

THE NEED FOR EMPIRICAL ASSESSMENTS OF COUNTER-DISINFORMATION LEGISLATION IN AN AGE OF POSTMODERN INFORMATION WARFARE

*Alexander Rojavin**

INTRODUCTION.....	1
I. BACKGROUND AND METHODOLOGY.....	4
II. CONCEPTUAL MODELS AND CAUSAL PATHWAYS.....	6
<i>A. Labeling Laws</i>	11
<i>B. Laws Criminalizing Social Media Content</i>	13
<i>C. Bans on Social Media Platforms</i>	15
III. DISCUSSION AND GAPS	16
<i>A. Labeling Laws</i>	16
<i>B. Content Criminalization Laws</i>	17
<i>C. Bans</i>	18
<i>D. Linking Laws to Macro Models of Disinformation</i>	18
CONCLUSION	19

INTRODUCTION

On January 21, 2015, Svetlana Davydova was arrested and charged with treason under Article 275 of the Criminal Code of the Russian Federation.¹ She was accused of divulging secret military operations that “could have been used against Russian national security.”² As it happened, in April 2014, Davydova had spotted a military convoy near her home in Smolensk Oblast’, reasoned that it was headed for Ukraine, and called the Ukrainian embassy to report the convoy’s whereabouts.³ After her neighbors tattled on

* Alexander Rojavin, J.D., Temple Beasley School of Law, 2020. The author is the Director of Counter Foreign Malign Influence Strategy at Deft9 Solutions, Inc., and a published translator. I extend a heartfelt thanks to Professor Scott Burris, both for allowing his students to contribute to real world research using an envelope-pushing method and for repeatedly taking the time to entertain a former student’s bushy-tailed attempts to apply that method in a thoroughly different context. Equally heartfelt gratitude goes out to Professor Nancy Knauer for always keeping her former students in mind.

1. Joshua Yaffa, *Why Was a Mother of Seven Arrested in Russia for Treason?*, NEW YORKER, <https://www.newyorker.com/news/news-desk/mother-seven-arrested-russia-treason> [<https://perma.cc/8G6K-A9C7>].

2. Grigory Tumanov, *Zhitel’niisa Smolenskoj oblasti obviniaetsia v gosizmene v pol’zu Ukrainy* [*Smolensk Oblast’ Citizen Accused of Treason in Ukraine’s Favor*], KOMMERSANT (Jan. 29, 2015), <https://www.kommersant.ru/doc/2656046> [<https://perma.cc/F5PJ-3SVQ>].

3. *Id.*

her, the Federal Security Service (FSB) came calling.⁴ However, the charge of treason created a metaphysical paradox, because at the time, in the world constructed by Kremlin-aligned media, *there were no Russian troops in Ukraine*.⁵ How, then, could Davydova have betrayed her country by revealing a military unit's position to the adversary . . . when *there was no adversary, and there was no unit*? The Kremlin, realizing the mutually exclusive nature of the charges vis-à-vis the party line, hastily released Davydova on February 3,⁶ after which it became known on March 13 that all charges against her were dropped.⁷

In November 2022, as we near the ninth month of all-out war between Russia and Ukraine and the ninth year since Russia first invaded Ukraine, this episode perfectly encapsulates the miasmatic senility within which the Kremlin wallows—and within which it has attempted to force the victims of its information war efforts to shamble for eight years. Since 2014, the Kremlin's strategy in the information war has been one of confounding polyphony. Gone is the painstaking, surgical precision of KGB disinformation campaigns, replaced by a tsunami of cheap, sloppy, easily disproven falsehoods designed not to take over the marketplace of ideas, but to burn it down, to erode trust in democratic institutions, and, critically, to keep audiences on the edge of hysteria.⁸ The purpose of the Kremlin's information war efforts domestically *and* abroad has been not simply to sow distrust, but to convince audiences to stop believing in the bare existence of truth—the apotheosis of postmodernism gone wrong.⁹ It has been a campaign to export a hallmark of authoritarianism wherein mutually exclusive ideas are peddled simultaneously—e.g., our enemies are simultaneously strong and weak, treacherously devious and laughably foolish, in accordance with what we need them to be on any given day.¹⁰

4. Liubov' Chizhova, "U vsekh est' bditel'nye sosedi" ["Everyone Has Vigilant Neighbors"], RFE/RL (Feb. 4, 2015), <https://www.svoboda.org/a/26829801.html> [https://perma.cc/72CE-LAR9].

5. *Rossia: mnogodetnuiu mat' obviniaiat v izmene rodine* [Russia: Mother of Several Charged with Treason], EURONEWS (Jan. 30, 2015), <https://ru.euronews.com/2015/01/30/russian-mother-of-seven-to-stand-trial-for-treason-over-ukraine-call> [https://perma.cc/C23Q-AEFA].

6. Grigory Tumanov, *Obviniayemaia v gosizmene Svetlana Davydova osvobozhdena iz SIZO* [Svetlana Davydova, Accused of Treason, Released from Detention], KOMMERSANT (Feb. 3, 2015), <https://www.kommersant.ru/doc/2659995> [https://perma.cc/ZHC6-576H].

7. *Advokat soobshchil o prekrashchenii dela obviniavsheisia v gosizmene Davydovoi* [Lawyer Says that Davydova Treason Case is Over], INTERFAX (March 13, 2015), <https://www.interfax.ru/russia/429569> [https://perma.cc/8QP5-N457].

8. See, e.g., Peter Pomerantsev, *Inside the Kremlin's Hall of Mirrors*, THE GUARDIAN (Apr. 9, 2015), <https://www.theguardian.com/news/2015/apr/09/kremlin-hall-of-mirrors-military-information-psychology> [https://perma.cc/NL9S-FZT7].

9. *Id.*; JUDIT BAYER, NATALIJA BITUIKOVA, PETRA BÁRD, JUDIT SZAKÁCS, ALBERTO ALEMANN & ERIK USZKIEWICZ, POL'Y DEP'T FOR CITIZENS' RTS. AND CONST. AFFS., DIRECTORATE GEN. FOR INTERNAL POL'YS OF THE EUR. UNION, DISINFORMATION AND PROPAGANDA – IMPACT ON THE FUNCTIONING OF THE RULE OF LAW IN THE EU AND ITS MEMBER STATES 64–65 (2019).

10. See Alexander Rojavin, *Dominant Rhetorical Themes in the Third Era of Post-Soviet Russian Media History*, TEMP. L. AND PUB. POL'Y BLOG (May 4, 2022), <https://www2.law.temple.edu/lppp/dominant-rhetorical-themes-in-the-third-era-of-post-soviet-russian-media-history> [https://perma.cc/ASY5-HRMB] (illustrating the mutually exclusive nature of the Kremlin's rhetoric with the "Provocation" and "State Weakness and Dysfunction" meta-narratives).

This modern (or postmodern) hydra of disinformation is a many-headed, self-regenerating beast requiring multipronged, interdisciplinary countermeasures. Democracies worldwide have been experimenting with different approaches, metaphorical vaccines and sera to disinformation, including state-run fact-checking operations, strategic communications, cross-sector funding and collaboration, and media literacy initiatives.¹¹ However, despite an array of regulatory steps, most nations have shied away from passing legislation targeting disinformation spread.¹² Only a few have implemented legal interventions, and they have done so to varying degrees of success, typically in the face of political headwinds and without a clear understanding of the social, technological, and cognitive causal chains the laws would trigger.¹³ Nevertheless, tailored legislation remains a viable countermeasure to disinformation and the information war efforts of the West's geopolitical adversaries. The issue is that the few enacted laws have not been appraised in an empirically sound manner, and countries will be reticent to consider further counter-disinformation legal interventions without understanding the causal chains of the legislation already in effect. Such (understandable) reticence would deprive the West of a potentially important counter-disinformation tool.

The solution to this dilemma may lie in international public health policy, where legal epidemiology—"the study of law as a factor in the cause, distribution, and prevention of disease and injury"—has been a burgeoning methodology for the past several years.¹⁴ Even without the obvious relationship between disinformation and public health that came to light during the COVID-19 pandemic,¹⁵ and ignoring the easy-fitting biological framing of the spread of (dis)information,¹⁶ legal epidemiology's bridging between lawyers, policymakers, practitioners, and scholars¹⁷ is precisely the remedy that the fledgling realm of counter-disinformation legislation needs.

In particular, scholars of counter-disinformation policy should adopt and implement the Identifying Data for the Empirical Assessment of Law (IDEAL) method to evaluate the effectiveness, causal pathways, and outcomes of implemented

11. See, e.g., PAUL BUTCHER, *DISINFORMATION AND DEMOCRACY: THE HOME FRONT IN THE INFORMATION WAR* (2019), https://wms.flexious.be/editor/plugins/imagmanager/content/2140/PDF/2019/190130_Disinformationdemocracy_PB.pdf [<https://perma.cc/X8TR-YUR7>] (surveying ongoing counter-disinformation efforts).

12. See *id.* at 11.

13. See *id.*

14. *Legal Epidemiology*, CTRS. FOR DISEASE CONTROL & PREVENTION (May 12, 2021), https://www.cdc.gov/dhbsp/policy_resources/legal_epi.htm [<https://perma.cc/3KTN-D27C>].

15. See, e.g., Erin Simpson & Adam Conner, *Fighting Coronavirus Misinformation and Disinformation*, CTR. FOR AM. PROGRESS (Aug. 18, 2020), <https://www.americanprogress.org/article/fighting-coronavirus-misinformation-disinformation> [<https://perma.cc/E2AD-T7MC>]; see also *GEC Counter-Disinformation Dispatches #5*, GLOBAL ENGAGEMENT CTR., U.S. DEP'T OF STATE (May 27, 2020), <https://e.america.gov/t/ViewEmail/i/B253AE94519376FD2540EF23F30FEDED/F2AB8F86DC5635A4AF060D655554232> [<https://perma.cc/Y3AP-MT7S>].

16. See, e.g., RICHARD DAWKINS, *THE SELFISH GENE* (1976) (framing the diffusion of ideas in biological terms).

17. *Legal Epidemiology*, *supra* note 14 ("Legal epidemiology is a field that unites public health lawyers, researchers, and practitioners toward a common pursuit to inform public health decision making that affects which laws work, for whom, and in what context.").

counter-disinformation laws. IDEAL, developed by a team of public health scholars and lawyers¹⁸ to evaluate health outcomes of abortion regulations, is used to “form . . . a causal logic model setting out events and outcomes that may plausibly occur *assuming* key facts that can and should be investigated in future research.”¹⁹ The models’ “value lies in identifying evidence that can be useful in making tentative inferences about legal effects in the absence of direct evidence, and in pointing to important research questions.”²⁰ “The IDEAL method attempts to create an objective framework for crystallising the various influences and consequences attributable to the impact of specific . . . restrictions, leading to the identification of untapped scientific evidence on plausible effects of the law.”²¹ Absent direct evidence, the IDEAL method “can also serve a precautionary role” by shedding light on “non-trivial” links in the causal chain that policymakers must heed when shaping a legal intervention.²² Just as IDEAL was originally applied to six types of laws regulating abortion access and provision, so should it be applicable to different types of laws designed to combat the spread of disinformation—laws whose construction would undeniably benefit from causal modeling and empirical evaluations, which have thus far been limited.

This Article presents an initial application of IDEAL to the counter-disinformation context, (1) highlighting links in multiple causal models that have been empirically studied, thereby allowing one to infer a causal path affected by relevant law, and (2) identifying empirical gaps in the models. The Article’s objective is to provide a concrete example of how to model and quantify the effects of counter-disinformation legislation, which may well be an effective instrument against the postmodern information war efforts of the Kremlin and the more traditional strategies of other adversaries. It need only be enabled by the proper methodology.

I. BACKGROUND AND METHODOLOGY

Implementation of the IDEAL method to counter-disinformation legislation is similar to its use in the abortion law context. As Burris et al. note, “[a]bortion laws, like other legal interventions, operate in a complex and context-dependent manner, with multiple components that may be non-linear in their effects,”²³ and the same principle applies to counter-disinformation laws. IDEAL was designed to assess laws from different countries and legal systems, which makes it a perfect analytical tool for a global legal landscape as uneven as the counter-disinformation one.²⁴ Moreover, IDEAL’s

18. Professor Scott Burris, Professor of Law and the Director of the Center for Public Health Law Research at Temple University’s Beasley School of Law, led this initiative. I was fortunate to be one of the students in his International Health Law and Policy class, during which we received hands-on experience with the IDEAL method, applying it to several kinds of laws restricting abortion access.

19. Scott Burris, Adrienne R. Ghorashi, Lindsay Foster Cloud, Rachel Rebouché, Patty Skuster & Antonella Lavelanet, *Identifying Data for the Empirical Assessment of Law (IDEAL): A Realist Approach to Research Gaps on the Health Effects of Abortion Law*, *BMJ GLOB. HEALTH*, 2 (2021).

20. *Id.* at 2.

21. *Id.* at 8.

22. *Id.* at 2.

23. *Id.*

24. *Id.* (explaining that though most research assessing the effects of abortion laws focuses on single jurisdictions, IDEAL was developed with a global population in mind).

purpose is “to support the development of evidence-based guidelines and practices by identifying . . . evidence . . . that does not explicitly address law, but can nonetheless enhance the understanding of legal effects and identify priority research topics.”²⁵ Additionally, as Burris et al. acutely point out, though “[h]uman rights approaches have galvanised abortion law reform across numerous countries . . . human rights analysis is not designed to empirically assess how legal provisions regulating abortion shape the actual delivery of abortion services and outcomes.”²⁶ It has become similarly vogue to lambast counter-disinformation efforts through a human rights lens,²⁷ and this criticism is equally ill-equipped to develop empirical guidelines for crafting such policies.²⁸ At the same time, current literature laments the limitations of extant methodologies to conduct empirical assessments of counter-disinformation laws.²⁹ Thus, IDEAL is exactly the method with which we can begin to fill the gaps in our understanding of the effects of counter-disinformation laws passed by democratic governments.

The IDEAL method unfolds in three steps: the identification of empirical research showing the effects of given legislation, the development of causal logic models for relevant legal interventions, and a second rapid scan informed by the developed causal models.³⁰

For the first step, a rapid scan was conducted for research on counter-disinformation legislation generally. A broad net was cast, and search terms included *counter-disinformation*, *law*, *disinformation*, *propaganda*, *misinformation*, *fake news*, and *information war*. These searches were run in Proquest’s Policy File Index, the *Harvard Misinformation Review*, and Proquest’s Politics Collection on the assumption that these were the databases most likely to have relevant, up-to-date research on the existing laws. Each search was supplemented with an additional search for gray literature, and search results’ references were reviewed for additional relevant material. The first step concluded with a data abstraction table summarizing the discovered empirical assessments.

For step two, causal models were developed for the three kinds of legal interventions revealed by the research in the first step. These models depict conceivable pathways from the intervention’s implementation to the outcomes discussed in the research.

25. *Id.*

26. *Id.* at 1.

27. See, e.g., Sara Dillon, *The Propaganda Conundrum: How to Control This Scourge on Democracy*, 23 OR. REV. INT’L L. 123, 125 (2022); see also Henning Lahmann, *Protecting the Global Information Space in Times of Armed Conflict*, 915 INT’L REV. OF THE RED CROSS 1227 (2020).

28. See, e.g., Evelyn Douek, *The Limits of International Law in Content Moderation*, 6 U.C. IRVINE J. INT’L., TRANSNAT’L. & COMPAR. L. 37 (2020).

29. HEIDI TWOREK & PADDY LEERSSEN, TRANSATLANTIC WORKING GRP., AN ANALYSIS OF GERMANY’S NETZDG Law 7 (2019) (“The law’s actual impacts on hate speech may be difficult to prove empirically, since this complex phenomenon is influenced by countless other factors as well, including political, cultural, demographic, and economic shifts.”); NINA JANKOWICZ & SHANNON PIERSON, WILSON CTR. FREEDOM AND FAKES: A COMPARATIVE EXPLORATION OF COUNTERING DISINFORMATION AND PROTECTING FREE EXPRESSION 11 (2020) (“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.”).

30. See Burris et al., *supra* note 19, at 2.

Step three relied on the models generated in step two to conduct a second rapid scan of nonlegal studies assessing “whether the processes and outcomes posited in the models do, in fact, occur, and with what frequency, severity or consequence.”³¹ It is this step that crystallizes IDEAL’s utility, enabling the inference of causality even when there is a lack of direct research on a law’s effects.³² It is this research of the intermediate links in the causal chain that can “support plausible inferences of causality for practical policy and guideline development purposes.”³³

For the purposes of this Article, ostensible counter-disinformation or counter-propaganda laws passed by authoritarian or quasi-authoritarian regimes were ignored. The legislative process and purpose of such laws are not congruent with the process and purpose of genuine—if poorly crafted—laws passed by democratic governments. Authoritarian regimes use the cover of “counter-propaganda” measures to suppress free speech, typically through the weaponization of the criminal code—there is a clear difference between authoritarian legislation designed to safeguard the power of the regime and democratic legislation designed (again, possibly poorly designed) to protect the marketplace of ideas.³⁴ For this reason, research on the mechanisms of sham legislation passed by Russian, Chinese, Iranian, and other authoritarian regimes was omitted, as was research on Singapore’s Protection from Online Falsehoods and Manipulation Act,³⁵ considering its quasi-authoritarian nature.³⁶

Also omitted was research into cybersecurity and electoral finance laws in favor of research on laws expressly designed to counter the spread of disinformation.³⁷ For this first implementation of IDEAL in the counter-disinformation context, the research net, though broadly cast, still had to be narrow enough to avoid findings that were overly sprawling.

II. CONCEPTUAL MODELS AND CAUSAL PATHWAYS

Just as there is a deficit of empirical studies in the abortion policy context, there are even fewer such studies assessing either the effects of counter-disinformation laws or

31. *Id.*

32. *Id.*

33. *Id.*

34. See, e.g., Amr Hamzawy, *Legislating Authoritarianism: Egypt’s New Era of Repression*, CARNEGIE ENDOWMENT FOR INT’L PEACE 6–19 (Mar. 16, 2017), <https://carnegieendowment.org/2017/03/16/legislating-authoritarianism-egypt-s-new-era-of-repression-pub-68285> [<https://perma.cc/5DMA-2EQC>] (discussing how an authoritarian regime passes intentionally vague criminal laws to enable broad implementation in its fight against speech it finds threatening to its hold on power).

35. Protection from Online Falsehoods and Manipulation Act 2019, GOV’T GAZETTE ACTS Supplement No. 26, June 28, 2019, (Sing.), <https://sso.agc.gov.sg/Acts-Supp/18-2019/Published/20190625?DocDate=20190625> [<https://perma.cc/KVY2-FYSX>].

36. See, e.g., Tyler Roylance, *Singapore and the Limits of Authoritarian Prosperity*, FREEDOM HOUSE (Mar. 26, 2015), <https://freedomhouse.org/article/singapore-and-limits-authoritarian-prosperity> [<https://perma.cc/FK4K-FVYE>].

37. However, the literature suggests that there may well exist a nexus between counter-disinformation efforts and cybersecurity and electoral finance laws, so research on how such measures could be explicitly tailored as part of counter-disinformation legislation would be welcome. See, e.g., BAYER, ET. AL, *supra* note 9, at 107–10.

counter-disinformation itself on society.³⁸ For some types of laws, like conscientious objection provisions, Burris et al. faced the problem that though there were multiple instances of the law's enactment, there were few empirical studies directly linking the law to public health outcomes.³⁹ In the counter-disinformation context, there are both few enacted laws *and* few empirical studies—including studies focusing on each causal chain's various links, which are what enables IDEAL. Though this makes the initial application of IDEAL to the new context somewhat frustrating, it simultaneously underscores its urgency and value, as “causal modelling is an expeditious way to identify data that measures the effects of processes that law requires or will influence.”⁴⁰

Research revealed three main types of legal interventions currently enacted by democratic regimes, as shown in Table 1. These interventions are (1) laws requiring an entity, typically one publishing communications on behalf of a foreign principal, to label any communications as being disseminated by a foreign agent, (2) laws criminalizing the publication of false information online, and (3) laws or executive acts banning a social media outlet, online print news publication, or television channel. There exists enough research to enable the creation of causal models for each of these types of law, and their causal pathways and outcomes are presented in Table 1, with representative examples of studies supporting them.

Table 1. Causal Pathways Linking Legal Regulations of Disinformation to Plausibly Related Outcomes and Relevant Research.

Legal Intervention	Sample Acts	Select Causal Pathways	Plausibly Related Outcomes	Examples of Relevant Research
---------------------------	--------------------	-------------------------------	-----------------------------------	--------------------------------------

38. BAYER ET AL., *supra* note 9, at 128 (“Empirical studies about the effect of propaganda on public opinion are rare . . .”).

39. Burris et al., *supra* note 19, at 23–25 (modeling the effects of conscientious objection laws largely on the basis of indirect empirical research rather than direct assessments of such laws).

40. *Id.* at 3.

Law requires affiliation labeling.	Foreign Agents Registration Act (United States).	<ul style="list-style-type: none"> • Entity ceases spreading information. • Entity self-labels. • Entity refuses to self-label. • Enforcement mechanisms. 	<ul style="list-style-type: none"> • Entity keeps spreading information. • Prevention of the further spread of false information. • Increased public awareness of false information. • Prosecution as a result of noncompliance. • Reputational costs for the labeled entity. • Streisand effect. 	TODD C. HELMUS ET AL., RUSSIAN PROPAGANDA HITS ITS MARK: EXPERIMENTALLY TESTING THE IMPACT OF RUSSIAN PROPAGANDA AND COUNTER-INTERVENTIONS (2020).
Law criminalizes publishing false information online.	NetzDG (Germany); Law Against the Manipulation of Information (France).	<ul style="list-style-type: none"> • Social media user behavior. • Hosting platform behavior. • Oversight effectiveness. • Enforcement mechanisms. 	<ul style="list-style-type: none"> • Removal of false information. • Removal of protected information. • Prevention of the further spread of false information. • No effect. • Streisand effect. 	JUDIT BAYER ET AL., DISINFORMATION AND PROPAGANDA – IMPACT ON THE FUNCTIONING OF THE RULE OF LAW IN THE EU AND ITS MEMBER STATES 96–107 (2019); HEIDI TWOREK & PADDY LEERSSEN, AN ANALYSIS OF GERMANY’S NETZDG Law (2019); NINA JANKOWICZ & SHANNON PIERSON, FREEDOM AND FAKES: A COMPARATIVE EXPLORATION OF COUNTERING DISINFORMATION AND PROTECTING

				FREE EXPRESSION 8–12 (2020).
Law or national security council bans social media platform, self-described news outlet, or TV channel spreading false information.	Law Against the Manipulation of Information (France); NSDC decisions (Ukraine); 2022 ban of RT and Sputnik (European Union).	<ul style="list-style-type: none"> • Platform, outlet, or channel shuts down. • Access to platform, outlet, or channel is restricted on given nation’s territory. 	<ul style="list-style-type: none"> • Removal of false information. Prevention of the further spread of false information. • Reputational costs among Western observers. 	RAPHAEL S. COHEN ET AL., COMBATING FOREIGN DISINFORMATION ON SOCIAL MEDIA: STUDY OVERVIEW AND CONCLUSIONS 69–70 (2021); Anton Dek et al., <i>The Effects of Banning the Social Network VK in Ukraine, in RESPONDING TO COGNITIVE SECURITY CHALLENGES</i> 38, 38–58 (Anna Reynolds & Mike Collier eds., 2019).

Prior to developing conceptual models for the laws, it was also necessary to create a model for the spread of disinformation itself. This is to enable the identification of where in the larger causal chain the modeled legal interventions actually intervene—and what links of the chain could be targeted by future legislation. A review of the literature reveals broad consensus on the societal, political, and technological mechanisms and purposes of disinformation.⁴¹ Synthesizing the literature could yield a conceptual model like in Figure 1, which focuses specifically on disinformation spread on social media (and not via television, radio, print media, or traditional online news outlets).

41. See, e.g., *Countering Russian Disinformation*, CTR. FOR STRATEGIC & INT’L STUD. (Sept. 23, 2020), <https://www.csis.org/blogs/post-soviet-post/countering-russian-disinformation> [<https://perma.cc/GP2L-E8GU>] (“Disinformation is a tool commonly used . . . to sow discord, undermine faith in governing institutions, stoke fear and anxiety . . .”); BAYER ET AL., *supra* note 9, at 16 (“Increasing polarisation is an outspoken purpose of the Kremlin’s information war ‘to destabilize a society and a state . . .’”).

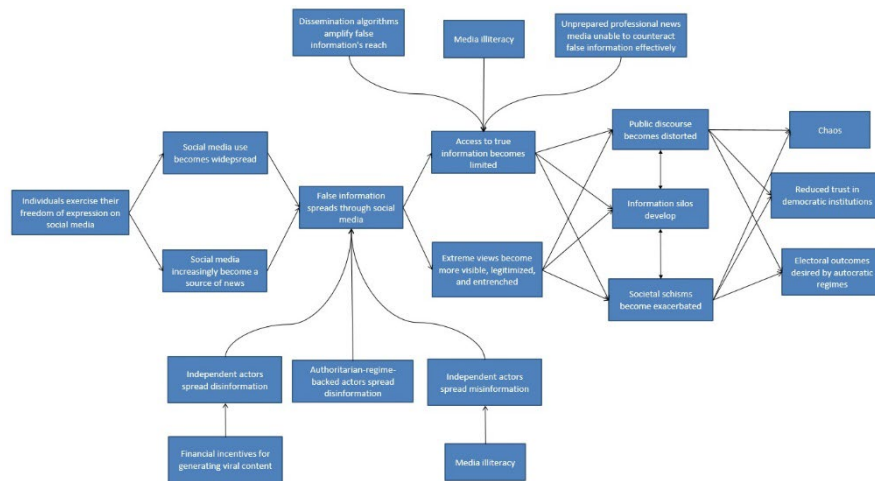


Figure 1. Causal Model of Disinformation Spread on Social Media

“At the core, disinformation is designed to destabilize.”⁴² With the post-2014 Kremlin as the strategy’s poster child, disinformation’s goal is to reduce Western audiences’ trust in democratic institutions,⁴³ prime audiences to vote in a way that favors electoral outcomes desired by the West’s adversaries,⁴⁴ and generally cause chaos to destabilize the domestic environment, possibly through violence.⁴⁵ The spread of disinformation achieves this goal through concrete causal pathways, taking advantage of audiences’ media illiteracy,⁴⁶ favorable social media algorithms,⁴⁷ an unprepared fourth estate,⁴⁸ and existing financial incentives for spreading viral content⁴⁹ to drown out true information and normalize fringe views. Unanswered, this results in hyper-partisan,

42. BEN RADERSTORF & MICHAEL J. CAMILLERI, *THE DIALOGUE*, ONLINE DISINFORMATION IN THE UNITED STATES: IMPLICATIONS FOR LATIN AMERICA 9 (2019), <https://www.thedialogue.org/wp-content/uploads/2019/06/OnlineDisinfoinUS.pdf> [<https://perma.cc/27WA-KMUJ>].

43. See, e.g., *Countering Russian Disinformation*, *supra* note 41.

44. See Andrew Weisburd, Clint Watts & JM Berger, *Trolling for Trump: How Russia Is Trying to Destroy Our Democracy*, *WAR ON THE ROCKS* (Nov. 6, 2016), <https://warontherocks.com/2016/11/trolling-for-trump-how-russia-is-trying-to-destroy-our-democracy/> [<https://perma.cc/W78K-N2HK>].

45. See e.g., Anne Applebaum & Peter Pomerantsev, *How to Put Out Democracy’s Dumpster Fire*, *THE ATLANTIC* (Mar. 8, 2021), <https://www.theatlantic.com/magazine/archive/2021/04/the-internet-doesnt-have-to-be-awful/618079/> [<https://perma.cc/XN2D-VBX5>].

46. See, e.g., ALICE HUGUET, GARRETT BAKER, LAURA S. HAMILTON & JOHN F. PANE, *MEDIA LITERACY STANDARDS TO COUNTER TRUTH DECAY* (2021), https://www.rand.org/content/dam/rand/pubs/research_reports/RR100/RR112-12/RAND_RRA112-12.pdf [<https://perma.cc/TZC4-N4RJ>].

47. THE ANTI-DEFAMATION LEAGUE ET AL., *TRAINED FOR DECEPTION: HOW ARTIFICIAL INTELLIGENCE FUELS ONLINE DISINFORMATION* (2021), https://d1y8sb8igg2f8e.cloudfront.net/documents/Trained_for_Deception_How_Artificial_Intelligence_Fuels_Online_Disinformation1.pdf [<https://perma.cc/ANN8-EDBM>].

48. BAYER ET AL., *supra* note 9, at 59–60.

49. See EUR. COMM’N, *ASSESSMENT OF THE CODE OF PRACTICE ON DISINFORMATION – ACHIEVEMENTS AND AREAS FOR FURTHER IMPROVEMENT* 4 (2020) (noting the need to disrupt ad monetization practices abetting the spread of disinformation).

mutually unintelligible public discourse,⁵⁰ ideological siloing,⁵¹ and societal schisms on the basis of affective ideologies⁵² that can, in turn, lead to the aforementioned strategic goals.

A. Labeling Laws

A labeling law, such as the United States' Foreign Agents Registration Act (FARA), signed in 1938 as an explicit measure to counter Nazi and Soviet propaganda efforts,⁵³ requires an entity operating on behalf of a foreign principal to accompany all public communications with conspicuous notice of the entity's foreign allegiance. A labeling law leverages public shaming and transparency to stymie such an entity's communication efforts.⁵⁴

The model in Figure 2 conveys the major inflection points in the causal chain set off by the implementation of a labeling law. One experiment tested the effectiveness of labeling in reducing Facebook users' vulnerability to Russian memes,⁵⁵ and there is a breadth of literature on labeling effectiveness in other contexts, e.g. warning labels in consumer products, but no studies yet link FARA directly to societal outcomes.⁵⁶

50. Alexander Rojavin, *Disinformation and Polarization: The Other Pandemic*, TEMP. L. AND PUB. POL'Y BLOG (Feb. 5, 2021), <https://www2.law.temple.edu/lppp/disinformation-and-polarization-the-other-pandemic/> [https://perma.cc/2ZU7-ZQMT].

51. See, e.g., Jonathan Reisher, *The Effect of Disinformation on Democracy: The Impact of Hungary's Democratic Decline*, XIV CES WORKING PAPERS 42, 48 (2022) (remarking how disinformation spread via social media reinforces and thrives in environments marked by information silos).

52. BAYER ET AL., *supra* note 9, at 48.

53. *Criminal Resource Manual 2062: Foreign Agents Registration Act Enforcement*, U.S. DEP'T OF JUST., https://web.archive.org/web/20150107230654/http://www.justice.gov/usao/eousa/foia_reading_room/usam/title9/crm02062.htm [https://perma.cc/ZZ2P-EXMB].

54. Alexander Rojavin, *Out of the Looking Glass: Reassessing the Foreign Agents Registration Act to Counter 21st-Century Foreign Information Warfare*, 83 PHILA. LAW. 4, 17 (2021).

55. See TODD C. HELMUS, JAMES V. MARRONE, MAREK N. POSARD & DANIELLE SCHLANG., RAND CORP., RUSSIAN PROPAGANDA HITS ITS MARK: EXPERIMENTALLY TESTING THE IMPACT OF RUSSIAN PROPAGANDA AND COUNTER-INTERVENTIONS 36–37 (2020).

56. *Id.* at 16–17 (highlighting the lack of publicly available statistics or studies appraising the effectiveness of FARA and recommending that such studies be conducted).

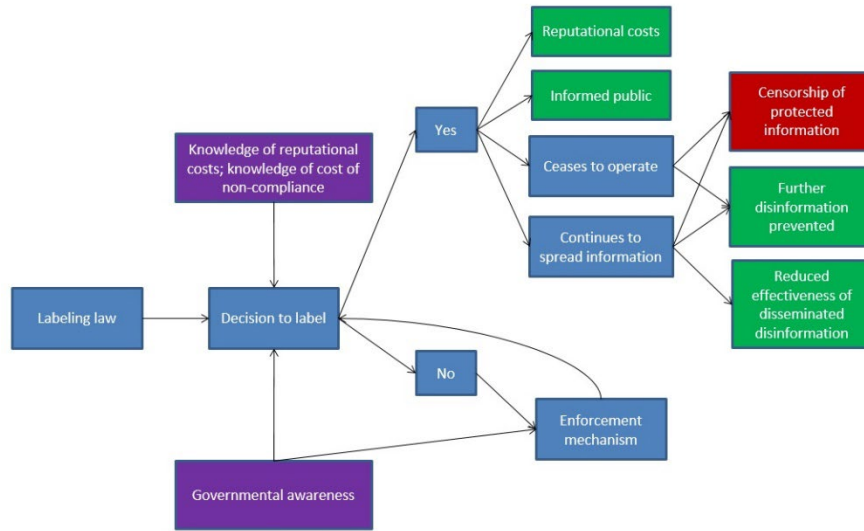


Figure 2. Labeling Law Causal Model

After such a labeling law’s enactment, an entity identified by the law must register with the relevant federal organ and then make plain its affiliation in every public communication.⁵⁷ The decision to self-label is largely contingent on the entity’s awareness of possible reputational costs labeling could incur and the potential cost of noncompliance.⁵⁸ The government’s assiduousness in monitoring affected entities and enforcing the law also influences this decision.⁵⁹ Should an affected entity decide to ignore the law, the enforcement mechanism is triggered successfully only if the government has a capable enough monitoring system, which is complemented by means of compelling an entity to register. Absent an effective monitoring operation, an entity can skirt the law indefinitely.

If, however, an entity chooses or is compelled to obey and begins to self-label, the limited research suggests that the outcomes may be largely positive. Labeling improves the audience’s awareness of the issue of source alignment generally, possibly improving the audience’s media literacy.⁶⁰ Labeling also results in reputational costs for entities

57. *Id.* at 16.

58. See Monica Romero, *How Far Will FARA Go? The Foreign Agents Registration Act and the Criminalization of Global Human Rights Advocacy*, 96 WASH. L. REV. 2, 695, 713–19 (2021) (explaining the behavior of a few entities that registered or were forced to register under FARA).

59. Robert Kelner et al., *The Foreign Agents Registration Act (“FARA”): A Guide for the Perplexed*, COVINGTON (July 26, 2019), https://www.cov.com/-/media/files/corporate/publications/2018/01/the_foreign_agents_registration_act_fara_a_guide_for_the_perplexed.pdf [<https://perma.cc/ST2F-VNMY>] (remarking that before the government began aggressively enforcing FARA, it was “a very still backwater” of U.S. law).

60. See TODD C. HELMUS, ET AL., *supra* note 55.

operating on behalf of the democracy's adversaries.⁶¹ After beginning to label, an entity may choose not to change the content of its communications whatsoever, or its communications may become more tempered.⁶² The entity may alternatively decide that it has been too effectively hamstrung and cease operations entirely, in which case one spout of disinformation is closed. However, in either case, the effectiveness of any disinformation peddled by the entity is diminished, though the risk of the law's being too broad remains. If too broad, it would inadvertently catch legitimate allies of democracy that happen to be operating on behalf of a foreign principal, thereby styming protected speech or communications that may legitimately be in the interest of public discourse.⁶³

B. Laws Criminalizing Social Media Content

Figure 3 represents the causal pathways in play when a law criminalizing select social media content goes into effect.

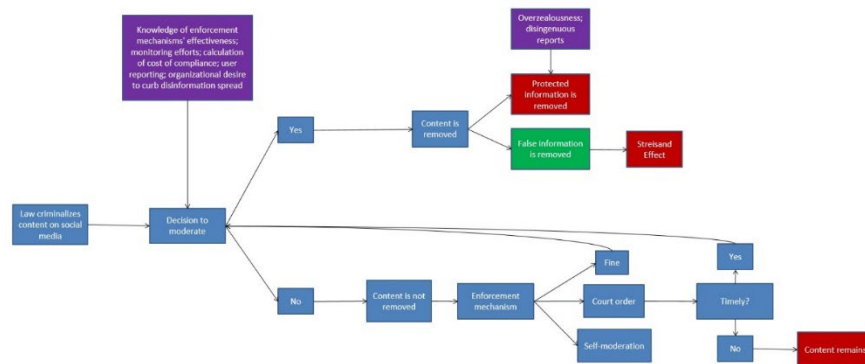


Figure 3. Content Criminalization Model

A law like Germany's NetzDG⁶⁴ or France's Law Against the Manipulation of Information⁶⁵ provides the characteristics of certain types of content and mandates the circumstances in which the relevant social media platform is responsible for removing such content.⁶⁶ These circumstances may include a time period during which the law is

61. See, e.g., Romero, *supra* note 58, at 715 (explaining that though the Department of Justice's FARA case against the Peace Information Center was dismissed, the reputational harm created by the investigation ultimately caused the Center to dissolve).

62. Rojavin, *supra* note 54, at 16–17.

63. BAYER ET AL., *supra* note 9, at 98 (discussing the risks of labeling content as disinformation).

64. See Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerkdurchsetzungsgesetz--NetzDG] [Network Enforcement Act], Sept. 1, 2017, BUNDESGESETZBLATT, Teil I [BGBL I] at 3352 (Ger.), https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/NetzDG_engl.pdf?__blob=publicationFile&v=2 [<https://perma.cc/W2B8-JWHT>].

65. Loi 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [Law 2018-1202 of December 22, 2018 on the fight against the manipulation of information], JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Dec. 23, 2018.

66. See Julian Jaurisch & Théophile Lenoir, *Disinformation: The German Approach and What to Learn from It*, INSTITUT MONTAIGNE (Feb. 28, 2020), <https://www.institutmontaigne.org/en/analysis/disinformation-german-approach-and-what-learn-it> [<https://perma.cc/W3S6-7DL3>].

in effect or the minimum number of users a social media platform must have before being affected by the law.⁶⁷

Empirical studies of such laws have been inhibited by a lack of uniformity in reporting and monitoring standards, which make affected social media platforms' operations rather opaque.⁶⁸ However, there is enough evidence to generate the model above.

Once a law criminalizing social media content is enacted, an affected social media platform faces a choice pertaining to how diligently it wishes to moderate its users' posts. This decision is a function of an array of externalities, including the platform's knowledge of the (in)effectiveness of the government's oversight and enforcement mechanisms, calculation of the cost of noncompliance, assiduousness of its users' flagging of suspect content, and the extent to which it genuinely wishes to curb the spread of disinformation.⁶⁹ If a platform chooses to act and remove content, then one of two things is possible: either it eliminates actual disinformation, or, due to a mix of overzealousness and disingenuous reporting on the part of bad-faith or ill-informed users, it removes innocuous or truthful content.⁷⁰ If content moderation is automated, the risk of overmoderation increases.⁷¹ If the removed content truly was disinformation, then there is the risk of the Streisand effect occurring, whereby removing or prohibiting certain content inadvertently publicizes it further.⁷²

If, on the other hand, a platform chooses not to fulfill its obligations under the law, a variety of enforcement mechanisms may be triggered. The platform could be fined, after which the causal chain circles back to making a decision—a loop that can go on indefinitely, so long as the platform is unfazed by the cost.⁷³ A judicial authority may be called to compel the platform to remove the offensive material, provided that the authority can do so in a timely manner.⁷⁴ However, this enforcement mechanism is accompanied with the risk of judicial overload.⁷⁵ Alternatively, the social media's internal oversight mechanisms could be triggered, invariably delaying any outcome.⁷⁶

67. The Law Against the Manipulation of Information is in effect in the three months directly preceding a nationwide election or referendum. *Id.* NetzDG affects social media platforms with over two million users. See Jenny Gesley, *Germany: Network Enhancement Act Amended to Better Fight Online Hate Speech*, LIBR. OF CONG. (2021), <https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-online-hate-speech/> [<https://perma.cc/KP3E-CKPL>].

68. See TWOREK & LEERSSEEN, *supra* note 29, at 8.

69. See *id.* at 4–7.

70. BAYER ET AL., *supra* note 9, at 99.

71. Céline Castets-Renard, *Algorithmic Content Moderation on Social Media in EU Law: Illusion of Perfect Enforcement*, 2020 U. ILL. J.L. TECH. & POL'Y 283, 313–16.

72. See TWOREK & LEERSSEEN, *supra* note 29, at 3; BAYER ET AL., *supra* note 9, at 101.

73. More research into platform behavior is necessary here; though the literature believes that platforms will not engage in such loops because of the steep fines involved, there have been but a few instances of the fines coming into play. See JANKOWICZ & PIERSON, *supra* note 29, at 10–11.

74. See BAYER ET AL., *supra* note 9, at 101.

75. *Id.*

76. See JANKOWICZ & PIERSON, *supra* note 29, at 10–11.

Hanging over the entire causal chain is the specter of delay. The longer that a platform neglects to remove transgressive content, the longer it has to spread.⁷⁷ If a time-bound judicial authority is empowered to order that content be taken down, there is the risk of a platform running out the clock until the deadline passes and the disinformation has already taken its toll, possibly achieving the strategic goal for which it was deployed.⁷⁸

C. Bans on Social Media Platforms

The model depicted in Figure 4 shows the causal chain effected by an outright ban on a social media platform.

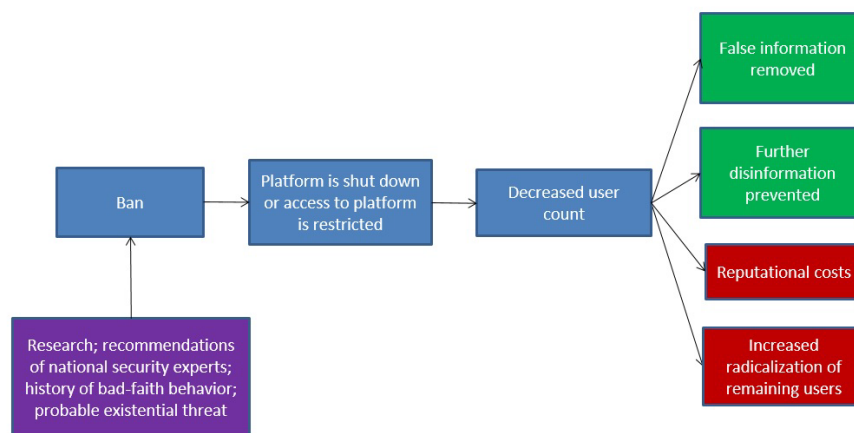


Figure 4. Ban Causal Model

Though there is limited empirical research directly linking bans to societal outcomes, there is enough to conclude that a ban, deemed “the most draconian defense countermeasure,” is the most effective step a government can take.⁷⁹ Once a nation restricts access to a foreign social media platform, the effect is instantaneous, with much fewer people having access to the disinformation spread via the platform.⁸⁰ Research shows that a ban precipitously cuts the number of users who continue to use (typically via a VPN) a banned platform, either because of their law-abiding nature or because using a VPN represents too great a hassle.⁸¹ However, users continuing to use the platform may become more radicalized, considering the exodus of much of the

77. See BAYER ET AL., *supra* note 9, at 101 (“[C]alling on an interim relief judge to act—the fastest form of serving justice—will always be too little too late.”).

78. *Id.*

79. RAPHAEL S. COHEN, NATHAN BEAUCHAMP-MUSTAFAGA, JOE CHERAVITCH, ALYSSA DEMUS, SCOTT W. HAROLD, JEFFREY W. HORNUNG, JENNY JUN, MICHAEL SCHWILLE, ELINA TREYGER & NATHAN VEST, *COMBATING FOREIGN DISINFORMATION ON SOCIAL MEDIA: STUDY OVERVIEW AND CONCLUSIONS* 69 (2021).

80. *Id.* (describing how Ukraine’s 2017 VK ban drastically cut the number of Ukrainian users).

81. *Id.*

opposition.⁸² A ban also comes with reputational costs for the government in the eyes of many Western observers, who currently find it difficult to square such a countermeasure with Western values.⁸³

In a democracy, a measure as radical and decisive as a ban is not made lightly, and it is unlikely that it will land on a good-faith actor. The purple node in the model represents the substantial number of externalities influencing the implementation of such a measure. A democracy issues a ban as a last resort, so the assumption is that considerable deliberation went into the decision, much informed discussion and research, and it was brought about by a consistent history of bad-faith behavior by the entity targeted by the ban. The few democratic regimes that have implemented a ban did so only in the context of war to ban entities controlled by an aggressor state.⁸⁴

III. DISCUSSION AND GAPS

Lamentably, “gaps” may be a slight misnomer, as the dearth of empirical research reveals that the entire field is just one massive gap.⁸⁵ There are frustratingly few empirical studies evaluating the effects of existing counter-disinformation laws. Though the state of research into causal relationships depicted in models like Figure 1 is somewhat more robust, there have been insufficient efforts to link societal outcomes to enacted counter-disinformation legislation.

Counter-disinformation legislation easily lends itself to human rights analysis, but such analysis does not provide actionable data on the laws’ effects and implementation.⁸⁶ Though existing empirical research is scarce, Section III shows that it is sufficient to create causal models of such laws’ effects and to identify the many lacunae in the literature which, if filled, would provide valuable evidence-based guidance for policymakers.

A. Labeling Laws

Researchers need more conclusive studies directly linking labeling laws to outcomes, including how audiences engage with labeled content, affected entity behavior

82. See Anton Dek, Kateryna Kononova & Tetiana Marchenko, *The Effects of Banning the Social Network VK in Ukraine*, in *RESPONDING TO COGNITIVE SECURITY CHALLENGES* 38, 44–48 (Anna Reynolds & Mike Collier eds., 2019) (measuring the decrease in the number of Ukrainian VK users after Ukraine’s ban of the platform).

83. See, e.g., *Freedom in the World 2022: Ukraine*, FREEDOM HOUSE, <https://freedomhouse.org/country/ukraine/freedom-world/2022> [<https://perma.cc/AH8L-W97H>] (last visited Mar. 1, 2023) (misguidedly identifying the Ukrainian government’s ban of Kremlin-controlled television channels as an unequivocally counter-democratic development).

84. Ukraine banned Kremlin-controlled social media platforms and self-styled news outlets during a state of war with Russia, and the European Union banned RT and Sputnik only after Russia’s full-scale invasion of Ukraine on February 24, 2022. see *Ukraine Bans Pro-Russian TV Stations*, DEUTSCHE WELLE (Feb. 3, 2021), <https://www.dw.com/en/ukraine-zelenskiy-bans-three-opposition-tv-stations/a-56438505> [<https://perma.cc/JP94-K9QX>]; Foo Yun Chee, *EU Bans RT, Sputnik Over Ukraine Disinformation*, REUTERS (Mar. 2, 2022), <https://www.reuters.com/world/europe/eu-bans-rt-sputnik-banned-over-ukraine-disinformation-2022-03-02/> [<https://perma.cc/78YU-8FXC>].

85. BAYER ET AL., *supra* note 9, at 128 (“Empirical studies about the effect of propaganda on public opinion are rare . . .”).

86. Burris et al., *supra* note 19, at 8 (asserting an analogous state of affairs in the abortion law context).

prior to labeling, and entity behavior after beginning to label. A labeling law combats disinformation indirectly by controlling the messenger rather than the message. The most visible existing law—FARA—has come under extensive criticism for its vague definitions and sporadic oversight and enforcement. Little direct research on the law’s effects exists, but IDEAL can be leveraged in this case, because labeling can be an effective deterrent, even in a hyper-partisan environment. Case studies and quantitative research on how entities have responded to being forced to label in the past would be invaluable for policymakers looking to improve FARA or to pass similar legislation.⁸⁷ Understanding the law’s effects in recent history would allow policymakers to tinker with it to improve its transparency of implementation and predictability of application.

Without effective oversight and enforcement, a labeling law relies on the law-abiding behavior of affected entities. Moreover, it is vital to understand that it only affects those entities disseminating white propaganda, and not even effective oversight and enforcement would make it a proper countermeasure to spreaders of gray and black propaganda.⁸⁸

B. Content Criminalization Laws

The most attempted counter-disinformation legal intervention in democratic countries has been content criminalization. To stay abreast of freedom of speech provisions, these laws narrowly determine the kind of content to be regulated not only by characterizing the content but via complementary content-neutral parameters. Two such laws have been the focus of most of the literature—France’s Law Against the Manipulation of Information and Germany’s NetzDG—but the variety in their implementation, oversight, and enforcement accords bountiful ground on which to conduct empirical studies.⁸⁹ Existing studies have attempted to track the raw amount of removed content, and social media platform reporting behavior, but much more empirical research is necessary to enable the use of IDEAL and to link the laws to societal outcomes.⁹⁰ Future studies should address not only social media platform behavior in policing forbidden content and reporting on their efforts but the accuracy of platform policing (i.e., identifying the proportion of removed content that is actually misinformative), the effectiveness of fining versus judicial intervention as an enforcement mechanism, probability of judicial overload, and conditions giving rise to the Streisand effect, to name but a few important data for policymakers.

87. See Burris et al., *supra* note 19, at 1 (explaining that legislators need empirical inputs to craft effective policy).

88. White propaganda is propaganda spread by a source whose identity and motives are clear, gray propaganda is propaganda whose true source is unclear, and black propaganda is propaganda deceptively and falsely attributed to a source other than the one disseminating it. Vanessa Molter & Renee DiResta, *Propaganda and Pandemics: How Chinese State Media Shapes Narratives on Coronavirus*, STAN. INTERNET OBSERVATORY, https://hai.stanford.edu/sites/default/files/6-renee_diresta_coronavirus.pdf [<https://perma.cc/34EJ-7KJV>].

89. See *supra* note 67.

90. See *supra* Parts II.B–C.

C. Bans

Democracies are beginning to entertain the idea of explicitly banning an entity from a nation's information space more and more now that Ukraine, the Baltic nations, and the collective European Union selectively activated the ban hammer in cases in which the banned entities' primary purpose was—beyond a reasonable doubt—the weakening of these nations' democratic firmament.⁹¹

Banning a social media platform via law, executive order, or National Security Council decision measurably reduces exposure to disinformation by lowering the number of users but has also triggered reputational costs among Western observers, at least as evidenced by the wealth of critiques based on human rights analysis. However, with the European Union's ban on RT and Sputnik, perhaps there is a growing realization among Western policymakers that a ban is nothing more than another instrument in the toolbox—one requiring exactly precise definitions and robust guardrails to keep us from the slippery slope, to be sure, but a viable and effective instrument, if deployed tactically.

However, interdisciplinary (including empirical) research is necessary to determine the best way to implement a ban. Among other priority topics, necessary research includes identifying the criteria (content-based, content-neutral, geopolitical, etc.) by which a ban would be appropriate for a democratic government, identifying the processes through which past bans were implemented, measuring their effects on a government's perceived trustworthiness, and tracking whether a platform's behavior changes after being banned. Putting in the hands of policymakers concrete cause-and-effect empirical data showing a ban's precision and side effects would also mitigate rights-based critiques.

D. Linking Laws to Macro Models of Disinformation

Counter-disinformation laws do not exist in a vacuum, and causal models must be contextualized within broader societal models of disinformation spread. The model depicted in Figure 1 is not the only possible visualization, and efforts to refine it are welcome. Furthermore, there are other models that must be generated, as social media are not the only vector for disinformation spread, nor the only vector addressed by existing laws. Though there are obstacles to measuring news consumption behavior in the twenty-first century,⁹² television remains a widely consumed source of information in many democratic nations,⁹³ along with online news publications, and though disinformation's strategic goal is constant, there are nuances in the tactical objectives of

91. Patrick Wintour, Jennifer Rankin, & Kate Connolly, *EU To Ban Russian State-Backed Channels RT and Sputnik*, THE GUARDIAN (Feb. 27, 2022), <https://www.theguardian.com/media/2022/feb/27/eu-ban-russian-state-backed-channels-rt-sputnik> [<https://perma.cc/MM2L-2AZH>].

92. Michael Barthel, Amy Mitchell, Dorene Asare-Marfo, Courtney Kennedy & Kirsten Worden, *Measuring News Consumption in a Digital Era*, PEW RES. CTR. (Dec. 8, 2020), https://www.journalism.org/wp-content/uploads/sites/8/2020/12/PJ_2020.12.08_News-Consumption_FINAL.pdf [<https://perma.cc/9EVD-WJRZ>].

93. See Katerina Eva Matsa, *Most Western Europeans Prefer TV News While Use of Print Outlets Lags*, PEW RES. CTR. (Sept. 27, 2018), <https://www.pewresearch.org/fact-tank/2018/09/27/most-western-europeans-prefer-tv-news-while-use-of-print-outlets-lags/> [<https://perma.cc/J3JE-N83A>].

disinformation's dissemination via television or legitimate online print news sources.⁹⁴ These nuances should be reflected in causal models demonstrating how disinformation spreads via television and print news outlets—our objective is competently tailored legislation, and it is only possible with sufficient conceptual and empirical data.

As demonstrated, Germany's NetzDG and France's Law Against the Manipulation of Information are designed to disrupt the central node ("False information spreads through social media") in Figure 1.⁹⁵ Something like Ukraine's ban of VKontakte and Odnoklassniki targets the same node, only more fundamentally, removing not just select content posted on the social media platform but the entire platform itself.⁹⁶ FARA, meanwhile, seeks to combat the media illiteracy fueling misinformation spread and the audience's inability to identify unreliable sources of information by publicly identifying a source's foreign principal.⁹⁷

Other legislation could conceivably be tailored to inhibit the spread of disinformation by targeting other links in the chain, like laws diminishing financial incentives for generating viral content⁹⁸ or laws setting standards for social media's dissemination algorithms.⁹⁹ Cohen et al. parsimoniously categorize counter-disinformation measures as targeting either production, distribution, or consumption,¹⁰⁰ but as the conceptual models show, production, distribution, and consumption can translate into a broad array of nodes in a causal chain. The possibilities are many, and having such a visualization—and being aware of the research gaps the visualization reveals—is a critical input for policymakers seeking to stem disinformation's corrosive effects.

CONCLUSION

Counter-disinformation legislation is a burgeoning field of law, one that could be an effective instrument in a democracy's toolbox. However, as the literature keeps reminding us, "legal measures to target [disinformation] may result in unexpected scenarios."¹⁰¹ This Article has presented a preliminary and admittedly limited application of IDEAL, using available empirical research to connect a law's implementation with its legal effects down the causal chain. However, just as there is limited direct research on the laws' effects, there is also dishearteningly little research that we could use via IDEAL to infer causality in intermediate links. Once researchers

94. See BAYER ET AL., *supra* note 9, at 52–59 (exploring the different channels through which disinformation is disseminated).

95. See *supra* Part II.B.

96. See *supra* Part II.C.

97. See *supra* Part II.A.

98. See EUROPEAN COMMISSION, *supra* note 49, at 7 (addressing the importance of demonetizing the spread of misrepresentative or misleading ads).

99. See BAYER ET AL., *supra* note 9, at 13 (discussing the desired manner in which social media algorithms should function).

100. COHEN ET AL., *supra* note 79, at 61.

101. GULIZAR HACIYAKUPOGLU, JENNIFER YANG HUI, V.S. SUGUNA, DYMPLIS LEONG & MUHAMMAD FAIZAL BIN ABDUL RAHMAN, S. RAJARATNAM SCH. OF INT'L STUDS., NANYANG TECH. UNIV. SING., COUNTERING FAKE NEWS: A SURVEY OF RECENT INITIATIVES 13 (2018), https://www.rsis.edu.sg/wp-content/uploads/2018/03/PR180416_Countering-Fake-News.pdf [<https://perma.cc/9TBS-53WP>].

begin to fill the empirical gaps in the counter-disinformation literature, IDEAL's application would be considerably more fruitful—it could well be the method that empowers us to make the unexpected expected, “provid[ing] plausible and actionable insights that can better inform guidance documents, as well as targeted strategies for research, policy and advocacy.”¹⁰²

Few democracies have deployed the law as a remedy to the plague of disinformation, but more and more democratic governments worldwide are debating legal measures as a viable possibility.¹⁰³ However, counter-disinformation law is woefully understudied, with little empirical evidence supporting policymakers, preventing us from understanding the precise mechanisms at work. The application of IDEAL could dispel this uncertainty.

102. Burris et al., *supra* note 19, at 9.

103. Amy Yee, *The Country Inoculating Against Disinformation*, BBC (Jan. 30, 2022), <https://www.bbc.com/future/article/20220128-the-country-inoculating-against-disinformation> [<https://perma.cc/V5FA-LPAY>] (surveying recent parliamentary debates in the United States, Germany, and the United Kingdom).