

Science and Technology Innovation Program



Freedom and Fakes:

A Comparative Exploration of Countering Disinformation and Protecting Free Expression

Authors

Nina Jankowicz

Shannon Pierson

December 2020





Executive Summary

In the United States in 2020, there is no clear plan of action to combat disinformation at a national level. No single body has fully acknowledged their role in the cleanup of the online environment, and disinformation has been politicized. While the American response to online disinformation has been marked by partisanship and gridlock, at least 50 other nations have taken action against it, often to the detriment of human rights and democratic standards.

In addition to briefly exploring the existing human rights-based framework for the protections of free speech, this paper compares the development and, where applicable, implementation of efforts to regulate social media platforms in response to disinformation. The four case studies—in Germany, Brazil, Singapore and Ukraine—outlined in this paper provide some basic guidelines and cautionary tales observed from democracies that have already begun their regulatory journeys.

Even governments with benign motivations behind their counter-disinformation regulations, such as Germany and Brazil, have run up against the dangerous normalization of restrictions on speech. Germany's NetzDG law has been cited in authoritarian countries as fodder for their own crackdowns on freedom of expression.¹ Brazil, the country that passed the world's first online bill of civil rights, flirted with passing harsh fines for the spread of disinformation. Two governments—in Singapore and Ukraine—have used the threat of disinformation to excuse anti-democratic actions against political opposition and dissenting opinions. This cross-continental, cross-cultural, and cross-contextual examination distills five guiding principles for any regulation aimed at countering online disinformation while protecting democratic ideals:

First, when defining what speech will be deemed harmful in counter disinformation regulation, precision is key. In both the Singaporean and Ukrainian cases, overbroad definitions contributed to fears that laws drafted under the guise of protecting national security and freedom of opinion would rather contribute to a chilling effect on free speech as well as empower the government to quash criticism.

Second, the case studies demonstrate the importance of mandating transparency and oversight—ideally from an apolitical, expert body—as part of any regulatory framework.

Third, the importance of establishing an independent body to enforce and adjudicate counter disinformation law,

Germany's NetzDG law has been cited in authoritarian countries as fodder for their own crackdowns on freedom of expression.

1 Jacob Mchangama and Joelle Fiss, "The Digital Berlin Wall: How Germany (Accidentally) Created a Prototype for Global Online Censorship," justitia.int.org, Justitia, October 2020.



ideally drawing on the existing structures and expertise of judicial authorities, cannot be understated. Any body overseeing these laws should be expert, politically insulated, and utilize the independent judiciary for adjudication.

Fourth, users must have recourse above the platform level in order to dispute takedowns of their content. They must be informed of its removal as well as of the opportunities they have to challenge the decision. These appeals should move through the aforementioned independent, expert commissions charged with overseeing and enforcing social media regulation.

Finally, the development of any social media regulation should be pursued in consultation with civil society and other democratic partners, and with the use of existing legal frameworks. The United States can and should learn from these case studies when considering adaptation of intermediary liability laws or seeking to compel social media platforms to take more action against harmful content online. The United States has a duty as the democracy in which most of the world's most popular social media platforms were founded and nurtured to set the standards for democratic, human rights-based regulation. Disinformation's ultimate victim is democracy, and it is up to democratic governments to protect the principles on which they run in crafting responses to the threat.





Introduction

“We do not want to be the arbiters of truth.”

This refrain—first voiced by Facebook’s Mark Zuckerberg—encapsulates the debate on regulating the social media space to counter disinformation. In an increasingly polluted online ecosystem, what structures are obligated to preserve democratic discourse, freedom of speech, or human rights? Who should stand up for the truth?

In the United States, there is no clear plan of action to combat disinformation at a national level. No single body has fully acknowledged their role in the cleanup of the online environment, which has left critical first legislative steps toward regulation unfulfilled. Further, the issue of disinformation has become politicized. Social media platforms, under pressure from policymakers and consumers, have settled into a mostly reactive policy process, driven by the latest events and scandals, rather than anticipation of future threats and exploits. These reluctant policies are unevenly enforced with little transparency into the platforms’ decision-making processes, sometimes with opaque exceptions.² Meanwhile, the Executive Branch has entered the fray, issuing executive orders against social media companies to correct what it perceives as bias against conservative voices in existing content moderation policies or national security threats; it is possible that the orders “[will not] withstand judicial scrutiny.”³

While the American response to online disinformation has been marked by partisanship and gridlock, at least 50 other nations have taken action against it, often to the detriment of human rights and democratic standards.⁴ In addition to briefly exploring the existing human rights-based framework for the protections of free speech, this paper compares the development and, where applicable, implementation of four efforts—in Germany, Brazil, Singapore and Ukraine—to regulate social media platforms in response to disinformation. It is not meant to be prescriptive, but cautionary. These case studies were chosen because they encapsulate the tension democracies face between fighting disinformation and protecting freedom of expression. Two governments—in Ukraine and Singapore—have used the threat of disinformation to excuse anti-democratic actions against political opposition and dissenting opinions. Even governments with benign motivation behind their counter disinformation regulations, such as Germany and Brazil, have run up against the dangerous normalization of restrictions on speech; Germany’s NetzDG law has been cited in authoritarian countries as fodder for their own crackdowns on freedom of expression, and Brazil, the country that passed the world’s first online bill of civil rights, flirted with passing harsh fines for the spread of disinformation.

2 Olivia Solon, [“Sensitive to claims of bias, Facebook relaxed misinformation rules for conservative pages,”](#) NBCNews.com, NBC News, August 7, 2020.

3 Charles Duan and Jeffrey Westling, [“Will Trump’s Executive Order Harm Online Speech? It Already Did.”](#) Lawfareblog.com, Lawfare, June 1, 2020.

4 Daniel Funke and Daniela Flamini, [“A guide to anti-misinformation actions around the world,”](#) Poynter.org, Poynter. Accessed September 3, 2020.



The United States can and should learn from these case studies; it has a duty as the democracy in which most of the world's most popular social media platforms were founded and nurtured to set the standards for democratic, human rights-based regulation. Put simply: American inaction means online freedom of expression is under threat.

Existing Frameworks

Though the phenomenon of online disinformation is relatively new, the structures and principles for countering it while simultaneously protecting freedom of expression already exist. International institutions, together with key civil society and rights organizations, have drafted and confirmed these documents. The United Nations itself has “affirmed that offline rights apply equally online, but it is not always clear that the [social media] companies protect the rights of their users or that states give companies legal incentives to do so.”⁵ This section describes a selection of the key frameworks policymakers should consider when crafting social media regulation.

The governing article for all worldwide freedom of expression discussions is Article 19 of the United Nations’ International Covenant on Civil and Political Rights (ICCPR). Ratified by 170 states, it affirms that “everyone shall have the right to hold opinions without interference,” and that “everyone shall have the right to freedom of expression.” Restrictions against these rights are only permissible “for the respect of the rights or reputations of others; [or] for the protection of national security, or of public order [...], or of public health or morals.”⁶ Article 19 also succinctly delineates the standards states must meet to enact restrictions on freedom of expression. Limitations may only be enacted if they are: legal, including a precise definition; are reached through democratic consultation; are implemented with the oversight of “independent judicial authorities;” demonstrate necessity and proportionality, as evidenced by the state itself; and are legitimate, protecting “only those interests enumerated in Article 19 [...]: the rights or reputations of others, national security or public order, or public health or morals.”⁷

Platforms, too, have a roadmap for their conduct towards their users in the United Nations’ Guiding Principles on Business and Human Rights. Though the Principles are non-binding, they establish a standard of conduct which should guide any private company. These include the expectation that companies will “avoid infringing on the human rights of others and should address adverse human rights impacts with which they are involved,” “conduct due diligence that identifies, addresses, and accounts for actual and potential human rights impacts of their activities,” consult with stakeholders about their effect on human rights, and allow users a course for remediation.^{8,9}

Social media regulation and oversight often addresses the concept of intermediary liability, or “the legal liability of intermediaries such as internet service providers, social networks, and search engines” for the content provided on their services.¹⁰ In the United States, this has manifested itself as scrutiny for Section 230 of the

5 David Kaye, “[Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression](#),” (A/HCR/38/35), ap.OHCHR.org, Office of the High Commissioner for Human Rights, April 4, 2018.

6 United Nations General Assembly, “[International Covenant on Civil and Political Rights](#),” UN.org, December 16, 1966.

7 Kaye, “[Report of the Special Rapporteur](#),” 2018.

8 UN General Assembly, “[Guiding Principles on Business and Human Rights](#),” OHCHR.org, Office of the High Commissioner for Human Rights, 2011.

9 Kaye, “[Report of the Special Rapporteur](#),” 2018.

10 Electronic Frontier Foundation, “[Manila Principles on Intermediary Liability](#),” manilaprinciples.org, Electronic Frontier Foundation, March 24, 2015.



Communications Decency Act, which reads: “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.” As disinformation in the United States has proliferated, presidential candidates have put CDA 230 in their sights, with both President Trump and former Vice President Joe Biden indicating they favor, at the very least, reimagining the law for the internet era, if not fully eliminating the intermediary liability exception for social media platforms. However, as this paper will explain, and as scholars and rights groups have underlined, increasing the liability of intermediaries has “an impact on users’ rights, including freedom of expression, freedom of association, and right to privacy.”¹¹ To address this, a group of civil society organizations developed the Manila Principles, a set of safeguards “with the aim of protecting freedom of expression and creating an enabling environment for innovation.”¹²

Based on the United Nations’ Universal Declaration on Human rights, the ICCPR, and the United Nation’s Guiding Principles on Business and Human Rights, the Manila Principles state:

1. *Intermediaries should be shielded by law from liability for third party content;*
2. *Content must not be required to be restricted without an order by a judicial authority*
3. *Requests for restrictions of content must be clear, be unambiguous, and follow due process*
4. *Laws and content restriction orders and practices must comply with the tests of necessity and proportionality*
5. *Laws and content restriction policies and practices must respect due process*
6. *Transparency and accountability must be built into laws and content restriction policies and practices*

The principles place a special emphasis on the necessity of the involvement of independent judicial authorities in intermediary liability cases, as well as the requirement that the ICCPR’s standards for clarity, necessity, and proportionality in requests for restrictions and in potential consequences. It also implores intermediaries themselves to respect human rights when drafting and enforcing their content moderation policies.¹³

International organizations have also reacted to the increased prevalence of disinformation itself, adopting the “Joint Declaration on Freedom of Expression and ‘Fake News,’ Disinformation and Propaganda” in 2017. Drafted by freedom of expression specialists from the United Nations, the Organization for Security and Cooperation in Europe, the Organization of American States, and the African Commission on Human and Peoples’ Rights, together with two civil society groups, Article 19 and the Center for Law and Democracy, the document draws together the standards described above in the specific context of online disinformation. As David Kaye, the former United Nations Special Rapporteur on Freedom of Opinion and Expression and one of the drafters of the document describes, speech experts were “worried that a rush to prohibit ‘fake news’ rather than finding a combination of technical, educational, and legal solutions, would almost certainly undermine the protections for free expression over the long term.”¹⁴ Kaye and his colleagues noted that “there was no specific legal

11 Ibid.

12 Ibid.

13 Ibid.

14 David Kaye, *Speech Police: The Global Struggle to Govern the Internet*, (New York: Columbia Global Reports, 2017), 95.



framework” for countering disinformation apart from the treaties above and “interpretations by courts and other monitoring bodies, so we decided to create one ourselves.” The result, he writes, is “a kind of soft law framework to deal with disinformation.”

In its preambles, the Joint Declaration expresses alarm “at instances in which public authorities denigrate, intimidate, and threaten the media, including by stating that the media is ‘the opposition’ or is ‘lying’ and has a hidden political agenda.”¹⁵ It further “deplor[es] attempts by some governments to suppress dissent and control public communications” by interfering in the practice of journalism through measures such as denying accreditation to critical outlets, using digital technologies to suppress information, and pressuring intermediaries to restrict politically sensitive content.¹⁶ The Declaration also singles out intermediaries themselves for their use of opaque automated processes “which fail to respect minimum due process standards” in content moderation.

The Declaration then establishes specific standards for states responding to disinformation. “Vague definitions and ambiguous ideas, including ‘false news’ or ‘non-objective information,’ are incompatible with international standards...and should be abolished,” it reads. It also underlines the responsibility state actors have to “not make, sponsor, encourage, or further disseminate statements which they know or reasonably should know to be false (disinformation) or which demonstrate a reckless disregard for verifiable information (propaganda).”¹⁷ Similar principles are laid out for intermediaries and journalists, in line with previous international agreements. Finally, it “urge[s] platforms to be transparent about their rules and support technical solutions to disinformation that would strengthen user autonomy and individual ability to separate fact from fiction.”¹⁸

Despite this fairly wide body of standards related to the protection of online speech and fight against disinformation, no government has yet developed regulations that fully respect these principles.

Despite this fairly wide body of standards related to the protection of online speech and fight against disinformation, no government has yet developed regulations that fully respect these principles. In fact, as this paper will demonstrate, some have used the threat of disinformation to justify draconian restrictions on freedom of expression and human rights that, when implemented, will pose a greater threat to democracy than the foreign disinformation that they were drafted to counter.

These existing standards also remind states and platforms of a critical component of any work countering disinformation, whether in the United States or any other nation; governments and intermediaries cannot credibly hold rogue actors like Russia and China accountable for violating principles of freedom of expression, truth, transparency, and trust if they do not project and adhere to and project them themselves.

15 OSCE Representative on Freedom of the Media, “[Joint declaration on freedom of expression and “fake news,” disinformation and propaganda](#),” OSCE.org, Organization for Security and Co-operation in Europe, March, 3 2017.

16 Ibid.

17 Ibid.

18 Kaye, *Speech Police*, 96.

Germany: NetzDG as Cautionary Standard for Social Media Regulation

In the heat of the European migrant crisis in 2015, Chancellor Angela Merkel made a landmark decision to open Germany's borders and admit over 1.3 million refugees into the country. The rapid influx of refugees sparked a rise in xenophobia, far-right extremism, and violence against immigrants and Muslims in Germany. The Bundestag and the German domestic intelligence agency's Federal Criminal Police Office (Bundeskriminalamt, or BKA), asserted that social media companies are partially responsible for this rise in radical extremism for permitting and amplifying hate speech on their platforms.¹⁹

Germany has some of the strictest speech laws among democratic nations, a product of the country's reconciliation with its World War II history.²⁰ "Incitement to hatred," "public incitement to crime," "defamation of religions, religious and ideological associations," and "dissemination of propaganda material of unconstitutional organizations" are some forms of speech illegal under the German Criminal Code.²¹ In summer 2016, federal police cracked down on online hate crime by conducting house raids across the country on individuals suspected to be behind xenophobic, anti-Semitic, and far-right extremist posts on social media.^{22, 23}

In 2016-2017, then-Minister of Justice and Consumer Protection Heiko Maas initiated a joint task force on illegal hate speech on social media platforms.²⁴ The "Task Force Against Illegal Online Hate Speech" convened Facebook, Google, Twitter, Microsoft, and German anti-hate advocacy groups to address hate speech circulating on the platforms. The task force quickly produced voluntary commitments from the social media companies toward a common Code of Conduct. The code encouraged platforms to make an effort to respond to reports of hate speech within 24 hours and make reporting easier for users,²⁵ but drew criticism from digital rights advocacy organizations for its call for mass removal of content that may not be illegal in countries outside of Germany.^{26, 27}

Maas pledged to hold platforms to their commitment, and commissioned an independent audit by Jugendschutz, a youth digital rights protection organization, to assess if social media companies kept their promises. The study found that platforms routinely failed to remove reported content within the 24-hour timeframe for deleting it. Twitter removed zero percent of reported illegal content within 24 hours, Facebook removed 31 percent, and YouTube removed 92 percent. Overall, Twitter removed just one percent of reported illegal content. Facebook

-
- 19 Ruth Bender, "German Police Carry Out Nationwide Crackdown on Internet Hate Speech," WSJ.com, *The Wall Street Journal*, July 13, 2016.
- 20 Flemming Rose, "Germany's Attack on Free Speech," Cato.org, Cato Institute, May 30, 2017.
- 21 Federal Parliament of Germany, "Network Enforcement Act (Netzdurchsetzungsgesetz NetzDG)," germanlawarchive.iuscomp.org, German Law Archive, September 1, 2017.
- 22 Bender, "German Police Carry Out National Crackdown," 2016.
- 23 Amar Toor, "German police raid homes over Facebook hate speech," TheVerge.com, The Verge, July 13, 2016.
- 24 "German justice minister to set up task force on Internet hate speech," DW.com, Deutsche Welle, September 14, 2015.
- 25 Task Force against illegal online hate speech, "Together against Hate Speech: Ways to tackle hateful content proposed by the Task Force against illegal online hate speech," EDRI.org, European Digital Rights, December 15, 2015.
- 26 AJ Dellinger, "New hate speech code of conduct wins support from Facebook, Twitter, Google, and Microsoft," DailyDot.com, The Daily Dot, June 1, 2016.
- 27 EDRI, "Guide to the Code of Conduct on Hate Speech," EDRI.org, EDRI, June 3, 2016.



removed 39 percent, and YouTube removed 82 percent.²⁸ Maas argued that the results conveyed that “the networks don’t take the complaints of their own users seriously enough.”²⁹

While Facebook disputed the results on the grounds that the company performed better in a study conducted by The German Association for Voluntary Self-Regulation of Digital Media Service Providers, the then-Minister of Justice leveraged the Jugendschutz report to make aspects of the code mandatory.³⁰ Maas introduced the “Act to Improve Enforcement of the Law in Social Networks,” the bill which would become known as NetzDG.

NetzDG obliges large social media companies with over two million registered users in Germany to enforce 21 statutes in the German Criminal Code related to hate speech in the online spaces they manage and delete any illegal content.³¹ Platforms must review and remove “manifestly unlawful” content within 24 hours of receiving a complaint, and have up to seven days to review cases in which legality is unclear.³² Thus, NetzDG does not create new categories of illegal content, but applies pre-existing German criminal hate speech law on online social spaces and pushes the responsibility onto the platforms to enforce those laws online.³³ Platforms serve as deputies by making judgement calls on whether content is legal.

The law also requires platforms to publish public transparency reports on their content moderation practices if they receive more than one hundred NetzDG complaints annually.³⁴ Reports must include details on the criteria used to decide whether to remove or block content, a timeline for action taken on complaints, as well as the number of complaints issued and that resulted in deletion or blocking, broken down by users and complaints bodies and reason for complaint.

NetzDG levies fines up to 50 million EUR (\$59,210,250) for platforms that fail to remove content within the 24-hour deadline or provide sufficient transparency reports.^{35, 36} However, fines are reserved only for “systematic” violators.

There are also several key liability caveats for platforms. They are not liable for content removal mistakes or over-removal. Moreover, platforms are not obliged to provide users a course to dispute content removal decisions. However, if removal is related to the falsity of the content, users may respond. The bill’s authors reasoned that removing hate speech from platforms facilitates free speech and expression on the platform by ensuring a “safe online environment” that permits free exchange of ideas.³⁷ Maas argued that hate speech infringes on citizens’ freedom of expression and opinion. “Incitements to murder, threats, insults, and incitement of the masses

28 Jugendschutz.net, “[Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter](#),” Jugendschutz.net, March 2017.

29 Mike Butcher, “[Unless online giants stop the abuse of free speech, democracy and innovation is threatened](#),” TechCrunch.com, TechCrunch, March 20, 2017.

30 Butcher, “[Unless online giants stop the abuse of free speech](#),” 2017.

31 Imara McMillan, *Enforcement Through the Network: The Network Enforcement Act and Article 10 of the European Convention on Human Rights*, CJIL.uchicago.edu, Chicago Journal of International Law, January 1, 2019.

32 Federal Parliament of Germany, 2017.

33 Heidi Tworek and Paddy Leerssen, “[An Analysis of Germany’s NetzDG Law](#),” IVIR.nl, Transatlantic Working Group, April 15, 2019.

34 Federal Parliament of Germany, “[Network Enforcement Act \(Netzdurchsetzungsgesetz NetzDG\)](#),” 2017.

35 Danielle Keats Citron, “[Extremist Speech Compelled Conformity, and Censorship Creep](#),” NDlawreview.org, Notre Dame Law Review, April 21, 2018.

36 Kate Brady, “[German justice minister calls for hefty fines to combat online hate speech](#),” DW.com, Deutsche Welle, April 6, 2017.

37 William Echikson and Olivia Knodt, “[Germany’s NetzDG: A Key Test for Combatting Online Hate](#),” SSRN.com, CEPS Policy Insight, November 2018.



or Auschwitz lies are not an expression of freedom of opinion but rather attacks on the freedom of opinion of others," he said.³⁸

Passage and Enforcement

NetzDG moved through the Bundestag rapidly. At a Parliamentary hearing on the bill, the Bundestag consulted ten civil society experts. Eight of ten experts expressed serious concern and criticism about the bill, and five of ten described the legislation as unconstitutional.³⁹ After some revisions, the bill passed just one month after its initial introduction. When NetzDG came into force, it was the first law to establish a framework that required platforms to remove content in a rapid timeframe.⁴⁰

The passage of NetzDG was also notable for its transparency requirements. Though it aimed to increase oversight and accountability of the platforms, the law lacks sufficient information for regulators and researchers to perform this function. NetzDG permits platforms to create their own company-specific transparency reporting formulas, which produce metrics that are difficult to compare. Consequently, there is no way to assess if the platforms are all enforcing the law in the same way without uniform data.

Further, though NetzDG was designed to obligate social media companies to apply German law online for users located in Germany, in practice, platforms may have imposed German law online for users globally by scaling up enforcement of their globally applicable community guidelines to remove content illegal under NetzDG and evade fines. Facebook's vice president for global policy solutions Richard Allan wrote in a 2018 blog post: "We have taken a very careful look at the German law. That's why we are convinced that the overwhelming majority of content considered hate speech in Germany would be removed if it were examined to see whether it violates our community standards."⁴¹ Google and Twitter assert that their community guidelines cover NetzDG rules around hate speech, although research suggests a severe lack of enforcement.

The platforms all employ a two-step screening process for content reported under NetzDG, in which they review complaints against their community guidelines before their compliance with NetzDG.^{42, 43} In response to Germany's new law, platforms applied greater enforcement of their broadened community guidelines to evade fines. Platforms are incentivized to over-delete on a mass scale, given both the risk of fines and the existing caveats in the law, which excuse platforms from liability for content removal mistakes. They have legal cover to remove legal content or partake in over-blocking without obligation to provide users a course to dispute content removal decisions.

Thus far, Germany has issued only one fine under NetzDG. In July 2019, Germany's Federal Office of Justice

38 Guardian service, "Tough new German law puts tech firms and free speech in Spotlight," IrishTimes.com, *The Irish Times*, January 5, 2018.

39 EDRI, "Germany: Will 30 June be the day populism killed free speech?," EDRI.org, EDRI, June 29, 2017.

40 Raphael Markert, "'Da ist Deutschland nicht ganz unschuldig,'" Sueddeutsche.de, Sueddeutsche Zeitung, March 12, 2020.

41 Hia Datta, "Facebook upholds Germany's hate-speech law by taking down several objectionable posts," TechGenyz.com, TechGenyz, July 27, 2018, 2020.

42 Amélie Heldt, "Reading between the lines and the numbers: an analysis of the first NetzDG reports," Policyreview.info, Internet Policy Review, June 12, 2019.

43 Tworek and Leerssen, "An Analysis of Germany's NetzDG Law," 2019.



fined Facebook 2 million EUR (\$2,367,540) for underreporting content complaints and violating its transparency obligations. According to the authority, “the published information does not provide a conclusive, transparent picture of the organization and the processes involved in handling complaints about illegal content.”⁴⁴ Facebook only counted NetzDG complaints and omitted complaints issued for violations of Facebook’s own community standards in its first NetzDG transparency report, resulting in far fewer reported complaints compared to Twitter and Google/YouTube. From January 1 to June 30 2018, Facebook reported only 1,704 complaints, while Google and Twitter reported upwards of 200,000. Facebook removed 2.5 million posts under its community standards for hate speech between January and June 2018, but removed only 362 posts under German law during the same timeframe.^{45, 46} The significant disparity in numbers is largely attributed to German users’ lack of ease in accessing Facebook’s NetzDG reporting mechanisms, as well as users’ lack of awareness of the availability of such reporting mechanisms. Facebook is appealing the decision.⁴⁷

While NetzDG achieved the Bundestag’s goal of forcing platforms to remove illegal content more quickly and consistently, it is difficult to draw conclusions from the transparency reports when their methodology varies and their respective community guidelines—which apply first—all differ.⁴⁸ More detailed reporting requirements, frameworks and quality standards are needed for researchers, regulators, and policymakers to obtain valuable information in transparency reports which they can use to assess enforcement trends across the platforms. As evidenced by Facebook’s decision to omit NetzDG complaints that fall under their Community Guidelines, platforms may not be forthcoming without quality standards for reports. It is difficult to analyze the meaning of these numbers and the reasons for their differences across reports without more access to more uniform data.

NetzDG demonstrates that social media regulation laws have effects that ripple out beyond its borders, as platforms may broaden the scope of their community guidelines to include its stipulations to avoid liability, duck fines, and respond at a global scale.

NetzDG demonstrates that social media regulation laws have effects that ripple out beyond its borders, as platforms may broaden the scope of their community guidelines to include its stipulations to avoid liability, duck fines, and respond at a global scale. Moreover, removing a piece of content globally rather than at a country level is simpler for platforms, especially given that content may be or later become illegal in other countries as well. In effect, platforms’ community guidelines may gradually become more conservative, constricting free speech and expression as a result.⁴⁹

In addition to concerns about platform enforcement, NetzDG and its amendments raise concerns about user privacy, free speech, and due process. The law does not require platforms to inform users if their post is

44 “Deutsche Behörde verhängt Millionenstrafe gegen Facebook,” Zeit.de, Zeit Online, July 2, 2019.

45 “Facebook deletes hundreds of posts under German hate-speech law,” July 27, 2018.

46 Sarah Perez, “Facebook’s new transparency report now includes data on takedowns of ‘bad’ content, including hate speech,” TechCrunch.com, TechCrunch, May 15, 2018.

47 The Associated Press, “Germany: Facebook to appeal fine under hate speech law,” ABCNew.go.com, ABC News, July 19, 2019.

48 “Deutsche Behörde verhängt Millionenstrafe gegen Facebook,” 2019.

49 “Facebook deletes hundreds of posts under German hate-speech law,” Reuters.com, Reuters, July 27, 2018.



reported by another user or to provide a path of recourse for users to challenge content platform moderation decisions. Failure to notify users and the lack of opportunity for appeal is problematic when reporting translates to law enforcement action in the real world. In recent years the Federal Criminal Police began cracking down on hate speech and “incitement to crime” on social media by launching home raids, confiscation of devices, and interrogations of users suspected of posting illegal content online.^{50, 51, 52} Further, in June 2020, the Bundestag passed the Law to Combat Right-Wing Extremism and Hate Crime, which contained a provision that amended NetzDG to require platforms to forward all content flagged with a NetzDG complaint directly to German authorities.⁵³ Authorities receive the content of posts and data about the user who posted it, including IP addresses or port numbers.⁵⁴ The objectives of the amendment were to begin prosecuting offenders of NetzDG and issuing offline punishments—not just blocking or deletion—to discourage and reduce hate speech online, as well as expedite the reporting timeline to police. Under this law, platforms are now obliged to send user information to federal authorities before violation claims are judged to be unlawful by the companies. Users who post legal content that does not violate NetzDG may have their personal information sent to the BKA if a complaint is filed against them—compromising their privacy. Lawmakers have raised concerns that this would force platforms to open a backdoor for the German government to stockpile user data about suspicious behavior.⁵⁵

Criticism and Copycat Laws

When NetzDG came into force in 2017, it was an unprecedented approach to internet governance that made waves globally and gave the green light to democracies and autocracies alike to consider legislating over hate and false content online.⁵⁶ Internationally, NetzDG has served as a regulatory blueprint for countries establishing policy to reign in misinformation, so-called “fake news,” and hate speech on social media platforms. However, NetzDG has also been used by authoritarian states as a template for laws that limit and repress free expression on the internet.⁵⁷

As of November 2019, 13 countries and the European Union introduced or proposed intermediary liability legislation modeled after NetzDG’s regulations. Nine of the countries—Venezuela, Vietnam, Russia, Belarus, Honduras, Kenya, India, Singapore, Malaysia, the Philippines, France, the United Kingdom, and Australia—credited NetzDG as their inspiration or justification in the text of their legislation.⁵⁸

50 “Germany: Dozens of raids over online speech,” DW.com, Deutsche Welle, June 6, 2019.

51 Bender, “German Police Carry Out National Crackdown,” 2016.

52 David Shimer, “Germany Raids Homes of 36 People Accused of Hateful Postings Over Social Media,” NYTimes.com, *The New York Times*, June 20, 2017.

53 Bundesministerium der Justiz und für Verbraucherschutz, “Gesetz zur Bekämpfung des Rechtsextremismus und der Hasskriminalität,” BMJV.de, Bundesministerium der Justiz und für Verbraucherschutz, February 19, 2020

54 Philipp Grill, “German online hate speech report criticised for allowing ‘backdoor’ data collection,” Euractiv.com, Euractiv, June 19, 2020.

55 Ibid.

56 Markert, “Da ist Deutschland nicht ganz unschuldig,” 2020.

57 Jacob Mchangama and Joelle Fiss, “Germany’s Online Crackdowns Inspire the World’s Dictators,” ForeignPolicy.com, Foreign Policy, November 6, 2018.

58 Ibid.



During the Bundestag’s hearings on NetzDG in May 2017, Reporters Without Borders (RSF) CEO Christian Mihr testified and expressed his concerns that the law was “reminiscent of autocratic states” and would be easily abused. These concerns materialized less than two weeks after the law’s passage, in July 2017, when Russia introduced a bill that was a “copy-and-paste” version of Germany’s law and lifted direct passages from it.⁵⁹ In March 2019, Russia passed an anti-fake news law that outlaws “unreliable socially significant information” and online content that “displays obvious disrespect for society, the state, the official state symbols of the Russian Federation, the Constitution of the Russian Federation or the bodies exercising state power in the Russian Federation.”⁶⁰ This is a law that grants the Russian government the ability to silence government critics in online spaces, and NetzDG was cited in the law’s explanatory report.⁶¹ Though many laws in countries like Russia credit NetzDG as inspiration for their own free speech crackdowns, they go far beyond its regulatory scope. Malaysia’s law, now repealed over its misuse by the government against opposition parties, made it illegal for individuals to create or share fake news at the risk of steep fines and jail time. Venezuela has shortened the removal timeline for illegal content, mandating the removal of illegal content within six hours instead of 24.

NetzDG is a fundamentally imperfect law, but Germany’s robust protections for political rights and civil liberties equip the country to mitigate its adverse domestic effects.⁶² However, implementation of laws similar to NetzDG in states with fewer protections could result in severe consequences for users’ digital rights. Although Bundestag set out to safeguard German citizen’s right of free speech and expression through NetzDG by ensuring a “safe online environment,” it inadvertently weakened global internet freedom.⁶³

NetzDG should serve as an important case study for lawmakers considering adapting intermediary liability laws or seeking to compel social media platforms to take more action against harmful content online. It shows the importance of transparency and oversight, as well as the necessity for regulation to establish clear, uniform reporting standards that provide granular data that are comparable across services in service of such goals. It also shows the necessity of empowering users affected by the content moderation process, including by notifying them that content has been removed and offering opportunity for recourse. Most importantly, NetzDG shows the ripple effect that powerful democratic governments have when crafting social media regulation. Not only do such laws compel platforms to remove content at a global scale under their own guidelines, they may give authoritarian states the green light to pass copy-cat social media regulation to limit and discourage online discourse, leveraging democratic seal of approval on legislation but lacking the democratic safeguards to defend against abuse.

59 “Russian bill is copy-and-paste of Germany’s hate speech law,” RSF.org, Reporters Without Borders, July 19, 2017.

60 Oreste Pollicino, “Fundamental Rights as Bycatch – Russia’s Anti-Fake News Legislation,” *Verfassungsblog.de*, *Verfassungsblog*, March 28, 2019.

61 Maria Vasilyeva and Tom Balmforth, “Russia’s parliament backs new fines for insulting the state online,” *Reuters.com*, *Reuters*, March 13, 2019.

62 “Freedom in the World 2020: Germany,” *FreedomHouse.org*, Freedom House, 2020.

63 Federal Government of Germany, “Answers to the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression in regard to the Act to Improve Enforcement of the Law in Social Networks (Network Enforcement Act),” *OHCHR.org*, OHCHR, June 1, 2017.

Brazil: Crafting Policy Through Multi-Stakeholder Consultation

In 2014, Brazil passed its landmark Marco Civil da Internet (MCI/The Brazilian Internet Bill of Rights) which guarantees Brazilians privacy, freedom of speech, and net neutrality. It was drafted based on years of collaborative efforts between the government and civil society. “This law modifies the country’s constitution to give citizens, the government and organizations rights and responsibilities with regard to the Internet.”⁶⁴ Tim Berners-Lee, inventor of the World Wide Web, celebrated the Marco Civil as “a very good example of how governments can play a positive role in advancing web rights and keeping the web open.”⁶⁵ Where other countries either attempt to legalize or encourage the government-forced takedown of user-generated content on platforms, the Marco Civil expressly prohibits this government control of the internet. “The law prevents the government from taking down or regulating content, or exercising any pressure on online platforms, or internet service providers. Only the courts, by means of due process, and limited by clear legal thresholds, can actually take action regarding online content when it infringes on other rights.”⁶⁶ Thanks to the MCI, when lower courts issued injunctions requesting the removal of online content, including the blocking of popular messaging platform WhatsApp, higher courts immediately reversed the decisions in line with the new law. The law’s success was due in part to a seven-year online consultative process. “Individuals, organizations, companies, government agencies and even other governments offered input...improving the pioneering online, open source system along the way.”⁶⁷ The law—and Brazil’s thinking about internet governance—was clearly ahead of its time.

In the intervening years, disinformation found fertile ground in Brazil, a highly-networked society that relies on social media, in particular the Facebook-owned encrypted messaging app WhatsApp. Anya Prusa, a Program Associate at the Wilson Center’s Brazil Institute, notes that “Facebook and WhatsApp are central to the way Brazilians engage with one another, from family group chats to professional communications.”⁶⁸ Furthermore, since the passage of the Marco Civil, cell phone coverage has become more ubiquitous in the country. More people are relying on internet-based services for all of their communications; over 83 percent of Brazilians made a video or voice call using an internet service in 2018.⁶⁹

Brazilians’ dependence on social media and internet-based communications has had serious consequences in politics and beyond. Disinformation ran rampant during the 2018 presidential elections. Jair Bolsonaro’s campaign, afforded little television time because of his party’s limited representation in parliament, relied on social media to communicate with voters. WhatsApp was particularly critical for Bolsonaro: *Folha de Sao Paulo* reported that his campaign used the service to launch informational attacks against his leftist opponent Haddad.⁷⁰ The government and private sector attempted to respond to the use of disinformation and bot networks in the lead up to the election:

64 Daniel Arnaudo, “Brazil, the Internet and the Digital Bill of Rights - Reviewing the State of Brazilian Internet Governance,” Igarape.org, Igarape Institute, Accessed September 2020.

65 Tim Berners-Lee, “We Need a Magna Carta for the Internet,” *Huffpost.com*, *Huffpost*, May 6, 2014.

66 Ronaldo Lemos, “Brazil’s Internet Law, the Marco Civil, One Year Later,” CFR.org, Council on Foreign Relations, June 1, 2015.

67 Arnaudo, “Brazil, the Internet and the Digital Bill of Rights,” 2020.

68 Nina Jankowicz, “Quitting Facebook is easier in rich countries,” *Medium.com*, *Illustrisima, Folha de Sao Paulo*, May 18, 2019.

69 Bruno Villas Bôas, “IBGE: Brasil ganha 10 milhões de internautas em apenas um ano,” *Valor.com*, *Valor*, December 20, 2018.

70 Tom Phillips, “Bolsonaro business backers accused of illegal Whatsapp fake news campaign,” *TheGuardian.com*, *The Guardian*, October 18, 2018.



In January of 2018, under the direction of the Superior Electoral Court, Brazil's Federal Police established a task force with the specific goal of preventing and limiting fake news during the election. In the private sector, Facebook removed almost 200 pages run by several individuals connected to the right-wing activist organization Movimento Brasil Livre, whose names were not disclosed. Facebook asserted that the pages were part of a larger network, which was operated with the intent to spread misinformation.⁷¹

These discrete actions had little impact; according to a *Guardian* report, on Bolsonaro's use of WhatsApp, the disinformation campaign "was designed to inundate Brazilian voters with untruths and inventions, by simultaneously firing off hundreds of millions of WhatsApp messages."⁷² Other studies found that "on Election Day, 86 percent of all voters heard claims that voting machines had been rigged in favor of Fernando Haddad; more than half of Bolsonaro voters believed the story. These were not isolated cases: a study of several hundred political WhatsApp groups found that among the fifty most shared images, more than half contained false information."⁷³

Two years into Bolsonaro's term, disinformation continues to plague Brazilian political discourse. As late as July 2020, Facebook removed Pages, Facebook and Instagram accounts, and Groups with over 883,000 followers linked to "individuals associated with the Social Liberal Party and some of the employees of the offices of Anderson Moraes, Alana Passos, Eduardo Bolsonaro, Flavio Bolsonaro and Jair Bolsonaro."⁷⁴ Facebook wrote that the network "relied on a combination of duplicate and fake accounts...to evade enforcement, create fictitious personas posing as reporters, post content, and manage Pages masquerading as news outlets."

As the country was embroiled in COVID-19-related misinformation, Brazil's Supreme Federal Court also honed in on the President's use of disinformation. It launched an investigation of President Bolsonaro for creating an alleged fake news network to coordinate digital smear campaigns to discredit his enemies by propagating conspiracy theories and misinformation systematically via a network of online allies. The Supreme Federal Court probe evaluated whether Bolsonaro's inner circle was responsible for waging coordinated attacks, inserting disinformation and conspiracy theories into the national discourse. It delivered search warrants and raided the homes of devout Bolsonaro allies in search of evidence of financing for a fake news network and ordered Facebook and Twitter to impose global bans on accounts operated by Bolsonaro allies. At publication, the investigation is ongoing.

Law of Freedom, Liability, and Transparency on the Internet Bill

In mid-May 2020, Senator Alessandro Vieira introduced a bill entitled "The Law of Freedom, Liability, and Transparency on the Internet," colloquially known as the "fake news law." The bill sought to reign in the unchecked proliferation of misinformation on social media and private messaging apps by imposing rules for companies

71 Andrew Allen, "[Bots in Brazil: The Activity of Social Media Bots in Brazilian Elections](#)," WilsonCenter.org, The Wilson Center, August 17, 2018.

72 Ibid.

73 Christopher Harden, "[Brazil Fell for Fake News: What to Do About It Now?](#)" WilsonCenter.org, The Wilson Center, January 21, 2019.

74 Nathaniel Gleicher, "[Removing Coordinated Inauthentic Behavior](#)," about.fb.com, Facebook, July 8, 2020.



to limit and identify inauthentic accounts leveraged by political entities, mandate transparency requirements from platforms, and provide due process and the right of appeal for users. Non-compliance with the bill would result in fines up to ten percent of companies' annual revenue generated in Brazil. However, early versions of the bill included provisions that posed a serious threat to privacy rights and freedoms of expression, press, and association. For example, posting or sharing content that threatens the "social peace or to the economic order" would be criminalized under initial drafts of the bill, punishable for up to five years in jail.⁷⁵

It also would have required the mass legal identification of users to root out inauthentic accounts. It obliged users to hand over identification information—including social security numbers, passport information, active phone numbers, and addresses—in order to create an account with social media and messaging apps. These data-gathering provisions would be retroactive. Rights groups and platforms objected, pointing out both privacy and accessibility concerns; given that identifying information in other countries has been used to crack down on citizens exercising their right to free speech, it posed a threat to privacy, and since the bill excluded users without phone numbers from accessing these services, it would limit access to social platforms.

The bill also contained a traceability requirement that required social media platforms and private messaging apps to keep traceability data of messages exchanged "through mass forwarding" for up to three months. "Mass forwarding" refers to messages that are sent to more than five people and reach more than a thousand people within a 15-day period. Providers retain message metadata, which includes sender and recipient information plus the date and time the message was sent—not the actual content of messages. Rights organizations warned that this provision would require private messaging services to monitor message flow, which would weaken encryption and endanger user privacy.⁷⁶

Both civil society organizations and platforms were concerned about these provisions. Forty-five human rights organizations issued a joint statement warning lawmakers about the bill's potential consequences for freedom of expression, press, and association.⁷⁷ Platforms, too, spoke out about the possible implications of the bill; Facebook, Twitter, and Google released a joint statement, describing it as "a project of mass collection of data from individuals, resulting in worsening digital exclusion and endangering the privacy and security of millions of citizens."⁷⁸

Current Version

After significant revision, the Brazilian Senate passed the bill on June 30, 2020. The vote was postponed three times due to pressure from civil society groups and disagreement among lawmakers about the contents of the bill.⁷⁹ Unlike in earlier versions of the bill, sharing fake news would no longer be considered a crime, the 1 million BRL (\$191,240) penalty for political candidates who distribute misleading information about their competition would be waived, and mass identification verification of users would be reigned in.⁸⁰

75 "Brazil: Reject 'Fake News' Bill," hrw.org, Human Rights Watch, June 24, 2020.

76 Greg Nojeim, "Update on Brazil's Fake News Bill: The Draft Approved by the Senate Continues to Jeopardize Users' Rights," cdt.org, Center for Democracy & Technology (CDT), July 24, 2020.

77 "Nota conjunta de organizações de Direitos Humanos contra propostas normativas que podem levar à criminalização e vigilância de movimentos sociais ao tratar de 'fake news,'" RSF.org, Reporters without Borders, June 6, 2020.

78 Eric Neugeboren, "Brazil's Draft 'Fake News' Bill Could Stifle Dissent, Critics Warn," VOAnews.com, Voice of America, July 10, 2020.

79 "Brazil's 'fake news' bill poses major threat to freedom to inform," RSF.org, Reporters Without Borders, June 25, 2020.

80 Angelica Mari, "Brazilian Senate passes fake news bill," ZDnet.com ZDNet, July 1, 2020.



As the bill stands, it primarily aims to curb malicious activity by bots by obliging social media platforms and private messaging services to ban inauthentic and automated accounts, implement technical measures to identify them and to detect them at account registration, and establish usage limits on the number of accounts controlled by the same user. It sets out to prevent mass forwarding by obligating private messaging apps to limit on the size of message groups, as well as the amount of times a single message can be forwarded to users and groups. In addition, platforms are required to ask users for consent before including them in message groups.

Amendments to the draft rolled back the mass identification requirement significantly, and now companies “may” require their users to hand over identification information, such as a valid ID, if they are reported or suspected of inauthentic behavior. Still, the provision grants platforms powers typically reserved for law enforcement, such as power to solicit government identification from users. Additionally, private messaging apps, where accounts are tied to users’ phone numbers, are tasked with communicating with phone service providers to identify numbers that have had their contracts rescinded and then to delete related accounts on their service. The purpose of this provision is to ensure messaging accounts are associated with authentic users and active numbers and to prevent the creation of inauthentic messaging accounts from inactive numbers, but it is unclear how this verifying process will work. Under the most recent draft, only private messaging apps must keep traceability data of messages exchanged “through mass forwarding” for up to three months—not social media platforms. Rights organizations warn that this provision would require private messaging services to monitor message flow, which would weaken encryption and endanger user privacy.⁸¹

The reports would include the number of automated and inauthentic accounts detected by platforms during the quarter, and data about user engagement with content identified as irregular, including number of views, shares, and reach.

The bill would require platforms provide mechanisms for appeal and due process for users whose content is removed. Users must also be notified if their post is reported by a user or acted upon by the platform, given an explanation of why their posts were reported or removed, as well as provided information about the process for evaluation, how decisions will be applied, and how and by when they can appeal. Users who report content are given the right of response.

The new bill would establish comprehensive transparency obligations for platforms, where platforms must provide transparency reports on a quarterly basis. It would require social media platforms to share data (including disaggregated data) with academic research institutions. The reports would include the number of automated and inauthentic accounts detected by platforms during the quarter, and data about user engagement with content identified as irregular, including number of views, shares, and reach. Reports also would include details on the “motivation and methodology used for the detection of irregularities and the type of measures adopted,” the numbers of account and content moderation actions taken to separately comply with company community guidelines, the fake news law, and court orders. Additionally, it would note the average amount of time between detection and action taken or reported or detected bad content. Access to such data would grant researchers visibility into the internal procedures of platforms around account and content moderation, enabling researchers

81 Nojeim, “[Update on Brazil’s Fake News Bill](#),” 2020.



to exercise oversight and scrutiny platforms' handling of inauthentic accounts and gauge compliance with the law, enforcement of company community guidelines, and the reach of coordinated disinformation campaigns.

Transparency obligations also apply to political sponsored content and advertisements. For ads and sponsored content, social media providers are required to first validate the identity of advertisers and content sponsors before accepting payment, possibly by providing a valid identification. They also must identify advertisers and provide their contact information for users who see the advertised or sponsored content. For political ads or sponsored content mentioning political candidates, parties, or coalitions, they must disclose much more to users: the amount spent, the duration of broadcast, identify the advertiser behind it, and the characteristics of the targeted audience.

The bill would establish a Council for Transparency and Responsibility on the Internet, a multi-stakeholder oversight body composed of 21 members from government agencies, private sector, and civil society. The Council would be responsible for monitoring compliance with the fake news law, reviewing transparency reports and assessing moderation practices, assessing the sufficiency of companies' terms of service, and developing and monitoring the application of a common Code of Conduct for platforms and messaging services. A key caveat of the provision is that the Code of Conduct and members must be approved by the Brazilian Senate.

Data localization requirements in the bill would require social media companies and private messaging services to appoint legal representatives in Brazil who have permissions to remotely access Brazilian user data. The bill explains that this may be necessary to comply with a court order. This poses serious privacy concerns and may violate the CLOUD Act and the Electronic Communications Privacy Act, which are laws that "require US providers to satisfy certain procedural safeguards before turning over private data to foreign law enforcement agents."⁸²

Rights groups approve of the bill's transparency reporting obligations, establishment of due process for users, and the disclosure of political ads.⁸³ However, some are dissatisfied with the provisions on the articles on user identification and traceability. In a blog post, the Electronic Frontier Foundation's International Rights Director Katitza Rodriguez wrote, "While we do not know how a service provider will implement any traceability mandate nor at what cost to security and privacy, ultimately, any implementation will break users' expectations of privacy and security, and would be hard to implement to match current security and privacy standards. Such changes move companies away from privacy-focused engineering and data minimization principles that should characterize secure private messaging apps."⁸⁴ Civil society organizations have succeeded in pressuring members of congress to reconsider portions of the bill to which they object. After the bill passed in the Brazilian Senate on June 30, a working group convened to discuss possible adjustments to the legislation before moving it on to the lower chamber, the House of Deputies, for a vote. The bill's authors announced they are working on removing the provisions about user identification and traceability data requirements.⁸⁵

82 Udbhav Tiwari and Jochai Ben-Avie, "[Mozilla's analysis: Brazil's fake news law harms privacy, security and free expression](#)," blog.mozilla.org, Mozilla, June 29, 2020.

83 Access Now, "[What is the Brazilian fake news bill?](#)," YouTube, Access Now, July 28, 2020

84 Katitza Rodriguez and Seth Schoen, "[FAQ: Why Brazil's Plan to Mandate Traceability in Private Messaging Apps will Break User's Expectations of Privacy and Security](#)," EFF.org, Electronic Frontier Foundation (EFF), August 7, 2020.

85 Access Now, "[No Fake News, More Rights](#)," 2020.



Although the bill looks set to pass the lower chamber, Bolsonaro does not support it, noting in a Facebook live broadcast in July that he would veto it if it did pass through Congress and expressed support for “total freedom of the media.”^{86, 87}

The Brazilian case showcases that the multi-stakeholder policy making approach to social media regulation can help lawmakers who may not be well-versed in the technology craft sustainable legislation that safeguards users’ rights while crafting productive responses to the threat. Brazilian lawmakers have engaged with civil society and rights organizations throughout the drafting process and incorporated their input when making revisions to the bill to produce legislation that will pass in both chambers of the Congress. By engaging with organizations steeped in digital rights issues, Brazilian politicians avoided social media regulation pitfalls and reigned in the more controversial aspects of the law to produce a bill that has the potential to set global precedent on transparency and to reign in inauthentic behavior in online social spaces.



Jair Bolsonaro, Photo courtesy of Shutterstock.com/ BW Press

86 Anthony Boadle, “Brazil’s Bolsonaro would veto bill regulating fake news in current form,” Reuters.com, *Reuters*, July 2, 2020.

87 Lisandra Paraguassu, “Brazil finalizing bill to target financiers of ‘fake news’ attacks,” Reuters.com, *Reuters*, September 8, 2020.



Singapore: Counter Disinformation as Cover for Censorship

In early October 2019, Singapore's Protection from Online Falsehoods and Manipulation Act (abbreviated as 'Pofma' in press), known as the "anti-fake news law," took effect. The law prohibits the communication of falsehoods over the internet and grants government ministers sweeping powers and discretion to correct and remove online content. Singapore's government pitched Pofma as a means to limit the dissemination of false information online and a way to protect Singapore's national security and democratic processes. However, it has instead been utilized as a pro-government spin tool by the parliament's supermajority, the People's Action Party (PAP), to police criticism and unfavorable coverage of their government from opposition parties, journalists, rights organizations, and individuals online.

Singapore's democracy has been electorally dominated by the People Action Party (PAP) party since it was established in 1959. PAP's one-party dominance has remained unrivaled by any opposition party for over 60 years, making PAP synonymous with government since the country's independence. While opposition parties do exist, they face overwhelming obstacles and pressure from the ruling party.⁸⁸ PAP often files sedition and defamation civil lawsuits against opposition party members running for public office to discredit them, issue criminal charges, and bankrupt candidates with damages fees.⁸⁹ Today, only ten of 93 elected seats in Parliament belong to opposition party members.

Singapore places strict regulations on speech, which de facto repress free expression and discourages speech critical of the government. While Singapore's constitution guarantees all citizens freedom of speech and expression, it also reserves Parliament's right to impose restrictions on it in certain circumstances.⁹⁰ The government may pass regulations on free expression it considers

"necessary or expedient" on eight grounds: to ensure the city-state's security, to maintain friendly relations with other countries, to preserve public order or public morality, to protect parliamentary privileges, and to prevent contempt of court, defamation, or incitement to any offense.⁹¹ Singapore's judges and ministers have interpreted these stipulations very broadly, erring on the side of deference to the government's assessment of what is necessary to maintain public order. This broad interpretation facilitated the creation and use of Singapore's Sedition Act, Public Order Act, Broadcasting Act, Administration of Justice (Protection) Act, as well as several other penal code provisions that restrict free expression.⁹² Outright dissent against the government is perceived as a challenge to the social order and political status quo in Singapore.

It also grants government powers to restrict international news outlets from operating in Singapore and to appoint members to the newspaper's board of directors.

88 "'Kill the Chickens to Scare the Monkeys': Suppression of Free Expression and Assembly in Singapore," hrw.org, Human Rights Watch, December 12, 2017.

89 Ibid.

90 The Government of the Republic of Singapore, "Constitution of the Republic of Singapore," sso.agc.gov.sg, Singapore Statutes Online, August 9, 1965.

91 Ibid.

92 "'Kill the Chickens to Scare the Monkeys,'" 2017.



Media is centralized and heavily controlled in the city-state. The Newspaper and Printing Presses Act (NPPA) governs Singaporean print media and requires news outlets to renew their licenses annually. It also grants government powers to restrict international news outlets from operating in Singapore and to appoint members to the newspaper's board of directors.⁹³ Two media companies dominate the media landscape in the city-state: Singapore Press Holdings Ltd. (SPH) and Media Corp. SPH owns all mainstream print newspapers in circulation—a near print media monopoly—with the exception of one newspaper owned by MediaCorp.⁹⁴ MediaCorp operates the majority of broadcast outlets, including most television and radio stations. Both media companies are private; however, both are state-affiliated and generally pro-government.^{95, 96} Few domestic independent news organizations exist as they lack the resources to compete in Singapore. Compounding their monetary woes, the government levies costly civil defamation lawsuits and contempt charges to reign in domestic and international media outlets as a containment tactic.⁹⁷ The lack of alternative and independent news outlets creates an uncritical media environment.

Consideration of Pofma

Singapore did not have a notable disinformation problem prior to Pofma's passage in April 2019.⁹⁸ Rather, the law was created after Singapore's Ministry of Law published a paper that described the "impact of falsehoods" on American, European, and Indonesian democracies. It discussed tactics and trends such as foreign interference in elections, disinformation campaigns, conspiracy theories, exploitation of ethnic divisions, and particularly emphasized how these phenomena undermine citizens' confidence in government. Germany's NetzDG law was cited in the report.

Although Singapore had not been victim to outside interference in its democracy, the Ministry of Law argued that Singapore was inherently vulnerable and an attractive target due to the country's internet connectivity, ethnic and religious diversity, and economic prowess in the region.⁹⁹ "We have to ensure that our national security is not compromised," the report states. Despite this focus, the few examples provided in the report regarding "foreign interference" and the "spread of falsehoods" in Singapore concern press coverage from "anti-Government" independent news outlets allegedly designed "to mislead Singaporeans."¹⁰⁰ In short, the Ministry of Law argued the press are perpetrators of interference in Singapore's democracy, necessitating government regulation and correction.

Parliament later held eight days of hearings under its Select Committee on Deliberate Online Falsehoods. Experts, journalists, and internet platform representatives discussed the likelihood of foreign disinformation in

93 Ibid.

94 Andrew T. Kenyon, Tim Marjoribanks, and Amanda Whiting, *Democracy, Media and Law in Malaysia and Singapore: A Space for Speech* (Routledge, 2014), 28.

95 Ibid.

96 "'Kill the Chickens to Scare the Monkeys,'" 2017.

97 "'Singapore's Fake News and Contempt Laws a Threat to Media, Journalists Say,'" VOAnews.com, Voice of America, May 6, 2020.

98 Kaye, "'Report of the Special Rapporteur,'" 2018.

99 The Ministry of Communication and Information, The Ministry of Law, "'Deliberate Online Falsehoods: Challenges and Implications,'" nas.gov.sg, National Archives of Singapore, January 5, 2020.

100 Ibid., 21.



Singapore and how the government should prepare for it.¹⁰¹ The Senior Minister of State for Law and Health, Mr. Edwin Tong Chun Fai and other People Action Party (PAP) politicians cited a report published by Human Rights Watch (HRW) on Singaporean press freedom as “an example of how false and misleading impressions can be created by a selective presentation of facts, designed to promote an underlying agenda.”¹⁰² They characterized HRW as a deceptive NGO with ulterior motives and alleged it peddled misinformation “to advocate political change in another country” and “to sow seeds of doubt on national issues and government relations with various institutions of the country.”¹⁰³ This conversation was a harbinger for the kind of content Pofma would police in online spaces.

Pofma in Practice

Pofma prohibits the communication of false statements of fact over the internet that are “against the public interest,” applying to all online content that is accessible in Singapore regardless of the location where the content was published. It grants expansive powers to government ministers and civil servants to independently—with no oversight—use Pofma in three ways: to require correction of online content, to designate Declared Online Locations (DOLs), and to block site access.

They may issue Correction Directions to individuals who post falsehoods online, or issue Targeted Correction Directions to social media platforms where posts are made to require them to display a correction note alongside their posts. The note includes a government-provided statement notifying users that the post contains falsehoods and a link to a government response on Factually.sg, which is a government fact-checking website that publishes government responses to false posts and articles and provides readers the “correct facts” in response to alleged “falsehoods.”¹⁰⁴ It resembles a website for a fact-checking organization. Each government statement looks like an online news article and features an image of the original post with the words “FALSE” stamped on it in red. The Singaporean government says this approach provides transparency to the public for ministers’ Pofma decisions. However, activist organizations warn this is a tactic for the government to manipulate public opinion.¹⁰⁵ Reporters Without Borders compared the government’s use of this procedure to an “Orwellian Ministry of Truth.”¹⁰⁶

They may designate websites or web pages that have distributed three or more falsehoods over the course of six months as Declared Online Locations (DOLs). When a website or web page is designated a DOL, the site’s operator is required to pin a government-written pop-up warning on the DOL website that notifies Singaporean site visitors that the site has historically hosted falsehoods. So far, just four Facebook pages owned by self-exiled dissident Alex Tan Zhi Xiang have been labeled as DOLs. Xiang runs the Australia-based Singaporean news outlet States Times Review, a news outlet that produces coverage critical of Singapore’s government.¹⁰⁷ They may order internet service providers to disable access to websites or webpages. This has only happened once: in January

101 Singapore Parliament, “[Report of the Select Committee on Deliberate Online Falsehoods – Causes, Consequences, and Countermeasures](#),” [sprs.parl.gov.sg](#), Singapore Parliament, September 19, 2018.

102 “[Singapore’s ‘fake news’ bill is bad news for Facebook](#),” CNN.com, CNN, April 3, 2019.

103 Ibid.

104 c. <https://www.gov.sg/factually>

105 “[Singapore uses ‘anti-fake news’ law to eliminate public debate](#),” RSF.org, Reporters Without Borders, December, 6, 2019.,

106 Ibid.

107 c. <https://www.pofmaoffice.gov.sg/registry/declared-online-locations/>



2020, when the Minister of Communications and Information ordered internet service providers to disable access to Lawyers for Liberty, a Malaysian-based human rights and law organization that reported on unlawful execution methods practiced in Singapore's Changi Prison.¹⁰⁸

Criticism

Rights organizations, journalists, and internet platforms voiced concern over its implications for free expression online given the law's broad application and Singapore's poor track record of prohibiting speech critical of the government and placing limits on free expression. Human rights experts predicted that the law would be used to censor critics of the government and encourage self-censorship of citizens.¹⁰⁹ Reporters Without Borders warned that the law threatened journalistic independence. David Kaye, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, expressed concerns about its "overbroad definition of falsehood" in a letter to the Singaporean government.¹¹⁰ It appears critics were right: Journalistic freedom decreased in Singapore as a result of Pofma and its censorship practices "chilling" effects on free speech. Singapore is ranked 158th of 183 on Reporters Without Borders' (RSF) 2020 World Press Freedom Index, dropping seven spots since 2019, the year Pofma was enacted.¹¹¹

A year after taking effect, Ministers have used Pofma more to target opposition parties and discredit critical coverage of the government than to halt the spread of 'fake news.' Since October 2019, the law has been invoked 72 times by the government—the highest volume of uses occurring during the month of July 2020 during Singapore's parliamentary elections.¹¹² The law has been used to target opposition parties sharing criticism and unfavorable news about PAP online, to 'debunk' and discredit unfavorable coverage of the government by independent news outlets and shared by individual users, and to control the narrative about the government's handling of the COVID-19 outbreak in Singapore. Roughly 85 percent of Pofma uses relate to posts critiquing government's decisions or policies, whereas a minority of Pofma uses actually addressed harmful misinformation about COVID-19.¹¹³ As of July 2020, 49 percent of Pofma uses relate to COVID-19.¹¹⁴

Prior to the COVID-19 outbreak, the law had been invoked primarily against political opposition parties. The first five uses of Pofma were to police political information from individuals of opposition parties, which the Minister

Public interest is defined as anything threatening Singapore's national security, relations with other countries, public services, elections, or the public's confidence in their government.

108 Ministry of Communications and Information, "[Minister for Communications and Information directs IMDA to issue access blocking orders](#)," pofmaoffice.gov, Ministry of Communications and Information, January 23, 2020.

109 "[Singapore: Free Expression Restrictions Tighten](#)," HRW.org, Human Rights Watch, January 14, 2020.

110 David Kaye, "[Mandate of the Special Rapporteur on the promotion of the right to freedom of opinion and expression](#)," OHCHR.org, OHCHR, April 24, 2019.

111 "[Singapore](#)," RSF.org, Reporters Without Borders, 2020.

112 "[Explainer: What is POFMA](#)," pofmaed.com, POFMA'ed, July 5, 2020.

113 Paul Meyer, "[Singapore's First Election Under the Fake News Law](#)," thediplomat.com, The Diplomat, July 7, 2020.

114 Ibid.



for Communication and Information described as merely a “coincidence.”¹¹⁵ The initial use of Pofma in this way by the government signaled potential for misuse of Pofma against opposition parties during Singapore’s July 2020 elections. Because election campaigns in Singapore are short and typically occur over the course of 9 days, experts expressed concern that a correction directive could be issued without sufficient time for an appeal to be resolved.¹¹⁶

In an election cycle where all campaign activities took place online due to COVID-19, civil servants—taking over Pofma powers from ministers for the duration of the election cycle—doled out correction orders to opposition parties for Facebook posts made about “government spending for foreign students, plans to grow the city-state’s 5.7 million population to 10 million, and government advisories that discouraged COVID-19 testing for foreign workers.”¹¹⁷ No corrections were delivered to PAP.

Another problematic stipulation within the law is its overly broad, vague definitions for what can be considered “false” or “against the public interest.” A statement is considered false if it is “false or misleading, wholly or in part, and whether on its own or in the context in which it appears.” Public interest is defined as anything threatening Singapore’s national security, relations with other countries, public services, elections, or the public’s confidence in their government.¹¹⁸ These broadly defined terms allow for plenty of grey area and interpretation, leaving the door open for abuse. The majority of correction orders target posts containing small inaccuracies in order to label the entire post as inaccurate and discredit entire arguments and the reputation of independent media outlets or opposition parties.^{119, 120}

Blurring the lines between government oversight of facts and opinions, Singapore’s High Court clarified in a ruling on an appeal case challenging a Pofma decision that the law applies to not only expressed meanings but to implied meanings.¹²¹ In January 2020, the Ministry of Manpower used Pofma to correct claims made by opposition party Singapore Democratic Party (SDP) about local unemployment statistics in posts and articles published on the website. SDP filed a case to appeal, arguing that an interpretation of statistics did not qualify as a deliberate falsehood, but the High Court ruled against SDP and ruled that their content did contain implied falsehoods.^{122, 123} This leaves the judgement of both what is true and untrue and the meaning of a statement open to a given minister’s own interpretation.

Furthermore, Pofma’s Correction Directions, threats of criminal prosecution, and fines intimidate government critics and citizens, discourage public dissent, and encourage self-censorship and conformity among those critical

115 Dewey Sim, “Singapore defends fake news law, saying it’s a ‘coincidence’ that politicians targeted,” *scmp.com*, South China Morning Post, January 6, 2020.

116 “Singapore’s Fake News and Contempt Laws,” *VOAnews.com*, Voice of America, May 6, 2020.

117 John Geddie, “Singapore’s fake news law trips up opposition as election looms,” *Reuters.com*, *Reuters*, July 6, 2020.

118 Parliament of Singapore, *Protection from Online Falsehoods and Manipulation Act 2019*, <https://sso.agc.gov.sg/>, Singapore Statutes Online, June 28, 2019.

119 Meyer, “Singapore’s First Election Under the Fake News Law,” 2020.

120 Thum Ping Tjin and Kirsten Han, “Singapore’s ‘Fake News’ Bill: The FAQ,” *newnaratif.com*, New Naratif, April 9, 2019.

121 Rei Kurohi, “Fake news law does cover matters of interpretation: AGC,” *StraitsTimes.com*, *The Straits Times*, January 18, 2020.

122 Kirsten Han, “Want to Criticize Singapore? Expect a ‘Correction Notice,’” *NYTimes.com*, *The New York Times*, January 21, 2020.

123 Kok Xinghui, “In Singapore, first court challenge against anti-fake news law fails,” *scmp.com*, South China Morning Post, February 5, 2020.



of the government—including journalists, members of the opposition parties, and civil society.^{124,125} Individuals who unknowingly spread a falsehood do not face criminal charges under Pofma, but are asked to comply with correction directions. Refusal to comply with a correction direction could result in fines up to 20,000 SGD (\$14,720) and 12 months in prison.

Pofma is designed to force social media companies to submit to Singaporean government's decisions at risk of fines and encourages them to moderate content in line with the Singaporean government regime.¹²⁶ Platforms voiced censorship concerns about Pofma when it was still under consideration, however, when the law took effect, platforms' concern for their bottom line and market share took precedence over concerns for free expression. Platforms face fines of up to 1 million SGD (\$736,075) for refusal to comply and could face additional daily fines of up to 100,000 SGD (\$73,601) for continued noncompliance.¹²⁷ Facebook, Youtube, and Hardwarezone—a Singaporean online tech forum—all eventually complied with Pofma requests. Facebook displays disclaimers on labeled posts reading "Facebook is legally required to tell you that the Singapore government says this post has false information" and includes a "learn more" tab that explains Facebook's legal obligation to abide by Pofma, stressing the company's role as neutral platform that "does not endorse truthfulness."¹²⁸ The majority of Pofma orders involved statements made on Facebook.¹²⁹

Overall, Singapore's Pofma law—used to denigrate opponents and quash government criticism—should be a cautionary tale to all democratic governments contemplating creating state-sponsored fact-checking services or endowing partisan officials with the hefty responsibility of protecting freedom of expression. Its broad definitions of prohibited content, politically-motivated implementation, and large fines amount to a chilling effect on ruling party criticism and free expression in the country. As in Germany, Pofma has also had ripple effects around the world, with Nigeria and Thailand citing it in their own draconian speech crackdowns.¹³⁰

124 "Singapore Government says Washington Post article," ChannelNewsAsia.com, Channel News Asia, December 17, 2019.

125 Kaye, "Mandate of the Special Rapporteur," 2019.

126 Meyer, "Singapore's First Election Under the Fake News Law," 2020.

127 "Singapore: Chilling fake news law will 'rule the news feed'," Amnesty.org, Amnesty International, May, 8, 2019.

128 "Facebook says 'deeply concerned' about Singapore's order to block page," Reuters.com, *Reuters*, February 18, 2020.

129 "Explainer: What is POFMA," 2020.

130 Stefania Palma, Neil Munshi, and John Reed, "Singapore 'falsehoods' law shows perils of fake news fight," FT.com, Financial Times, February 3, 2020.



Ukraine: Counter disinformation policies with unclear guidelines pose a threat to freedom of expression

In 2014, when the Russian Federation illegally annexed Ukraine's Crimean peninsula and began a military conflict in the country's east and thousands of peaceful protestors demanded change across the country, Ukraine became not only the site of Europe's only hot war; it became the front line of the online information war, as well. Russian disinformers, including intelligence services and Kremlin-affiliated non-government entities like the Internet Research Agency launched "an all-out propaganda offensive against the new government in Kyiv and pro-Western demonstrators," according to a classified Russian intelligence report obtained by The Washington Post.¹³¹ Long before the term "fake news" became well-known in the United States, StopFake News, a Ukrainian fact-checking organization founded during the beginning of the Russian information onslaught, built awareness about the phenomenon in Ukraine and Europe, and the tools and tactics the Russian Federation used to attempt to undermine Ukraine's sovereignty and support for the country in the international community.

Russia's disinformation campaigns in Ukraine have critical implications for national security; the Kremlin has targeted Ukrainian soldiers with antagonistic text messages, encouraging them to desert their positions or intended to lower morale.¹³² It has used Facebook pages claiming to represent local news outlets to disguise its anti-NATO, anti-Western, anti-Kyiv narratives.¹³³ Anti-Ukrainian disinformation also spread on Russian-owned social media networks vKontakte (VK) and Odnoclassniki. In response, then-President Petro Poroshenko blocked these networks, as well as Russian search engine Yandex and email provider Mail.Ru, within Ukraine beginning in May 2017. "Massive Russian cyberattacks around the world, including the recent interference in the French election campaign, indicate that it is time to act differently and more decisively," Poroshenko wrote in an explanation of his decision. "I urge all my compatriots to leave Russian servers immediately for security reasons."¹³⁴

Both Ukrainian citizens and international human rights groups roundly criticized the decision. At the time of the ban, at least 78 percent of Ukrainians had an account on VK.¹³⁵ Ukrainian lawmaker Serhiy Leshchenko compared Poroshenko to Turkey's Raci Tayyip Erdogan, who had overseen an online crackdown in his own country.¹³⁶ Tanya Cooper, a researcher at Human Rights Watch, said of the ban: "This is yet another example of the ease with which President Poroshenko unjustifiably tries to control public discourse in Ukraine. Poroshenko may try to justify this latest step, but it is a cynical, politically expedient attack on the right to information affecting millions of Ukrainians, and their personal and professional lives."¹³⁷

131 Ellen Nakashima, "Inside a Russian disinformation campaign in Ukraine in 2014," washingtonpost.com, *The Washington Post*, December 25, 2017.

132 Raphael Satter and Dmytro Vlasov, "Ukraine soldiers bombarded by 'pinpoint propaganda' texts," APnews.com, *Associated Press*, May 11, 2017.

133 Nathaniel Gleicher, "Removing Coordinated Inauthentic Behavior from Russia," about.fb.com, Facebook Newsroom, January 17, 2019.

134 Petro Poroshenko, "Гібридна війна вимагає адекватних відповідей на виклики," m.vk.com, VK Post, May 17, 2017.

135 Christopher Miller, "Public Sharply Divided Over Ukraine's Ban On Russian Social Networks," RFERL.org, RFERL, May 17, 2017.

136 c. <https://twitter.com/Leshchenkos/status/864755600886038528>

137 "Ukraine: Revoke Ban on Dozens of Russian Web Companies," HRW.org, Human Rights Watch, May 16, 2017.



Along with Russian social networks, search engines, and mail servers, Ukraine's government also imposed sanctions on thousands of Russian websites and utilized other decrees to eliminate "anti-Ukrainian propaganda" from its informational ecosystem. Freedom House reports that "in 2018 alone [Ukraine's security services] identified and blocked 360 cyber incidents, convicted 49 administrators of social network groups for 'anti-Ukrainian propaganda,' and began proceedings against an additional 29 individuals."¹³⁸ Despite criticism, the attitude across Ukrainian government—that national security concerns should trump concerns about protecting freedom of expression—prevailed. Ivanna Klympush-Tsintsadze, Ukraine's former deputy prime minister for Euro-Atlantic and European integration, told *The Atlantic* in 2019: "We had a really hard time explaining to our partners...don't forget that we are a country at war. We are losing people every other day, if not every single day."¹³⁹ But simply banning access to the websites where false information proliferated was also not a cure-all; while both vKontakte and Odnoklassniki became distinctly less popular in Ukraine, they both remain among the top fifteen most accessed websites.¹⁴⁰

Throughout the 2019 presidential campaign, challenger Volodymyr Zelenskyy was critical of the Poroshenko administration's internet governance tactics. He "suggested users be given alternatives prior to considering wholesale blocking, and noted that because the authorities had failed to explain the real rationale behind the sanctions, they had not secured enough public support."¹⁴¹ More than a year after Zelenskyy's inauguration, however, the sanctions remain active, and two draft laws on the media and disinformation were introduced in early 2020, continuing Ukraine's trend towards restricted online expression.

Draft Law on Disinformation

On November 8, 2019, the nascent Zelenskyy administration issued a presidential decree "On urgent measures to reform and strengthen the state." Among these urgent reforms, Zelenskyy instructed the parliament to fast-track a bill on "mechanisms to prevent the dissemination of inaccurate, distorted information...and strengthen accountability for violation of information legislation."¹⁴² Before the end of the calendar year, the Ministry of Culture, Youth, and Information Politics began discussing its draft of the Law on Disinformation with members of civil society and international organizations, and was presented publicly on January 20, 2020.

According to the Ministry's presentation, Ukraine faces several challenges in its information space that the draft law sought to address. First, the Ministry claimed that "currently no one is endowed with [the] powers" to adjudicate decisions related to disinformation.¹⁴³ The Ministry laments the state of Ukrainians' media literacy skills and the difficulty of determining media ownership in Ukraine, where all major media are controlled by oligarchic and political elites. Finally, the Ministry asserts that the "status of journalists [in Ukraine] is diminished" because journalists do not have high enough professional standards, and because there is no liability for spreading disinformation.¹⁴⁴

138 Olga Kyryliuk, "Should Ukraine Drop Sanctions against Russian Tech Companies?" *freedomhouse.com*, Freedom House, 2019.

139 Nina Jankowicz, "Ukraine's Election is an All-Out Disinformation Battle," *theatlantic.com*, *The Atlantic*, April 17, 2019.

140 Игорь Бурдыга, "В контакте» и «Одноклассников» в Украине: действуют ли санкции?" *DW.com*, Deutsche Welle, May 16, 2018.

141 Kyryliuk, "Should Ukraine Drop Sanctions against Russian Tech Companies?" 2019.

142 Presidential Administration of Ukraine, "УКАЗ ПРЕЗИДЕНТА УКРАЇНИ №837/2019," *president.gov.ua*, Presidential Administration of Ukraine, November 8, 2019.

143 Ministry of Culture, "Про протидію дезінформації, Презентація законопроекту," *mkip.gov.ua*, Ministry of Culture, Youth, and Information Politics, January 2020.

144 Ibid.



The draft law seeks to address these challenges, and includes government oversight of both traditional and social media, creating a new government-appointed “Information Commissioner” to coordinate and enforce work relating to falsified information and disinformation. According to a Ministry of Culture presentation in late 2019, the government would define “falsified information” as “false information about a person, facts, events and phenomena that did not exist at all or that existed, but the details are incomplete or distorted.” Disinformation would be defined as “inaccurate information of public interest, including in relation to the national security, territorial integrity, sovereignty, and defense of Ukraine, the rights of the Ukrainian people to self-determination, the life and health of [Ukrainian] citizens, and the state of the environment. Disinformation is not: opinions, including criticism; satire and parody; or unfair advertising.”¹⁴⁵

Under these definitions, the new information commissioner would:

- “Have the power to fine media outlets and individual journalists, bring criminal charges against them, remove published materials, and ask the courts to shut down media outlets;”¹⁴⁶
- Create an online “trust index” on media outlets and information providers; only “trusted” media would be allowed to cooperate with the government through events like press briefings;
- Oversee the collection and storage of data on users and owners of information platforms and messenger services;
- Hold responsible all social network users and organizations for the accuracy of the information they post and disseminate.

The draft law also would create an Association of Professional Journalists in Ukraine with whom the government would cooperate; meanwhile, journalists found to “deliberately spread[...] disinformation would face a minimum fine of 4.7 million UAH (\$195,000).”¹⁴⁷ The charge would be added to the journalist’s criminal record, and repeat offenders could face up to five years in prison.

Criticism

Reaction to the draft law among rights organizations was negative and swift. Ukraine’s Commission on Journalistic Ethics demanded that the draft be withdrawn as it “[saw] no possibility of finalizing or improving it.”¹⁴⁸ The draft, they wrote, “violate[d] the basic principles of media independence.” Among the commission’s main concerns were the combination of regulatory standards and law enforcement functions in the role of the Information Commissioner. It also strongly opposed the introduction of criminal liability for the spread of disinformation: “World experience shows that the use of criminal liability is effective only in extreme cases, but it cannot be a systemic weapon against disinformation.” Instead, the commission supported more targeted

145 Ibid.

146 Diana Dutsyk and Marta Dyczok, “Ukraine’s New Media Laws: Fighting Disinformation or Targeting Freedom of Speech?,” WilsonCenter.org, The Wilson Center, February 10, 2020.

147 Dutsyk and Dyczok.

148 Commission on Journalistic Ethics, “Заява Комісії з журналістської етики щодо законопроекту про дезінформацію,” CJE.org.ua, Commission on Journalistic Ethics, January 27, 2020.



liability laws related to hate speech, as well as the development of and investment in a more robust media environment, including through public media.

Freedom House's Ukraine Project Director, Matthew Schaaf, sent a letter to the Ministry of Culture, Youth, and Sport outlining its criticisms of the proposals. He noted that, "enacting restrictions on distribution of information...may directly conflict with international standards" including Article 19 of the International Covenant on Civil and Political Rights.¹⁴⁹ Rather than target specific content or ideas, Schaaf suggested Ukraine seek to target "harmful behavior" such as "inauthentic behavior" or the use of bots. He also underlined concerns relating to the establishment of an Information Commissioner that would act as judge and jury for all matters relating to disinformation oversight in Ukraine. "The Commissioner's quasi law-enforcement and adjudicatory function, as well as the selection process as a direct appointment by the government, could negatively affect the position's perceived and real independence, as well as its ability to serve as a neutral defender of Ukrainians' interests and human rights," he wrote.¹⁵⁰

The United States or other countries will move to regulate social media to protect against disinformation and other online harms, but a question of when, and whether those eventual regulations will serve to protect or denigrate human rights, democracy, and freedom of expression.

The United Nations Special Rapporteur on Freedom of Opinion and Expression made a similar argument in a 2018 report to the Human Rights Council. Though outside of the Ukrainian context, he wrote:

In the light of legitimate state concerns such as privacy and national security, the appeal of regulation is understandable. However, such rules involve risks to freedom of expression, putting significant pressure on companies such that they may remove lawful content in a broad effort to avoid liability....Complex questions of fact and law should generally be adjudicated by public institutions, not private actors whose current processes may be inconsistent with due process standards and whose motives are principally economic.¹⁵¹

The United Nations Human Rights Monitoring Mission in Ukraine raised public concerns specifically about the draft law. "We share the concerns of the media community and civil society that the current draft law on disinformation is not in line with international human rights standards, may undermine media freedom and trigger self-censorship," based on the harsh fines and threat of jail time for those found to disseminate disinformation, the mission wrote.¹⁵² "An attempt to introduce criminal responsibility for disseminating disinformation puts journalists at risk of prosecution simply for doing their work. Furthermore, the government should refrain from

149 Matthew Schaaf, Freedom House, letter to First Deputy Minister Maksymchuk, December 17, 2019. Provided to The Wilson Center by Matthew Schaaf.

150 Ibid.

151 Kaye, "Report of the Special Rapporteur," 2018.

152 United Nations Human Rights Monitoring Mission in Ukraine, "Закликаємо українську владу утриматися від встановлення непотрібних обмежень для роботи медіа," subjectguides.library.american.edu, UN Human Rights Monitoring Mission in Ukraine, January 31, 2020.



establishing bodies in charge of monitoring or filtering content.”

These criticisms caused the Ministry of Culture to extend the comment period on the draft law, but as the COVID-19 pandemic hit Ukraine, progress on legislation regulating the country’s information environment stalled. But the draft law and the public reaction to it should serve as a warning to governments considering regulation. The concepts underpinning Ukraine’s law on disinformation were vague and ill-defined, leaving room for the politically-appointed Information Commissioner to interpret them as he sees fit, potentially using the cover of “national security imperatives” to support politically-motivated prosecutions against journalists or government critics. Though it remains in draft format, Ukraine’s attempt to crack down on disinformation paints a worrisome picture of the future of freedom of expression in the budding democracy.



RT, formerly Russia Today, is an state-backed news outlet used by Russia to disseminate divisive disinformation and fake news stories to audiences beyond the Russian context and around the world. Photo courtesy of Shutterstock.com



Conclusion: Five Principles for Democracy and Human Rights Based Social Media Regulation

At this writing, as the 2020 U.S. election nears, social media platforms, their content moderation decisions, and their effect on politics and public life are in the news on a weekly basis. It is no longer a question of if the United States or other countries will move to regulate social media to protect against disinformation and other online harms, but a question of when, and whether those eventual regulations will serve to protect or denigrate human rights, democracy, and freedom of expression. The case studies outlined in this paper provide some basic guidelines and lessons observed from the democracies that have already begun their regulatory journeys. This cross-continental, cross-cultural, and cross-contextual examination distill five guiding principles for any regulation aimed at countering online disinformation while protecting democratic ideals.

First, when defining what speech will be deemed harmful in counter disinformation regulation, precision is key. In both the Singaporean and Ukrainian cases, overbroad definitions contributed to fears that laws drafted under the guise of protecting national security and freedom of opinion would rather contribute to a chilling effect in society as well as empower the government to quash criticism. Even in the German case, in which robust German anti-hate speech laws were invoked as the framework for NetzDG takedowns, platforms themselves tended toward aggressive removal of reported content, with implications for content that reached far outside of Germany.

Second, the case studies demonstrate the importance of mandating transparency and oversight – ideally from an apolitical, expert body—as part of any regulatory framework. The necessity of these mandates is evident from Germany’s attempts to require enforcement disclosures from platforms, which all interpreted the tasking differently and reported varied data, making a cross-platform comparison of enforcement metrics nearly impossible. Germany’s NetzDG also shows that standardized data disclosures are not only necessary for reasons of basic compliance—it was through these disclosures that civil society groups and the German government first understood that platforms were under-reporting their takedowns—but to understand the effects a given law is having in the domestic context and beyond. Germany is now considering amendments to NetzDG based on early oversight metrics that would increase platform disclosures, specifically around automated procedures for identifying and removing illegal content. Brazil seems to have observed lessons from NetzDG and has included robust reporting provisions in its own bill to counter online disinformation.

Third, the importance of establishing an independent body to enforce and adjudicate counter disinformation law, ideally drawing on the existing structures and expertise of judicial authorities, cannot be understated. In Singapore, Ministers from the ruling party issue correction directions to the detriment of the fourth estate. Ukraine had similar plans for its Information Commissioner, who would be appointed by the President. Even if neither of these governments had questionable records of protecting freedom of speech, this structure would still be problematic. Changes in government happen frequently, and a new administration with less respect for fundamental freedoms might abuse the powers of the office against political opponents and critics. Any body overseeing these laws should be expert, politically insulated, and utilize the independent judiciary for adjudication.

Fourth, and related, users must have recourse above the platform level in order to dispute takedowns of their content. They must be informed of its removal, unlike under Germany’s NetzDG, as well as of the opportunities they have



to challenge the decision. These notifications should be easily accessible, written in plain language, and not deliberately obfuscated by platforms or governments. Users have the right to know why their content has been removed and what action they can take if they disagree with a platform or government's assessment. These appeals should move through the aforementioned independent, expert commissions charged with overseeing and enforcing social media regulation.

Finally, the development of any social media regulation should be pursued in consultation with civil society and other democratic partners, and with the use of existing legal frameworks including Article 19 of the ICCPR, the Manila Principles, and the 2017 Joint Declaration on Freedom of Expression and Fake News. Brazil's Marco Civil and its nascent counter disinformation legislation all benefited enormously from robust cooperation with civil society, which pushed the legislature to remove provisions that would have restricted speech and affected personal privacy. Ukraine's legislation, though still worrisome in its stalled state, also benefited from civil society input and pressure on relevant ministries.

As these cases have made clear, countering online disinformation through social media regulation is a delicate, nuanced issue with implications for governments, politicians, activists, the media, and the way ordinary citizens interact with their democratic institutions. It may be a delicate issue, but it ought not to be partisan one. Bipartisan leadership is needed in this space, and clear action. A politicized approach to crafting such legislation would only result in protections for the few, not the many. Disinformation's ultimate victim is democracy, and it is up to democratic governments to protect the principles on which they run in crafting responses to the threat.

The United States has a global responsibility to democracies around the world to hold accountable the social media companies headquartered in California's Silicon Valley. Inaction and abdication of responsibility threatens both democracy and freedom of expression not only at home, but abroad.

About the Authors

Nina Jankowicz studies the intersection of democracy and technology as the Wilson Center's Disinformation Fellow within the Science and Technology Innovation Program. She is the author of [How To Lose the Information War: Russia, Fake News, and the Future of Conflict](#) (Bloomsbury/IBTauris, July 2020). Ms. Jankowicz has advised the Ukrainian government on strategic communications under the auspices of a Fulbright-Clinton Public Policy Fellowship. Her writing has been published by *The New York Times*, *The Washington Post*, *The Atlantic*, and others. She is a frequent television and radio commentator on disinformation and Russian and Eastern European affairs. Prior to her Fulbright grant in Ukraine, Ms. Jankowicz managed democracy assistance programs to Russia and Belarus at the National Democratic Institute for International Affairs. She received her MA in Russian, Eurasian, and East European Studies from Georgetown University's School of Foreign Service, and her BA from Bryn Mawr College.

Shannon Pierson is a Research Assistant Intern for the Wilson Center's Science and Technology Innovation Program (STIP). Prior to working at the Wilson Center, Ms. Pierson worked as a Cybersecurity Research Consultant on multiple projects for the Microsoft Corporation's Defending Democracy Program—specifically on election security and internet governance legislation. She has also engaged in research related to Internet of Things (IoT) infrastructure security in smart buildings, law enforcement surveillance technology, and artificial intelligence (AI) technology policy. Her writing has been published by the Wilson Center, Disability & Society, and the Jackson School of International Studies' International Policy Institute. She received her BA in International Studies from the University of Washington in March 2020.








WOODROW WILSON INTERNATIONAL CENTER FOR SCHOLARS

The Woodrow Wilson International Center for Scholars, established by Congress in 1968 and headquartered in Washington, D.C., is a living national memorial to President Wilson. The Center's mission is to commemorate the ideals and concerns of Woodrow Wilson by providing a link between the worlds of ideas and policy, while fostering research, study, discussion, and collaboration among a broad spectrum of individuals concerned with policy and scholarship in national and international affairs. Supported by public and private funds, the Center is a nonpartisan institution engaged in the study of national and world affairs. It establishes and maintains a neutral forum for free, open, and informed dialogue. Conclusions or opinions expressed in Center publications and programs are those of the authors and speakers and do not necessarily reflect the views of the Center staff, fellows, trustees, advisory groups, or any individuals or organizations that provide financial support to the Center.





Woodrow Wilson International Center for Scholars
One Woodrow Wilson Plaza
1300 Pennsylvania Avenue NW
Washington, DC 20004-3027

The Wilson Center

-  www.wilsoncenter.org
-  wwics@wilsoncenter.org
-  facebook.com/woodrowwilsoncenter
-  [@thewilsoncenter](https://twitter.com/thewilsoncenter)
-  202.691.4000



Science and Technology Innovation Program

-  www.wilsoncenter.org/program/science-and-technology-innovation-program
-  stip@wilsoncenter.org
-  [@WilsonSTIP](https://twitter.com/WilsonSTIP)
-  202.691.4321

