An Analysis of Germany’s NetzDG Law†

Heidi Tworek, University of British Columbia
Paddy Leerssen, Institute for Information Law, University of Amsterdam

April 15, 2019

Contents
Introduction: What is the NetzDG? ................................................................. 1
Freedom of Expression Implications ................................................................. 2
The NetzDG in Practice: What Does the Evidence Show? ................................. 4
Outlooks and Next Steps .................................................................................... 7
  Transparency and Research.............................................................................. 7
  Due Process and Design Structure ................................................................. 8
  Multi-Stakeholder Relationships .................................................................. 9
Appendix: Links to transparency reports .......................................................... 10
Notes .................................................................................................................. 10

Introduction: What is the NetzDG?

Germany’s Network Enforcement Act (Netzwerkdurchsetzungsgesetz or NetzDG) entered into full force on January 1, 2018. Known colloquially as a “hate speech law,” it is arguably the most ambitious attempt by a Western state to hold social media platforms responsible for combating online speech deemed illegal under the domestic law. Other countries, like France, are using NetzDG as a basis for proposed legislation, making analysis of NetzDG urgent and important.²

While NetzDG has encouraged accountability and transparency from large social media platforms, it also raises critical questions about freedom of expression and the potential chilling effects of legislation. During a first meeting at Ditchley Park, UK (February 28-March 3, 2019), the Transatlantic Working Group (TWG) analyzed NetzDG based on an earlier draft of this document. The current version was revised to incorporate crucial insights from those discussions. This updated document introduces NetzDG’s content and political context and discusses its implications for freedom of

expression. It closes with next steps and recommendations for more research and transparency, while suggesting how to mitigate the troubling elements of the law.

At the outset, it is important to clarify that NetzDG does not actually create new categories of illegal content. Its purpose is to enforce 22 statutes in the online space that already existed in the German criminal code and to hold large social media platforms responsible for their enforcement. The 22 statutes include categories such as “incitement to hatred,” “dissemination of depictions of violence,” “forming terrorist organizations,” and “the use of symbols of unconstitutional organizations.” NetzDG also applies to other categories, such as “distribution of child pornography,” “insult,” “defamation,” “defamation of religions, religious and ideological associations in a manner that is capable of disturbing the public peace,” “violation of intimate privacy by making photographs,” “threatening to the commission of a felony” and “forgery of data intended to provide proof.”

NetzDG targets large social network platforms, with more than 2 million users located in Germany. It requires these platforms to provide a mechanism for users to submit complaints about illegal content. Once they receive a complaint, platforms must investigate whether the content is illegal. If the content is “manifestly unlawful,” platforms must remove it within 24 hours. Other illegal content must be taken down within 7 days. Platforms that fail to comply risk fines of up to €50 million.

NetzDG also imposes transparency requirements. If a platform receives more than 100 complaints per year, it is required to publish semi-annual reports detailing its content moderation practices. The act stipulates in some detail what types of information must be included. The first round of reports was published in June 2018; the second round appeared in early 2019. Their results are discussed in further detail below.

**Freedom of Expression Implications**

Keywords:
- Over-removal
- Privatized enforcement
- Definition of “unlawful content”
- The Streisand effect (counterproductive outcomes of censorship)
- Inspiration for authoritarian regimes around the world to restrict speech

During its development and implementation, NetzDG triggered fierce debate and widespread concern about its implications for freedom of expression. Then Justice Minister (and current Foreign Minister) Heiko Maas from the SPD presented it as a means to tackle online hate speech and viral deception: “The freedom of expression also protects offensive and hateful statements. But it is not an excuse to commit crimes.” The measure seems popular with German voters; one poll in March 2018 showed an approval rate of 87%, including 67% who “strongly approved” of the law, and only 5% disapproval. NetzDG’s proponents included Stefan Heumann, co-director of the digital policy think tank Stiftung Neue Verantwortung, who emphasized Germany’s constitutional tradition of minority protection, the rule of law online, and German public support.
Criticisms from the tech industry, activists, and academics seemed to outweigh support. Although the law excludes journalistic platforms, the German Journalists Association joined civil rights activists, academics, and lawyers in signing a joint statement warning that the law “jeopardizes the core principles of free expression.” The Global Network Initiative (GNI), a multi-stakeholder self-regulatory body funded by social media companies, e.g., Google and Facebook, asserted that the law “poses a threat to open and democratic discourse.” Other critics included Wikimedia Deutschland, the Internet Society, and the German Startups Association. Despite their widespread objections, the law was drawn up swiftly and passed before the German elections in October 2017 with little time to consult civil society organizations or experts.

The first concern surrounding freedom of expression was that NetzDG would encourage the removal of legal content, also known as “over-removal.” Online platforms, it was argued, would not have the expertise or time to assess every complaint in detail. Making such legal assessments typically requires significant expertise in German language and jurisprudence, as well as complex case-by-case analysis and investigation. Given these costs as well as NetzDG’s tight deadlines and heavy fines, platforms would have a strong incentive simply to comply with most complaints, regardless of their actual merits. This would lead to over-removal.

Relatedly, critics objected to NetzDG as an instance of “privatized enforcement” because, rather than courts or other democratically legitimized institutions, platforms assess the legality of content. The NetzDG process does not require a court order prior to content takedowns nor does it provide a clear appeals mechanism for victims to seek independent redress. Wenzel Michalski, Germany Director at Human Rights Watch, argued that the law “turns private companies into overzealous censors to avoid steep fines, leaving users with no judicial oversight or right to appeal.”

Many critics also disagree with the substance of Germany’s speech prohibitions as too broadly defined or simply wrong on principle. As Article 19, a free speech advocacy group, put it, several provisions “should not be criminal offences in the first place,” including blasphemy, broad definitions of “hate speech,” criminal defamation and insult. There were broader worries about the law’s potential effects on German democracy: prominent cases of deletion might fuel anti-government sentiment or publicize the deleted material far more widely, a phenomenon often known as the Streisand effect.

Critics of NetzDG were typically positive about NetzDG’s transparency requirements. Although Article 19 requested the repeal of the law itself, it has asked that transparency requirements be maintained in a separate act.

The law symbolized a deeper disagreement on the role of free speech in democracy. Many West German politicians — conservatives and Social Democrats alike — believed in a “militant democracy” (wehrhafte Demokratie), where free speech could be constrained to protect democratic norms. As historian Udi Greenberg has put it, “limiting rights became synonymous with the democratic order.” West Germany was the only country in postwar Europe to ban both nationalist parties and the Communist Party in the early 1950s. During the creation of NetzDG, Heiko Maas drew explicitly on Germany’s Nazi past and the tradition of militant democracy to assert that “freedom of speech has boundaries.”
Beyond Germany, there were concerns that NetzDG would serve as a blueprint or precedent for authoritarian regimes to repress online speech. The Global Network Initiative claimed that NetzDG “posed unintended but potentially grave consequence for free expression in Germany, across the EU, and worldwide” and expressed concerns that it may “empower authoritarian leaders.” For example, Russia copied passages from NetzDG in mid-2017 for an anti-terror law that required internet and telecoms providers to save the contents of all communications for six months.

Finally, a striking aspect of the NetzDG debate is that the law’s implementation and subsequent transparency reports do not seem to have changed many minds. Proponents see the law as a vital measure to rein in platforms and implement German law online, often with an implied broader goal of preserving German democracy. Critics worry about the freedom of expression implications. The battle lines drawn before the law’s implementation remain entrenched.

The NetzDG in Practice: What Does the Evidence Show?

A few days after NetzDG came into force in January 2018, prominent AfD (Alternative für Deutschland) politician Beatrix von Storch saw one of her social media posts removed from Twitter and Facebook under the law. Widespread media coverage of this incident, including the post’s content and its potential illegality, seemed to confirm fears of the Streisand effect, or what one journalist dubbed the Storch effect. The AfD has marshalled NetzDG as part of a broader argument that its voice and opinions are being silenced.

Germany’s liberal party, the FDP, has claimed that NetzDG’s restrictions on freedom of expression violated the German constitution. Senior FDP politicians stated that they refrain from posting on social media because of NetzDG. After their parliamentary initiative to revoke the law failed, the FDP sued the government in November 2018 seeking repeal of NetzDG. The case is currently pending before the Cologne Administrative Court (Verwaltungsgericht), though some experts give it little chance of success because the FDP has not produced specific examples.

The best evidence to date about the specific effects of NetzDG comes from the law’s transparency requirements. Four major online platforms released their first transparency reports in June 2018: Google (i.e., YouTube), Facebook, Twitter, and Change.org. This provoked another round of debate about the law’s impact and efficacy. Perhaps unsurprisingly, opinion remains divided.

Before delving into the transparency reports, it is important to note that these data only cover removal decisions arising from NetzDG complaints, and do not account for other removals based on other types of complaints, referrals, or injunctions. Furthermore, the metric of takedowns does not reveal whether NetzDG has achieved its purpose of combating hate speech and other online excesses. The differences between complaint mechanisms and the reports themselves make certain types of comparison difficult. It is also hard to know how the volume of content removal compares to the overall volume of illegal speech online. Some key results are summarized below:
Despite Facebook’s size, it received significantly fewer reports than YouTube and Twitter. More than 100 times fewer, in fact, because Facebook’s NetzDG complaint form was relatively hard to access. YouTube and Twitter integrated NetzDG complaints into their regular “flagging” interface, which can be accessed through direct links next to every piece of content. Facebook, by contrast, placed their complaint form on a separate, less prominent page, requiring multiple clicks to access. The report data suggest that this design choice had massive impacts on the actual uptake of NetzDG complaints.

It is important to note that most of the takedowns resulting from NetzDG complaints removals appear to have occurred under the companies’ community guidelines (or “terms of service”), rather than the German speech laws which NetzDG is intended to enforce. Google, Facebook, and Twitter all prioritize compliance checks with their community guidelines; with each complaint, they first consider whether it violates their community standards. Any content that fails this check is removed. Only the content that passes is then considered for removal under one of the 22 statutes of the German criminal code enforced by NetzDG. Accordingly, as Google’s transparency report shows, a majority of removal decisions are based on the platform’s private standards, and not on German speech laws. Facebook and Twitter do not specify this data in their reports, but they do review complaints in the same order, prioritizing community guidelines.

Table 1: Overview of reported numbers by platform

<table>
<thead>
<tr>
<th>Platform</th>
<th>Total items reported</th>
<th>Total Removal Rate</th>
<th>Removal within 24 hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>1,704</td>
<td>362 (21.2%)</td>
<td>76.4%</td>
</tr>
<tr>
<td>Google (YouTube)</td>
<td>241,827</td>
<td>58,297 (27.1%)</td>
<td>93.0%</td>
</tr>
<tr>
<td>Twitter</td>
<td>264,818</td>
<td>28,645 (10.8%)</td>
<td>93.8%</td>
</tr>
<tr>
<td>Change.org</td>
<td>1,257</td>
<td>332 (26.4%)</td>
<td>92.7%</td>
</tr>
</tbody>
</table>

Source: Echikson & Knodt 2018

Community Guideline enforcement versus NetzDG statutes (Source: Google)
In this light, it may be that NetzDG’s most important effect was to ensure swifter and more consistent removal of content within Germany under the companies’ community guidelines.

The transparency reports also break down complaints by content type. The data (see Appendix) show that cases involving hate speech and defamation/insult were the most common. For Google, complaints related to hate speech and political extremism were most common (75,892 items), followed by defamation or insult (45,190 items), and sexual content (27,308 items). Facebook and Twitter break down their data differently than Google, making direct comparison difficult. Google aggregates data for comparable offences, such as “insult” and “defamation” or “incitement to hatred” and “propaganda for unconstitutional organizations,” whereas Facebook and Twitter list the data for each specific criminal offence separately. Furthermore, Facebook and Twitter’s data focus on the number of complaints, whereas Google’s focus on the number of content items (these do not correspond since one complaint may refer to multiple content items). Still, it is clear that insult-related offences were the most common on Facebook; with 460 complaints, insult is cited in over half of all complaints. Twitter provides the most detailed overview, since it also breaks down compliance rates for individual content types. Its data also show incitement to hatred (Volksverhetzung) leading with 82,095 complaints, followed by insult (Beleidigung), with 75,925.

The major platforms consulted outside counsel with comparable frequency (for fairly few cases): 40 cases for Google, 54 for Facebook, and 107 for Twitter. Each of these platforms was also a member of the self-regulatory advisory body on protecting youth online, Freiwillige Selbstkontrolle Multimedia-Dienstanbieter e.V. (FSM), though the platforms do not seem to have consulted FSM up to this point. FSM and eco, another advisory body, have hotlines for consumer complaints, which they then forward to the companies for review. Google provides feedback about its decisions back to the hotline.

The transparency reports also outline the companies’ procedures for notifying uploaders and submitters. They follow the same pattern: submitters receive a confirmation once the complaint is received, and another update once a decision has been made. Uploaders are notified only if and when a content removal decision is taken; they are not notified when a complaint is received about them, or when such complaints are dismissed.

Google’s report provides another interesting data point: interactions with uploaders and submitters. In some cases, the platform might require additional information from the uploader to assess whether an alleged falsehood is indeed untrue (e.g., for defamation cases). Google received 2,380 incomplete requests, for which it needed more information from the requester. Remarkably, Google seems to have reached out to uploaders zero times, ostensibly because “the majority of legal NetzDG complaints are unsubstantiated (even after YouTube explicitly asks for further information).” Twitter and Facebook do not address this issue in their reports.

A second round of reports was released in late January 2019. They do not reveal any major changes. The three major platforms all registered decreases in their total number of complaints received. Google & Twitter posted relatively minor changes (15% or less), whereas Facebook, already an outlier, declined more significantly. Removal rates also appear to have dropped for most types of complaints. Perhaps the most significant changes occurred at Twitter, where the number of complaints dropped by over 50% for cases involving child pornography and certain forms of hate speech (Volksverhetzung).
Overall, the inconsistencies among companies’ complaint forms make comparison of these numbers difficult. A problematic side effect of the reports could be that the number of takedowns or number of complaints become a metric to measure the law’s efficacy; these takedowns might be ineffective or even counterproductive in combating the overall prevalence of hate speech. The law’s actual impacts on hate speech may be difficult to prove empirically, since this complex phenomenon is influenced by countless other factors as well, including political, cultural, demographic, and economic shifts. The Federal Office for the Protection of the Constitution claims that NetzDG is hindering recruitment efforts by far-right extremists, since it has led to the shutdown of several influential Facebook pages and forced their organizers to move to other less visible platforms. However, the office has not published data to support its claim. Furthermore, these same censorship measures may well have other unintended consequences for counter-extremism, such as radicalization of the censored parties and interference with counter-speech and surveillance efforts. In short, it will require much more research — and greater access to data — to determine whether NetzDG is achieving its aim, and whether any benefits outweigh the harms to free speech.

Outlooks and Next Steps

There is little evidence that the data published in NetzDG’s transparency reports has changed anyone’s mind about the law. Opinion remains divided, along the lines of deeply held beliefs about the constitutional merits of criminal prohibitions in areas like hate speech and the role of nation-states in content regulation. Beyond repeal or narrowing the scope of the law to remove some of the categories deemed unlawful speech, NetzDG suggests some general lessons for policymakers about the challenges of regulating social media. The recommendations below focus specifically on NetzDG, though the High Level Working Group will release further papers over the next few months that examine some of these suggestions in a broader, transatlantic context.

Transparency and Research

A major challenge is that we need further research on the impact of NetzDG, such as potential chilling effects, but researchers do not have the data to conduct that research. Takedown metrics are problematic as a measure of success because they may simply encourage over-removal. There are multiple fundamental empirical questions that require further investigation. First, how do moderation practices affect the underlying problems and harms? Second, how does a law like NetzDG affect moderation practices and what metrics would enable us to measure that effect? Third, what are the broader societal effects or unintended consequences of such laws?

NetzDG’s transparency requirements offer a starting point to enable more robust research. Transparency is the one part of the law that has received almost universal support. Some of the most useful aspects of the transparency reports were about companies’ procedures, rather than the raw numbers of complaints or takedowns. The transparency reports showed that NetzDG is more a “community guidelines enforcement law” than anything else. Future transparency reports might contain more details on how companies train moderation staff, operate appeals mechanisms, and conduct quality control.
Still, NetzDG’s transparency mechanisms show room for improvement. One common complaint is the lack of standardized reporting formats; this hinders direct comparisons between different reports. Another area of improvement would be to offer more granular data. For instance, the reports specify compliance rates and the number of takedowns by content type, but they do not always break down compliance rates by content type. In other words, it is difficult to assess whether platforms were more likely to comply with hate speech claims than defamation claims.

A more fundamental complaint is that the reports offer little insight into individual cases, and make it hard for third parties to evaluate the merits of platform’s decisions. For this reason, the Green Party has proposed to develop a “Clearing House” (Clearingstelle) which would publish complaints from users who disagreed with platforms’ handling of their NetzDG cases.

A more ambitious approach would be for the German government to encourage and support the creation of research repositories that would combine data from multiple platforms. The repository could start by collecting all deleted content or all complaints under NetzDG. Both platforms and enforcement authorities within governments could be more transparent in this regard. German ministries could lead the way by making transparent their complaints bodies’ submissions. The repository could serve multiple purposes simultaneously. First, it could facilitate appeals processes. Second, it could provide data to researchers and more broadly enable regulators, researchers, or other watchdogs to review platform decisions. These types of disclosure could reveal important information about whether and how the law harms freedom of expression. Eventually, the repository might also expand beyond NetzDG.

There are already some suggestions about how to create such a repository. U.S. Senator Mark Warner, for instance, suggested a Public Interest Data Access Bill in August 2018. Any such repository would have to address social media companies’ legitimate legal and business concerns about sharing proprietary information or users’ private data. This could follow similar access procedures to researching confidential medical or statistical records, something that many governments already facilitate.

Due Process and Design Structure

Within the constraints of the current NetzDG, there are several ways to improve users’ rights and ensure due process both on platforms and more generally. First, NetzDG highlights the importance of design thinking for user-facing interventions, such as complaint mechanisms. YouTube and Twitter’s user-friendly implementations of NetzDG complaint forms led to significantly higher uptake than Facebook’s. The German Green Party, the Counter Extremism Project, and the Center for European Policy Studies are currently advocating for additional rules about accessibility in NetzDG complaint forms. If implemented, such rules could be an interesting case study in creating tools that are not just accessible, but visible and easy to use.

Second, NetzDG offers little recourse to uploaders who believe their content has been wrongfully deleted. The Green Party proposes a reinstatement procedure (Wiedereinstellungsverfahren), where uploaders could appeal to the platform to have their content reinstated. Others have suggested an appeals procedure before an independent oversight body or a judge.
A third way to safeguard the freedom of expression would be to improve the notification rights of uploaders; NetzDG’s transparency reports show that none of the major platforms notifies uploaders about possible complaints submitted about their content. Only the complaint issuer is informed throughout the process, and the uploader who is the subject of the complaint is only informed if and when the platform decides to remove their content. The current system only allows for uploaders to appeal after their content is taken down. It may be worth considering whether to notify uploaders immediately and not to wait until a decision is reached to enable those people to defend their post as lawful. Improving notice rights could enhance procedural fairness, and adjust the playing field in favor of free speech.

Multi-Stakeholder Relationships

More broadly, the question remains how and whether outside stakeholders can and should be involved in platform content moderation processes or their regulation. Like the European Commission’s Code of Conduct on Hate Speech, the NetzDG law itself was drawn up in 2017 with little to no input from civil society organizations. A law is also a rather static instrument for addressing such a swiftly changing ecosystem.

Beyond the specifics of NetzDG, the law also suggests wider lessons about regulation and platforms. Platforms may consider how better to contribute to a positive framework of cooperation among social media companies, governments, and civil society organizations. Germany’s approach has seemed to illustrate that, currently, the only way countries outside the U.S. receive sustained attention from social media companies is if they are a massive market (like China or the European Union) or journalists uncover significant human rights violations or they threaten companies with significant fines. Real financial liability commanded platform companies’ attention. Germany had tried a voluntary compliance system with the companies since 2015 but found it ineffective. The German government chose the path of law only after it deemed the companies insufficiently compliant. A government study in early 2017 found that YouTube had deleted 90% of criminal content, Facebook 39%, and Twitter only 1% within the requested 24 hours (though there are broader questions from a freedom of expression perspective about whether takedown compliance is an appropriate metric). Since the introduction of the law, transparency reports indicate that compliance rates are far higher.

There are many other potential approaches to redefine the relationships between platforms, governments, civil society, and users. One is to design robust transparency mechanisms that enable research on the pressing questions about social media and their broader societal effects before governments undertake any regulation. Another approach is the creation of independent appeals bodies for takedown decisions. A third is to consider rapid-response judicial mechanisms to adjudicate complaints. A fourth is the creation of Social Media Councils that regularly convene platforms, government, and civil society organizations to share information and debate possible new approaches.

Overall, it is hard to predict and measure the full effects of legal policies. The narrow focus on the number of complaints and narrow, problematic categories of illegal speech tell us little about any potential larger effects on Germany’s information ecosystem or political discourse. German politicians drew lessons from history to try to protect democracy by curtailing free speech. In the long run, however, they must be careful not to undermine the freedoms embedded in the political system that they seek to protect.
Appendix: Links to transparency reports

Google:

- Transparency Report: Removals under the Network Enforcement Law

Facebook:

- Facebook NetzDG Transparency Report, July 2018
- Facebook NetzDG Transparency Report, January 2019

Twitter (NB: Only available in German):

- Twitter Netzwerkdurchsetzungsgesetzbericht: Januar-Juni 2018
- Twitter Netzwerkdurchsetzungsgesetzbericht: Juli-Dezember 2018

Notes

1 Dr. Heidi Tworek (heidi.tworek@ubc.ca) is assistant professor at the University of British Columbia as well as a non-resident fellow at the German Marshall Fund of the United States and the Canadian Global Affairs Institute. Her book, News from Germany: The Competition to Control World Communications, 1900-1945, was published in March 2019 by Harvard University Press. Paddy Leerssen is a doctoral candidate at the Institute for Information Law (IViR) at the University of Amsterdam and a non-resident fellow at the Stanford Center for Internet & Society.

2 https://www.buzzfeednews.com/article/lesterfeder/fraud-in-the-europe-unclear...

3 e.g. the platform’s content removal procedures; the number of complaints received and their source (users or authorities); the number of content removal decisions based on these complaints as well as the reason for removal; and their personnel and other resources dedicated to content moderation.


7 Critics include local digital rights organizations such as Netzpolitik and Digitale Gesellschaft, global international groups such as CDT, Access Now, Article 19, and European Digital Rights (EDRi), Reporters Without Borders, and Human Rights Watch. One prominent German academic critic is Wolfgang Schulz, Director of the Hans Bredow Institute for Media Research. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3216572

8 http://deklaration-fuer-meinungsfreiheit.de/en/

9 http://globalnetworkinitiative.org/proposed-german-legislation-threatens-free-expression-around-the-world/

10 http://deklaration-fuer-meinungsfreiheit.de/en/


15 “Die Meinungsfreiheit hat auch Grenzen,” Medienpolitik.net, 09.01.17.


17 https://www.heise.de/tp/features/Russland-kopiert-deutches-Netzwerkdurchsetzungsgesetz-3773642.html


Opposing:

Supporting:
https://www.n-tv.de/politik/NetzDG-erschwert-Rechten-Rekrutierung-article20597308.html
https://www.educsu.de/themen/innen-recht-sport-und-ehrenamt/netzwerkdurchsetzungsgesetz-wirkt
https://www.spdfraktion.de/presse/pressemitteilungen/netzdg-wirkt

20) For takedowns unrelated to NetzDG, many platforms offer additional transparency reports on a voluntary basis.
22) See Appendix for complete data. Source: https://transparencyreport.google.com/netzdg/overview?hl=en
24) See Appendix. (Note: these figures for Twitter are the sum of “Beschwerden von Nutzern” + “Beschwerden von Beschwerdestellen”).
25) See Appendix. (Note: these figures for Twitter are the sum of “Beschwerden von Nutzern” + “Beschwerden von Beschwerdestellen”).
26) See Appendix for complete data. Source: https://transparencyreport.google.com/netzdg/overview?hl=en
28) See Appendix. (Note: these figures for Twitter are the sum of “Beschwerden von Nutzern” + “Beschwerden von Beschwerdestellen”).
29) See Appendix for complete data. Source: https://transparencyreport.google.com/netzdg/overview?hl=en
30) See Appendix for complete data. Source: https://transparencyreport.google.com/netzdg/overview?hl=en
31) See Appendix for complete data. Source: https://transparencyreport.google.com/netzdg/overview?hl=en