacy

The different framings illustrate a controversy that ensued over whether NetzDG’s regulatory interventions into corporate content moderation protected, nourished, and complied with – or endangered and undermined – democratic values and principles such as freedom of speech, the rule of law, and public democratic discourse. Different framings based their evaluations on distinct interpretations of how such democratic principles and values should be understood on social media platforms and how they should be upheld there. I suggest that this contestation can be understood *as a quest for the democratically legitimate way to govern content moderation in accordance with fundamental democratic principles and values*.

### 5.1 The concept of democratic legitimacy

“Legitimacy” refers to the conditions under which some actors, such as governments, law enforcement agencies, and state institutions, are (morally) justified in *wielding power* over others, such as citizens (Buchanan, 2002).

---

<sup>33</sup> Cited in: “Titanic bleibt gesperrt: Tweets der Wahrheit,” *Frankfurter Allgemeine Zeitung* ([faz.net](http://faz.net)), January 4<sup>th</sup>, 2018.

In *descriptive* terms, legitimacy refers to when people *accept* laws and governmental practices as legitimate because they believe that they comply with the demands of democracy. In *normative* terms, legitimacy describes when it *should* be acceptable for the state to wield such power, to exercise force and coercion, or to restrict individuals' autonomy and liberty. Legitimacy provides reasons for accepting rules or governance practices and for allowing the state to "exercise a monopoly on the making, application, and enforcement of laws" (Buchanan, 2002, p. 695). Therefore, democratic legitimacy refers to the circumstances under which the wielding of governmental power is compliant with democratic rules, where citizens' acceptance of governance structures and laws pertains not to the state, but to one another (Bekkers & Edwards, 2016, p. 41; Buchanan, 2002, p. 714).

Alan Buchanan suggests that democratically legitimate governmental practices must be based on a principle of equality that guarantees that all citizens have an equal say in deciding on the government and fundamental laws and that recognizes and credibly protects human rights (Buchanan, 2002, p. 710). This ought to ensure the right to self-governance or *popular sovereignty* – the reign of people over themselves – which is a foundation for democratic thinking and its justification for governmental practices (Bhagwat & Weinstein, 2021, p. 83). In its function of ensuring that everyone can participate in public discourse, make their voices heard, and access the facts and information necessary to form opinions, freedom of speech is intimately related to democratic legitimacy (Bhagwat & Weinstein, 2021, pp. 83, 90; Restrepo, 2013, p. 380). It ought to establish *political equality* and guarantee that all citizens can contribute equally to public opinion, persuade others of their views, and participate "as political equals in making binding decisions, enforced by the state, on matters that have important consequences for their individual and collective interests" (Bhagwat & Weinstein, 2021, p. 88). This should bring about a public discourse by which a public opinion is formed that exercises *popular control* and holds the government *accountable*.

Although legitimacy has traditionally been discussed in relation to states' and governments' power, social media platforms' capacities to set communicative norms and regulate users' behavior has more recently raised the question of whether legitimacy requirements can and should apply to platforms' private governance (Cowls et al., 2022; Taylor, 2021). By exercising power over people's right to express themselves and to share and access information, and by shaping the forms that communication and discourse take online, content moderation policies and practices raise questions of democratic legitimacy for both governmental and corporate practices; it is to these questions that the framings responded. They concerned the principles on which to base policies and practices, but also addressed the processes by which rules can be legitimately set and the actors that can enforce them. Consequently, the quest for democratic legitimacy in content moderation is also a struggle over the distribution of agency, power, and responsibility.

## 5.2 On the permissibility of speech regulation

The debate surrounding the right regulations, policies, and practices for content moderation addresses the value of democratic principles and their implications for how to deal with content on social media. This raises difficult questions about the democratic governance of public discourse, the meaning of free speech, and the permissibility of regulatory or governmental interference with speech. These questions have been hotly debated in democratic theory, where vastly different views exist about what freedom of speech entails and the extent to which democracy is served by legal limitations on speech and governmental interventions into public discourse. They put forward contrasting understandings of how discourse operates, how discourse participants act and process information, and how the state relates to this discourse.

Advocates against state interference and speech limitations argue that governments cannot be trusted to decide what is true or right speech; instead, they must leave it up to citizens as mature and rational individuals (Loewy, 1993, p. 430). The *marketplace of ideas* concept suggests that truths will prevail when all ideas are proposed, tested, opposed, and defended by autonomous participants in a marketlike structure to which everyone has access (Marshall, 2021, p. 44, Blasi, 2021, p. 29). Limitations on speech are characterized as unacceptable because they are viewed as compromising the autonomy citizens need “to form their own opinions about their beliefs and actions” and express themselves (Stone & Schauer, 2021, p. xiii). In this view, denying people access to information is akin to controlling thought processes and behaviors, and citizens require such autonomy to enact their democratic right to self-government (Mackenzie & Meyerson, 2021, p. 64, 66).

Habermas’s deliberative approach to democracy also advocates for far-reaching speech protections so that public discourse can be held without state influence and with the participation of as many citizens as possible (Calhoun, 1992, pp. 7–8, 13–15; Rehg, 1996, p. xi ff.). In the free and open discourse thus imagined, critical rationality and the exchange of arguments would facilitate the formation of a public will that would act as a counterbalancing force to the state (Calhoun, 1992, pp. 9, 17; Habermas, 1996a, pp. 106, 108, 119). Discourse participants would act as *citizens* in search of *common* interest and would enter public discourse detached from their personal interests, letting the quality and rationality of arguments be decisive (Calhoun, 1992, p. 13). The media, according to the deliberative approach, should then facilitate rational deliberation through the circulation of accurate facts and information rather than catering to individual interests, providing entertainment, or seeking attention (Habermas, 1996b, pp. 368, 379).

Other perspectives have argued that state interventions into speech are justified because hate speech diminishes the autonomy of *its targets* and deters people from participating in discourse (Mackenzie & Meyerson, 2021, p. 62).

They see such interventions as necessary to *prevent* the subversion of democratic values and the distribution of false, hateful, and conspiratorial content that may harm individuals, groups, and democratic systems, impinge on equality and freedom for everyone, and have adverse effects on public discourse. Restrictions may also be democratically permissible when speech leads to false beliefs and harmful acts that endanger autonomy, as well as when it promotes fraud and deceit that hinders people from making the free, informed, and wise political choices that democracy requires of them (Restrepo, 2013).

Where relations of domination put equal rights at risk and speech may perpetuate the domination of certain people, groups, or interests over others, appeals to equality and human dignity further provide justification for regulatory interventions. These interventions ought to ensure that public discourse produces a “polity and policy which demonstrates tolerance, mutual respect, and an embrace of diversity” (Bhagwat & Weinstein, 2021, p. 102). In this vein, Nancy Fraser has argued that state engagement in the sphere of public discourse may be permissible or even necessary to reveal and eliminate existing inequalities that prevent equal participation (Fraser, 1990, p. 63 ff.). Similarly, the agonistic approach to democracy which Chantal Mouffe has developed can justify regulatory engagement where it is necessary to channel inevitable political agonisms, emotions, and conflicts into productive democratic debates; in Mouffe’s view, failing to recognize and adequately account for the role of such agonism and affects in politics could allow forms of domination and violence to go unrecognized (Mouffe, 2000, p. 26 ff.). Taking responsibility for choices on how “the just political order” is to be constituted (Mouffe, 2000, p. 62) and actively shaping the channels of democratic communication then becomes a central task of democratic politics (Mouffe, 2016, p. 22).

Finally, there is another argument for governmental intervention, which centers on preventing privatized economic interests from impeding on open and free discussion (Fraser, 1990, p. 74). According to this position, states may step in where markets fail to enable a democratic, plural media or fail to counter oppression (Benson, 2009, p. 188 ff.). In this view, state engagement to ensure a diversity of news outlets ultimately *enables* freedom of speech, whereas a strictly privatized media system may not bring about free-floating and disinterested public debate. This is supported by a view of freedom of speech as a *positive right* that sustains a plurality of public speech, enables structural diversity, and counters unequal economic power (Kenyon, 2021).

### 5.3 Grounding NetzDG in democratic legitimacy

The different views on the acceptability of NetzDG’s intervention into content moderation and its effects on democracy – presented by the framings – reflected such arguments about the acceptability of governmental intervention to

public speech. This happened, for instance, where they argued that regulatory intervention was necessary to channel public discourse into democratic forms or that it was dangerous because it could lead to democratically unacceptable infringements on speech and a plurality of views. However, these arguments were not part of a general debate about speech regulations; rather, they were employed to provide concrete assessments of the situation. Much of the debate left the substantive content of speech rules untouched; existing speech laws were largely accepted. Instead, the *consequences of NetzDG's enforcing* of these laws on *the internet* were discussed. This shifted the focus from the substantive permissibility of speech to the systems and mechanisms that determine and apply rules and the institutional logics behind content moderation systems. It invited a closer look at the structural effects of state involvement and the establishment of *public accountability* for content moderation. It also addressed the *territory of the law*, debating whether legal judgments needed to be made in courts or on platforms.

In the controversy, NetzDG was characterized as either supporting or threatening democratic principles through its *practical application of the law* and *its interaction with social media platforms*. The framings differed in how they described the interactions between NetzDG's regulatory interventions and social media platforms. On the one hand, there was the view that NetzDG would incentivize, reinforce, and even exacerbate problematic and anti-democratic tendencies in corporate platform governance or even governmental regulation. Fears over the negative consequences of overblocking arose from the interaction between NetzDG's incentive structures and platforms' financial interests and the consequences of entrusting companies with quasi-legal decisions under precarious labor conditions. On the other hand, there was also the idea that NetzDG would address problematic tendencies on social media and help to make communication and discourse more democratic. Here, NetzDG was described as a democratically necessary intervention to counter the negative effects of platforms' practices. These differences draw attention to how underlying *structures* of speech systems are envisioned, which is consistent with Evelyn Douek's proposition to shift the attention of governance efforts away from the content of moderation rules and decisions toward the institutional logics and incentives behind them (Douek, 2022). The question then becomes not what kind of speech falls within democratic boundaries and whether the outcome of an individual decision lies within them, but rather how the decision-making *process* can be governed in a publicly accountable and democratically legitimate manner.

Furthermore, it is paramount to examine the framings' underlying vision of the relations between users, state institutions, and platforms; who is trusted with judging and evaluating information; and which capacities are required of internet users. In framings that described NetzDG as a potentially dangerous infringement of freedom online, users were envisaged as needing more capacities to cope with undesirable content, hate, and discrimination; to judge

information; and to engage actively in public discourse. They were viewed as being able to use their engagement to exercise power over platforms once transparency was established and, at the same time, to utilize platforms to exercise their democratic rights against the state. This mostly bracketed structural effects, such as hate speech's disproportional targeting of groups already discriminated against. On the other hand, in positive framings of NetzDG it appeared as acceptable – and even desirable – to let the state perform some of this work, for instance by actively guiding how content was curated and moderated or by ensuring law and order, despite the potential to infringe on individuals' liberties. These framings described such interventions as protecting platform users against unruly private corporations or malicious and antidemocratic actors, but also as ensuring everyone's dignity and equal opportunity to participate in public discourse.

In the controversy, the argument that there was nothing special about NetzDG – that it was “simply” applying laws to platforms just as they are applied offline – faced serious pushback. This meant that NetzDG was unsuccessful in avoiding the need to draw new boundaries for democratic discourse but also changed the *modalities* by which such boundaries were drawn. In the NetzDG controversy, the negotiation of boundaries did not concern the content of permissible speech; instead, they addressed the right mechanisms and actors to govern content moderation systems and the ways in which the state ought to bring laws to platforms in accordance with democratic principles.

Thus, it makes sense that voices outside mainstream political discussion or who seek to expand the bounds of what is generally accepted as democratically legitimate speech may be especially vociferous in their opposition to a law like NetzDG. For those who benefit from platforms and can use or exploit their algorithmic logics, the realm of content moderation and the permissibility of state intervention into it can open a space to renegotiate the boundaries of democratic speech without attacking the substance of speech laws. This may be done by actors seeking to shift public discourse to the political right or to normalize problematic or discriminatory speech. However, the debate over content moderation practices can also offer a chance for those seeking to eradicate racist or sexist speech norms or change communicative and social norms. There may be greater room to negotiate such norms between platforms and users than there is with democratic states, against whose monopoly of force wide-ranging and defensive speech rights often need to be strongly upheld.

## 6 Democratic legitimacy as a framework for content moderation

My research has shown that the a great part of the public (media) controversy over NetzDG in Germany revolved around questions of democratic legitimacy raised by the governance of content moderation. Beyond discussing what content to delete, this controversy drew attention to the values, principles, and institutional structures that ought to govern moderation systems. When assessing different perspectives on how to govern content moderation, I suggest it is important to pay critical attention to how these perspectives and their respective policies and practices envision the distribution of powers and responsibilities between platforms, state institutions, and users/citizens. Scrutinizing different legitimation strategies for the regulation of content moderation, in terms of how they prefigure social interactions and power relationships, can make content moderation more democratically accountable and sensitive to its implications for social and political order. In the final section of this paper, I reflect on the merits, drawbacks, and further questions of democratic legitimacy as a lens for discussing and approaching the governance of content moderation. This concerns which aspects the debate over NetzDG in terms of its democratic legitimacy made visible, which ones it made invisible, and what we can learn from this debate about the governance of content moderation on social media platforms.

The lens of democratic legitimacy directs attention toward how the appropriate framework for governing content moderation can be derived from the right selection, interpretation and application of democratic principles and values. However, it also raises the question of *what* content moderation governs, especially given that this lens suggests approaching content moderation as governing a public discourse in the service of democracy. This, in turn, draws attention to the challenges of running this discourse on a private infrastructure within the internet economy. It also prompts careful reflection on the kind of content and speech that is regulated by moderation practices and the circumstances under which they function as part of public discourse. Finding the right framework to appropriately respond to the myriad of human interactions on social media seems a daunting task, and upholding fundamental human rights and the rule of law may appear to be the lowest common denominator. However, the controversy over NetzDG illustrates that even this is not a straightforward or uncontested task.

The question of how to govern content moderation also concerns the capacities that users bring to platforms and the kind of collective of which they are part. For instance, a public discourse perspective requires social media users to act as discourse participants who understand themselves as citizens who take on civic responsibilities and act as members of a shared democratic community endeavoring to organize itself. Applying national laws may be one strategy to create such a community, as such laws correspond to the social and political



structures that created them. However, as the NetzDG controversy has demonstrated, their interaction with the terrain of social media platforms may nevertheless prompt new challenges for their democratically legitimate enforcement.

Vice versa, the way in which content moderation is enacted may also *create* specific types of users, online communities and discourses, and the rules implemented may lead to new social norms. Strong prohibitions on hate speech are one way to establish collective and inclusive norms based on principles of equality, autonomy, and respect. The consequences of content moderation requirements and of employing algorithms force us to confront how communication structures are shaped and prevent us from imagining that “neutral,” free-floating discourses can magically work themselves out in the name of democracy. Consequently, the challenges created by content moderation practices and their regulation require us to actively shape collective boundaries and make visible the principles by which we do so.

A second interesting point prompted by the NetzDG controversy concerns how to practice law enforcement and the rule of law online – the difficulties that were debated arose where “traditional” law enforcement approaches collided with the functionalities and workings of online platforms. One reason for this potential mismatch is the volume and speed of posts that appear to make thorough judicial evaluation difficult or even impossible. Furthermore, the impacts of algorithmic orderings and complex dynamics are hard to tackle through speech laws that judge individual instances and may thus lead to calls for new institutions. At the same time, framing content moderation in terms of the rule of law inspires us to think about content moderation as a public task and compare it to public administrations and civil services. It holds up a promise of impartiality, democratic procedures, clear and transparent rules, accountability, and equal treatment. It also raises questions about *who* is making content moderation decisions, in what capacity, with what expertise, under which circumstances –and, importantly, with what justification.

Third, understanding the controversy as a quest for how to govern content moderation in a democratically legitimate way illustrates how platforms and regulatory interventions can *bring about a renegotiation of the meaning of democracy*. Here, the specific technological and political conditions at hand, as well as platforms’ ways of operating, have transformed existing democratic discussions, calling on us to reevaluate what democratic principles and values mean when enacted online. The NetzDG controversy showed that although speech laws may have remained unchanged, *corporate and regulatory practices of content moderation* can bring forth new, technologically situated meanings for democratic values and principles. These practices are also constitutive of the capacities with which state institutions, platforms, and users are acting online. They do so in interaction with preexisting structures and local contexts – a law like NetzDG was possible because of the preexisting form of German jurisprudence and its understanding of democracy.

Finally, the use of democratic legitimacy as a framework for governing content moderation casts the overall discussion in terms of citizen–state relations and principles of democratic rule and public discourse. This can, however, obscure other aspects of content moderation. For instance, platforms may open up important alternative spaces that can be governed by rules and norms more restrictive – but perhaps also more just or inclusive – than those that the speech regulations of democratic governments allow for, given their need to uphold far-reaching protections of free speech against the state and its monopoly of force. Such alternative – that is, privately governed – spaces of communication may then do justice to different people’s and groups’ identities, histories and needs, and they may allow for sources of community outside legal boundaries. Beyond hardwired and forceful laws, they may give people and communities the space to contest and create *social norms* and acceptable ways of communicating and interacting. The question that remains is how this could happen in an equitable manner outside of governmental or legal frameworks.

## 7 Conclusion

This paper started with the question of how regulatory interventions into content moderation were made sense of in the public controversy surrounding NetzDG. It found that a major point of contention was the relationship between content moderation and democracy; different framings struggled over how to ground content moderation on social media in principles of democratic legitimacy. Thinking about content moderation as a quest for democratic legitimacy points to some of the most difficult challenges the governance of content moderation on social media platforms faces. It helps to sharply articulate what the fundamental questions are that need to be collectively addressed. This perspective encourages us to pinpoint the exact societal and political consequences inherent in various approaches to content moderation and empowers us to actively design our democracies rather than taking the notion of an immutable “democracy” as the conversation’s starting point. It also asks us to carefully consider how democratic values and principles are transformed when they meet platform practices and, consequently, what the right frameworks for establishing politically accountable content moderation systems might look like. The questions concerning the rule of law that were raised in the NetzDG controversy contributed an interesting perspective to this discussion because they drew attention away from the substance of moderation decisions to the underlying systems that govern their rules and practices. The framework of democratic legitimacy, therefore, calls on us to not only clarify the statuses, roles, and responsibilities of different actors online, but also to determine exactly what is regulated by content moderation.

## 8 Acknowledgements

I wish to thank two anonymous reviewers, as well as Judith Simon, Gijs van Maanen, Sarah Carter, and Lena Ulbricht for their constructive feedback on earlier drafts. I would also like to thank Roisin Cronin for support in language editing.

## References

- Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act). (2017). [https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG\\_engl.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2)
- Bekkers, V., & Edwards, A. (2016). Legitimacy and democracy: A conceptual framework for assessing governance practices. In V. Bekkers, G. Dijkstra, A. Edwards, & M. Fenger (Eds.), *Governance and the Democratic Deficit: Assessing the Democratic Legitimacy of Governance Practices* (pp. 35–60). Routledge.
- Benson, R. (2009). Shaping the public sphere: Habermas and beyond. *The American Sociologist*, 40, 175–197. <https://doi.org/10.1007/s12108-009-9071-4>
- Bernau, P. (2018, March 13). Digitalkonferenz SXSXW - Drohung mit deutschen Gesetzen. *Frankfurter Allgemeine Zeitung*. <https://www.faz.net/aktuell/wirtschaft/digitalkonferenz-sxsw/sxsw-londoner-buergermeister-droht-mit-dem-deutschen-netzdg-15491236.html>
- Bhagwat, A., & Weinstein, J. (2021). Freedom of expression and democracy. In A. Stone & F. Schauer (Eds.), *The Oxford Handbook of Freedom of Speech* (pp. 82–105). Oxford University Press.
- Buchanan, A. (2002). Political legitimacy and democracy. *Ethics*, 112(4), 689–719.
- Bundesministerium der Justiz und für Verbraucherschutz. (1949). *Basic Law for the Federal Republic of Germany*. [https://www.gesetze-im-internet.de/englisch\\_gg/](https://www.gesetze-im-internet.de/englisch_gg/)
- Calhoun, C. (1992). Introduction: Habermas and the public sphere. In C. Calhoun (Ed.), *Habermas and the Public Sphere* (pp. 1–48). The MIT Press.
- CDU/CSU; SPD. (2017). *Gesetzentwurf der Bundesregierung – Entwurf eines Gesetzes zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*. [https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RegE\\_NetzDG.pdf?\\_\\_blob=publicationFile&v=2](https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RegE_NetzDG.pdf?__blob=publicationFile&v=2)

- Cowls, J., Darius, P., Santistevan, D., & Schramm, M. (2022). Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance. *New Media & Society*, 1–25. <https://doi.org/10.1177/14614448221085559>
- Der Bundeswahlleiter. (2022). *Bundestagswahl 2021*. <https://www.bundeswahlleiter.de/bundestagswahlen/2021/ergebnisse/bund-99.html#zweitstimmen-prozente12>
- Douek, E. (2022). Content moderation as systems thinking. *Forthcoming Harvard Law Review*, 136, 1–82. <https://doi.org/10.2139/ssrn.4005326>
- Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Fraser, N. (1990). Rethinking the public sphere: A contribution to the critique of actually existing democracy. *Social Text*, 25(26), 56–80.
- Gesetz zur Änderung des Netzwerkdurchsetzungsgesetzes. (2021). *Bundesgesetzblatt*, 1(29), 1436–1443.
- Gillespie, T. (2018). *Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Habermas, J. (1996a). A reconstructive approach to law I: The system of rights. In *Between Facts and Norms. Contributions to a Discourse Theory of Law and Democracy* (pp. 82–131). MIT Press.
- Habermas, J. (1996b). Civil society and the political public sphere. In *Between Facts and Norms. Contributions to a Discourse Theory of Law and Democracy* (pp. 329–387). MIT Press.
- Hülsen, I., & Müller, P. (2018, January 19). EU-Justizkommissarin zweifelt am Maas-Gesetz. *Spiegel Netzwelt*. <https://www.spiegel.de/netzwelt/netzpolitik/vera-jourova-und-das-netzdg-eu-justizkommissarin-zweifelt-am-deutschen-gesetz-a-1188703.html>
- Jasanoff, S. (2004). The idiom of co-production. In S. Jasanoff (Ed.), *States of Knowledge* (pp. 1–12). Routledge.
- Kenyon, A. T. (2021). Positive free speech: A democratic freedom. In A. Stone & F. Schauer (Eds.), *The Oxford Handbook of Freedom of Speech* (pp. 231–248). Oxford University Press.
- Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598–1670.
- Loewy, A. H. (1993). Freedom of speech as a product of democracy. *University of Richmond Law Review*, 27(3), 427–439.

- Mackenzie, C., & Meyerson, D. (2021). Autonomy and free speech. In A. Stone & F. Schauer (Eds.), *The Oxford Handbook of Freedom of Speech* (pp. 61–81). Oxford University Press.
- Marshall, W. P. (2021). The truth justification for freedom of speech. In A. Stone & F. Schauer (Eds.), *The Oxford Handbook of Freedom of Speech* (pp. 44–60). Oxford University Press.
- Mouffe, C. (2000). *The democratic paradox*. Verso.
- Mouffe, C. (2016). Democratic politics and conflict: An agonistic approach. *Política Común*, 9, 1–20. <https://doi.org/10.3998/pc.1232227.0009.011>
- Rehg, W. (1996). Translator’s introduction. In *Between Facts and Norms. Contributions to a Discourse Theory of Law and Democracy* (pp. ix–xxxvii). MIT Press.
- Restrepo, R. (2013). Democratic freedom of expression. *Open Journal of Philosophy*, 3(3), 380–390. <http://dx.doi.org/10.4236/ojpp.2013.33058>
- Roberts, S. T. (2019). *Behind the screen*. Yale University Press.
- Schmidt, A. (2015). Moralvorstellungen in der öffentlichen Debatte: Konzeptionelle und methodische Überlegungen zu Relevanz und empirischer Untersuchung. *Studies in Communication Media*, 4(2), 69–134. <https://doi.org/10.5771/2192-4007-2015-2-69>
- Stone, A., & Schauer, F. (2021). Introduction. In A. Stone & F. Schauer (Eds.), *The Oxford Handbook of Freedom of Speech* (pp. xi–xxiii). Oxford University Press.
- Taylor, L. (2021). Public actors without public values: Legitimacy, domination and the regulation of the technology sector. *Philosophy & Technology*, 34, 897–922. <https://doi.org/10.1007/s13347-020-00441-4>
- van Hulst, M., & Yanow, D. (2016). From policy “frames” to “framing”: Theorizing a more dynamic, political approach. *American Review of Public Administration*, 46(1), 92–112. <https://doi.org/10.1177/0275074014533142>

---

**Date received:** September 2021

**Date accepted:** September 2022