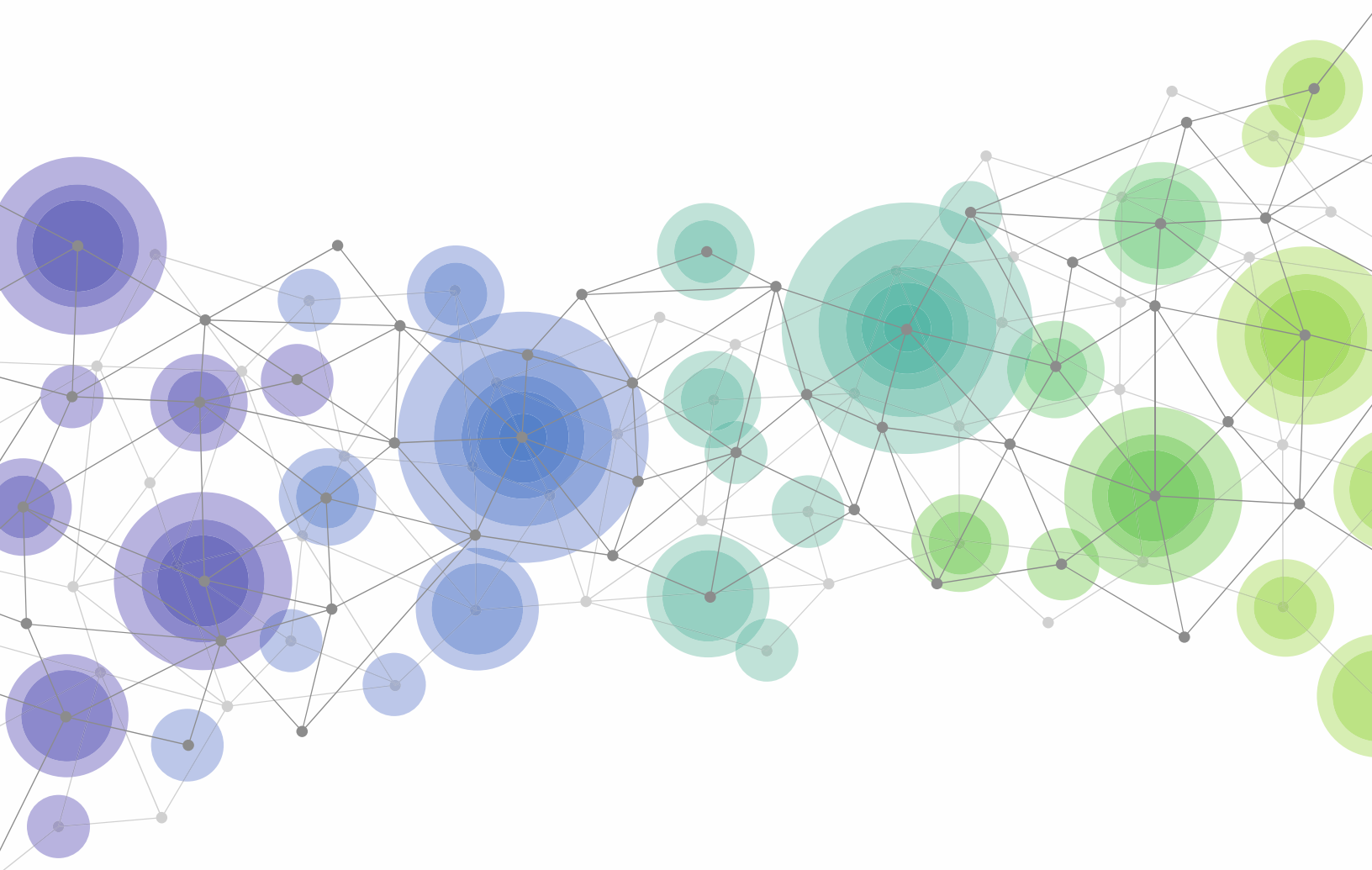# Tackling the "Fake" Without Harming the "News"

A Paper Series on Regulatory Responses to Misinformation

Edited by Michael Karanicolas

Acknowledgements

Table of Contents

Foreword

Amanda Keton, General Counsel, Wikimedia Foundation

At the scale of the internet, which now connects more than 4.5 billion people,[1] the rapid spread of disinformation is one of the foremost challenges of our time. Disinformation erodes trust in institutions and weakens democratic processes. Provocative and outrageous headlines travel quickly and misleading information is freely available, while nuanced reporting is often more difficult to understand and buried behind paywalls. Major internet platforms are structured to amplify this type of disinformation, which is further encouraged through their incentive models which prioritize engagement over accuracy and can be gamed by bad actors. This phenomenon is not just a technical problem, but also a social and economic one.

Assessing the quality of information is at the heart of Wikipedia, the free encyclopedia that anyone can edit. Over the past 20 years of Wikipedia's history,[2] the hundreds of thousands of volunteer editors that contribute to the site have embraced radical approaches to transparency and collaborative decision making in order to ensure the most accurate information possible. They discuss and debate and eventually reach a consensus. Wikipedia's model emphasizes time for deliberation over the speed of breaking news and social media or "clicks" that eventually lead to profit. It also gives a lot of agency to the individual, while relying on the law and regulation to leave room for such community-based decision making. Editors have become so skilled at this work, sometimes with support from the Wikimedia Foundation, that other internet platforms now depend on Wikipedia to fight conspiracy theories. Its collaborative model for content moderation has been celebrated by commentators and piqued the interest of lawmakers around the world.

As supporters of the Wikimedia Initiative on Intermediaries and Information at Yale ISP, we are delighted to see this white paper series come to fruition. Disinformation is a complex challenge without a single culprit, and this series suggests regulatory approaches that take a more holistic view of the entire ecosystem of information. The Wikimedia Foundation applauds Michael

---

[1] Statista, *Global digital population as of October 2020*, STATISTA, https://www.statista.com/statistics/617136/digital-population-worldwide/ (last visited Feb. 23, 2021).
[2] Wikimedia Foundation, *Celebrating 20 years of Wikipedia*, WIKIMEDIA, https://wikimediafoundation.org/wikipedia20/ (last visited Feb. 23, 2021).

Karanicolas for facilitating and coordinating this important work. It is inspiring to see people in the field of internet and technology policy propose innovative tools and solutions that help society move closer to reaching a shared understanding of knowledge and truth again. Rebuilding trust in the online knowledge ecosystem will require bold thinking and collaboration between all stakeholders and sectors.

Introduction – Understanding the Challenge

Michael Karanicolas, Wikimedia/Yale Law School Initiative on Intermediaries and Information

Over the past decade, optimism over the Internet's potential as a force for democratization has given way to a more nuanced understanding of its disruptive impact. The decentralized nature of online discourse, which can circumvent traditional gatekeepers, has led to challenges in promoting integrity and veracity in the civic discourse. These issues are compounded by the emergence of new, intentionally manipulative technologies.[1] Where internet access was once perceived as a threat to autocrats, it is now also seen as a potential threat to democracy.[2] Over the past five years, governments around the world have become particularly vocal regarding the dangers of online misinformation.[3]

The COVID-19 crisis accelerated these concerns, crystalizing the harms flowing from online rumors and lies.[4] In the context of a global pandemic, trust in

---

[1] Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753 (2019); Chengcheng Shao et al, *The Spread of Low-credibility Content by Social Bots*, 9 NATURE COMMUNICATIONS 1 (2018); Lisa-Marie Neudert, *Future Elections May Be Swayed by Intelligent, Weaponized Chatbots*, MIT TECH. REV. (Aug. 22, 2018), https://www.technologyreview.com/s/611832/future-elections- may-be-swayed-by-intelligent-weaponized-chatbots/.

[2] Anamitra Deb, Stacy Donohue, & Tom Glaisyer, *Is Social Media a Threat to Democracy?*, OMIDYAR GROUP (Oct. 1, 2017), https://fronteirasxxi.pt/wp-content/uploads/2017/11/Social-Media-and-Democracy-October-5-2017.pdf.

[3] *See, e.g.,* Ishaan Tharoor, *'Fake news' threatens Germany's election, too, says Merkel*, WASHINGTON POST (Nov. 23, 2016), https://www.washingtonpost.com/news/worldviews/wp/2016/11/23/fake-news-threatens-germanys-election-too-says-merkel/. Terms like "misinformation", "disinformation", and "fake news" can be difficult and controversial to define. There have been some efforts at developing more widely accepted definitions, notably by Claire Wardle and Hossein Derakhshan, who authored a framework whereby "information disorder" is categorized as a broader term which encompasses "mis-information", "dis-information", and "mal-information". Claire Wardle & Hossein Derakhshan, *Information Disorder: Toward and Interdisciplinary Framework for Research and Policymaking*, COUNCIL OF EUR. 20-21 (Sept. 27, 2017), https://rm.coe.int/information-disorder-toward-aninterdisciplinary-framework-for-researc/168076277c. However, while this framework has gained some traction, there remains considerable variation as to the common usage of these terms, which are still used interchangeably. See, e.g., *Who Starts Viral Misinformation?*, BBC NEWS (May 4, 2020) https://www.bbc.com/news/av/world-52454129. In order to avoid a definitional morass, in this paper series terms like "misinformation" and "fake news" are used in this broader sense, rather than to refer to a particular subset of information disorder.

[4] Michael Karanicolas, *Even in a Pandemic, Sunlight is the Best Disinfectant: COVID-19 and Global Freedom of Expression*, OREGON REVIEW OF INTERNATIONAL LAW (Forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3726892.

---

government institutions is of paramount importance, and false narratives which undermine this confidence, or which otherwise sow confusion about proper safety protocols, are especially dangerous. The pandemic spawned a host of conspiracy theories related to its origins and treatments, which prompted global governments to crack down on the so-called "infodemic".[5] Throughout 2020, a number of governments either passed new laws criminalizing misinformation or expanded enforcement of their existing misinformation laws.[6]

This legislative trend taps into a long-running debate around the appropriateness of rules restricting the spread of false information.[7] Broadly structured laws prohibiting the publication of "fake news", "false news", or "misinformation" have a long provenance as tools for autocrats to target journalists, opposition figures, or anyone else who espouses a narrative which runs counter to official government positions.[8] Around the world, there are numerous examples of "fake news" laws being used to stifle anti-corruption investigations, or reporting which uncovers government incompetence or mismanagement.[9]

One core problem with criminal prohibitions targeting "fake news" is that they grant authorities a broad power to establish and enforce their own centralized

---

[5] *UN tackles 'infodemic' of misinformation and cybercrime in COVID-19 crisis*, UNDGC (Mar. 31, 2020), https://www.un.org/en/un-coronavirus-communications-team/un-tackling-'infodemic'-misinformation-and-cybercrime-covid-19.

[6] See, e.g., *Public Health (COVID-19 Prevention, Containment and Treatment) (National Lockdown) Order, Statutory Instrument 83 of 2020*, s. 14 (Zim.), https://www.veritaszim.net/sites/veritas_d/files/SI%202020-083%20Public%20Health%20%28COVID-19%20Prevention%2C%20Containment%20and%20Treatment%29%20%28National%20Lockdown%29%20Order%2C%202020.pdf; *Amendment Of State Of Emergency Covid-19 Regulations: Namibian Constitution (Proclamation No. 13 of 2020),* s. 15(1)(e) (Nam.), www.lac.org.na/laws/2020/7180.pdf; Emergency Powers (Covid-19) Regulations, 2020 (Statutory Instrument No. 61 of 2020), s. 30(3) (Bot.), https://bw.usembassy.gov/wp-content/uploads/sites/125/Emergency-Powers-COVID-19-Regulations-2020.pdf;

[7] Annual General Assembly Report of the Human Rights Committee, UN Doc. A/50/40, 3 October 1995, para. 89; Concluding Observations of the Human Rights Committee: Mauritius, UN Doc. CCPR/C/79/Add.60, 4 April 1996, para. 19, https://www.refworld.org/pdfid/3f4746ae4.pdf.

[8] *See, e.g.,* Concluding Observations of the Human Rights Committee: Cameroon, CCPR/C/79/ Add.116, 4 November 1999, para. 24.

[9] *See, e.g., Egypt puts former auditor on trial over corruption claim*, BBC NEWS (Jun. 7, 2016), https://www.bbc.com/news/world-middle-east-36469408; *"Journalism Is Not a Crime" - Violations of Media Freedoms in Ethiopia*, HUMAN RIGHTS WATCH (Jan. 21, 2015), https://www.hrw.org/report/2015/01/21/journalism-not-crime/violations-media-freedoms-ethiopia; *Russian journalists face fines and probes amid creeping censorship*, INTERNATIONAL PRESS INSTITUTE (Jun. 25, 2020), https://ipi.media/russian-journalists-face-continued-fines-and-probes-amid-creeping-censorship/; *Iranian Journalists Hounded for Disputing Official Coronavirus Figures*, REPORTERS WITHOUT BORDERS (Mar. 13, 2020), https://rsf.org/en/news/iranian-journalists-hounded-disputing-official-coronavirus-figures.

---

version of the truth.[10] While there are other forms of legislation which penalize false speech, particularly defamation, these typically do so in a narrowly circumscribed way, and with robust safeguards to prevent their abuse.[11] In the absence of such protections, laws targeting "fake news" not only lend themselves to abusive applications, but also tend to create a chilling effect against legitimate forms of speech. This impact may be particularly pronounced among journalists who report stories that are rapidly developing, or which are otherwise challenging to verify beyond a shadow of a doubt.[12]

As a consequence, academics and human rights defenders, as well as formal human rights mechanisms, have traditionally adopted a skeptical attitude towards the legitimacy of criminal prohibitions targeting misinformation. In a 2017 Joint Declaration, the United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights Special Rapporteur on Freedom of Expression and Access, stressed "that the human right to impart information and ideas is not limited to 'correct' statements… and that prohibitions on disinformation may violate international human rights standards".[13]

Although there is relatively little international jurisprudence specifically assessing the legitimacy of misinformation laws as a class, there have been some adverse findings against particular laws, such as the the Community Court of Justice of the Economic Community of West African States' 2018 finding against the Gambia, and the East African Court of Justice's 2019 decision against Tanzania.[14] Likewise, a number of domestic courts have struck down criminal prohibitions targeting

---

[10] Michael Karanicolas, *Subverting Democracy to Save Democracy: Canada's Extra-Constitutional Approaches to Battling 'Fake News'*, 17 CAN. J.L. & TECH. 201 at 223 (2019).

[11] *See, e.g.*, New York Times Co. v. Sullivan, 376 U.S. 254, 283 (1964); Grant v. Torstar Corp., [2009] 3 SCR 640 (Can.).

[12] *See* Sunday Times v. United Kingdom, 30 EUR. CT. H.R. (ser. A) at 49 (1979).

[13] Office of the Special Rapporteur for Freedom of Expression, *Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda*, ORG. AM. STS. (Mar. 3, 2017), www.oas.org/en/iachr/expression/showarticle.asp?artID=1056&lID=1.

[14] Federation of African Journalists (FAJ) and others v. The Gambia, ECOWAS judgment no. ECW/CCJ/JUD/04/18, https://globalfreedomofexpression.columbia.edu/wp-content/uploads/2016/04/FAJ-and-Others-v-The-Gambia-Judgment.pdf; Media Council of Tanzania & 2 others v. the Attorney General of the United Republic of Tanzania, EACJ reference no. 2 of 2017 (para. 49). There is some ambiguity to the latter decision, since the Court invalidated the provisions despite also claiming that they appear to be "largely unobjectionable".

misinformation, finding them to be incompatible with the right to freedom of expression.[15]

Despite the inherent human rights risks, since 2018 a number of relatively progressive democracies have passed new restrictions targeting fake news.[16] Germany's NetzDG has been particularly influential, and has spawned copycat proposals in a number of countries, including Kyrgyzstan, Singapore, Brazil, and Honduras.[17] Canada's law reforms to criminalize false information in an election context are also noteworthy, insofar as their previous Supreme Court decision striking down a false news law was influential among other courts considering similar questions.[18] The COVID-19 crisis has significantly accelerated the move towards imposing harsh penalties for misinformation.[19] Predictably, COVID-era misinformation laws have been abused to target journalists, opposition figures, and civil society.[20]

---

[15] *See* R. v. Zundel, [1992] 2 S.C.R. 731 (Can.); Onyango-Obbo and Mwenda v. The Attorney General of Uganda, Constitutional Appeal No. 2 of 2002 (Feb. 11, 2004) (Uga.); Chavunduka and Choto v. Minister of Home Affairs & Attorney General of Zimbabwe (May 22, 2000), Judgment No. S.C. 36/2000 (Zim.); Chipenzi and Others v. The People, (HPR/03/2014) [2014] ZMHC 112, p. J23-J25 (Zam.).

[16] Loi relative à la lutte contre la manipulation de l'information, JO, 24 December 2018, no 0297 (Fra.); An Act to amend the Canada Elections Act and other Acts and to make certain consequential amendments, S.C. 2018, c. 31, s. 61 (Can.); 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.

[17] Jacob Mchangama and Natalie Alkiviadou, *The Digital Berlin Wall: How Germany (Accidentally) Created A Prototype For Global Online Censorship - Act Two*, JUSTITIA (Sep. 2020), https://justitia-int.org/wp-content/uploads/2020/09/Analyse_Cross-fertilizing-Online-Censorship-The-Global-Impact-of-Germanys-Network-Enforcement-Act-Part-two_Final-1.pdf. It is worth noting that, although NetzDG was presented at least in part as a measure to combat misinformation, the law does not expressly target misinformation as a general category, except insofar as it references offences which are related to misinformation, such as Holocaust denial. However, a number of the laws which were modeled on NetzDG do include broad categorical prohibitions targeting misinformation.

[18] Karanicolas, *supra* note 10. It is worth noting that the new Canadian rules are the subject of a constitutional challenge: Michael Karanicolas, *Canada's fake news laws face a Charter challenge. That's a good thing*, OTTAWA CITIZEN (Oct. 14, 2019), https://ottawacitizen.com/opinion/columnists/karanicolas-canadas-fake-news-laws-face-a-charter-challenge-thats-a-good-thing.

[19] Karanicolas, *supra* note 4.

[20] Ani Mejlumyan, *Armenia takes hard line against media reporting on COVID-19*, EURASIANET (Mar. 23, 2020), https://eurasianet.org/armenia-takes-hard-line-against-media-reporting-on-covid-19; Sesupo Rantsimako, *BPF Spokesman Arrested For Fake News*, THE BOTSWANA GAZETTE (Apr. 16, 2020), https://www.thegazette.news/news/bpf-spokesman-arrested-for-fake-news/30891/; *Myanmar journalist jailed for two years for reporting Covid-19 death*, REPORTERS WITHOUT BORDERS (May 25, 2020), https://rsf.org/en/news/myanmar-journalist-jailed-two-years-reporting-covid-19-death.

While the COVID-19 crisis will eventually abate, changes in our collective consciousness regarding the potential harms flowing from misinformation are likely to be permanent, particularly as they represent the culmination of a longer-term shift in attitudes. It is not uncommon for temporary measures, which have been adopted in times of emergency, to become normalized.[21] However, if the current approach to regulating misinformation, criminalizing false news, becomes a global standard, it would represent a great gift to the world's authoritarians, validating their efforts to impose a single, State-sanctioned version of the truth.

This is not intended to minimize the very real challenge that misinformation poses, and the importance of integrity and veracity in the civic discourse. However, the human rights harms of criminalizing misinformation suggest that there is a need to think more broadly, and explore innovate regulatory measures which promote integrity in the discourse without adversely impacting the right to freedom of expression.

The purpose of this series is to address that regulatory gap and provide a set of innovative proposals for regulatory responses to misinformation which do not revert to the use of criminal sanctions targeting false speech. The authors were drawn from an open call for proposals which expressly called for novel solutions to the global challenge of misinformation.[22] The winners were selected primarily based on the strength of their abstracts, with a particular focus on creative or under-researched approaches to the problem. However, the organizers also sought to guarantee that a diversity of perspectives would be reflected in the series, given the global and multifaceted nature of the challenge. The authors were given a free hand to develop their ideas. The Wikimedia/Yale Law School Initiative on Intermediaries and Information supported this process by offering multiple rounds of commentary and editorial support to the papers, and by facilitating presentations to the Information Society Project community to solicit further feedback, though final editorial discretion remained with the individual authors.

The first paper, by Nathalie Maréchal, Elizabeth M. Renieris, and Jan Rydzak of Ranking Digital Rights, proposes two changes to the corporate governance framework of social media companies, which support efforts to promote more

---

[21] Richard Burchill, *When does an Emergency Threaten the Life of the Nation - Derogations from Human Rights Obligations and the War on International Terrorism*, 8 Y.B. N.Z. JURIS. 99, 115 (2005).
[22] *Call for Papers: Regulatory Responses to Misinformation*, YALE INFORMATION SOCIETY PROJECT (Jul. 9, 2020), https://law.yale.edu/isp/initiatives/wikimedia-initiative-intermediaries-and-information/wiii-blog/call-papers-regulatory-responses-misinformation.

responsible behaviour among online platforms in mitigating the harmful impacts of the speech they host. The second paper, by Akriti Gaur, proposes a new model of accountable self-regulation for platforms which grants users and civil society groups greater opportunity to meaningfully impact how content decisions around misinformation are set and applied. The third paper, by Lisa Macpherson of Public Knowledge, considers the issue through an analogy to environmental pollution, and proposes the creation of a new "superfund" to support fact-checking initiatives, modeled on the Environmental Protection Agency's Superfund to clean up toxic waste sites. The fourth paper, by Ivar Hartmann of the Insper Learning Institution, proposes increasing the transparency around online funding flows through improved ad libraries. The fifth paper, by Jonathan Obar of York University, aims to boost information literacy, linking these efforts with existing initiatives aimed at expanding internet access.

Although each of the papers was written with a particular national context in mind, all five were chosen due in part to their potential scalability across different jurisdictions. None of these proposals purports to provide a silver bullet solution to the challenge posed by misinformation. Moreover, all five are ripe subjects for additional research and consideration including, in some cases, to clarify potential constitutional or regulatory hurdles to their implementation. Misinformation is an iterative threat which is constantly evolving, including in response to regulatory efforts aimed at combating it. However, while none of these proposals are a panacea, they are also not mutually exclusive. To the contrary, they are designed to be complementary, and to target different aspects of the misinformation ecosystem, in line with the complex and multi-faceted nature of this phenomenon. Together, the papers provide a menu of alternative policy proposals for regulators and public officials seeking new avenues to combat the harm of misinformation.

We hope that this series will provide a useful set of ideas for policy makers, as well as civil society activists who are working to challenge abusive misinformation laws or proposals. Such debates naturally put pressure on opposition voices to offer some alternative vision to the use of the criminal law. It is also our intention for these proposals to serve as a conceptual baseline for researchers, both in academia and in civil society, to develop these ideas further, including potentially adapting them to different national and cultural contexts.

There are no quick and easy solutions to promoting integrity in the public discourse, and no shortcuts to bootstrap democratic institutions. Where countries have been able to establish a strong democratic culture, it has typically come

through decades, if not generations, of work to build and enshrine respect for foundational democratic rights, including the right to freedom of expression. The current debate around misinformation is, at its core, a debate about how best to preserve democracy, which is under threat from the evolution of competing and contradictory understandings of reality. However, any solution to the present misinformation crisis which comes at the cost of freedom of expression will be a pyrrhic victory. With this series, we hope to demonstrate that there are avenues to promoting integrity in the public discourse without undermining its vibrancy and diversity.

Better Processes Lead to Better Outcomes: Corporate Governance as a Tool to Address Misinformation

Nathalie Maréchal, Elizabeth M. Renieris, & Jan Rydzak,
Ranking Digital Rights*

I. Introduction

The popularization of the internet has been a boon for journalists, activists, and individuals whose ability to access and impart information had been stymied by traditional gatekeepers in the public and private sectors. The emergence of social media further democratized access to online expression and facilitated the growth of decentralized networks, including a range of technologically-mediated social movements.[1] As naive as the previous decade's enthusiasm for these "liberation technologies" may seem in hindsight, their benefits to human rights and the democratic project should not be discounted.[2] Nonetheless, the same ability to circumvent traditional gatekeepers, including editors and fact-checkers, has also enabled the spread of misinformation and disinformation, with grave consequences for democracy, human rights, and public health.[3]

Misinformation is only one manifestation of structural flaws in a complex socio-technical system for data collection and information dissemination, governed by the market logic that Shoshana Zuboff terms "surveillance capitalism."[4] To date,

---

* Nathalie Maréchal is Senior Policy Analyst at Ranking Digital Rights, Elizabeth Renieris is the Founding Director of the Notre Dame-IBM Technology Ethics Lab, and Jan Rydzak is Company Engagement Lead at Ranking Digital Rights.

[1] *See* ZEYNEP TUFEKCI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST (2018).

[2] *See* Larry Diamond, *Liberation Technology*. 3 J. OF DEMOCRACY 69 (2010).

[3] In this paper, we use Caroline Jack's definitions: misinformation is "information whose inaccuracy is unintentional, while disinformation is "information that is deliberately false or misleading." As Jack notes, "whether a given story or piece of content is labeled as misinformation or disinformation can depend as much on a speaker's intent as on the professional standards of who is evaluating." This distinction can be exceedingly challenging to parse, and while germane to many approaches to combating misinformation, is not particularly relevant to the interventions discussed in this paper. Therefore, for the purpose of this paper, we will refer to misinformation as including disinformation. *See* Caroline Jack. *Lexicon of Lies: Terms for Problematic Information* (Data & Society report, 2017), https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf.

[4] *See* SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019). *See also* Nathalie Maréchal & Ellery Roberts Biddle, *It's Not Just the Content, It's*

---

many proposed and implemented interventions target symptoms that manifest themselves at the content layer of the internet rather than root causes concentrated at the corporate level.[5] Symptomatic efforts targeting misinformation and other harmful or dangerous content are necessary but not sufficient.

Addressing misinformation alongside other harms (notably surveillance and discrimination), while respecting international human rights (particularly freedom of expression and information), requires a regulatory toolkit that includes antitrust and competition law, data protection, privacy, and consumer protection law. It also requires a careful balancing of intermediary liability with intermediary immunity.[6] This paper examines good corporate governance as a prerequisite for effective content moderation, dealing with the root causes of misinformation in a manner which avoids threatening freedom of expression.

First, this paper will introduce the relationship between corporate governance and misinformation. The paper will then describe two categories of corporate governance reform that could help platforms improve content moderation while enhancing respect for human rights, namely structural reforms of company leadership, and enhanced investor oversight. For each category, the paper outlines the industry status quo, including company behavior demonstrating the need for the reforms in question, before recommending best practices for companies and considering the legal and regulatory levers that could incentivize or even compel companies to alter their practices. The paper draws on prior coverage and analysis of several companies to illustrate our argument, though for consistency it uses Facebook as the primary example.

II. The Relationship Between Corporate Governance and Misinformation

"Corporate governance" refers to the set of policies, processes, and procedures through which the entities and individuals comprising a corporation make decisions impacting the company and hold one another accountable for following

---

the Business Model: Democracy's Online Speech Challenge (Ranking Digital Rights report, 2020), http://newamerica.org/oti/reports/its-not-just-content-its-business-model/. See also Nathalie Maréchal et al., Getting to the Source of Infodemics: It's the Business Model (Ranking Digital Rights report, 2020), https://www.newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/.

[5] See LAURA DENARDIS. THE GLOBAL WAR FOR INTERNET GOVERNANCE (2015).

[6] See Jack M. Balkin, How to Regulate (and Not Regulate) Social Media, KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY (Mar. 25, 2020), https://knightcolumbia.org/content/how-to-regulate-and-not-regulate-social-media.

through on them.[7] In the case of public U.S. companies listed on a major stock exchange, including large social media and technology companies, these entities and individuals may include the Board of Directors and its Chair; the Chief Executive Officer and other high-ranking executives; managers and employees throughout the corporate hierarchy; and individual and institutional shareholders. Non-company actors may also influence corporate governance; the most notable among them are governments and regulatory bodies, the media, and civil society organizations, including consumer protection and human rights groups.

The Board of Directors determines the company's priorities with an eye to both long-term profitability and environmental, social, and governance (ESG) concerns.[8] It appoints the Chief Executive Officer and holds that individual accountable for carrying out those priorities. The CEO, executives, senior management, and other employees execute the Board's vision and produce regular financial reports and other documentation for the Board and company shareholders to demonstrate progress toward pre-established goals such as revenue targets, environmental impact, control over its supply chain, and other functions of the company.

Public companies that host large volumes of user-generated content (UGC) and paid advertising, including Alphabet (Google's parent company), Microsoft, Facebook, and Twitter, long treated content moderation as a relatively low priority, particularly in terms of consistently enforcing rules and effectively responding to contextual nuances.[9] Despite the dire consequences of rampant mis- and disinformation on their platforms in other parts of the world,[10] it is only since the well-documented foreign influence operations into the U.K.'s Brexit referendum and the 2016 U.S. general election that the major social media companies have

---

[7] *See* California Public Employees' Retirement System, *Global Principles of Accountable Corporate Governance*, CALIFORNIA PUBLIC EMPLOYEE'S RETIREMENT SYSTEM (Nov. 14, 2011), https://www.iccr.org/sites/default/files/resources_attachments/2011-11-14-global-principles-of-accountable-corp-gov.pdf. Numerous definitions of corporate governance exist, reflecting varying stakeholder priorities. Taking the simplest approach, corporate governance can be defined as "the system of rules, practices, and processes by which a firm is directed and controlled." James Chen & Margaret James, *Corporate Governance*, INVESTOPEDIA (Apr. 12, 2020), https://www.investopedia.com/terms/c/corporategovernance.asp.
[8] *See* Olivier Jan, *The Board and ESG*, HARVARD LAW SCHOOL FORUM ON CORPORATE GOVERNANCE (Feb. 25, 2019), https://corpgov.law.harvard.edu/2019/02/25/the-board-and-esg/.
[9] *See* Kate Klonick, *The New Governors: The People, Rules and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018). *See also* SARAH T. ROBERTS. BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA (2019).
[10] *See* Amanda Taub & Max Fisher, *Where Countries Are Tinderboxes and Facebook Is a Match*, THE NEW YORK TIMES (Apr. 21, 2018), https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html.

begun to devote significant resources into grappling with these challenges.[11] Steeped in a maximalist interpretation of the First Amendment and cyberlibertarianism, technology companies have traditionally been resistant to calls to take a stronger hand towards combating misinformation.[12] Even now, a strong discursive current in Silicon Valley and Washington, D.C. insists that free speech is paramount to all other values in the digital sphere.[13]

But the internet is not a parallel dimension and is very much part of the "real world."[14] Governments have long sought to pressure social media platforms into doing more to combat child sexual abuse material (CSAM), copyrighted materials, and other problematic content. Simultaneously, civil society has been fighting for due process and transparency around content removals, both in response to government demands and in cases where content violates company rules.[15] Forced to confront the ugly realities of "lawful but awful" expression, companies are gradually and somewhat haphazardly developing increasingly sophisticated content-related rules and enforcement processes.[16]

Given the multiplicity of companies, governments, languages, and sociocultural contexts involved, what has emerged is a hodgepodge of international norms, national laws, co- and self-regulation, and seemingly capricious decisions by company leadership. Despite repeated assurances to the contrary, companies seem unable to exercise meaningful control over the (mis)information that circulates on their platforms. In this context, it is little surprise that even democratic

---

[11] *See* YOCHAI BENKLER ET AL. NETWORK PROPAGANDA: MANIPULATION, DISINFORMATION, AND RADICALIZATION IN AMERICAN POLITICS (2018).

[12] *See* David Golumbia, *Cyberlibertarianism: The Extremist Foundations of 'Digital Freedom.'* Presented at Clemson University, Clemson, SC (Sept. 5, 2013), http://www.uncomputing.org/wp-content/uploads/2014/02/cyberlibertarianism-extremist-foundations-sep2013.pdf.

[13] See, for example, Mark Zuckerberg's 2019 speech at Georgetown University, and various proposals from the Trump Administration and congressional Republicans to strip platforms of Section 230 when they moderate or fact-check misinformation and other kinds of "lawful-but-awful" content. *See* Mark Zuckerberg, *Standing for Voice and Free Expression*, Presented at the Speech at Georgetown University, Washington D.C. (Oct. 17, 2019). *See also* Daphne Keller, *Trump-Backed Rules for Online Speech: No to Porn, Yes to Election Disinformation and Hate Speech*, SLATE MAGAZINE (Oct. 1, 2020), https://slate.com/technology/2020/10/cda-section-230-graham-doj-reform-content-moderation.html.

[14] John P. Barlow, *A Declaration of the Independence of Cyberspace*, ELECTRONIC FRONTIER FOUNDATION (Feb. 8, 1996), https://www.eff.org/cyberspace-independence.

[15] Kate Klonick, *supra* note 8.

[16] Daphne Keller, *If Lawmakers Don't Like Platforms' Speech Rules, Here's What They Can Do About It. Spoiler: The Options Aren't Great*, TECHDIRT (Sept. 9, 2020, 12:00 PM), https://www.techdirt.com/articles/20200901/13524045226/if-lawmakers-dont-like-platforms-speech-rules-heres-what-they-can-do-about-it-spoiler-options-arent-great-daphne-keller.shtml.

---

governments that are ostensibly committed to free expression are tempted to turn to intermediary liability and restrictions on online speech to hold technology companies accountable. However, for reasons that are detailed in a recent report from Ranking Digital Rights, such measures raise serious freedom of expression concerns and, at least in the United States, would be unlikely to comply with the First Amendment.[17]

Governments should not dictate the processes for, or outcomes of, content governance decisions. Nevertheless, there is a role for governments to encourage better corporate governance, which in turn will translate into better processes and outcomes for content moderation. Many of the social media platforms where the dangers of misinformation are most severe (due in part to their large user bases) are publicly traded on the New York Stock Exchange and NASDAQ and therefore subject to regulation by the Securities and Exchange Commission (SEC) and other federal agencies. Requiring these companies to follow recognized best practices, rules, and regulations for corporate governance will improve internal decision-making and oversight mechanisms and generate better outcomes, while avoiding government intervention into content moderation, which poses serious challenges under both the First Amendment of the Constitution and Article 19 of the Universal Declaration of Human Rights.

Two specific corporate governance reforms that would be particularly impactful are structural reforms of company leadership and mechanisms for enhanced investor oversight. Ideally, these measures would be accompanied by other changes such as mandating company-wide human rights due diligence. To combat misinformation, it is critical to normalize strong oversight while holding companies accountable for their due diligence based on a common set of standards, parts of which already exist in the UN Guiding Principles on Business and Human Rights and guidance published by organizations such as the Danish Institute for Human Rights.[18] The interventions we explore below are only two of the top-level changes needed to stem misinformation, but they would unlock important changes further downstream.

---

[17] Nathalie Maréchal & Ellery Roberts Biddle, *supra* note 4.
[18] DANISH INSTITUTE FOR HUMAN RIGHTS, HUMAN RIGHTS IMPACT ASSESSMENT GUIDANCE AND TOOLBOX (2020).

III. Reforming Company Leadership Structures

The case for reform

In August 2020, the *Wall Street Journal* reported that Ankhi Das, Facebook's top public policy executive for India as well as South and Central Asia, had overruled her staff to prevent the removal of hate speech and incitement to violence posted on Facebook by politicians belonging to the ruling Hindu nationalist Bharatiya Janata Party (BJP).[19] Das, who herself had previously posted anti-Muslim hate speech on Facebook, reportedly justified her decision by saying that "punishing violations by politicians from Mr. Modi's party would damage the company's business prospects in the country." This was not the first time that Das had interfered in content moderation decisions in the BJP's favor.

According to the company's former Chief Information Security Officer, "a core problem at Facebook is that one policy org is responsible for both the rules of the platform and keeping governments happy."[20] Indeed, its content policy and government relations functions are merged from the C-suite down. As of late 2020, VP of Global Public Policy Joel Kaplan, hired to strengthen the company's ties to the GOP, is in charge of lobbying the U.S. government and overseeing the team in charge of content rules.[21] There is evidence of Kaplan torpedoing efforts to increase civility on the platform, fact-check political ads, and generally enforce rules on hate speech, incitement to violence, and voter suppression against Donald Trump and other Republican politicians.[22] In many cases, political considerations and a perceived need to kowtow to those in power influenced Facebook's content rules and enforcement decisions.[23]

Weak top-level human rights oversight undermines platforms' ability to halt the global spread of false information. In particular, Facebook is the world's dominant

---

[19] Newley Purnell & Jeff Horwitz, *Facebook's Hate-Speech Rules Collide with Indian Politics*, THE WALL STREET JOURNAL (Aug. 14, 2020), https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346.

[20] *Id.*

[21] Hayley Tsukayama, *Facebook Taps D.C. Office Head to Manage Global Policy*, THE WASHINGTON POST (Oct. 6, 2014), https://www.washingtonpost.com/news/the-switch/wp/2014/10/06/facebook-taps-d-c-office-head-to-manage-global-policy.

[22] Jeff Hortwitz & Deepa Seetharaman, *Facebook Executives Shut Down Efforts to Make the Site Less Divisive*, THE WALL STREET JOURNAL (May 25, 2020), https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499.

[23] *See* Jack M. Balkin, *supra* note 6.

platform for organized social media manipulation, in part due to its near-ubiquitous reach and widespread use as a news source.[24] Its problems are exacerbated by a proclivity to underreport political manipulation efforts in the "Global South," which whistleblowers have partially attributed to a lack of interest in "non-priority" markets among executives.[25] Meanwhile, domestically, Mark Zuckerberg has personally intervened to bend Facebook's own Community Standards to retain posts by Donald Trump that glorified violence, triggering strong internal and external criticism.[26]

The frequency and scale of such back-room content decisions at the executive level is unknown. Researchers, journalists, and the public all rely on scattered, leaked reports that likely capture a fraction of such actions. Additionally, Facebook does not disclose any data on "reporter appeals"—content that is left up after being reported, which shrouds its decision-making in further obscurity.[27] Facebook's opaque organizational chart and corporate governance processes impede transparency on how individual decisions are made, including on how to respond to misinformation that originates from high-profile individuals.

A company's top executives are responsible for managing conflicts of interest between the government relations and content policy functions, identifying the human rights risks involved, and mitigating them. If they fail to do so, the Board of Directors should exercise oversight and compel reform. While Facebook claims it oversees human rights at the Board level, it has fiercely resisted efforts to incorporate relevant expertise into its highest echelons of leadership and persuaded its shareholders to reject a 2020 proposal calling for the appointment of a recognized human and civil rights expert to its Board.[28]

---

[24] *See* Samantha Bradshaw & Philip N. Howard, *The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation.* UNIVERSITY OF OXFORD PROJECT ON COMPUTATIONAL PROPAGANDA (Sept. 26, 2019), https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf.

[25] Craig Silverman et al. *"I Have Blood on My Hands": A Whistleblower Says Facebook Ignored Global Political Manipulation.* BUZZFEED NEWS (Sept. 14, 2020), https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo.

[26] Mike Isaac et al., *Zuckerberg Defends Hands-Off Approach to Trump's Posts*, THE NEW YORK TIMES (June 2, 2020), https://www.nytimes.com/2020/06/02/technology/zuckerberg-defends-facebook-trump-posts.html.

[27] Facebook, *Understanding the Community Standards Enforcement Report*, FACEBOOK, https://transparency.facebook.com/community-standards-enforcement/guide (last visited Sept. 24, 2020).

[28] The proposal garnered 3.71% of the shareholder vote and was not successful. *See* Securities and Exchange Commission, *Schedule 14A Information: Proxy Statement Pursuant to Section 14(a) of the Securities Exchange Act of 1934,* SECURITIES AND EXCHANGE COMMISSION (Apr. 10, 2020),

---

Securing an institutional commitment from top leadership to fundamental human rights and oversight requires years of sustained pressure supported by research that exposes crucial shortcomings. While some technology companies have recently disclosed evidence of oversight over human rights risks at the Board and executive levels, this remains the exception.[29] For instance, since its reorganization under Alphabet in 2015, Google's Board of Directors has not exercised formal, systematic oversight over freedom of expression issues, and only recently brought privacy under the purview of the Board's Audit Committee.[30] Similarly, Twitter has not made any public statements about Board-level oversight of free expression and only recently announced that its Data Protection Officer provides regular updates to the Board as part of Twitter's push toward greater transparency on privacy.[31]

Recommended best practices

Companies that are serious about combating misinformation should give themselves the tools to do so, starting at the top of the hierarchy by ensuring that the Board of Directors has the relevant expertise and independence. According to the Business Roundtable, the composition of a Board should reflect a diversity of thought, backgrounds, skills, experiences, and expertise, and a range of tenures that are appropriate given the company's current and anticipated circumstances.[32] Together, these features should enable the Board to perform its oversight function effectively.

For large social media companies whose platforms are rife with misinformation, the Board should include at least one independent expert on civil and human rights, notably freedom of expression, privacy, freedom from discrimination, and the relationship between these values. The Board should also designate and appoint a separate committee responsible for overseeing the company's human rights

---

https://www.sec.gov/Archives/edgar/data/1326801/000132680120000037/facebook2020definitiveprox.htm#s0d2a815be7c34c01823d89c2956ef1e3.

[29] *See* Ranking Digital Rights, *2019 RDR Corporate Accountability Index: Governance* at 3.2, RANKING DIGITAL RIGHTS (May 15, 2019), https://rankingdigitalrights.org/index2019/report/governance/#section-32.

[30] Ranking Digital Rights, *2019 RDR Corporate Accountability Index* at 26-27, RANKING DIGITAL RIGHTS (May 15, 2019), https://rankingdigitalrights.org/index2019/assets/static/download/RDRindex2019report.pdf.

[31] Damien Kieran & Kayvon Beykpour, *What We Have Been Doing to Protect Your Privacy and Data*, TWITTER BLOG (Dec. 2, 2019), https://blog.twitter.com/en_us/topics/company/2019/privacy_data_protection.html.

[32] *See* Business Roundtable, *Principles of Corporate Governance*, HARVARD LAW SCHOOL FORUM ON CORPORATE GOVERNANCE (Sept. 8, 2016), https://corpgov.law.harvard.edu/2016/09/08/principles-of-corporate-governance/.

policies and practices and ensure that similar committees and teams are embedded throughout the company.

In addition to expertise, it is a well-established principle that an independent chairperson of the Board is an essential element of good corporate governance, necessary to ensure proper checks and balances within the company. For example, the Council of Institutional Investors, an authoritative voice on corporate governance, supports the separation of CEO and chairperson roles, cautioning that a "CEO who also serves as a chair can exert excessive influence on the board and its agenda."[33] This is precisely the case with Facebook, whose CEO's dual role as Chairman of the Board effectively renders him accountable to himself. If companies are serious about Board independence and oversight, they must avoid this kind of CEO duality and move away from having company executives and insiders serve on the Board altogether.

Companies that concentrate power in one person are also likely to focus on a single set of interests, rather than carrying out a careful balancing of the competing values at play where human rights questions are engaged. Managers, in this context, may prioritize the company's immediate short-term interests over conflicting ones, even to the detriment of shareholders' own financial interests.[34] Where one individual wields excessive influence over a Board lacking human rights expertise, incentives for human rights-based decision-making can be further diminished, particularly where they are framed as considerations which do not impact a company's financial health. All of these factors undermine companies' ability to combat systemic problems like misinformation, particularly since this challenge manifests differently across different local contexts. The disconnect between First Amendment values and the companies' global user base further exacerbates the problem.

---

[33] *See* Council of Institutional Investors, *Independent Board Leadership*, COUNCIL OF INSTITUTIONAL INVESTORS, https://www.cii.org/independent_board (last visited Oct. 1, 2020). *See* Council of Institutional Investors, *Policies on Corporate Governance* at 2.4, COUNCIL OF INSTITUTIONAL INVESTORS (Sept. 22, 2020), https://www.cii.org/corp_gov_policies#2.4. ("The board should be chaired by an independent director. The CEO and chair roles should only be combined in very limited circumstances; in these situations, the board should provide a written statement in the proxy materials discussing why the combined role is in the best interests of shareowners, and it should name a lead independent director who should have approval over information flow to the board, meeting agendas and meeting schedules to ensure a structure that provides an appropriate balance between the powers of the CEO and those of the independent directors.")

[34] Jena Martin, *Business and Human Rights: What's the Board Got to Do with It*, 3 U. ILL. L. REV. 959, 970 (2013).

Social media companies with an independent Board, including Board members with relevant human rights expertise, would be much better equipped to ensure that internal policies and procedures are applied consistently, as opposed to implementing rules in a way that is reactive to political influence and pressure. This includes separating the company's government relations and content governance functions. A strong "human rights by design" approach would bolster good practices from the top, with transparency into the circumstances in which senior executives may get involved in content decisions, the actual cases in which they did, and the chain of command enabling such decisions to be made.

Legal and regulatory levers for reform

Going beyond best practices, Board expertise, independence, and oversight should also be supported by legal and regulatory measures relating to corporate governance.

There are a number of sources for corporate governance requirements, including laws of the state of incorporation; federal securities laws, notably the Securities Act of 1933 (the "1933 Act") and the Securities Exchange Act of 1934 (the "1934 Act"); other federal laws, including the Sarbanes-Oxley Act of 2002 (Sarbanes-Oxley Act) and the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act); SEC regulations, rules, and guidance; and common law rules. In addition, listing standards published by registered stock exchanges such as the New York Stock Exchange Listed Company Manual (NYSE Listing Manual) and the National Association of Securities Dealers Automatic Quotation System Marketplace Rules (Nasdaq Marketplace Rules) require listed companies to maintain specified corporate governance practices.[35] Corporations are also subject to their own bylaws.

In the wake of financial crises, governments often enact more stringent laws and regulations related to corporate governance. For example, the Sarbanes-Oxley Act, which emerged in the wake of numerous large corporate fraud scandals such as Enron, WorldCom, and Tyco, was designed to implement sweeping reforms to corporate disclosures, financial reporting, and other corporate governance practices to restore investor confidence in the U.S. financial markets. Similarly, Dodd-Frank, enacted in response to the collapse of Lehman Brothers and the sub-

---

[35] *See* Stephen Giove & Robert Treuhold, *Corporate Governance and Directors' Duties in the United States: Overview*, THOMSON REUTERS WESTLAW (Feb. 1, 2013), https://content.next.westlaw.com/9-502-3346.

prime mortgage crisis of 2008, was designed to restore public confidence by restructuring the financial regulatory system and improving corporate governance, including by strengthening shareholder power.

Due to the nature of these crises, the emphasis of their resulting reform efforts was on strengthening corporations' expertise, oversight, disclosures, and reporting on financial matters. For example, SEC rules promulgated per Section 407 of Sarbanes-Oxley require a public company to disclose in its annual report whether or not its Board of Directors has at least one "financial expert" on its audit committee to properly understand and assess the company's financial risks, and, if not, why not.

In the wake of numerous civil and human rights related harms and abuses flowing from the policies adopted by large social media companies,[36] the SEC should adopt similar rules regarding civil and human rights expertise. Specifically, the rules should require companies that pose substantial civil and human rights-related risks to disclose whether they have at least one qualified civil and human rights expert on their Board of Directors and relevant oversight committees—and, if not, why not. This would help shareholders better assess the sufficiency of a company's oversight over human rights-related risks.

Expertise alone is insufficient without sufficient independence to apply it. While the separation of the roles of Board chairperson and CEO is mandated by regulation in some countries,[37] U.S. federal and state laws typically do not dictate or restrict board composition. However, both the NYSE Listing Manual and the Nasdaq Marketplace Rules do require a majority of Board members to be independent. Per the NYSE Listing Manual, anyone who owns company shares does not qualify as independent,[38] while ownership of company stock is a factor but not dispositive

---

[36] For a U.S. example, *see* Russell Brandom, *Facebook Chose Not to Act on Militia Complaints Before Kenosha Shooting*, THE VERGE (Aug. 26, 2020), https://www.theverge.com/2020/8/26/21403004/facebook-kenosha-militia-groups-shooting-blm-protest. For an international example, *see* Jennifer Whitten-Woodring et al., *Poison If You Don't Know How to Use It: Facebook, Democracy, and Human Rights in Myanmar*, 3 INT'L J. OF PRESS/POLITICS.

[37] *See, e.g.,* India: KPMG, *SEBI Defers the Timeline for Separation of the Roles of Non-Executive Chairperson and MD/CEO by Two Years*, KPMG (Jan. 17, 2020), https://home.kpmg/in/en/home/insights/2020/01/firstnotes-sebi-chairperson-md-ceo-defer.html. *See also* Organisation for Economic Co-operation and Development, *OECD Corporate Governance Factbook 2019* at 119, ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (June 8, 2019), https://www.oecd.org/corporate/Corporate-Governance-Factbook.pdf.

[38] *See* New York Stock Exchange, *NYSE Rule Guide* at 303A.02 (Independent Tests) (a)(i), NEW YORK STOCK EXCHANGE (Nov. 25, 2009),

---

under the Nasdaq Marketplace Rules[39] (note that Facebook chose Nasdaq over the NYSE for its IPO).[40]

Where companies nevertheless insist on CEO duality, they are required to disclose their rationale for doing so. Per 2009 amendments to Regulation S-K of the 1933 Act, which lays out the reporting requirements for the SEC filings of public companies, companies must disclose their reasons for combining or separating the roles of CEO and chairperson of the board in their proxy statement.[41] The amendments were designed to provide investors with more transparency over corporate governance matters. As the form and nature of these disclosures is not

---

https://nyseguide.srorules.com/listed-company-manual/document?treeNodeId=csh-da-filter!WKUS-TAL-DOCS-PHC-%7B0588BF4A-D3B5-4B91-94EA-BE9F17057DF0%7D--WKUS_TAL_5667%23teid-69. ("No director qualifies as "independent" unless the board of directors affirmatively determines that the director has no material relationship with the listed company (either directly or as a partner, shareholder or officer of an organization that has a relationship with the company).")

[39] *See* Nasdaq, *Nasdaq Rulebook* at IM-5605. Definition of Independence — Rule 5605(a)(2), NASDAQ LISTING CENTER, https://listingcenter.nasdaq.com/rulebook/nasdaq/rules/nasdaq-5600-series. ("It is important for investors to have confidence that individuals serving as Independent Directors do not have a relationship with the listed Company that would impair their independence. The board has a responsibility to make an affirmative determination that no such relationships exist through the application of Rule 5605(a)(2). Rule 5605(a)(2) also provides a list of certain relationships that preclude a board finding of independence. These objective measures provide transparency to investors and Companies, facilitate uniform application of the rules, and ease administration. Because Nasdaq does not believe that ownership of Company stock by itself would preclude a board finding of independence, it is not included in the aforementioned objective factors. It should be noted that there are additional, more stringent requirements that apply to directors serving on audit committees, as specified in Rule 5605(c).")

[40] *See* Jessica Guynn, *Facebook Selects Nasdaq over NYSE for IPO*, THE LOS ANGELES TIMES (Apr. 5, 2012), https://www.latimes.com/business/la-xpm-2012-apr-05-la-facebook-ipo-nasdaq-nyse-20120405-story.html.

[41] The SEC amended Item 407 of Regulation S-K and Schedule 14A to require disclosure of whether and why a company has chosen to combine or separate the principal executive officer and Board chairperson positions, and the reasons why it believes that this Board leadership structure is the most appropriate structure for the company at the time of filing. If these positions are combined, and a lead independent director is designated to chair meetings of the independent directors, the amendments require disclosure of why this leadership structure was chosen, as well as the specific role the lead independent director plays in the company's leadership structure. *See* 17 CFR § 229.407 (h). ("*Board leadership structure and role in risk oversight.* Briefly describe the leadership structure of the registrant's board, such as whether the same person serves as both principal executive officer and chairman of the board, or whether two individuals serve in those positions, and, in the case of a registrant that is an investment company, whether the chairman of the board is an "interested person" of the registrant as defined in section 2(a)(19) of the Investment Company Act (15 U.S.C. 80a-2(a)(19)). If one person serves as both principal executive officer and chairman of the board, or if the chairman of the board of a registrant that is an investment company is an "interested person" of the registrant, disclose whether the registrant has a lead independent director and what specific role the lead independent director plays in the leadership of the board. This disclosure should indicate why the registrant has determined that its leadership structure is appropriate given the specific characteristics or circumstances of the registrant. In addition, disclose the extent of the board's role in the risk oversight of the registrant, such as how the board administers its oversight function, and the effect that this has on the board's leadership structure.").

---

standardized, firms use widely divergent wording to describe their rationale, making it difficult to study the true impact of such disclosures on corporate governance and the market.[42] According to at least one analysis, the two most commonly cited reasons for separating the CEO and Chairperson roles are the differences between the roles (and its associated tasks) as well as enhanced oversight and monitoring of management.[43] While there is evidence that CEO duality has implications for shareholder value,[44] more guidance or standardization about the nature of these disclosures would advance research and analysis in the ongoing debate around CEO duality.

Even with these reforms, transparency alone on the part of company leadership is not enough to ensure good corporate governance. Armed with this improved transparency, shareholders must be empowered to exercise meaningful oversight, which is the subject of the next section.

IV. Empowering Investor Oversight

The case for reform

Independent analyses suggest a growing rift between executives and investors with regard to the time and attention that Boards should place on human rights and Environmental, Social, and Governance (ESG) concerns. In PwC's 2019 Annual Corporate Directors Survey, only 29% of Directors felt their Board needed more reporting on ESG measures, and a similar percentage felt that public ESG disclosure should be a priority.[45] Directors were much more inclined to declare the Board already had a strong understanding of ESG issues. Additionally, nearly half of those surveyed felt that institutional investors dedicate too much attention to corporate social responsibility (CSR) issues (up from 29% the prior year) and only one in five viewed human rights as a topic that could credibly impact company strategy. In contrast, a majority of institutional investors favors the use of ESG

---

[42] *See* Marc Goergen et al., *On the Choice of CEO Duality: Evidence from a Mandatory Disclosure Rule* (CFR Working Paper No. 18-06, Centre for Financial Research (CFR), University of Cologne, 2018), https://www.econstor.eu/bitstream/10419/191061/1/1045535931.pdf.

[43] *Id.* at 11.

[44] *See id.* at 23.

[45] PwC, *The Collegiality Conundrum: Finding Balance in the Boardroom – PwC's 2019 Annual Corporate Directors Survey*, PwC, at 31 (October 2019), http://www.circulodedirectores.org/wp-content/uploads/2019/12/pwc-2019-annual-corporate-directors-survey-full-report-v2.pdf.

frameworks to assess non-financial disclosures and a rapidly growing number believe companies do not adequately disclose ESG risks.[46]

These results suggest that boardrooms and investors have increasingly disparate visions of companies' responsibility toward the people who use their services and their impact on civil and human rights. The rift is compounded by the common view among executives that it is difficult to express dissenting perspectives in the boardroom—a trend that appears to be markedly stronger in companies that merge the roles of CEO and Board chair, demonstrating another downside of CEO duality in large technology companies.[47] These trends also illustrate why reforms at the Board level alone are insufficient to improve corporate governance in ways that will meaningfully mitigate ESG risks, and specifically human rights-related risks that manifest themselves through widespread mis- and disinformation and hate speech.

In addition to the Board of Directors, shareholders—individuals and institutions that own stock in a company—also play an oversight role. One vehicle for that oversight is the shareholder resolution, a formal proposal subject to a vote at the company's annual meeting. [48] Most proposals are non-binding and are typically opposed by management: if management agreed, there would be no need for a resolution. Resolutions typically focus on corporate governance, and increasingly involve human rights as well as environmental and social responsibility issues.[49]

Shareholder activism is a growing influence vector for activists seeking to change corporate behavior. However, this activism is often hampered by structural limitations such as dual-class or multi-class voting structures that dilute the influence of public investors relative to company insiders. Approximately 90% of all public companies in the U.S. have a 'one share one vote' structure, and IPOs in

---

[46] EY, *How Will ESG Performance Shape Your Future? Why Investors Are Making ESG an Imperative for COVID-19 and Beyond – Climate Change and Sustainability Services (Ccass) Fifth Global Institutional Investor Survey*, EY (July 22, 2020), https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/assurance/assurance-pdfs/ey-global-institutional-investor-survey-2020.pdf.

[47] PwC, *supra* note 45, at 29.

[48] *See also* Rebecca MacKinnon & Melissa Brown, *Digital Rights Outlook for 2021: Geopolitical Risks Are Rising—and Regulation Is Coming (Fall 2020 Investor Update)* (2020), https://rankingdigitalrights.org/wp-content/uploads/2020/10/Fall-2020-Investor-Update.pdf.

[49] *See* Diana Kearney & Sharmeen Contractor, *Investors Embrace Human Rights in the Age of Corona*, RESPONSIBLE INVESTOR (Sept. 4, 2020), https://www.responsible-investor.com/articles/investors-embrace-human-rights-in-the-era-of-corona.

recent years have largely mirrored the broader market in this regard.[50] But large technology companies subvert this trend: in the first half of 2019, seven out of the ten largest tech IPOs (including Lyft, Pinterest, and Zoom) adopted a dual-class or multi-class structure.[51] Dual-class shares grant differential voting rights to a favored class of founders and insiders while retaining standard voting power (or none at all) for an inferior class of public investors.

Typically, the disparity between the voting power of the two classes is ten to one, as in the case of Facebook and Zoom. However, the superior class in companies such as Pinterest holds twenty times the voting power that ordinary shareholders do, while Snap and Alphabet both maintain a triple-class structure in which the lowest tier is given no right to vote whatsoever. The superclass skews the playing field further in its favor when unequal voting rights are supplemented with an imbalance between the company's ownership and its voting hierarchy. Alphabet, for instance, has a voting superclass that controls more than 60% of the vote but retains less than 7% of the company's total outstanding equity; Snap's superclass controls nearly the entirety of the vote with an equity stake of less than 17%.[52] None of these companies have adopted sunset provisions for their dual-class structures, thus reducing the locus of decision-making power to a narrow group of insiders in perpetuity.

---

[50] Council of Institutional Investors, *Dual-Class IPO Snapshot: 2017—2019 Statistics*, COUNCIL OF INSTITUTIONAL INVESTORS, https://www.cii.org/files/Jan%202020%20Dual%20Class%20Update%20for%20Website.pdf (last visited Oct. 2, 2020).

[51] *See* Kosmas Papadopoulos, *Dual-Class Shares: Governance Risks and Company Performance*, HARVARD LAW SCHOOL FORUM ON CORPORATE GOVERNANCE (June 28, 2019), https://corpgov.law.harvard.edu/2019/06/28/dual-class-shares-governance-risks-and-company-performance/.

[52] Council of Institutional Investors, *Dual-Class Companies List: June 2020*, COUNCIL OF INSTITUTIONAL INVESTORS (June 2020), https://www.cii.org/files/Formatted%20Dual%20Class%20List%206-22-20.pdf.

	Superclass voting power	Standard class voting power	Superclass vote control	Superclass equity stake	Wedge (vote control - equity stake)	Sunset clause?
Alphabet*	10	1	60.85%	6.69%	54.16%	No
Amazon	N/A	N/A	N/A	N/A	N/A	N/A
Apple	N/A	N/A	N/A	N/A	N/A	N/A
Facebook	10	1	65.29%	15.83%	49.46%	No
Microsoft	N/A	N/A	N/A	N/A	N/A	N/A
Pinterest	20	1	99.06%	84.1%	14.96%	No
Snap*	10	1	97.89%	16.89%	81%	No
Twitter	N/A	N/A	N/A	N/A	N/A	N/A

**Alphabet and Snap both have a third tier of shareholders with no voting rights.*

Table 1. Voting rights and ownership figures for eight prominent U.S. technology companies. Source: adapted from Council of Institutional Investors, *Dual-Class Companies List: June 2020*, Council of Institutional Investors (June 2020), https://www.cii.org/files/Formatted%20Dual%20Class%20List%206-22-20.pdf.

A structural manifestation of the disproportionate power reserved for top executives and founders, dual-class structures both fossilize existing policy approaches and virtually shut down the prospect of anchoring them in international human rights standards. This further erodes companies' ability to manage complex and geographically nuanced problems like misinformation.

Recommended best practices

Retaining power through lopsided voting configurations is a choice, not an inevitability. Apple, Amazon, Microsoft, and Twitter all operate without a dual-class

share structure.[53] While this approach has not immunized them from human rights shortcomings, it nonetheless demonstrates that balancing the pursuit of profit with accountability to shareholders is possible in the sector without skewing the playing field. While tech giants that are entrenched in the dual-class model view it as a way to insulate themselves against activist investors and initially adopted it to maintain founders' control,[54] board members of companies across various industries are recognizing the reputational harm of this structure. According to industry surveys, nearly two thirds of corporate directors believe dual-class share structures reflect negatively or very negatively on the Board.[55] Continued pressure from civil society and responsible investors is likely to further this tendency and feed into critical public sentiment about companies operated as fiefdoms.

Early-stage companies should either consider avoiding dual-class structures altogether, or at least consider how to phase them out as the company grows and evolves, particularly to avoid negative public sentiment and reputational harm. Later-stage companies that already have dual-class or multi-class structures should reassess the potential reputational harm and increased scrutiny they create and consider restructuring ownership or else implementing countermeasures to address the risks. For example, where a company cannot easily restructure voting rights, they may consider whether other positive corporate governance measures, such as the separation of the CEO and Chairperson roles, more openness and receptivity towards shareholder resolutions, or voluntary enhanced corporate disclosures about the rationale for introducing or retaining differential voting structures, may help to improve investor or market confidence and trust in company leadership. While founders are unlikely to voluntarily relinquish control, companies should recognize that government intervention becomes more likely as public pressure mounts and that voluntary measures are in their own self-interest.

In addition, companies should disclose information about the "wedge" between ownership and control. At present, investors cannot easily quantify the difference between insider control of the vote and their equity stake—or the ways in which that difference can change. Companies have a range of options to disclose ownership information and are not required to disclose how the power differential between the

---

[53] Rani Molla, *More Tech Companies Are Selling Stock That Keeps Their Founders in Power*, Vox (Apr. 11, 2019), https://www.vox.com/2019/4/11/18302102/ipo-voting-multi-dual-stock-lyft-pinterest.

[54] Vijay Govindarajan & Anupt Srivastava, *Reexamining Dual-Class Stock*, 3 Business Horizons 461 (2018).

[55] PwC, *Turning Crisis into Opportunity: PwC's 2020 Annual Corporate Directors Survey*, PwC, at 18 (Sept. 2020), https://www.pwc.com/us/en/services/governance-insights-center/assets/pwc-2020-annual-corporate-directors-survey.pdf.

superclass and the other class or classes can change in their IPO registration statement. As the SEC's Investor Advisory Committee remarks, this means that the co-founder of a company like Snap can reduce their stake to below 1% without relinquishing total voting control, and no mechanism compels them to notify investors of the resulting risk.[56] Without a suite of disclosures on these topics, any ill-conceived approaches to issues like platform content governance will be harder to reverse by investor activism.

Legal and regulatory levers for reform

As with reforming Board composition and expertise, empowering shareholders to exercise adequate oversight and hold company leadership accountable for their social impact, including with respect to the human rights of their users and other impacted stakeholders, may require going beyond corporate best practices to include certain legal and regulatory interventions.

One option may be to limit the ability of publicly traded companies to create and maintain dual-class or multi-class share structures. Typically designed to allow the founders of emerging companies to experiment and innovate as they grow, differential voting schemes effectively decouple voting rights from one's economic interest in a company's future success. In fact, research suggests that companies with dual-class share structures are more likely to exhibit problematic corporate governance practices and, in turn, face more corporate governance challenges than companies without such structures in place.[57] As a result, many countries, including Australia, Germany, Italy, Spain, and the United Kingdom, have regulations that prohibit or restrict dual-class shares. For example, the London Stock Exchange's listing regime prohibits dual-class or differential voting structures within premium-listed companies, a category that spans many of the U.K.'s largest and most influential companies.[58] As with mandatory disclosures about the rationale for CEO duality or non-duality per amendments to Regulation S-K, the SEC may also

---

[56] Securities and Exchange Commission Investor Advisory Committee, *Recommendation of the Investor Advisory Committee Dual Class and Other Entrenching Governance Structures in Public Companies*, SECURITY AND EXCHANGE COMMISSION INVESTOR ADVISORY COMMITTEE (Mar. 2018), https://www.sec.gov/spotlight/investor-advisory-committee-2012/recommendation-on-dual-class-shares.pdf.

[57] *See* Kosmas Papadopoulos, *supra* note 48.

[58] *See* Marc T. Moore, *Designing Dual Class Sunsets: The Case for a Transfer-Centered Approach*, UNIVERSITY OF OXFORD BUSINESS LAW BLOG (Jan. 16, 2020), https://www.law.ox.ac.uk/business-law-blog/blog/2020/01/designing-dual-class-sunsets-case-transfer-centered-approach. ("The general tolerance shown towards DCS by US regulators puts the United States in stark contrast to numerous other jurisdictions including Australia, Belgium, Brazil, Germany, Italy, Spain and the United Kingdom.")

consider promulgating rules that require companies with dual- or multi-class share structures to explain their rationale for doing so.

Another option would be for the SEC to promulgate rules that enable shareholder activism by way of shareholder resolutions and other means, while rolling back rules that put up hurdles to this kind of activity. Unfortunately, in recent years, the SEC has chipped away at some of the core tools of activist investors concerned with ESG matters. For example, in September, the SEC enacted a rule that required shareholders to hold at least $25,000 of stock for a year (up from $2,000) before filing a shareholder resolution, increased the re-submission thresholds of votes required to get proposals included in the company's annual proxy materials, and included a variety of amendments to existing rules making it harder to vote by proxy.[59] The SEC should reverse these rule changes, which make it more difficult for shareholders to file proposals and to get them on proxy ballots.

Finally, the SEC should promulgate rules requiring publicly traded companies to disclose what percentage of their revenue comes specifically from targeted advertising, as opposed to advertising more generally, as targeted advertising creates perverse incentives that feed the proliferation of mis- and disinformation, polarization, and hate speech on large social media platforms. The SEC should also require companies to disclose non-financial information about their ESG impact, including information about the social impact of targeted advertising and algorithmic systems. Such enhanced disclosures help to empower investors by providing them with the information they need to make informed voting and investment decisions on areas of risk, including the kind of rampant misinformation that undermines and threatens the human rights of the company's user base.

V. Conclusion

This paper argues that the spread of misinformation is a downstream manifestation of upstream corporate governance and management failures. It is settled doctrine that a Board of Directors' fiduciary duties include establishing that management has an effective corporate compliance program in place, exercising

---

[59] *See* Security and Exchange Commission, *SEC Adopts Amendments to Modernize Shareholder Proposal Rule*, SECURITY AND EXCHANGE COMMISSION (Sept. 23, 2020), https://www.sec.gov/news/press-release/2020-220. Under the new rules, shareholders must own $25,000 of the company's securities for at least one year, $15,000 for two years, or $2,000 for three years before they can submit a proposal.

oversight of that program, and taking regular steps to stay informed of the program's content and operation.[60] The solutions we propose—reforming corporate leadership structures and enabling investors to exercise greater oversight—both address some of the root causes of misinformation and other societal ills that platforms propel. Neither requires an intervention into the actual content that proliferates on platforms, but rather into the systems that enable this proliferation.

At the same time, both measures are internationally scalable. Phasing out dual-class share structures would render Boards less prone to cherry-picking their priority markets and relying on scandal-driven due diligence for markets they relegate to a lower tier of importance.[61] Implementing clear rules to prevent the strong-arming of content decisions by executives and upper management could be possible in every country of operation if executives at the highest level demonstrate the willingness to change the status quo. Until top-level incentives are reformed through measures such as these, platforms will continue to address individual conflagrations reactively while ignoring the forest fires they help kindle.

---

[60] *See* Robert Biskup et al., *Board Oversight of Corporate Compliance: Is It Time for a Refresh?* DELOITTE LLP (Oct. 15, 2019), https://corpgov.law.harvard.edu/2019/10/15/board-oversight-of-corporate-compliance-is-it-time-for-a-refresh/.

[61] *See* Kendyl Salcito, *Company-Commissioned HRIA: Concepts, Practice, Limitations and Opportunities*, IN HANDBOOK ON HUMAN RIGHTS IMPACT ASSESSMENT 32 (2019).

Towards Policy and Regulatory Approaches for Combating Misinformation in India

Akriti Gaur

I. Introduction

Misinformation and hate speech on WhatsApp, Twitter, and Facebook have been contributing to violence and social unrest in India.[1] Currently, the government and users rely on private platforms to address the misinformation challenge on their own. This has resulted in parallel regulatory structures for hundreds of millions of Indians through the content moderation policies of online platforms, but without the legitimacy, oversight, or formal codification which bind traditional institutions. Recent reports of politically biased moderation of speech on platforms such as Facebook,[2] opacity in how platforms address misinformation across regions and user bases, and a general lack of public engagement by 'Silicon Valley' corporations in their largest market in the world have all served to rapidly erode public trust.[3] Platforms have introduced some measures to mitigate harmful content in India, while also acknowledging the need to "do more".[4] The Indian Government response has focused on changes to the intermediary liability regime,

---

[1] Chinmayi Arun, *On WhatsApp, Rumours, Lynchings, and the Indian Government*, ECONOMIC & POLITICAL WEEKLY (Feb. 2019), https://www.epw.in/journal/2019/6/insight/whatsapp-rumours-and-lynchings.html; Timothy McLaughlin, *How WhatsApp Fuels Fake News and Violence in India,* WIRED (Dec. 2018), https://www.wired.com/story/how-whatsapp-fuels-fake-news-and-violence-in-india; Archit Mehta, *Delhi Riots" The Link Between Misinformation and* Radicalisation, THE WIRE (Aug. 2020), https://thewire.in/communalism/delhi-riots-misinformation-radicalisation-social-media.

[2] Newley Purnell and Jeff Horwitz, *Facebook's Hate-Speech Rules Collide With Indian Politics*, THE WALL STREET JOURNAL (Aug. 14, 2020), https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346; Billy Perrigo, *Facebook's Ties to India's Ruling Party Complicate Its Fight Against Hate* Speech, TIME (Sept. 27, 2020), https://time.com/5883993/india-facebook-hate-speech-bjp/.

[3] Out of an estimated 1.5 billion people who use WhatsApp across the world, around 200 million users reside in India. India is also the largest market for Facebook, with an estimated 346 million users, *WhatsApp now has 1.5 billion monthly active users, 200 million users in India*, THE FINANCIAL EXPRESS (Feb. 1, 2018), https://www.financialexpress.com/industry/technology/whatsapp-now-has-1-5-billion-monthly-active-users-200-million-users-in-india/1044468/.

[4] Sandeep Soni, *Progress made but need to do more, says Facebook India's Ajit Mohan on tackling hate speech*, THE FINANCIAL EXPRESS (Aug. 22, 2020), https://www.financialexpress.com/industry/progress-made-but-need-to-do-more-says-facebook-indias-ajit-mohan-on-tackling-hate-speech/2062425/.

---

which are not only ill-designed to address online misinformation but also threaten the fundamental rights of free speech and privacy.[5]

As platforms emerge as "modern public squares"[6] and make speech decisions for billions of people online, there is a need for efficient governance of the ecosystem in which misinformation thrives. Thus far, neither platforms nor the government have effectively performed this function. This paper argues for a new participative and responsive rule-making structure, and an accountable and transparent platform governance framework for India.

The paper proceeds in Part II with a brief description of India's misinformation crisis and the role played by social media platforms and the Indian Government, and introduces the need to shift from traditional forms of rule-making to alternatives such as induced self-regulation. Part III explores possible measures to bolster regulatory capacity to address evolving platform harms and argues for a model of platform governance which involves regulatory nudges towards accountable self-regulation for India and similarly situated jurisdictions. Forms of co-regulation[7] which involve legislative backing and greater government involvement are not immediately desirable owing to jurisdictional challenges and repercussions on free speech and user rights. Accordingly, this part focuses on self-regulation through greater cooperation between platforms, governments, and civil society. Part IV discusses steps that can be taken by social media platforms to tailor platform policies and meaningfully engage with Indian users to address domestic misinformation challenges.

---

[5] Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, https://www.meity.gov.in/writereaddata/files/Intermediary_Guidelines_and_Digital_Media_Ethics_Code_Rules-2021.pdf, [Hereinafter New Intermediary Liability Rules], See also Apar Gupta, *Centre's IT Rules bring answerability in digital ecosystem. But they also increase political control*, THE INDIAN EXPRESS (Feb. 26, 2021), https://indianexpress.com/article/opinion/columns/it-act-social-media-govt-control-digital-ecosystem-7204972/.

[6] Packingham v. North Carolina, 137 S. Ct. 1730 (2017), See also Amber Sinha, *Beyond Public Squares, Dumb Conduits, and Gatekeepers: The Need for a New Legal Metaphor for Social Media*, IT FOR CHANGE (2020), https://itforchange.net/digital-new-deal/2020/11/01/beyond-public-squares-dumb-conduits-and-gatekeepers-the-need-for-a-new-legal-metaphor-for-social-media/.

[7] Various contemporary legislative proposals for regulating social media platforms advocate for a co-regulatory model as a starting point, *for e.g.*, THE REPUBLIC OF FRANCE, CREATING A FRENCH FRAMEWORK TO MAKE SOCIAL MEDIA PLATFORMS MORE ACCOUNTABLE: ACTING IN FRANCE WITH A EUROPEAN VISION – MISSION REPORT 23-24 (2019), https://www.numerique.gouv.fr/uploads/Regulation-of-social-networks_Mission-report_ENG.pdf [hereinafter FRENCH INTERIM MISSION REPORT]; for a comparison of co-regulation models with voluntary compliance and self-regulation, *see* Reeve's Regulatory Pyramid Model, B. Reeve, *The Regulatory Pyramid Meets the Food Pyramid: Can Regulatory Theory Improve Controls on Television Food Advertising to Australian Children?*, JLM 128, 133 (2011).

II. India's misinformation crisis

The role of platforms

In India, platform responses to misinformation have been inconsistently applied across regions and political groups, and suffer from a lack of transparency. Misinformation challenges are continuously evolving and, as a result, solutions are difficult to implement. However, technological changes that have been introduced thus far have been inadequate towards the fundamental goal of empowering users to identify fake news and restrict its circulation.

While platforms such as Facebook and WhatsApp have undertaken awareness campaigns and partnered with local law enforcement officials and policymakers to combat fake news, such measures are often ad-hoc, and not tailored to unique challenges faced by the Indian user base.[8] Many concerns have been raised regarding the efficacy of existing fact-checking programs of platforms. For instance, Facebook's third-party fact checking program was rolled out in Karnataka in 2018 and later expanded to cover other Indian States.[9] More recently, it launched a partnership with independent entities to identify misinformation across 11 Indian languages.[10] However, some experts have expressed concern over the inefficiency of Facebook's partner fact-checking organizations[11] and the lack of action taken by the platform to remove flagged content, rendering the fact-checking process ineffective.[12]

---

[8] Rasmus Kleis Nielsen, *US elections vs Bihar polls: Are all social media users created equal?*, SCROLL.IN (Nov. 10, 2020), https://scroll.in/article/978068/us-elections-vs-bihar-polls-are-all-social-media-users-created-equal.

[9] Daniel Funke, *In one month, Facebook doubled the countries using its fact-checking tool - all outside the West*, POYNTER.ORG (April 18, 2018), https://www.poynter.org/fact-checking/2018/in-one-month-facebook-doubled-the-countries-using-its-fact-checking-tool-%c2%97-all-outside-the-west/, PTI, *Facebook teams up with AFP to expand fact-checking programme in India,* BUSINESS TODAY (Nov. 6, 2018), https://www.businesstoday.in/current/economy-politics/facebook-partners-with-afp-to-expand-fact-checking-programme-in-india/story/288651.html.

[10] PTI, *Facebook teams up with 8 third-party fact checkers, covering 11 Indian languages, to flag Covid-19 fake news*, THE ECONOMIC TIMES (April 22, 2020), https://economictimes.indiatimes.com/magazines/panache/facebook-teams-up-with-8-third-party-fact-checkers-covering-11-indian-languages-to-flag-covid-19-fake news/articleshow/75284845.cms.

[11] Gopal Sathe, *Fact-Checkers Fight Fake News On Facebook. But Who Fact-Checks Them?*, HUFFPOST (Jan. 1, 2020), https://www.huffpost.com/archive/in/entry/fact-checking-fake-news-facebook-how-does-ifcn-work_in_5ca0fd29e4b00ba6327eb726.

[12] Anumeha Chaturvedi, *It's up to Facebook to respond on its fact checking programme, say FB fact-checkers after IT minister raises concerns*, THE ECONOMIC TIMES ET PRIME (Sept. 2, 2020),

---

Likewise, WhatsApp has introduced voluntary third-party fact-checking systems[13] and restrictions on bulk-messaging[14] to make it tedious for users to forward content to hundreds of users in one go, in order to dampen the virality of harmful content. However, it has been argued that these changes are ineffective and superficial.[15] For instance, the restrictions on bulk-messaging can be bypassed through affordable "clone applications" used by political party workers to forward content to a large number of people, or the use of anonymous phone numbers to send bulk messages.[16]

Platforms have also failed to engage with communities and civil society in collaboratively addressing the misinformation challenge in India. More broadly, platforms are not adequately accountable to State and local governments, as exemplified by Facebook's refusal to appear before a legislative committee investigating the platform's role in riots in India's capital.[17]

The role of government

At present, India's government does not have the tools to effectively govern the ecosystem in which misinformation thrives. India, and similarly situated countries, face significant jurisdictional challenges in attempting to regulate American social media platforms, which often refuse to engage with governments or the regulations they impose. However, India's rules are not fit for purpose in other ways as well.

https://economictimes.indiatimes.com/tech/internet/its-up-to-facebook-to-respond-on-its-fact-checking-programme-says-fb-after-it-minister-raises-concerns/articleshow/77895695.cms.

[13] Scroll Staff, *WhatsApp launches new fact-checking service to fight fake news ahead of elections*, SCROLL.IN (Apr. 02, 2019), https://scroll.in/latest/918725/whatsapp-launches-new-fact-checking-service-to-fight-fake-news-ahead-of-elections.

[14] The Wire Staff, *To Combat Fake News in India, WhatsApp to Limit Forwarding of Messages*, THE WIRE (July 20, 2018), https://thewire.in/tech/to-combat-fake-news-in-india-whatsapp-to-limit-forwarding-of-messages.

[15] Prashant Reddy T., *If WhatsApp Doesn't Regulate Itself, Parliament May Have to Step In*, THE WIRE (July 18, 2018), https://thewire.in/tech/if-whatsapp-doesnt-regulate-itself-parliament-may-have-to-step-in.

[16] Munsif Vengattil, Aditya Kalra and Sankalp Phartiyal, *INSIGHT – In India election, a $14 software tool helps overcome WhatsApp controls*, REUTERS (May 15, 2019), https://www.reuters.com/article/india-election-socialmedia-whatsapp-idINL4N22R3G3.

[17] Varun Thomas Mathew, *The Arrogance of Being Facebook, a Serious Tragedy for the Rule of Law*, THE WIRE (Oct. 8, 2020) https://thewire.in/law/facebook-delhi-assembly-summons-rule-of-law-riots.

The Information Technology (Intermediaries Guidelines) Rules, 2011 ('Intermediary Rules'),[18] enacted under the Information Technology Act, 2000,[19] regulate all intermediaries by the same set of regulations, including ISPs, search engines, DNS providers, mainstream social media platforms, and even cyber cafes. The regulations provide a "safe harbor" to a wide range of intermediaries including social media platforms. However, this regulatory approach is problematic because in applying the same set of obligations to all intermediaries, it fails to acknowledge the diversity of technological characteristics and functions across these services.[20]

More recently, the Indian Government has notified amendments to the Intermediary Liability Rules,[21] which remain broad-brushed. The new Intermediary Liability Rules create differentiated obligations for a new category of intermediaries called "significant social media intermediaries" which are likely to include Big Tech platforms.[22] However, the criteria for their classification is vague and, according to the Internet Freedom Foundation, these Rules provide wide discretion to the government to impose "discriminatory" obligations.[23] The amended Rules place an obligation on platforms to proactively identify and disable access to content based on vaguely defined criteria. This, as argued by experts, poses a threat to user privacy and end-to-end encryption on platforms such as

---

[18] The Information Technology (Intermediaries Guidelines) Rules, 2011, (India) https://www.meity.gov.in/writereaddata/files/GSR314E_10511%281%29_0.pdf [Hereinafter Intermediary Liability Rules].

[19] The Information Technology Act, 2000, (India), https://www.meity.gov.in/content/view-it-act-2000 [Hereinafter IT Act].

[20] Pritika Rai Advani, *Intermediary Liability in India*, EPW, Vol. 48, Issue No. 50, (Dec 14, 2013), https://www.epw.in/journal/2013/50/special-articles/intermediary-liability-india.html; For a discussion on the Indian Intermediary Liability framework, see Chinmay Arun and Sarvjeet Singh, *NOC Online Intermediaries Case Studies Series: Online Intermediaries In India*, (Feb 18, 2015), NATIONAL LAW UNIVERSITY, DELHI, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2566952.

[21] New Intermediary Liability Rules, *supra note 5*; Prashant Reddy T., *New IT Rules: The Great Stretching of 'Due Diligence' Requirements Under Section 79*, THE WIRE (Feb. 27, 2021), https://thewire.in/tech/new-it-rules-the-great-stretching-of-due-diligence-requirements-under-section-79. In 2018 the Indian Government released a draft of the amendments to the Intermediary Rules which have been heavily criticized by experts, academics and civil society. *See e.g.* Siddharth, *Intermediary Liability Amendment: Civil Society Counter Comments On Unlawful Content,* MEDIANAMA, (March 22, 2019), https://www.medianama.com/2019/03/223-intermediary-liability-amendment-civil-society-counter-comments-on-unlawful-content/, Agnidipto Tarafder and Siddharth Sonkar, *Attempt to Curb 'Unlawful Content' Online is Chilling to Free Speech*, THE WIRE (Dec. 25, 2018), https://thewire.in/government/it-ministry-attempt-to-curb-unlawful-content-online-chilling-blow-free-speech.

[22] Rule 2(5), New Intermediary Liability Rules, *supra note 5*.

[23] Internet Freedom Foundation, *Deep dive : How the intermediaries rules are anti-democratic and unconstitutional*, (Feb. 27, 2021) https://internetfreedom.in/intermediaries-rules-2021/.

WhatsApp.[24] The Rules also provide wide powers to the government to force platforms to take down user content, an obligation which poses a significant threat to the free speech of platform users.[25]

The need for new forms of rule-making

At present, self-regulation by platforms has failed to generate satisfactory solutions to address misinformation in India. There are a few instances where such efforts have been initiated, but their functioning is largely opaque. The Information Trust Alliance ('ITA') is one example of a recent collaborative effort by platforms, digital publishers, industry bodies, fact-checking organizations, civil society and academics.[26] The aim of the ITA is to work together to develop standardized procedures for resolving complaints regarding false or disputed content.[27] The ITA has reportedly drafted a Code of Practice, which is on hold due to a lack of consensus among participating social media platforms.[28] As of early 2021, little is known about the deliberations of the ITA beyond this.

Further, as seen above, government efforts to address misinformation also face a number of challenges including jurisdictional tensions and the complex nature of imposing new rules on freedom of expression. Vague, top-down rules, such as the new Intermediary Liability Rules, which force platforms to police content or prejudice users' freedom of speech, are not only harmful to users but also potentially unconstitutional.[29]

In order to address the misinformation challenge in India, rather than prescriptive intermediary liability laws, Indian lawmakers should work with platforms to develop self-regulation or co-regulation that is responsive and collaborative, while still sufficiently accountable to domestic policymakers. This is where "induced"

---

[24] *Id.*
[25] Rule 3(1)(d), New Intermediary Liability Rules, *supra note 5*.
[26] Megha Mandavia, *Social Media to Join Hands to Fight Fake News, Hate Speech*, THE ECONOMIC TIMES (Feb. 19, 2020), https://economictimes.indiatimes.com/tech/internet/social-media-to-join-hands-to-fight-fake-news-hate-speech/articleshow/74200542.cms.
[27] *Id.*
[28] *Id.*
[29] Internet Freedom Foundation, *supra note* 23.

self-regulation[30] or "regulated" self-regulation[31] may be an approach worth considering. In its most basic form, "induced" self-regulation refers to the creation of an enabling regulatory environment which incentivizes or induces industry actors to create rules for self-regulation, while also guarding against state interference or a compromise of the independence of entities.[32] It creates a controlled framework in which industry actors have flexibility to devise norms, but at the same time must be accountable for the enforcement of such norms.

III. Bolstering regulatory capability to combat misinformation in India

This part discusses two possible solutions for India to strengthen its regulatory capability to combat misinformation. First, developing effective standards to combat misinformation through "induced" self-regulation. Second, creating an independent and representative national platform oversight body to work with platforms, ensure their transparency, and co-create codes for evolving platform harms.

A. Induced self-regulation

This paper proposes that India consider a voluntary "outcomes-based" code for misinformation, including key norms for social media platforms to adhere to. Such a code should outline outcomes for platforms to achieve, including design-duties[33] and product features to empower users and fight misinformation.[34] The standards should be developed collaboratively through transparent and participative processes, and governments should evolve efficient metrics for assessing their performance.

---

[30] European Parliament, *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States* (Feb. 2019) at p. 104, https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf, See also OXFORD HANDBOOK OF REGULATION 541 (Robert Baldwin, Martin Cave, and Martin Lodge eds., 2010).
[31] *Id.;* Article 19, *The Social Media Councils: Consultation Paper* (June, 2019) at p. 7, https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.
[32] *Id.*
[33] Olivier Sylvain, Intermediary Design Duties, 50 CONN. L. REV. 203 (2018), https://ir.lawnet.fordham.edu/cgi/viewcontent.cgi?article=1892&context=faculty_scholarship.
[34] *See generally* EU Code of Practice on Disinformation, https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation (last visited Nov. 20, 2020), Daphne Keller, *Who Do You Sue? - State and Platform Hybrid Power Over Online Speech* , Hoover Institution, Aegis Series Paper No. 1902 https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf at p. 24.

The outcomes should be grounded in human rights principles and developed through a collaborative process involving the platforms, civil society, and other stakeholders. Some examples may include protecting the free speech and privacy of users, bolstering user autonomy and access to remedial measures and grievance redressal, and greater transparency and accountability around content moderation and efforts to combat misinformation. It is critical to ensure that the "outcomes-based" code is not vague or tilted to serve state interests, and does not incentivize platforms to adopt an overly heavy approach to removing content. The outcomes should be built around common objectives, and should provide flexibility for platforms to develop protocols and technological tools to achieve them.

Several sectors in India have successfully adopted self-regulation mechanisms. A notable example of 'induced' and outcomes-based self-regulation is the Securities and Exchange Board of India ('SEBI'), which regulates the securities market and protects investor interests.[35] In doing so, SEBI promotes 'self-regulatory organizations' ('SRO'),[36] investor awareness,[37] training of intermediaries in the securities market,[38] and has set up an Investor Protection and Education Fund.[39] The purpose of the SRO framework is to equip a particular group of entities in the securities market (such as investment advisors and fund managers) to act as 'first-level regulators' and provide a guiding framework for their registration, composition, and self-regulatory objectives. This framework also empowers SEBI with tools of inspection and audit to assess the functioning of the SROs and their compliance with high-level objectives to protect investors.[40] The SRO framework thereby allows sectoral entities to set up self-regulatory codes while following a broad set of duties and obligations outlined in law. Unlike the SEBI framework, India does not have a similar regulatory structure around platforms. While a top-down regulatory structure presents challenges to ensuring direct accountability to platforms' users, there are nonetheless certain aspects of the SRO framework that may be useful in guiding collaborative self-regulation for platforms.

---

[35] Preamble, Securities and Exchange Board of India Act, 1992 (SEBI Act) (India).
[36] § 11(2)(d), SEBI Act, 1992 (India).
[37] § 11(2)(f), SEBI Act, 1992 (India).
[38] § 11(2)(f), SEBI Act, 1992 (India).
[39] Securities and Exchange Board of India (Investor Protection and Education Fund) Regulations, 2009 (India), https://investor.sebi.gov.in/ipef.html (last visited Nov. 27, 2020).
[40] See generally, Securities and Exchange Board of India (Self Regulatory Organizations) Regulations, 2004 (India). (last visited Nov. 27, 2020).

Evolving practice globally also indicates a greater push towards accountable and collaborative self-regulation to address misinformation. For instance, the Australian Communications and Media Authority ('ACMA') has recommended a self-regulatory approach[41] to enable platforms to identify risks or "acute harms from misinformation"[42] and devise their own design solutions to empower users. The outcomes outlined by ACMA include reduced exposure of users to harmful misinformation on the platform, a robust system of misinformation reporting, and access to an effective "complaints handling process".[43] The proposal argues for the need to maintain consistency in measures to tackle misinformation, using "industry-wide" inputs for risk-assessment.[44] It also seeks to push platforms to develop common objectives to curb misinformation,[45] but give them the flexibility to develop their own solutions to achieve the objectives,[46] and provide a system of "performance reporting".[47] This mechanism aligns with the "positive state approach" where governments create institutions with incentives for private actors to tackle a challenge, rather than adopting a system which relies on fines and penalties.[48] This may be contrasted against a "negative state" approach in countries like Germany[49] and Singapore,[50] where restrictive regulation poses a threat to the free speech of users.

Some regions have developed codes to combat misinformation, such as the EU Code of Practice on Disinformation[51] ('EU Code') and NATO's approach to address

---

[41] "Misinformation and news quality on digital platforms in Australia - A position paper to guide code development", AUSTRALIAN COMMUNICATIONS AND MEDIA AUTHORITY (June 2020), https://www.acma.gov.au/sites/default/files/2020-06/Misinformation%20and%20news%20quality%20position%20paper.pdf.

[42] *Id.* at p. 11, These risks or acute harms should include— health and safety of vulnerable groups, threat of harm to public and private property, threat to elections or democratic processes, and imminent financial harm.

[43] *Id.* at. p.1.

[44] *Id.* at. p.31-32.

[45] *Id.*

[46] *Id.*

[47] *Id.*

[48] *Fighting Fake News – Workshop Report*, INFORMATION SOCIETY PROJECT (Mar. 17, 2017), p. 7, https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf.

[49] Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Netzwerkdurchsetzungsgesetz—NetzDG] [Network Enforcement Act], Sept. 1, 2017, Bundesgesetzblatt, Teil I [BGBl I] at 3352 (Ger.).

[50] Protection from Online Falsehoods and Manipulation Act, 2019, https://sso.agc.gov.sg/Acts-Supp/182019/Published/20190625?DocDate=20190625#:~:text=An%20Act%20to%20prevent%20the,to%20be%20taken%20to%20enhance (Singapore).

[51] EU Code of Practice on Disinformation, *supra note 34.*

misinformation related to COVID-19.[52] However, such collaborative efforts lack necessary measures to monitor progress. For instance, the EU Code lacks key performance indicators, or measures to conduct efficient monitoring, and does not provide adequate information for researchers.[53] These gaps should be explored further when thinking about a similar model for India.

To avoid excessive state control or "content cartel creep",[54] development and implementation of induced self-regulation or co-regulation should be guided by an independent body with adequate tools for oversight. In India, this might be carried out by an independent platform oversight body which operates at an arm's length from both the government as well as platforms. The platform oversight body should ensure that the development of these codes are participative and promote dialogue among state and local governments, platforms, civil society, academics, and other stakeholders.[55] The codes should provide platforms the flexibility to develop consensus and devise self-regulatory norms. Further, such codes should outline key performance indicators and lay down a metric for the independent body to monitor compliance. As seen with the SRO framework, this may be done through non-financial reporting norms, transparency reports, and regular audits. In order to ensure accountability and oversight, codes which are independently developed by platforms may be approved or accredited by the oversight body.

Since India does not have an existing legal framework around platform governance, an outcomes-based self-regulatory framework can serve as a necessary first step for stakeholders to work together to combat online

---

[52] NATO's approach to countering disinformation: A focus on COVID-19, (July 17, 2020) https://www.nato.int/cps/en/natohq/177273.htm.

[53] European Commission, Staff Working Document titled "Assessment of the Code of Practice on Disinformation – Achievements and areas for further improvement" (Sept. 10, 2020) https://ec.europa.eu/digital-single-market/en/news/assessment-code-practice-disinformation-achievements-and-areas-further-improvement.

[54] A term coined by Evelyn Douek, which means that the "relationship between economic competition in the platform market and a healthy public sphere is complicated concentration of platform power", Evelyn Douek, *The Rise of Content Cartels*, KNIGHT FIRST AMENDMENT INSTITUTE (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels.

[55] It is critical to ensure that the oversight body working with platforms to develop self-regulatory codes is transparent and collaborative. In India especially, a culture of opacity shrouds law-making by state actors and regulators like the Ministry of Electronics and IT are often criticized for the lack of transparency and public participation in framing rules for Internet governance. *See, e.g.*, Ayesha Khan, *Submission to the Ministry of Electronics and Information Technology, Government of India, on the Draft Non-Personal Data Governance Framework*, WIKIMEDIA/YALE LAW SCHOOL INITIATIVE ON INTERMEDIARIES AND INFORMATION BLOG (Sept. 16, 2020), https://law.yale.edu/submission-ministry-electronics-and-information-technology-government-india-draft-non-personal-data.

misinformation. In the absence of enforcement tools, its implementation hinges on platforms' willingness to work together. A potential oversight body may also face hiccups in terms of conducting audits or effectively monitoring implementation. However, this approach would be useful in bridging the gap between the government and social media platforms, and potentially avoid prescriptive regulations such as the amendments to the Intermediary Liability Rules which pose a larger threat to user rights. In the short-term this framework should, at the very least, prod platforms to take holistic action in curbing misinformation and creating an environment of greater transparency around their actions in this space.

B. Creating a national platform oversight body

In India, efforts to combat misinformation have been ad-hoc or siloed, and regulators have been unable to craft regulation which effectively tackles misinformation while also preserving user rights. In order to efficiently implement self-regulatory or co-regulatory efforts discussed in the previous section, there is a need for an alternative structure of governance where governments and platforms can develop a regulatory system which is responsive to evolving platform harms. An initiative of this nature should respect the central role that platforms play in facilitating the right to freedom of expression.

While ad-hoc measures have been taken to address certain aspects of the misinformation problem, there is no systematic approach to engage with platforms to collaboratively address the issue. Several government agencies, such as the Ministry of Home Affairs, the Ministry of Electronics and Information Technology ('MEITy'), the Department for Promotion of Industry and Internal Trade, and the Election Commission of India, have on different occasions issued advisories and policies for social media platforms to curb misinformation. Government officials have also attempted to engage directly with platforms like Facebook, WhatsApp and Twitter. However there is no metric to assess the success of these engagements. The advisories and calls to action are often not implemented by platforms. In some instances, the Indian Government has also attempted to tackle misinformation through repressive policies which may be used to unfairly target credible journalism.[56]

---

[56] *After widespread criticism and PMO directive, I&B ministry withdraws press release on fake news*, TIMES OF INDIA (April 3, 2018), https://timesofindia.indiatimes.com/india/after-widespread-criticism-pmo-says-withdraw-press-release-on-fake-news/articleshow/63593207.cms, Kai Schultz and Suhasini Raj, *India's 'Fake News' Crackdown Crumbles Over Journalists' Outrage,* THE NEW YORK TIMES (April 3, 2018) https://www.nytimes.com/2018/04/03/world/asia/india-fake-news.html.

Globally, institutions to support collaborative action around misinformation are evolving gradually. They include government task forces,[57] expert groups,[58] and public-private initiatives around issue-specific misinformation such as election interference and information related to the COVID-19 pandemic.[59] There are several proposals to establish more formal structures to address misinformation and platform harms. For instance, the Global Network Initiative advocates for a multi-stakeholder and collaborative approach which would involve "business, industry associations, civil society organizations, investors and academics".[60] The French Interim Mission Report recommends the creation of an independent authority created by elected representatives, with government and platform actors, civil society, citizens, domain experts, academics and researchers.[61] According to the Report, this model is similar to "banking supervisory authorities" which incentivize financial institutions to comply with prescribed disclosures and information sharing.[62] Nathaniel Persily and Alex Stamos also recommend the creation of an independent body akin to the Financial and Industrial Regulatory Authority in the USA.[63] They propose a self-regulatory body "blessed by the Congress"[64] where platforms can work together to address common challenges such as "foreign sponsored advertising" or voter suppression.[65] The U.K. White Paper on Online Harms recommends the creation of an "independent regulatory body" which will also be responsible for creating codes for social media platforms

---

[57] Rachel Aiello, *Feds unveil plan to tackle fake news, interference in 2019 election,* CTV NEWS (Feb. 27, 2019), https://www.ctvnews.ca/politics/feds-unveil-plan-to-tackle-fake-news-interference-in-2019-election-1.4274273.
[58] *Minister De Croo bindt strijd aan met fake new*s (2018)
https://www.standaard.be/cnt/dmf20180502_03492898, PTI, *India, with 12 nations, leads initiative at UN to counter misinformation on COVID-*19, THE HINDU (June 14, 2020),
https://www.thehindu.com/news/national/india-with-12-nations-leads-initiative-at-un-to-counter-misinformation-on-covid-19/article31825371.ece.
[59] Such as the "EU v. Disinformation" initiative which brings together online platforms, civil society, fact-checking organisations and state actors, https://euvsdisinfo.eu/ (last visited Jan. 20, 2021).
[60] Global Network Initiative, *GNI Principles on Freedom of Expression and Privacy*, https://globalnetworkinitiative.org/wp-content/uploads/2018/04/GNI-Principles-on-Freedom-of-Expression-and-Privacy.pdf.
[61] FRENCH INTERIM MISSION REPORT, *supra* note 7, at 23-24.
[62] "This intervention approach is designed to create targeted incentives for platforms to participate in achieving a public interest objective without having a direct normative action on the service offered." *Id.* at 16.
[63] The Financial and Industry Regulatory Authority is "a not-for-profit organization that – working under the supervision of the SEC – actively engages with and provides essential tools for investors, member firms and policymakers.", https://www.finra.org (last visited Sept. 20, 2020).
[64] Nathaniel Persily and Alex Stamos, *Regulating Online political Advertising by Foreign Governments and Nationals, in* SECURING AMERICAN ELECTIONS 34 (Michael McFaul ed., 2019).
[65] *Id.*

---

to address hate speech and harms.[66] These proposals have largely advocated for the establishment of independent bodies to govern several aspects of social media platform functioning. In the specific context of misinformation, the Australian Competition and Consumer Commission advocates for an independent regulator to monitor voluntary measures taken by platforms to ensure authentic and trustworthy content.[67]

At present, India lacks a body which is responsible for developing tools to curb online misinformation. Further, the present regulatory architecture is not equipped with the necessary tools or resources to oversee standard-setting and compliance of some of the measures discussed in this paper. While MEITy is the primary body responsible for developing law and policy associated with technology and the Internet, the Ministry has so far been unable to effectively work with platforms to address the misinformation challenge. In general, MEITy's rule-making around social media platforms has been opaque, and lacks insight from experts, academic stakeholders, and civil society. Further, in the context of platform governance, it is unclear if MEITy will have sufficient regulatory capacity to govern platforms and effectively monitor efforts to combat misinformation, given the size of this industry, and the importance of the platforms to facilitating freedom of expression among their hundreds of millions of users. The MEITy's role in the blocking of social media accounts linked to the 2020–2021 Indian farmers' protest is further cause for concern in the context of safeguarding the expressive rights of users.[68]

In order to model the potential structure and functions of a platform oversight body for India, it may be useful to briefly examine existing models, such as the Press Council of India ('PCI'),[69] the News Broadcasters Association,[70] the Advertising Standards Council of India[71] and existing or emerging statutory regulators in the field of telecommunications and data protection. None of these agencies present a perfect template of a body which is independent from both state control and

---

[66] DEPARTMENT FOR DIGITAL, CULTURE, MEDIA AND SPORT & HOME OFFICE, ONLINE HARMS WHITE PAPER, 2019, Cm. 57, https://www.gov.uk/government/consultations/online-harms-white-paper/online-harms-white-paper (UK).

[67] Recommendation 14, Australian Competition and Consumer Commission (ACCC) Digital Platforms Inquiry (DPI), https://www.accc.gov.au/system/files/Digital%20platforms%20inquiry%20-%20final%20report.pdf (last visited Nov. 15, 2020).

[68] Danya Hajjaji, *Twitter Rejects India's Order to Block Accounts Amid Farmer Protests*, NEWSWEEK (Feb 10, 2021), https://www.newsweek.com/twitter-rejects-india-order-block-accounts-farmer-protests-1568278.

[69] The Press Council of India (statutory body), https://www.presscouncil.nic.in/.

[70] News Broadcasters Association (private self-regulatory body), http://www.nbanewdelhi.com/about-nba.

[71] Advertising Standards Council of India (voluntary self-regulatory body), https://www.ascionline.org/.

industry bias. However, a proposed platform oversight body for India might consider adopting certain regulatory strengths of existing bodies, namely those where standard setting is efficient and public/stakeholder consultation processes are collaborative and transparent.[72] Generally, these bodies function best where representation from industry and other stakeholder groups is mandatory,[73] and where their functions are subject to parliamentary or external/independent oversight,[74] with a representative council empowered to draw up "codes of practice" for industries.[75] The PCI, and proposals regarding the creation of a Data Protection Authority for India,[76] also empower sectoral regulators to act as primary standard setting bodies or initiate co-regulatory codes. It should be noted that while certain structural features of the PCI may serve as a useful examples to consider, there has been significant criticism of the body and its functioning.[77] For instance, the PCI does not have sufficient power to enforce its guidelines,[78] and its members have been accused of conflicts of interest.[79]

In order to support proactive efforts to curb online misinformation across different platforms, India needs an independent and representative platform oversight body to structure self-regulatory initiatives around misinformation. It is recommended that the platform oversight body consist of government and platform

---

[72] When compared to existing law-making processes in the realm of technology and digital rights, the Telecom Regulatory Authority of India has adopted an efficient framework for public consultations on rule-making which is transparent and stakeholder friendly, Anirudh Burman & Bhargavi Zaveri, *Measuring Regulatory Responsiveness in India: A Framework for Empirical Assessment*, CARNEGIE INDIA (Apr. 2, 2019), https://carnegieindia.org/2019/04/02/measuring-regulatory-responsiveness-in-india-framework-for-empirical-assessment-pub-78871.

[73] For instance, The Press Council of India is a statutory standard-setting body for print newspapers and news agencies. Members of Parliament, eminent academics and experts, journalists, newspaper editors, managers of news agencies, and individuals who manage the newspaper business are adequately represented in the Council; See Press Council Act, No. 37 of 1978, INDIA CODE, Preamble (India), Press Council Act, No. 37 of 1978, INDIA CODE, § 5 (India).

[74] Press Council Act, No. 37 of 1978, INDIA CODE, § 20 (India).

[75] The Press Council of India is empowered to develop codes of practice for newspapers, define journalistic standards, and help "newspapers and news agencies to maintain their independence", Press Council Act, No. 37 of 1978, INDIA CODE, § 13(2) (India). Evolving discussions on the functions of the Data Protection Authority (under the Personal Data Protection Bill, 2019) also provide for codes of practice to be jointly developed by industry actors and the government.

[76] Personal Data Protection Bill, 2019, § 41 (India).

[77] Nandita Singh, *Media Experts Say Press Regulation Doesn't Work in India, Press Council is a Joke*, THE PRINT (Apr. 9, 2018), https://theprint.in/india/governance/media-experts-say-press-regulation-doesnt-work-in-india-pci-is-a-joke/48213/.

[78] *Regulation of media in India – A brief overview*, PRS INDIA (Nov. 16, 2011), https://www.prsindia.org/tags/press-council-india.

[79] *Id.*

---

representatives, policymakers, civil society, academics, and other subject-matter experts to ensure efficient and accountable functioning. Such a body may be created after adequate consultation and feasibility assessments, in order to develop the necessary buy-in from social media platforms. The scope of the body's powers would need to be carefully considered to ensure that the platform oversight body is not placed in the role of defining the legal limits of acceptable speech. The functioning of the body should also be subject to appropriate judicial oversight.[80]

The challenges in setting such a body up suggest that a gradual approach to implementation may be best. This may include progressive development of certain responsibilities. Standards for effective non-financial disclosures,[81] the creation of a standardized metric for performance reporting on these transparency obligations, and the creation of resources for community awareness of misinformation would be a good place to start, with particular resources dedicated to supporting mid-level or smaller platforms which are struggling to comply with industry best practice due to resource constraints.[82] If consensus is achieved on these relatively simple issues, it could provide a baseline for further and more ambitious collaboration.

IV. Promote responsive platform responses to misinformation in India

Every country's misinformation challenge is unique, and requires platforms to work collaboratively and tailor responses through meaningful engagement with the communities that they serve. The following section explores ways in which responsive platform behavior can be achieved in India.

A. Meaningful engagement with communities and stakeholders in India

The need for a "bottom-up approach"

Speech policies of platforms have traditionally been modelled on American understandings of freedom of speech, implemented in a manner which attempts to

---

[80] GLOBAL NETWORK INITIATIVE, POLICY BRIEF ON CONTENT REGULATION AND HUMAN RIGHTS 30 (2020), https://globalnetworkinitiative.org/content-regulation-policy-brief/.

[81] Mark MacCarthy, *Transparency Requirements for Digital Social Media Platforms: Recommendations for Policy Makers and Industry*, *in* TRANSATLANTIC WORKING GROUP 11 (2020) https://www.ivir.nl/publicaties/download/Transparency_MacCarthy_Feb_2020.pdf.

[82] Alex Stamos et. al, *Combatting State-Sponsored Disinformation Campaigns from State-Aligned Actors*, *in* SECURING AMERICAN ELECTIONS 48 (Michael McFaul ed., 2019).

harmonize enforcement around the world according to this standard.[83] While there have been some efforts to move away from a core focus on American speech standards and towards more international approaches, the Community Standards enforcement process is still a "top-down" model of governance.[84] Platform users have no influence over the "definition and enforcement of rules" which are placed in the hands of platform "officials".[85]

User policies across Big-Tech platforms do not adequately appreciate the religious, linguistic and cultural diversity of the Indian user base or, for that matter, of their users in many other parts of the world. The complexity of language in their user agreements, including their Community Standards, in many instances does not correspond to the literacy level of users, including those who are using Internet services for the first time. The need to make platform policies available in local or regional languages of platform users has been emphasized by the Forum on Information and Democracy in its Report on Infodemics.[86] Further, platforms and content moderators are often not equipped to appreciate critical contexts of online speech which are unique to a geographical area or community. In her work, Chinmayi Arun has emphasized the need to appreciate the distinction between "global" and "local" elements of hate speech, and highlighted the failure of platforms to address "hyper-local harmful speech".[87] Arun cautions that "hyper-local" elements of speech which are specific to a village or local area may be completely ignored by global corporations, which underscores the need for platform initiatives which are tailored to local realities.[88]

While platforms like Facebook have been working with certain experts and fact-checking organizations to identify false news, the level of community involvement in developing speech rules and supporting a contextual interpretation of these rules is not consistent. In order to ensure responsive speech rules which are better

---

[83] Kate Klonick, "The New Governors: The People, Rules, and Processes Governing Online Speech", 131 HARV. L.REV. 1598, 1625–29 (2018), https://harvardlawreview.org/wp-content/uploads/2018/04/1598-1670_Online.pdf at p. 1621.

[84] The Justice Collaboratory, *Report of the Facebook Data Transparency Advisory Group* (April 2019), https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf.

[85] *Id.* p. 31.

[86] Forum on Information & Democracy, *Working Group on Infodemics – Policy Framework*, p. 23, https://informationdemocracy.org/wp-content/uploads/2020/11/ForumID_Report-on-infodemics_101120.pdf; Mark MacCarthy, *A dispute resolution program for social media companies*, BROOKINGS (Oct. 2020), https://www.brookings.edu/research/a-dispute-resolution-program-for-social-media-companies.

[87] Chinmayi Arun, "Rebalancing Regulation of Speech: Hyper-Local Content on Global Web-Based Platforms," (Jan. 23, 2018), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3108238.

[88] *Id.*

tailored to these local contexts, some have advocated for a "bottom-up" approach to developing platform policies,[89] and the creation of "nested communities" to involve large and diverse groups in the rule-making process.[90] This would mean that governments, platforms and other stakeholders would work together to formulate responsive and accessible platform rules— akin to an "elected Parliament of users"[91] empowered to define community standards.[92] Echoing the need for greater engagement, a White Paper on online content regulation released by Facebook argues for the introduction of "procedural accountability regulations" which would require platforms to empower users by giving them the tools to challenge speech decisions.[93] John Bowers and Jonathan Zittrain have also emphasized the need for platforms to meaningfully "externalize" important policymaking obligations.[94] Platforms have started relying on independent third parties for certain aspects of their functioning. However, presently, these efforts are focused on outsourcing of content moderation appeals to independent entities, particularly in Facebook's case through the creation of their Oversight Board.[95] Other models have been suggested, particularly ARTICLE 19's proposal to establish "Social Media Councils" as a tool for "multi-stakeholder accountability" to address content moderation challenges with respect to international human rights standards.[96] However, models like the Oversight Board are largely focused on decision-making around online speech and not at the stage of actually setting the relevant standards.

The highly local and contextual nature of misinformation creates a stronger impetus to work with users and communities and develop responsive policies.

---

[89] The Justice Collaboratory, *supra note 84,* p. 30.

[90] *Id.* at 33, Based on Elinor Ostrom's principle of "nested enterprises" for governing "information commons".

[91] Laurence Dodds, *Facebook audit panel suggests creating an elected 'parliament'*, THE TELEGRAPH (May 23, 2019), https://www.telegraph.co.uk/technology/2019/05/23/facebook-audit-panel-suggests-creating-elected-parliament-users/.

[92] The Justice Collaboratory, *supra note 84,* p. 43.

[93] Monika Bickert, *Charting a Way Forward – Online Content Regulation*, FACEBOOK (Feb. 2020) at p. 10, https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf.

[94] John Bowers and Jonathan Zittrain, *Answering impossible questions: content governance in an age of misinformation*, MISINFORMATION REVIEW (Jan. 14, 2020), https://misinforeview.hks.harvard.edu/article/content-governance-in-an-age-of-disinformation/.

[95] The Oversight Board, https://www.oversightboard.com (last visited Nov. 21, 2020); Chinmayi Arun, *The Facebook Oversight Board: An Experiment in Self-Regulation*, JUST SECURITY (May 6, 2020) https://www.justsecurity.org/70021/the-facebook-oversight-board-an-experiment-in-self-regulation/.

[96] Article 19, *Regulating social media: we need a new model that protects free expression* (Apr. 25, 2018) https://www.article19.org/resources/regulating-social-media-need-new-model-protects-free-expression/, Article 19, *The Social Media Councils: Consultation Paper* (June, 2019), https://www.article19.org/wp-content/uploads/2019/06/A19-SMC-Consultation-paper-2019-v05.pdf.

Platforms can initiate "bottom-up" approaches to developing community standards by involving citizens, experts, and civil society. For issue-specific misinformation challenges, platforms might also create cross-platform initiatives in collaboration with civil society and industry bodies to oversee the development of community standards and their enforcement. This may be done in two ways.

First, specific voluntary groups may be created for high-risk or sensitive misinformation challenges for India. For instance, online misinformation pertaining to State and local elections poses a great threat to voter interests and election integrity in India. While platforms have taken measures to combat election misinformation,[97] when compared to platform efficiency in tackling election misinformation in the US, a comprehensive effort to tackle this in India remains largely unaddressed.[98] In the context of election information, specialized voluntary groups consisting of civil society members, election officials, fact-checking entities, human rights experts, and platform representatives can work together to address specific challenges for different types of elections. The ACMA classifies risks from misinformation into two categories— short term or imminent risks, and long term or systemic risks.[99] Imminent or 'acute' risks would include instances such as a public health outbreak, financial fraud, or a threat to election integrity.[100] Voluntary groups across these categories of misinformation can ensure effective and targeted policy and platform responses. Further, delegating platform policy-making to external entities is likely to support independence of thought and judgment, rather than having policies driven by the platforms' own direct interests. A structured approach will ensure that entities such as civil society members, experts, and academics providing independent perspectives on platform decisions are made a part of the standard setting process from the beginning.

A second, generalized approach would be to form regional or 'hyperlocal platform communities' which consist of a diverse panel of platform users and residents, policy-makers at the local levels, journalists, researchers and academics, and other stakeholders. This concept is akin to recommendations for the creation of "nested

---

[97] Ajit Mohan, *Preparing for Upcoming Indian Elections*, Facebook Newsroom (April 8, 2019) https://about.fb.com/news/2019/04/preparing-for-indian-elections/.

[98] In the context of electoral integrity, Nikhil Pahwa highlights the need for measures to address "the gap between the responsibility and accountability" of significant platforms and the urgency for democracies to address the lack of platform accountability and transparency, Nikhil Pahwa, *Can Regulation Douse Populism's Online Fires?*, FOREIGN POLICY (Jan. 8, 2021), https://foreignpolicy.com/2021/01/08/capitol-mob-social-media-right-wing-violence-regulation/, See also Rasmus Kleis Nielsen, *supra note 8*.

[99] Australian Communications and Media Authority, *supra note 41*, p.11.

[100] *Id.*

communities" which can work with platforms to address local speech challenges.[101] These communities could work with platforms to audit community standards, co-develop guidelines on fake news based on local contexts, and provide inputs on speech rules. An elected and inter-platform 'hyperlocal platform community' might also interface with local state actors in fact-finding, investigations, and developing policies around platform governance. Procedures pertaining to their constitution, election, independence, funding, and jurisdiction might be co-developed by platforms, state actors and civil society.

It is critical that such bodies are representative and unbiased, and measures to ensure their independence should be a primary focus for researchers. An open and transparent selection procedure undertaken by inter-platform bodies like the platform oversight body may be useful in ensuring this. Buy-in from both governments and from the platforms themselves is essential to their function, as the challenges faced by the Delhi Assembly Committee in investigating Facebook's role in February 2020 communal riots illustrates. The Committee has been largely unsuccessful in investigating the role of Facebook due to inadequate cooperation by the platform and concerns about its lack of jurisdiction in questioning Facebook.[102] A 'hyperlocal platform community' could have played a crucial role in assisting the Committee in reviewing and auditing Facebook's response to specific instances of fake news locally, structuring accessible surveys of residents to understand what could be done better, and providing inputs on building effective and contextualized strategies to combat misinformation and hate speech.

These recommendations require substantial additional development to operationalize, and should be the subject of analysis through all stages of their evolution. As a result, although there is no reason why work to develop 'hyperlocal platform communities' need be delayed, and requirements for platforms to engage with communities should not be hardcoded in legislation until more is known about their utility and efficacy.

---

[101] The Justice Collaboratory, *supra note 84*, p. 33.
[102] Staff Reporter, *Attempt to hide crucial facts on Facebook's role in Delhi riots, says Assembly Committee*, THE HINDU (Sept. 15, 2020), https://www.thehindu.com/news/cities/Delhi/attempt-to-hide-crucial-facts-on-facebooks-role-in-delhi-riots-says-assembly-committee/article32608982.ece.

V. Conclusion

Regulation of misinformation is a complex endeavor, and countries across the world are struggling to adapt laws to address evolving platform harms. In the absence of a robust regulatory framework or timely conversations around platform governance, India not only lacks the bargaining power to negotiate with foreign platforms, but also the basic tools necessary to hold them accountable to Indian law. This paper has so far outlined critical areas of reform which can be prioritized in the regulatory conversation. The objective of responsive and participatory rule-making, for state actors and platforms, forms the core of all the proposals discussed. The proposals for regulatory reform through a dedicated platform oversight body and tools for community engagement are significant objectives which will require state-platform consensus and the will to alter the regulatory landscape. It is hoped that the recommendations outlined in this paper will not only be explored for transformational action for India, but assist similarly situated countries in combating misinformation online.

Addressing Information Pollution With a "Superfund for the Internet"

Lisa H. Macpherson*

I. Introduction

Current policy approaches to combat the spread of misinformation in the United States are largely focused on the content moderation practices of digital platforms. But our contemporary news and information ecosystem is complex and interconnected, with quality information and misinformation flowing across *both* legacy and digital media. Often, misinformation originates on fringe websites, gets amplified by partisan media, and then spun back out online.[1] Platforms are not the source of the problem, but thanks to their advertising-driven business model, which rewards emotional engagement and time-on-device, they benefit from the systems that create and amplify misinformation.[2] Policy strategies exclusively focused on controlling speech on platforms will also inevitably encounter First Amendment hurdles while doing little to actively promote a healthy and robust civic discourse.

This paper offers a creative, evidence-based policy proposal that addresses the twin problems in our information ecosystem: the virulent spread of misinformation on digital platforms and the crisis in local journalism. It considers the dominance of the major digital platforms in political and social discourse and mandates that they adopt an approach in their content moderation that serves the public interest. Since toxic content on digital platforms is sometimes compared to toxic chemicals dumped by industrial companies into fresh water, our proposal is modeled on the Environmental Protection Agency's 1980 Superfund to clean up toxic waste sites: a "superfund for the internet." It would create demand for news analysis and fact-checking services among the major information distribution platforms, and reward supply of such services among reputable news organizations, creating a new

---

[1] YOCHAI BENKLER, ET AL., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*, (BERKMAN KLEIN CTR. FOR INTERNET & SOC'Y, 2020), https://cyber.harvard.edu/publications/2017/08/mediacloud.
[2] NATHALIE MARÉCHAL ET AL., *Getting to the Source of Infodemics: It's the Business Model*, OPEN TECH. INST. AND NEW AMERICA (May 27, 2020), https://www.newamerica.org/oti/reports/getting-to-the-source-of-infodemics-its-the-business-model/.

---

revenue stream for the local news organizations that support healthy civic information and discourse.

This paper calls for the platforms, accountable to an independent expert body established through legal mandate, to master the process of identifying, minimizing, and helping the public navigate misinformation—without interfering with constitutionally-protected speech rights. In doing so, it is possible to provide an essential new revenue stream to local journalism organizations that also help protect our public and democratic institutions.

II. The Essential Role of Local Journalism in Mitigating Misinformation

Traditionally, local journalism was the primary source of information that is helpful and truthful for communities. A Poynter Media Trust Survey in 2018 found 76 percent of Americans across the political spectrum have "a great deal" or "a fair amount" of trust in their local television news (compared to 55 percent trust in national network news), and 73 percent have confidence in local newspapers (compared to 59 percent in national newspapers).[3] A Gallup survey in 2019 found 74 percent of Americans trust the accuracy of the news and information they get from local news stations (compared to 54 percent for nightly network news programs), and 67 percent trust their local newspapers (compared to 49 percent for national newspapers).[4] A 2019 study from the Knight Foundation's Trust, Media, and Democracy Initiative with Gallup found that 45% of Americans trust reporting by local news organizations "a great deal" or "quite a lot," while 15% have "very little" or no trust at all. But in the same survey the public's view of *national* news organizations was more negative than positive: only 31% expressed "a great deal" or "quite a lot" of trust, and 38% "very little" or no trust in national news.[5] In aggregate, the data suggests that the viability and availability of local news are important components of a trustworthy information ecosystem, the vital link between informed citizens and a healthy democracy.[6]

---

[3] ANDREW GUESS, ET AL., ALL MEDIA TRUST IS LOCAL? FINDINGS FROM THE 2018 POYNTER MEDIA TRUST SURVEY (2018).

[4] Megan Brenan, *In U.S., 40% Trust Internet News Accuracy, Up 15 Points*, GALLUP (Aug. 22, 2019), https://news.gallup.com/poll/260492/trust-internet-news-accuracy-points.aspx.

[5] THE JOHN S. & JAMES L. KNIGHT FOUND., STATE OF PUBLIC TRUST IN LOCAL NEWS 12 (2019), https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/440/original/State_of_Public_Trust_in_Local _Media_final_.pdf.

[6] STEVEN WALDMAN, FEDERAL COMMUNICATIONS COMMISSION, THE INFORMATION NEEDS OF COMMUNITIES: THE CHANGING MEDIA LANDSCAPE IN A BROADBAND AGE (July 2011), https://www.fcc.gov/sites/default/files/the-information-needs-of-communities-report-july-2011.pdf.

Perhaps because of how much they are able to rely on their reporting, 71 percent of Americans think their local news organizations are "doing either somewhat or very well financially."[7] However, due to years of losses in circulation and advertising revenue to digital platforms (compounded by their own failure to address the changes in readership brought about by digital technology, plus consolidations and cost-cutting brought about by financially-motivated owners), local journalism now faces what has been described as an existential threat.[8] Over the past 15 years, the United States has lost over 2,100 or one-fourth of its local newspapers.[9] About 1,800 of the communities that have lost a paper since 2004 now do not have easy access to *any* local news source – such as a local online news site or a local radio station.[10] The policy solution presented here serves to create a potential new funding source for local journalism.

III. The Regulatory Context for Content Moderation

The "superfund for the internet" sits within a broader legal and regulatory framework for content moderation. There is precedent in the United States for industries to be mandated to take particular actions in the public interest, either to guarantee the availability of certain goods and services, or to address externalities that are not taken into consideration when unregulated firms make their decisions. For example, since the passage of the Communications Act of 1934, it has been recognized that because of the essential role of radio and television licensees in public discourse, the Federal Communications Commission has a duty to use its licensing authority to promote "the public interest, convenience and necessity."[11] This requires that each station licensee identify the needs and problems of its community of license, and air programming (news, public affairs, etc.) that is responsive to them. This is partially due to the position of telecommunications as an industry essential to democratic self-governance.

---

[7] Pew Rsch. Ctr., *For Local News, Americans Embrace Digital but Still Want Strong Community Connection* (Mar. 26, 2019), https://www.journalism.org/2019/03/26/for-local-news-americans-embrace-digital-but-still-want-strong-community-connection/.

[8] PEN AMERICA, LOSING THE NEWS: THE DECIMATION OF LOCAL JOURNALISM AND THE SEARCH FOR SOLUTIONS 24 (2019), https://pen.org/wp-content/uploads/2019/12/Losing-the-News-The-Decimation-of-Local-Journalism-and-the-Search-for-Solutions-Report.pdf.

[9] PENELOPE MUSE ABERNATHY, THE CENTER FOR INNOVATION AND SUSTAINABILITY IN LOCAL MEDIA, HUSSMAN SCHOOL OF JOURNALISM AND MEDIA, UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, NEWS DESERTS AND GHOST NEWSPAPERS: WILL LOCAL NEWS SURVIVE? 11 (2020) https://www.usnewsdeserts.com/wp-content/uploads/2020/06/2020_News_Deserts_and_Ghost_Newspapers.pdf.

[10] *Id.*

[11] Telecommunications Act of 1934, Pub. L. No. 74-415, 48 Stat. 1064 (1934) (codified as amended at 47 U.S.C. §§ 151-646).

The FCC has very limited jurisdiction over digital platforms, and this paper does not recommend its authority be broadened, or that the FCC should oversee the proposal. But the dominance of the platforms in our political and social discourse qualifies them for this same public interest mandate. In other words, *the dominant information distribution platforms should be mandated to adopt an approach in their content moderation that serves the public interest.* Critically, this must be accomplished in a way that passes First Amendment scrutiny, especially if any government body is involved in the creation or administration of their content moderation approach.

Adoption of a public interest standard for content moderation, whether voluntary or mandated, may represent a significant shift from the platforms' historical emphasis on free expression. But a new framework is *already* required: As the platforms have become more active in content moderation and exercise more control over what content their users see, platforms are beginning to migrate from frameworks which are primarily concerned with facilitating the unfettered exchange of ideas to more carefully managing the content they host. This paper proposes the goal be "to create an online information environment that serves the public interest" and that the companies "welcome a political and public debate as to how the public interest is defined."[12]

Additionally, the optimal content moderation scheme transparently defines community standards that govern content moderation; focuses on conduct relative to the community standards, not content *per se,* as the basis of decision-making; describes methods of enforcing the standards; and includes an accessible process for appeal. It should also address the highest-priority sources and harms of misinformation, and the fact that some groups are more likely to be the target of, and suffer harm from, misinformation.

IV. A Proposal for an Internet Superfund

Toxic content on digital platforms has been compared to toxic chemicals dumped by industrial companies into fresh water, and "paranoia and disinformation" dumped in the body politic has been described as "the toxic byproduct of

---

[12] Joel Simon, *For Big Tech, What Follows 'Free Expression'?*, COLUM. JOURNALISM REV. (Nov. 5, 2020), https://www.cjr.org/opinion/big-tech-senate-hearing-free-expression.php.

[Facebook's] relentless drive for profit."[13] This proposal is modeled on the Environmental Protection Agency's 1980 Superfund to clean up toxic waste sites: a "superfund for the internet" (hereinafter "Internet Superfund"). The proposal is to use public policy tools to create a market mechanism to clean up information pollution. This market mechanism would consist of demand and payment for news analysis services from the major information distribution platforms, and creation and compensation for supply of these services among qualified news organizations. The payment would be collected from dominant platforms and administered to qualified news organizations by an independent "trust fund" established and administered by the government. *The independent government body administering the trust fund would have no role in the actual identification, review, analysis, or action on content.*

The Internet Superfund reflects our interconnected information ecosystem as well as a growing body of research on what actually works, in practice, to counter the spread of misinformation. One important concept from the research is that countering misinformation—and more importantly, countering its devastating effects on individual well-being, societal cohesion, and trust in institutions—is as much about producing and elevating accurate and authoritative information as it is about detecting, countering, or removing misinformation.[14] This "information curation" (which encompasses actively finding what is helpful, contextual, and truthful as well as controlling and restraining what is false and harmful) must be accomplished without violating the United States' constitutional standards.[15] In the best case, it fosters and promotes speech, including from traditionally marginalized groups and voices.

In addition to research on strategies for countering misinformation, this proposal is supported by information gathered from extensive tracking and reporting on the

---

[13] Roger McNamee, *Facebook Cannot Fix Itself. But Trump's Effort to Reform Section 230 is Wrong*, TIME (June 4, 2020), https://time.com/5847963/trump-section-230-executive-order/; Jamelle Bouie, *Facebook Has Been a Disaster for the World*, N.Y. TIMES (Sept. 18, 2020), https://www.nytimes.com/2020/09/18/opinion/facebook-democracy.html.

[14] STEPHAN LEWANDOWSKI & JOHN COOK, THE CONSPIRACY THEORY HANDBOOK (2020), https://www.climatechangecommunication.org/wp-content/uploads/2020/03/ConspiracyTheoryHandbook.pdf; Kathleen Hall Jamieson & Dolores Albarracin, *The Relation Between Media Consumption and Misinformation at the Outset of the SARS-CoV-2 Pandemic in the US*, HARV. KENNEDY SCHOOL (HKS) MISINFORMATION REV. (2020), https://misinforeview.hks.harvard.edu/article/the-relation-between-media-consumption-and-misinformation-at-the-outset-of-the-sars-cov-2-pandemic-in-the-us/.

[15] Joan Donovan, *You Purged Racists From Your Website? Great, Now Get to Work*, WIRED (July 1, 2020), https://www.wired.com/story/you-purged-racists-from-your-website-great-now-get-to-work/.

platforms' efforts to address misinformation related to the COVID-19 pandemic.[16] Even the platforms that had been most resistant to moderation of political content took a harder line on misinformation about the pandemic, on the belief that health information would be objective and politically neutral. But that's not how it played out: Information about the novel coronavirus pandemic became every bit as politicized as what would normally be considered highly partisan topics, subject to the same patterns of creation and distribution as other content designed to sow division and undermine democratic institutions. That makes the pandemic an appropriate model for how platforms can manage other types of misinformation, including overtly political misinformation. Recent research also indicates that citizens concerned about contracting COVID-19 are more likely to support fact-checking, including of politicians.[17]

Research from Public Knowledge showed that — in a situation in which the potential for harm is high and a finite number of sources for authoritative information are in place — platforms can and will set new standards and develop new solutions to problems previously positioned as insoluble. Many of these standards and solutions were replicated or expanded to mitigate misinformation about the 2020 U.S. presidential election. One common strategy enabled their most effective approaches: partnering with authoritative sources of information, news analysis and fact-checking. These partnerships allowed the platforms to evaluate sources, verify the accuracy of claims, up- and down-rank content, label posts, direct people who have experienced misinformation to debunking sites, and understand what kinds of misinformation may create the greatest harm. They also began to demonstrate that the power of "fact-checking" lies not in the fact checks themselves, but in how they are used to change the user experience in ways that deter the spread of misinformation.

But COVID-19 also highlighted the limitations of platforms' efforts. At best, they can be considered as a model for a new baseline duty of care.[18] Most notably, the

---

[16] LISA MACPHERSON & KATHLEEN BURKE, PUBLIC KNOWLEDGE, HOW ARE PLATFORMS RESPONDING TO THIS PANDEMIC? WHAT PLATFORMS ARE DOING TO TACKLE RAMPANT MISINFORMATION (2020), https://misinfotrackingreport.com/wp-content/uploads/2020/06/FinalMisinfoReport.pdf.

[17] Timothy S. Rich, et. al, *Research Note: Does the Public Support Fact-Checking on Social Media? It Depends on Who and How You Ask*, HARV. KENNEDY SCHOOL (HKS) MISINFORMATION REV. (2020), https://misinforeview.hks.harvard.edu/article/research-note-does-the-public-support-fact-checking-social-media-it-depends-who-and-how-you-ask/.

[18] In a privacy context, "duty of care" has been proposed to refer to the prohibition of covered entities from causing foreseeable injuries to individuals. *See* CAMERON F. KERRY ET AL., BROOKINGS, BRIDGING THE GAPS: A PATH FORWARD TO FEDERAL PRIVACY LEGISLATION (2020), https://www.brookings.edu/wp-

---

platforms' efforts to counter misinformation about COVID-19 and the presidential election were highly discretionary. That is, it shows that the platforms only apply these disinformation mitigation practices to topics for which they feel the political consequences of doing nothing are greater than the financial losses that would result from doing something. Overall, the existing incentives of digital platforms (including the desire to avoid bad publicity and the desire of advertisers to avoid association with harmful content) are insufficient to address the individual, social and political harms associated with misinformation. Aligning platforms' incentives with those of the public interest *requires* policy mechanisms that lower the cost of good behavior and/or raise the cost of bad behavior while not mandating censorship of permissible speech.[19]

The proposal for an Internet Superfund calls for the platforms themselves, accountable to an independent body established through legal mandate, to master the process of identifying, minimizing, and helping the public navigate misinformation in general—without interfering with constitutionally-protected speech rights. The proposal is able to accomplish this by ensuring that the independent body has no role in the selection of content for review, in the news analysis itself, or in the action taken by platforms on content. In doing so, the proposal provides an essential new revenue stream to local journalism organizations, in the form of a new product or service offering that is consistent with their existing journalism and fact-checking skill set. This, in turn, helps promote a healthy and robust civic discourse, which protects trust in, and effectiveness of, our public and democratic institutions.

V. An Evidence-Based Approach to Content Moderation

Fact-checking (in the context of information pollution) is the process of evaluating the truthfulness and accuracy of published information by comparing an explicit claim against trusted sources of facts.[20] The most prominent fact-checking

---

content/uploads/2020/06/Bridging-the-gaps_a-path-forward-to-federal-privacy-legislation.pdf. *See also* Data Care Act of 2019, S. 2961, 116th Cong. (2019) (proposed Senate bill establishing "duties for online service providers with respect to end user data that such providers collect and use"). Here, the idea of a "duty of care" refers to the obligation of covered digital platforms to apply a standard of reasonable care in regard to content that could foreseeably harm others. The goal is similar to that of a fiduciary model; that is, to change how digital platforms think about their end users and their obligations to their end users.

[19] HAROLD FELD, THE CASE FOR THE DIGITAL PLATFORM ACT: MARKET STRUCTURE AND REGULATION OF DIGITAL PLATFORMS 135 (2019).

[20] Alexios Mantzarlis, *Will Verification Kill Fact-Checking?*, POYNTER (Oct. 21, 2015), https://www.poynter.org/fact-checking/2015/will-verification-kill-fact-checking/.

---

organization is the non-partisan International Fact-Checking Network (IFCN), which certifies organizations that successfully apply to be signatories to a Code of Principles. The principles are a series of commitments organizations abide by to promote excellence in fact-checking. They comprise what is essentially a good journalistic process, encompassing principles related to fairness, sourcing, transparency, methodology, and corrections.[21] Most importantly, fact-checking is the foundational enabler to the platforms' ability to apply labels, make changes in user experience to reduce sharing, demonetize, and when appropriate, "cleanse" or remove false or harmful content or accounts. It is an appropriate role for policymakers and regulators to encourage the development of such services, provide opportunities for platforms and service providers to share information necessary to develop these services, and ensure a competitive market in their development.[22]

Research has shown that flagging of false news may have an effect on reducing sharing of deceptive information on social media.[23] A recent study describing "the state of the art" in measuring the impact of fact-checking demonstrated that fact-checking has a positive impact in reducing misinformation about specific claims related to COVID-19, and that there are ways to extend and optimize fact-checking's impact.[24] A "Debunking Handbook" written by 22 prominent scholars of misinformation summarized what they believe "represents the current consensus on the science of debunking." It described the essential role of fact-checking in debunking and unsticking misinformation.[25] Importantly, some pitfalls that came out in earlier research, like the so-called "backfire effect" in which corrections actually strengthened misperceptions, or the "repetition effect" in which false information is reinforced when it is repeated within the correction, have been shown not to exist as robust, measurable empirical phenomena.[26]

---

[21] Poynter, *IFCN Code of Principles*, https://www.ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles (last visited Sept. 21, 2020).
[22] FELD, *supra* note 19, at 168.
[23] Paul Mena, *Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook*, 12 POLICY & INTERNET 165 (2019), https://onlinelibrary.wiley.com/doi/abs/10.1002/poi3.214.
[24] Tracie Farrell et al., Assessment of the online spread of coronavirus information, The HERoS Project, Assessment of the Online Spread of Coronavirus Information (September, 2020), https://www.heros-project.eu/wp-content/uploads/Assessment-of-the-online-spread-of-coronavirus-misinformation.pdf (last visited Dec 18, 2020).
[25] STEPHEN LEWANDOWSKY, ET AL., DEBUNKING HANDBOOK 2020, https://sks.to/db2020.
[26] Briony Swire-Thompson, et al., *Searching for the Backfire Effect: Measurement and Design Considerations*, 9 J. APPLIED RSCH. MEMORY & COGNITION 286 (2020), https://doi.org/10.1016/j.jarmac.2020.06.006.

The dominant information distribution platforms, including Facebook, Instagram, Google and YouTube, already use fact-checking services.[27] Twitter uses internal systems to monitor content and relies on "trusted partners" to identify content that may cause offline harm.[28] But today, the user experience for fact checks varies widely by platform. Researchers say it is impossible to know how successful or comprehensive the companies have been in removing bogus content because the platforms often put conditions on access to their data. Even the platforms' own access tools, like CrowdTangle, do not allow filtering for labeled or fact-checked posts.[29] The platforms control virtually every aspect of their interaction with fact-checking organizations, and those organizations have complained that their suggestions for improvements to the process and requests for more information on results go unheeded.[30]

There are strong signs that the platforms' efforts to mitigate misinformation through the use of fact-checking could be made more effective through an independent oversight process. The author of this paper has been unable to find any academic or social science study that exactly replicates what actually occurs when the results of fact-checking are used to label content, downrank it, create friction in the sharing experience, notify users of designations after exposure, and enact other strategies that are embedded in the actual user experience.[31] However, there are indications of a critical multiplier effect. Facebook's website notes that a news story that has simply been labeled false sees its future impressions on the

---

[27] *Fact-Checking on Facebook*, FACEBOOK, https://www.facebook.com/business/help/2593586717571940?id=673052479947730 (last visited Sept. 21, 2020); *Why is a Post on Instagram Marked as False Information?*, INSTAGRAM, https://www.facebook.com/business/help/2593586717571940?id=673052479947730 (last visited Sept. 21, 2020); *Find Fact Checks in Search Results*, GOOGLE, https://support.google.com/websearch/answer/7315336 (last visited Sept. 21, 2020); *See Fact Checks in YouTube Search Results*, YouTube, https://support.google.com/youtube/answer/9229632 (last visited Sept. 21, 2020).

[28] Yoel Roth & Nick Pickles, *Updating Our Approach to Misleading Information*, TWITTER (May 11, 2020), https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html.

[29] Tommy Shane, *Searching for the Misinformation 'Twilight Zone'*, MEDIUM (Nov. 24, 2020), https://medium.com/1st-draft/searching-for-the-misinformation-twilight-zone-63aea9b61cce

[30] Daniel Funke & Alexios Mantzarlis, *We Asked 19 Fact-Checkers What They Think of Their Partnership with Facebook. Here's What They Told Us*, POYNTER (Dec. 14, 2018), https://www.poynter.org/fact-checking/2018/we-asked-19-fact-checkers-what-they-think-of-their-partnership-with-facebook-heres-what-they-told-us/.

[31] Emily Saltz, et al., *It Matters How Platforms Label Manipulated Media. Here are 12 Principles Designers Should Follow*, MEDIUM (2020). https://medium.com/swlh/it-matters-how-platforms-label-manipulated-media-here-are-12-principles-designers-should-follow-438b76546078.

platform drop by 80%.[32] Facebook has also claimed that warnings on COVID-19 misinformation deterred users from viewing flagged content 95% of the time.[33] Twitter reported a 29% decrease in "quote-tweeting" of 2020 election information that had been labeled as refuted by fact-checkers.[34] The proposal for an Internet Superfund would require the qualifying digital platforms to be more transparent in their practices as well as the results associated with them, to share best practices, to share appropriately privacy-protected data with researchers, and to try alternatives from researchers and civil society groups to improve results.[35,36]

Because of their association with their certifying bodies, as well as their own journalistic brands, fact-checking organizations have collectively "proven themselves as professionally legitimate, trustworthy, and, to some extent, beyond the charges of interest and bias that are often leveled at Facebook."[37] Despite that, some partisan groups have claimed fact-checking is arbitrary, or an extension of the liberal-leaning editorial bias of the organization doing the checking.[38] This is demonstrably untrue. The fact-checkers themselves come from a range of backgrounds, including journalism but also political science, economics, law, and public policy.[39] In fact, some of the organizations certified by IFCN lean right, such

---

[32] Craig Silverman, *Facebook Says Its Fact Checking Program Helps Reduce the Spread of a Fake News Story by 80%*, BUZZFEED NEWS (Oct. 11, 2017), https://www.buzzfeednews.com/article/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its.

[33] Elizabeth Culliford, *TikTok Faces Another Test: Its First U.S. Presidential Election*, REUTERS (Sept. 17, 2020), https://uk.reuters.com/article/uk-usa-election-tiktok-misinformation-in-idUKKBN2682PP.

[34] Vijaya Gadde and Kayvon Beykpour, *An Update on Our Work Around the 2020 US Election,* TWITTER (Nov. 12, 2020), https://blog.twitter.com/en_us/topics/company/2020/2020-election-update.html.

[35] Irene V. Pasquetto, et al., *Tackling misinformation: What researchers could do with social media data*, HARV. KENNEDY SCHOOL (HKS) MISINFORMATION REV. (2020), https://misinforeview.hks.harvard.edu/article/tackling-misinformation-what-researchers-could-do-with-social-media-data/

[36] Nick Statt, *Major Tech Platforms Say They're 'Jointly Combating Fraud and Misinformation' About COVID-19*, THE VERGE (Mar. 16, 2020), https://www.theverge.com/2020/3/16/21182726/coronavirus-covid-19-facebook-google-twitter-youtube-joint-effort-misinformation-fraud.

[37] Mike Ananny, *The Partnership Press: Lessons for Platform-Publisher Collaborations as Facebook and News Outlets Team to Fight Misinformation*, COLUM. JOURNALISM REV. (Apr. 4, 2018), https://www.cjr.org/tow_center_reports/partnership-press-facebook-news-outlets-team-fight-misinformation.php.

[38] Craig Timberg & Andrew Ba Tran, *Facebook's Fact-Checkers Have Ruled Claims in Trump Ads Are False – But No One is Telling Facebook's Users*, WASH. POST (Aug. 5, 2020), https://www.washingtonpost.com/technology/2020/08/05/trump-facebook-ads-false/.

[39] LUCAS GRAVES & FEDERICA CHERUBINI, DIGITAL NEWS PROJECT, THE RISE OF FACT-CHECKING SITES IN EUROPE (2016), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/research/files/The%2520Rise%2520of%2520Fact-Checking%2520Sites%2520in%2520Europe.pdf.

as CheckYourFacts, which is part of the conservative publication *The Daily Caller*, as well as *The Dispatch*, which says on its website it is "informed by conservative principles."[40,41] Fact-checking is a fast-growing and diverse industry with organizations taking different approaches to fighting misinformation.[42] It is inevitable that some of the mistrust in the media as a public institution has migrated to the fact-checkers.

There may be ways to enhance or supplement the role of fact-checking, such as the addition of media literacy tools that help consumers evaluate the news for themselves.[43] This proposal calls for fact-checking organizations to be independently certified, then allows the platforms to select those that are compatible with their own content moderation standards and their audiences.

VI. First Amendment Considerations for the Internet Superfund

The First Amendment does not prevent any and every legislative effort to protect individuals, or society as a whole, from harassing or fraudulent content or content that seeks to undermine democracy and civic discourse.[44] However, both the First Amendment and general concerns for freedom of expression require exercising a good deal of care in crafting legislative solutions that have the potential to impact speech.[45] This area of law requires the balancing of many competing policies, and

---

[40] *Poynter Accepts The Daily Caller's 'Check Your Fact' Site To International Fact-Checking Network,* THE DAILY CALLER, https://dailycaller.com/2018/02/15/poynter-accepts-daily-caller-check-your-fact-international-fact-checking-network/.

[41] *About*, THE DISPATCH, https://thedispatch.com/about.

[42] EMILY BELL, *THE FACT-CHECK INDUSTRY* (2019), https://www.cjr.org/special_report/fact-check-industry-twitter.php.

[43] Harold Feld & Jane Lee, *Part V: We Need to Fix the News Media, Not Just Social Media — Part 3*, PUBLIC KNOWLEDGE (Oct. 20, 2018),
https://www.publicknowledge.org/blog/part-v-we-need-to-fix-the-news-media-not-just-social-media-part-3/.

[44] Illustratively, the existence of the "Public Forum Doctrine" and specific cases arising from the regulation of electronic media by the FCC make clear that viewpoint-neutral federal regulation designed to promote access to competing sources of news and a diversity of views and opinions is permitted by the Constitution. *See* Turner v. FCC (*Turner I*), 512 U.S. 622 (1994) (finding support for local news was "content neutral" because the government did not favor particular content, but simply continued a long federal policy of promoting access to news and important community perspectives). *See also* Daniels Cablevision v. FCC, 835 F.Supp 1, 21 (D.D.C. 1993) (holding PEG requirement content neutral and thus survives intermediate scrutiny), *aff'd on other grounds sub nom.* Time Warner Entertainment Co., L.P. v. FCC, 93 F.3d 957 (D.C. Cir. 1996).

[45] FELD, *supra* note 19, at 6. To be clear, the fact that certain speech is protected by the First Amendment does not eliminate the possibility of legal liability for the speech under various alternative theories. For example, speakers may still be held liable for libel.

---

often requires detailed, fact-specific analysis. First Amendment scholars may come to different conclusions when assessing theoretical proposals such as this one.

As a general rule, the First Amendment applies to government initiatives that control or suppress speech, not the role of private companies like social media platforms. Social media platforms are free to establish and enact their own content moderation standards pursuant to their terms of service and community guidelines and to label or demonetize content as they see fit. Fact-checking itself, as well as resultant actions like warnings, labels, and adding of interstitial content adjacent to posts are examples of the platforms themselves exercising their own speech rights.

One potential challenge stemming from the creation of an Internet Superfund is that it could alter this dynamic, creating a nexus between the platforms' decision-making and the government, therefore changing how the First Amendment applies. One avenue to safeguard against this challenge is to ensure that the governing body for the Internet Superfund is given no role in the selection, review, evaluation of accuracy, or any other actions on content that the intermediaries might choose to levy. This also significantly limits the potential for government abuse over this process. Rather than requiring the adoption of any particular viewpoint, this mechanism simply requires there be some fact-checking process in place by major information distribution platforms, collects the fees from platforms, and makes the payments to fact-checking organizations.

Another challenge may be in regard to whether the Internet Superfund represents a form of compelled speech, or otherwise places a burden on the platforms' speech rights. In this proposal, platforms may partner with qualified news analysis organizations that support their own terms of service and content moderation standards, and they retain discretion about whether or how to act on the outcomes of these fact checking processes. The requirement that a fact-checking *process* be in place is content-neutral, and the involvement of the government in establishing this requirement does not suddenly transform the platform into a state actor.[46]

An assessment of the constitutional fit of this proposal would also require considering whether a court might apply an intermediate or a strict standard of scrutiny. Here, again, the fact that the proposal is not content-based, insofar as it applies without regard to viewpoint or subject matter of the content, argues in favor

---

[46] *See* Manhattan Community Access Corp. v. Halleck, 139 S. Ct. 1921 (2019) (holding private operators of public access cable channels are not state actors).

of an intermediate standard.[47] In defending the proposal against constitutional challenge, the government would need to demonstrate that the statute is narrowly tailored to serve a substantial government interest. Combating misinformation and elevating the availability of quality civic information through local journalism has been shown to be a substantial government interest in and of itself.[48] Depending on the topic, it may also engage government interests in public health (e.g., coronavirus and vaccine misinformation), public safety (e.g., hate speech designed to invoke violence against individuals or groups), or national security (e.g., according to the former general counsel of the National Security Agency, misinformation designed to "sow discord... or undermine confidence in our democratic institutions").[49] Lastly, the level of scrutiny would depend on whether the court finds that there are still "ample alternative channels for communication of information."[50] This would involve an inquiry focused on "methods of communication." This proposal impacts only the largest and most dominant platforms, and it calls for independent content moderation policy and decision-making by platforms in accordance with their public rules.

VII. Administration of the Internet Superfund

One of the key mechanisms for the Internet Superfund is determining which platforms it addresses and how to assign the financial contributions required from each platform. Several past proposals have called for imposing various forms of taxes—usually based on advertising revenue—on digital platforms for the purpose of creating trust funds to support local journalism.[51] For the market mechanism

---

[47] *See* Ward v. Rock Against Racism, 491 U.S. 781, 791–94 (1989).

[48] *See* Turner Broadcasting Systems, Inc. v. F.C.C., 512 U.S. 622 (1994) (finding "promoting the widespread dissemination of information from multiple sources" to be an important government interest).

[49] TACKLING DISINFORMATION IS NATIONAL SECURITY ISSUE SAYS FORMER NSA GENERAL COUNSEL, CBS NEWS (Dec. 16, 2020) https://www.cbsnews.com/news/tackling-disinformation-is-national-security-issue-says-former-nsa-general-counsel.

[50] *See* Ward, 491 U.S. at 791; *See also* Turner Broadcasting Systems, Inc. v. F.C.C., 512 U.S. 622 (1994) (finding the FCC's must-carry provisions to be content-neutral and the "editorial discretion" of the broadcasters to be a speech interest that permits intermediate scrutiny).

[51] *See* Victor Pickard, *Journalism's Market Failure is a Crisis for Democracy*, HARVARD BUSINESS REVIEW (Mar. 12, 2020), https://hbr.org/2020/03/journalisms-market-failure-is-a-crisis-for-democracy (proposing taxation of platforms like Facebook and Google to fund public media); Frank Blethen, *Save the Free Press Initiative Seeks Solutions for Local Journalism*, SEATTLE TIMES (May 1, 2020), https://www.seattletimes.com/opinion/save-the-free-press-initiative-seeks-solutions-for-local-journalism/ (proposing a "Free Press" trust or superfund funded by a fee on the ad revenue of major internet platforms); Karen Kornbluh & Ellen P. Goodman, *Five Steps to Combat the Infodemic*, The German Marshall Fund of the U.S.: Transatlantic Takes (Mar. 26, 2020), https://www.gmfus.org/sites/default/files/Kornbluh%20Goodman%20_5%20steps%20to%20combat%20the%

created by the Internet Superfund, a more appropriate tool is a federal user fee: a "fee assessed to users for goods or services provided by the federal government."[52] User fees generally apply to federal programs or activities that provide special benefits to identifiable recipients above and beyond what is normally available to the public.[53] Examples of federal user fees include tobacco fees collected by the Food and Drug Administration, filing fees collected by the Securities and Exchange Commission, and the motor vehicle and engine compliance program fee collected by the EPA.[54] Although the public as a whole absolutely benefits from fact-checking and the flourishing of local news media, it is the platforms who receive special benefits. Thus, the platforms would be the payers of the federal user fee.

The standards for platforms' obligations under the Internet Superfund should be based on their dominant role in communications and the incentives associated with their business models. This proposal suggests that digital platforms should be required to contribute a user fee based on their total number of monthly active users provided that that they:

- Are based in the United States;
- Rely predominantly on locating, indexing, linking to, displaying or distributing third-party or user-generated content for their commercial value, in the form of an advertising business model that generates more than 85% of the company's total annual revenue;
- Have advertising-based revenue exceeding $1 billion annually; and
- Have a total global monthly active base of at least 1 billion users.

20infodemic.pdf (proposing a fund for local journalism); Ethan Zuckerman, *The Case for Digital Public Infrastructure*, KNIGHT FIRST AMENDMENT INSTITUTE (Jan. 17, 2020), https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure (proposing a 1 percent tax on "highly surveillant advertising" to create an endowment used as a subsidy for independent journalism); Paul Romer, *A Tax That Could Fix Big Tech*, N.Y. TIMES (May 6, 2019), https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html (proposing a tax on revenue from sales of targeted digital ads to "encourage platform companies to shift toward a healthier, more traditional model"); Berggruen Institute, Renewing Democracy in the Digital Age (Mar. 2020), https://www.berggruen.org/activity/renewing-democracy-in-the-digital-age/ (proposing a tax based on profit from advertising revenue to expand newsroom revenue streams); Russell Brandom, *Bernie Sanders Endorses a Targeted Advertising Tax to Fund Local Journalism*, THE VERGE (Aug. 27, 2019), https://www.theverge.com/2019/8/27/20835018/bernie-sanders-targeted-advertising-tax-google-facebook-journalism (proposing a targeted ad tax used to fund "civic-minded media").

[52] D. ANDREW AUSTIN, CONG. RSCH. SERV., ECONOMICS OF FEDERAL USER FEES (2019) [available at https://fas.org/sgp/crs/misc/R45463.pdf].

[53] *Id.*

[54] U.S. GOV'T ACCOUNTABILITY OFF., FEDERAL USER FEES: KEY CONSIDERATIONS FOR DESIGNING AND IMPLEMENTING REGULATORY FEES 4 (2015) [available at https://www.gao.gov/assets/680/672572.pdf].

This would mean that Google (based on its estimated 1.0 billion monthly active search users), Facebook (2.7 billion), YouTube (2.0 billion) and Instagram (estimated at 1.1 billion) would currently qualify for the fund.[55] Assuming a purely illustrative fee of $1 annually per monthly active user, the Internet Superfund would start as an annual fund of $6.8 billion for information analysis services to clean up the internet. In that case, the total user fees assigned to each platform would represent just 1.6% (Google, including YouTube) or 4.4% (Facebook, including Instagram) of total global corporate revenue.[56] Even a fee of $0.10 per monthly active user, collected from the leading information distribution platforms, would allow over half a billion dollars for information cleanup.

A calculation method based on the number of monthly active users avoids the need to know in advance the quantity of information that will need to be fact-checked, or what proportion of it is "good" or "bad"—it can be assumed that the quantity of misinformation and the potential for harms associated with it increases based on the number of users of each platform. This assumption also avoids the complexity associated with the different quantities of information represented by video, images, or text, as a so-called "bit tax" would introduce. Monthly active users is a nonfinancial performance indicator often used to assess online user growth and engagement among the platforms.[57] The fees could be assessed on a monthly or quarterly basis, accounting for any fluctuations in the number of active users. Use of this metric would provide a strong incentive to standardize its calculation and clean up fraudulent, unidentified, and corrupting accounts on each platform.[58]

---

[55] Christo Petrov, *The Stupendous World of Google Search Statistics*, TECHJURY (July 27, 2020), https://techjury.net/blog/google-search-statistics/#gref; *Facebook Reports Second Quarter 2020 Results*, FACEBOOK (July 30, 2020), https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Second-Quarter-2020-Results/default.aspx; YouTube for Press, YOUTUBE, https://www.youtube.com/about/press/ (last visited Sept. 10, 2020); Jenn Chenn, *Important Instagram Stats You Need to Know for 2020*, SPROUT SOCIAL (Aug. 5, 2020), https://sproutsocial.com/insights/instagram-stats/.

[56] Calculated using the monthly active users noted for each platform, the illustrative fee of $1 per monthly active users, and F'20 year-end revenue for Google (which includes YouTube) and Facebook (which includes Instagram). See *Facebook Reports Fourth Quarter and Full Year 2020 Results*, FACEBOOK (January 27, 2021), https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx; *Alphabet Announces Fourth Quarter and Fiscal Year 2020 Results*, GOOGLE, February 2, 2021, https://abc.xyz/investor/static/pdf/2020Q4_alphabet_earnings_release.pdf

[57] The measure of monthly active users remains in wide use despite concerns that it is an unaudited measure, does not consistently correlate with future financial performance, and is calculated differently among platforms. *See* Theresa F. Henry, David A. Rosenthal & Rob R. Weitz, *Socially awkward: social media companies' nonfinancial metrics can send a mixed message*, 218(3) J.A. 52 (2014), https://link.gale.com/apps/doc/A381838689/LT.

[58] Illustratively, Facebook reports it disabled 1.7 billion fake accounts in the first quarter of 2020, and estimates that as much as 5% of its worldwide monthly active users during Q4 2019 and Q1 2020 were also fake. See

---

Lastly, using a fee instead of a tax on advertising avoids having to address fluctuations in spending by advertisers, which may have no correlation with the amount of misinformation flowing across platforms.

The proposed framework for the collection, allocation, and distribution of payments under the Internet Superfund is modeled after the system put in place for the collection of regulatory user fees by the EPA and the allocation and distribution of payments from the EPA Superfund.[59] Payments from platforms would be housed in a federal trust fund, administered by an independent body, and allocated among qualified news organizations certified to cleanse toxic information. Platforms subject to the fees will each have member accounts tied to an Internet Superfund website used to calculate and assess costs. This account will be connected to the federal agency designated to house the Internet Superfund (again, ideally, one with other regulatory authority over platforms).[60]

In order to avoid politicization of the allocations to fact-checking organizations, the funds would be disbursed as fees per hour, amount of content, or number of fact checks completed for the platforms. In other words, a general rate and basis of payment would be established by the federal agency (potentially derived from the payment schedules already in place between platforms and fact-checking organizations), but the actual amount each news organization receives would be dependent on the amount of service exchanged between the platforms and the news organization. Regular audits would be conducted by an independent committee of stakeholders to prevent organizations from taking advantage of the payment system and inflating their fees. In addition to funding news organizations, money from the Internet Superfund will go to administration of the fund itself.

Fact-checking is a fast-growing and diverse industry with organizations taking different approaches to fighting misinformation, so it is necessary platforms are given some discretion to partner with organizations best suited to their individual

---

Community Standards Enforcement Report, Facebook: Transparency, FACEBOOK (Aug. 2020), https://transparency.facebook.com/community-standards-enforcement#fake-accounts.

[59] Bryan Anderson, *Taxpayer Dollars Fund Most Oversight and Cleanup Costs at Superfund Sites*, WASH. POST (Sept. 20, 2017), https://www.washingtonpost.com/national/taxpayer-dollars-fund-most-oversight-and-cleanup-costs-at-superfund-sites/2017/09/20/aedcd426-8209-11e7-902a-2a9f2d808496_story.html.

[60] *See* Tom Wheeler, Phil Verveer & Gene Kimmelman, *New Digital Realities; New Oversight Solutions in the U.S.: The Case for a Digital Platform Agency and a New Approach to Regulatory Oversight,* THE SHORENSTEIN CENTER ON MEDIA, POLITICS, AND PUBLIC POLICY, HARVARD KENNEDY SCHOOL 6 (Aug. 2020), https://shorensteincenter.org/wp-content/uploads/2020/08/New-Digital-Realities_August-2020.pdf (discussing the need for a Digital Platform Agency to conduct oversight of digital platform market activity).

content moderation policies.[61] Diversity of platform content moderation standards and practices better ensures that all kinds of misinformation can be effectively and efficiently dealt with.[62] However, this does not mean the platforms will be completely without guidance. The designated federal agency will provide the platforms with assistance in choosing qualified fact-checking organizations to partner with and guidelines designed to ensure there is still transparency, autonomy and equity between them.

VIII. Conclusion

The 2020 presidential election and its aftermath have provided more vivid examples of the volume, velocity and potential for harm associated with misinformation on digital platforms—and of the potential power of strategies rooted in fact-checking, such as labeling, addition of interstitial content, deamplification, friction, and changes in user experience.[63] Given the scale and complexity of the problem, it is important to acknowledge that the Internet Superfund would only be one of a system of solutions to create content moderation that supports the public interest. Other proposals may engage competition policy, address reforms to Section 230 of the Communications Act of 1934, or seek to regulate the platforms' internal mechanisms, like algorithmic amplification or improved training or working conditions for human content moderators.

---

[61] Bell, *supra* note 42 (discussing the increase in fact-checking organizations between 2016 and 2019).

[62] Evelyn Douek, *The Rise of Content Cartels*, KNIGHT FIRST AMENDMENT INSTITUTE (Feb. 11, 2020), https://knightcolumbia.org/content/the-rise-of-content-cartels (examining some of the costs associated with collaboration and standardized content moderation practices).

[63] Kurt Wagner, Sarah Frier & Mark Bergen, *Social Media's Election Report Card*, BLOOMBERG (Nov. 5, 2020), https://www.bloomberg.com/news/newsletters/2020-11-05/social-media-s-election-misinformation-report-card.

Combining Ad Libraries with Fact Checking to Increase Transparency of Misinformation

Ivar A. Hartmann, Insper Learning Institution, Brazil

I. Introduction

While most people might associate fake news with politicians, a substantial component of the current misinformation crisis threatening democracies is websites producing misinformation for commercial reasons.[1] The pandemic has exacerbated this problem, as it provides fertile ground for conspiracy theories and those willing to publish them for ad profit. A responsible press should not indulge their readers' desire to project fanciful theories onto complex global problems. However, unscrupulous players profit by emulating the communication style of traditional news outlets,[2] but without going through the effort of applying journalistic guidelines to produce accurate information.[3]

The reason this business model is viable is that an increasingly large share of advertising money is allocated by algorithms. Shady publishers would not attract substantial revenue if companies were making individualized choices of *where* to place their ads. But microtargeting has removed advertisers' incentive to associate their brand solely with reputable news sources. Instead, companies are served neatly carved demographics, and make decisions based on *who* to target, rather than which website to sponsor. Automated decisions on ad placement can be obscure even to advertisers themselves. A regulatory response which introduced more transparency to programmatic advertising, thus realigning the incentives of advertisers, would deny ad revenue to harmful misinformation publishers. The best way to achieve this is to provide companies, and the public, with access to large scale, detailed data on advertising spending that is cross-referenced against databases of fact-checked content.

Fact-checking has become popular in recent years as a central journalistic defense against misinformation. While its usage to directly counter fake news remains

---

[1] Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31(2) JOURNAL OF ECONOMIC PERSPECTIVES, 211 (2017) at 218.
[2] David M. Lazer et al., *The science of fake news*, 359 SCIENCE 1094 (2018) at 1094
[3] Hunt Allcott & Matthew Gentzkow, supra note 1, at 213.

crucial, combining fact-checks with enhanced advertising transparency can serve to illuminate the money trail which drives the production of misinformation. Transparency could help to increase accountability and oversight around programmatic advertising, to track how resources are channeled to fake news outlets. This paper proposes a framework of enhanced transparency obligations by digital advertising platforms to shed light on the profile, size and impact of ad-subsidized online misinformation.[4] Creating open access repositories of information regarding the resources that fund fact-checked content would increase public knowledge about the problem and its patterns, providing a strong disincentive for companies who currently sponsor misinformation through microtargeted advertising.

Part II of this paper presents the problem: how exploitative misinformation funded by programmatic advertising jeopardizes the health of the public sphere. Part III describes how the platforms' voluntarily published ad libraries, including lists of political advertisements and basic information about them, are one of the most effective responses so far to this problem. Part IV offers a proposal to mandate an increase in the scope of ad libraries, cross-referencing these libraries with the results of fact-checking processes. Part V concludes.

II. The Problem

Across all constitutional democracies, material published on regular websites and blogs, and funded by advertising in the form of banners, is a crucial piece of the online misinformation puzzle.[5] Financial gain is as much a motivation for fake news production as political gain.[6] For-profit misinformation or exploitative

---

[4] Commentary surrounding the German Network Enforcement Act (NetzDG) aptly illustrates the chilling effects of command-and-control regulation of online speech and its disproportionate restriction of the right to freedom of expression. *See, e.g.,* Gerald Spindler, *Internet Intermediary Liability Reloaded The New German Act on Responsibility of Social Networks and its (In-) Compatibility with European Law*, 8 JIPITEC 166 (2017). On the other hand, the section of the NetzDG creating transparency obligations for social media regarding their moderation practices is in line with the broader focus of transparency in this proposal. So far, it seems the main obstacle in its enforcement is companies directing users to register complaints with grounds different from those in the law, such that a large share of moderation decisions fall outside of the scope of the transparency reports. Ben Wagner et al., *Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act*, *in* FAT* '20: Conference on Fairness, Accountability, and Transparency (2020), https://dl.acm.org/doi/abs/10.1145/3351095.3372856 (last visited Feb 1, 2021).
[5] Richard Graham, *Google and advertising: digital capitalism in the context of Post-Fordism, the reification of language, and the rise of fake News*, 3(45) PALGRAVE COMMUNICATIONS 1 (2017), at 12.
[6] Giandomenico Di Domenico et al., *Fake News, social media and marketing: A systematic review*, 124 JOURNAL OF BUSINESS RESEARCH 329 (2021) at 337. There is also a third motivation, based on more anarchic or

---

misinformation is particularly detrimental to a healthy public forum because it is produced at a faster pace and disseminated in an organized, profit-driven manner. Indeed, "the primary driver behind the recent explosion of fake news online is advertising income."[7] This may be distinguished from fake news posted for purely ideological or electoral reasons, even if political actors play an important role in the spread of for-profit fake news produced by third parties.[8] Despite experiments suggesting people are able (to some extent) to identify fake news, they are also more likely to believe misinformation aligned with their political affiliations.[9]

Misinformation which masquerades as legitimate journalism is not a new phenomenon. What has tipped the scales in favor of misinformation in recent years is the new funding system of programmatic advertising, whereby advertisers pick the profile of people who they would like to be exposed to their ad, and algorithms then choose the venue regardless of the integrity of the underlying content.[10] This means that companies no longer consciously choose ad space based on the reputation of a media outlet.

Part of the problem lies in the way that Google and Facebook's business models interact with one another.[11] User behavior in relation to paid content on Facebook

---

trollish instincts, and which is more resistant to traditional fact-checking. See Brooke Borel, *Fact-Checking Won't Save Us From Fake News*, FIVETHIRTYEIGHT (Jan 4, 2017), https://fivethirtyeight.com/features/fact-checking-wont-save-us-from-fake-news/.

[7] Adam J. Mills, Christine Pitt & Sarah Lord Ferguson, *The Relationship between Fake News And Advertising Brand Management in the Era Of Programmatic Advertising and Prolific Falsehood*, 59(1) JOURNAL OF ADVERTISING RESEARCH 3 (2019), at 4.

[8] "Carroll estimated that a fake-news share from within the Trump campaign could earn the lucky hoaxer as much as $10,000 in extra revenue, provided they have taken full advantage of the ad services available to them. That's a "huge economic incentive to create stories that they want to distribute." Abby Ohlheiser, *This is how Facebook's fake-news writers make money*, THE WASHINGTON POST (Nov 18, 2016), https://www.washingtonpost.com/news/the-intersect/wp/2016/11/18/this-is-how-the-internets-fake-news-writers-make-money/.

[9] Andrew Guess et al., *Fake news, Facebook ads, and misperceptions. Assessing information quality in the 2018 U.S. midterm election campaign* (2019), http://www.dartmouth.edu/~nyhan/fake-news-2018.pdf, at 21.

[10] "When brands can track desirable users across the web, they not only have less incentive now than they once did to think about editorial context, they may see a benefit in reaching those users in the cheapest possible spaces—spaces that tend not to belong to reputable publishers, but to clickbait artists." Joshua A. Braun, John D. Coakley & Emily West, *Activism, Advertising, and Far-Right Media: The Case of Sleeping Giants*, 7(4) MEDIA AND COMMUNICATION 68 (2019), at 70.

[11] "Combined with the media- specific characteristics of search engine results and social media feeds that decontextualize individual articles and present a diverse range of content, Google and Facebook encourage a logic that incentives clickbait headlines. (⋯) As long as profits are tied directly to how much an article is shared or viewed then very particular kinds of media content will continue to be incentivized over others. Fake news is just one example to consider when investigating how Google creates avenues for profit and how

---

is a determinant of the flow of readers to off-platform misinformation on websites and blogs, which in turn are funded by programmatic ads courtesy of Google.[12] The two companies are deeply intertwined in the ecosystem of fake news[13] and even after the public focus on curbing misinformation since 2016, they continue to exploit it for profit.[14]

As a result, the essential funding infrastructure for media is no longer one of open and accountable decisions regarding ad placement, where companies seek to associate their image with that of a reputable news source. Instead, advertising resources are guided by automated decision-making performed by obscure, platform-driven algorithms. This is precisely the "model required for fake news to flourish".[15] Algorithms are trained to push engaging news content regardless of its integrity, and in a manner which is not subject to public scrutiny. As hoax websites gradually increase their revenue, swallowing a larger share of a revenue stream that emanates largely from Google Ads, their economic model also competes with legitimate news media, crowding the latter out.[16]

Programmatic advertising is driven by microtargeting, which creates a lack of broader accountability. People know the ad they were shown but have no cost-effective way of learning what ads the digital platforms offered others.[17] Microtargeting also threatens the health of the public discourse by fragmenting the electorate into groups whose understanding of reality may be mutually

---

Google's economics co-depend on other online institutions, in particular, Facebook." Richard Graham, *supra* note 5 at 17.

[12] Emily Bell & Taylor Owen, *The Platform Press: How Silicon Valley reengineered journalism*, TOW CENTER FOR DIGITAL JOURNALISM (Mar 29, 2017), https://www.cjr.org/tow_center_reports/platform-press-how-silicon-valley-reengineered-journalism.php. The story is now famous of fake news producers from Macedonia who earned a proportionally tremendous sum of money: "He claims to have earned up to $2,500 a day from advertising on his website, while the average monthly income in Macedonia is just $426." Florence Davey-Attlee & Isa Soares, *The Fake News Machine: Inside a Town Gearing up for 2020*, CNN (Sept 13, 2017), https://money.cnn.com/interactive/media/the-macedonia-story/.

[13] Adam J. Mills, Christine Pitt & Sarah Lord Ferguson, *supra* 7 note at 5. *See also* Richard Graham *supra* note 5, at 17.

[14] Jeremy B. Merrill & Marshall Allen, *"Trumpcare" Does Not Exist. Nevertheless Facebook and Google Cash In on Misleading Ads for "Garbage" Health Insurance*, PROPUBLICA (Oct. 20, 2020), https://www.propublica.org/article/trumpcare-does-not-exist-nevertheless-facebook-and-google-cash-in-on-misleading-ads-for-garbage-health-insurance.

[15] Richard Graham, supra note 5 at 14.

[16] Joshua A. Braun & Jessica L. Eklund, *Fake News, Real Money: Ad Tech Platforms, Profit-Driven Hoaxes, and the Business of Journalism*, 7(1) DIGITAL JOURNALISM 1 (2019), at 19.

[17] Andrew Guess et al., supra note 9, at 2.

incompatible.[18] Additionally, microtargeting for political advertising "has a unique risk of harm not associated with traditional political advertising: interference with the rights to privacy and data protection."[19]

Transparency matters because it affects citizen behavior towards political advertising, and therefore reader attitudes toward publishers that profit from using fake news to tap into political controversies. In an experiment to test the extent to which current ad disclosure features on Facebook were effective at drawing the attention of users, Matthew T. Binford and several coauthors found that the transparency measures implemented by the social media giant are not enough to ensure user awareness.[20] However, if readers know that a message is a paid advertisement, they attribute less credibility to it, share it less often, and also view the brand more negatively.[21]

Much of the current flow of misinformation is supplied by publishers whose incentive structure is opaque. The programmatic advertising system rewards editors whose publishing is built around maximizing attention. Transparency could illuminate the resources financing specific instances of fake news, exposing disingenuous publishers. Over time, this would change reader attitudes towards such publishers and negatively impact the brand image of their advertisers, thus shaming the companies feeding the continuous exploitative misinformation system.

III. Existing Ad-Transparency Initiatives

Ad Libraries as a Driver of Accountability

The most effective tool to support greater transparency in the advertising ecosystem has been the publication of ad libraries. These are repositories curated by digital platforms listing all promoted content, including relevant metadata, such as the money paid to boost the reach of a post and basic information about the

---

[18] Abby K. Wood & Ann M. Ravel, *Fool Me Once: Regulating Fake News and Other Online Advertising*, 91 S. CAL. L. REV. 1223 (2018), at 1236.

[19] Tom Dobber, Ronan Ó Fathaigh & Frederik J. Zuiderveen Borgesius, *The regulation of online political micro-targeting in Europe,* 8(4) INTERNET POLICY REVIEW 1 (2019), at 13.

[20] Matthew T. Binford et al., *Invisible transparency: Visual attention to disclosures and source recognition in Facebook political advertising*, JOURNAL OF INFORMATION TECHNOLOGY & POLITICS 1 (2020, ahead of print) at 11. A lesser-known politician in the ad seems to render people unable to remember it was even an ad in the first place, meaning that the risk of explicitly marked ads not being recognized as such is high.

[21] Id., at 4.

---

people who actually saw it.[22] Facebook and Instagram launched an ad library in 2018,[23] followed by Twitter[24] and then Google, though the latter only includes candidate ads.[25]

There are a number of key benefits flowing from these libraries, particularly the fact that they reveal the resources funneled into political marketing through digital advertising platforms. The libraries uncover the profile and patterns of the non-organic social media operations of political campaigns, allowing the press, academia, and NGOs to digest this information and share their analyses with the public. If a specific post misinforms constituents, there is significant public interest in releasing data about its promotion and targets – data which was wholly unavailable before ad libraries. Ad libraries provide valuable knowledge on the content, authors and audience of artificially spread fake news. They also allow researchers to track the impact of particular campaigns after the fact, by looking for changes in behavior by those exposed to misinformation.

In addition to voluntary efforts by digital platforms, many jurisdictions have pushed for increased transparency in the advertising space by regulating programmatic advertising disclosures, especially connected to elections. The 2018 European Union Code of Practice on Disinformation requires adhering companies to disrupt exploitative misinformation models, restricting advertising revenue through measures which include partnerships with fact-checking organizations.[26] In 2019,

---

[22] Juan Carlos Medina Serrano et al., *Exploring Political Ad Libraries for Online Advertising Transparency: Lessons from Germany and the 2019 European Elections*, *in* SMSociety'20: International Conference on Social Media and Society (2020), https://dl.acm.org/doi/abs/10.1145/3400806.3400820 (last visited Feb 2, 2021), at 111.

[23] "The labeling won't just apply to candidate and election ads, but those dealing with political issues such as "abortion, guns, immigration or foreign policy." John Constantine, *Facebook and Instagram launch US political ad labeling and archive*, TECHCRUNCH (May 24, 2018), https://techcrunch.com/2018/05/24/facebook-political-ad-archive/. The library (initially named Ad Archive) is available at https://www.facebook.com/ads/library/ (last visited Feb 2, 2021).

[24] Anthony Ha, Twitter launches its Ads Transparency Center, where you can see ads bought by any account, TECHCRUNCH (June 28, 2018), https://techcrunch.com/2018/06/28/twitter-ads-transparency-center/. The library is available at https://business.twitter.com/en/help/ads-policies/product-policies/ads-transparency.html (last visited Feb 2, 2021)

[25] Taylor Hatmaker, *Google releases a searchable database of US political ads*, TECHCRUNCH (Aug 15, 2018), https://techcrunch.com/2018/08/15/google-political-ad-library/. The library is available at https://transparencyreport.google.com/political-ads/home (last visited Feb 2, 2021).

[26] "Relevant Signatories commit to deploy policies and processes to disrupt advertising and monetization incentives for relevant behaviors, such as misrepresenting material information about oneself or the purpose of one's properties. These policies and processes can include, for example, the restriction of advertising services or limiting paid placements, and could potentially take place in partnership with fact-checking

---

the European Commission extolled the ad libraries put up by Facebook, Google and Twitter, but criticized their shortcomings in design (lacking data on issue-based ads) and implementation (technical difficulties).[27]

In the United States, a bill introduced in 2019, named the Honest Ads Act, would require digital platforms under specific circumstances to publish "a public file of all electioneering communications purchased" that would include "a digital copy of the advertisement, a description of the audience the advertisement targets, the number of views generated, the dates and times of publication, the rates charged, and the contact information of the purchaser."[28] The bill is part of a list of legislation that the 2021 US Congress is considering in order to curtail the digital platforms' power and lack of accountability.[29]

In Brazil, a draft known as "the fake news bill" was quickly approved in the Senate in 2020 and is under consideration in the House of Deputies as of February 2021. Among a number of reforms, including some very problematic ones,[30] the bill creates two major obligations for social media companies with respect to their advertising disclosures. First, the companies are required to publish an open database of all advertisements which mention a candidate, coalition or political

---

organizations." European Union, *EU Code of Practice on Disinformation* (2018), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54454 (last visited Feb 3, 2021), at 5.

[27] "In particular, Facebook and Twitter have made political advertisement libraries publicly accessible, while Google's library has entered a testing phase. This provides the public with more transparency around political ads. However, further technical improvements as well as sharing of methodology and data sets for fake accounts are necessary to allow third-party experts, fact-checkers and researchers to carry out independent evaluation. At the same time, it is regrettable that Google and Twitter have not yet reported further progress regarding transparency of issue-based advertising, meaning issues that are sources of important debate during elections." European Commission, *Code of practice against disinformation: Commission welcomes the commitment of online platforms ahead of the European elections*, European Commission (April 23, 2019), https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_19_2174.

[28] Senator Mark R. Warner, *The Honest Ads Act*, https://www.warner.senate.gov/public/index.cfm/the-honest-ads-act (last visited Feb 3, 2021).

[29] Lauren Feiner, *Democratic Senate control could breathe new life into net neutrality and other tech regulations*, CNBC (Jan 9, 2021), https://www.cnbc.com/2021/01/09/democratic-senate-control-could-breathe-new-life-into-tech-policy.html.

[30] Previous versions of the bill tried to produce a viewpoint specific concept of fake news that would invite State censorship while also creating exacting requirements for people subscribing to social networks. See Raphael Tsavkko Garcia, *Brazil's "fake news" bill won't solve its misinformation problem*, MIT Technology Review (Sept 10, 2020), https://www.technologyreview.com/2020/09/10/1008254/brazil-fake-news-bill-misinformation-opinion/. The draft eventually approved by the Senate was "considerably dehydrated" but still contained a few controversial elements. Angelica Mari, *Brazilian Senate passes fake news bill*, ZDNet (July 1, 2020), https://www.zdnet.com/article/brazilian-senate-passes-fake-news-bill/.

party, regardless of whether these ads take place during a campaign. The database must include the identity of the advertiser, amount spent, and duration of the ad campaign. Second, drawing inspiration from the German Network Enforcement Act, the companies are required to publish quarterly content moderation reports, including figures for the total amount of content published by bots and artificial content distribution networks, the so-called unidentified sponsored and advertising content detected by the company, as well as the corresponding action taken.[31]

Shortcomings of Ad Libraries

Ad libraries, as they currently exist, have a number of shortcomings. First, they are limited in their scope. They only encompass ads by politicians, parties, or their representatives. Selection criteria are limited to election-related advertising, and not all libraries include "issue ads".

Limiting the ads available in these libraries to those that are classified as "political" fails to provide an adequate solution to track misinformation flows.[32] Even during elections, exploitative misinformation does not originate solely from websites and profiles explicitly linked to candidates and parties. The identification and classification mechanisms used to sort political ads are also prone to error.[33] Using automation to define which ads are political in nature – because their content is political or because the website in which they are displayed is political – creates a severe risk of inaccuracy, particularly when combined with a lack of reliable data about how the algorithms from different companies actually perform. Furthermore, this method of separating political ads from the rest can be more easily gamed.[34]

A second problem is that some libraries only cover advertising spending during election season. Fake news is a challenge to the public discourse year round, as it

---

[31] See articles 13, V and 15. Senado Federal do Brasil, *Projeto de Lei nº 2630, de 2020*, https://www25.senado.leg.br/web/atividade/materias/-/materia/141944. The Center for Technology and Society at FGV Law School in Rio de Janeiro published a translation of the entire bill, available at https://bit.ly/br_fakenewsbill (last visited Feb 3, 2021).

[32] "[T]he drawback of 'issue ads' as a scoping device, is that the concept of a political 'issue' is broad and subjective, and makes it difficult for archive operators to develop actionable definitions and enforce these in practice." Paddy Leerssen et al., *Platform ad archives: promises and pitfalls,* 8(4) INTERNET POLICY REVIEW 1 (2019), at 8.

[33] The press was able to prove certain political ads get through on Facebook without proper listing in the ad library. See Jennifer Baker, *Facebook Ad Library: Only 3 'Brexit' ads for whole month. Yeah, right!,* TNW (Nov 1, 2018), https://thenextweb.com/contributors/2018/11/01/facebook-ad-library-only-3-brexit-ads-for-whole-month-yeah-right/.

[34] Juan Carlos Medina Serrano et al., supra 23 note at 119.

influences what citizens do not only as voters, but also as communicators who guide the public discourse. Fortunately, approximately a year after their ad library was first created, Facebook decided to increase its scope to encompass any ad run by pages, instead of only those classified as political ads.[35] Twitter went one step further and banned political ads altogether, though it did not provide information on how it will apply this rule to issue ads.[36] Twitter still does not provide data for ads it considers non-political.

A third problem relates to access conditions and usability. The Facebook ad library, for example, is riddled with bugs that make it impossible for information to be extracted.[37] Its requirement for researchers to confirm their identity prior to obtaining access is also problematic.[38] Public interest information should be openly accessible, rather than restricted through demands for personal information and a need to provide statements regarding one's purpose for access.[39] Companies currently deny access to ad data in bulk, either directly, as Twitter does,[40] or

---

[35] Andrew Hutchinson, *Facebook Expands Ad Archive to All Ads and Pages*, SOCIALMEDIATODAY (March 29, 2019), https://www.socialmediatoday.com/news/facebook-expands-ad-archive-to-all-ads-and-pages/551584/. The problematic "political" criterion made a comeback, though, as the platform moved to *ban* all political ads the week before election in 2020. In addition to false negatives, the policy is ineffective because it overlooks posts that were already being promoted before the final week. Davey Alba & Sheera Frenkel, *Why Facebook's Blocking of New Political Ads May Fall Short*, THE NEW YORK TIMES (Sept 4, 2020), https://www.nytimes.com/2020/09/04/technology/facebooks-political-ads-block-election.html.

[36] Amrita Khalid, *Twitter's political ad ban won't stop politicians getting their messages out*, QUARTZ (Nov 6, 2019), https://qz.com/1742852/campaigns-arent-worried-about-twitters-political-ad-ban/.

[37] These technical issues were uncovered by researchers from two separate initiatives. "In general, we encountered three categories of issues with the Facebook Ad Library API. First, software programming errors that cripple a user's ability to complete even a single search (⋯) Second, technical or data issues that affect a user's ability to reliably retrieve data from multiple searches". Mozilla Foundation, *Data Collection Log — EU Ad Transparency Report*, Mozilla Foundation (2019), https://adtransparency.mozilla.org/eu/log/ (last visited Feb 2, 2021). See also French Ambassador for Digital Affairs, *Facebook Ads Library Assessment*, French Ambassador for Digital Affairs (2019), https://disinfo.quaidorsay.fr/en/facebook-ads-library-assessment (last visited Feb 2, 2021).

[38] "We'll need to confirm your identity and location. If you haven't already confirmed your ID, it typically takes 1-2 days to complete this step." Facebook Help Center, *What is the Facebook Ad Library and how do I search it?*, FACEBOOK, https://www.facebook.com/help/259468828226154 (last visited Feb 2, 2021).

[39] "The real genius of FOIA, however, is its provision permitting anyone to seek access to government-held information upon request. (⋯) It allows "any person" -including corporations, nonprofit entities, and even foreign nations -to request any record from any federal agency or government-controlled entity on any subject, without saying why the record was requested or what purpose disclosure would serve." David C. Vladeck, *Information Access - Surveying the Current Legal Landscape of Federal Right-to-Know Laws*, 86 TEX. L. REV. 1787 (2008), at 1797.

[40] Matthew Rosenberg, *Ad Tool Facebook Built to Fight Disinformation Doesn't Work as Advertised*, THE NEW YORK TIMES (July 25, 2019), https://www.nytimes.com/2019/07/25/technology/facebook-ad-library.html.

---

indirectly, as Facebook does by limiting the number of search results in queries.[41] While technical glitches are likely to be fixed eventually, accessibility problems which stem from deliberate design choices are more concerning.

A fourth problem relates to the inconsistent geographic scope of ad libraries. Google provides information for ads based only in Australia, the European Union, the United Kingdom, India, Israel, New Zealand, Taiwan, and the United States. Facebook includes a much larger number of countries from more continents.[42] The Twitter library, discontinued since the social media company banned political ads entirely in 2019, offers data for older political ads that targeted most countries, but the information on issue ads is limited to the United States.

A fifth problem concerns inconsistency in the data catalogued by ad libraries. Given the crucial role that microtargeting plays in the programmatic advertising funding system for misinformation, data on the characteristics of the audience chosen by the advertiser is vital for ad libraries to fulfill their role. Twitter includes all information about how the ad was targeted. Google provides information on all the targeting options available for political ads, namely, geographic location, age and gender. Facebook does not provide any targeting information, only data about the profile of users who actually viewed the ads.[43]

Of these five issues, the thematic scope of ad libraries is the most pressing. By limiting their libraries to "political" or "issue" ads, Google and Twitter deny the public access to vital data about exploitative misinformation that often does not fit neatly into such categories. Fake news is not a problem exclusively *about* elections or *during* elections, it is a malady of the networked public sphere that, in turn, hinders the functioning of democracy.[44] Coronavirus hoax messages and websites,

---

[41] "The USA, for instance, counts 3.8 million ads and we observed that requests cannot ask for more that 2000 ads each. To reach the last page, one has to successfully execute about 1900 requests in order, which we found impossible to achieve in the two weeks we tried." French Ambassador for Digital Affairs, supra note 38.

[42] Though the dropdown in the report tool page lists 83 countries, the source code for the page includes a much larger number, perhaps as a sign of plans to increase to more countries in the future. https://www.facebook.com/ads/library/report/ (last visited Feb 2, 2021).

[43] Juan Carlos Medina Serrano et al., supra note 23 at 118.

[44] "The information imbalance between a voter who has little time to inform herself on politics and the political or media speaker is often profound. This imbalance gives the voter reason to trust the veracity of a piece of false campaign speech." Alvin I. Goldman & Daniel Baker, *Free Speech, Fake News, and Democracy*, 18 FIRST AMENDMENT LAW REVIEW 66 (2019), at 129.

for example, have proven to be a nightmare for public health officials in 2020, with knock-on effects in the political realm.[45]

Furthermore, misinformation is not produced and spread solely by politicians or parties.[46] Much of the content published by hoax websites has no direct connection to specific politicians. Antivaxx misinformation impairing COVID-19 inoculation initiatives is both a more pressing problem than libel of candidates for office, and one that runs much deeper.[47]

A significant amount of the money which drives misinformation is absent from the Google and Twitter ad libraries because of the design decision to limit transparency to "political" or "issue" ads. This prevents civil society from accessing and analyzing valuable data about how fake news is promoted, or scrutinizing the broader role of advertising in the misinformation ecosystem. Likewise, the Facebook ad library amasses all ads together without classification options that would make it possible to isolate fake news ads. The following section lays out a proposal for remedying these deficiencies, by mandating that ad libraries should be raised to a higher common standard, to facilitate better oversight of the misinformation ecosystem.

IV. The Proposed Response

This paper proposes that that all major platforms[48] be mandated to set up ad libraries of all paid or sponsored content, populated according to universal

---

[45] Jesselyn Cook, *A Toxic 'Infodemic': The Viral Spread Of COVID-19 Conspiracy Theories*, HUFFPOST (July 4, 2020), https://www.huffpost.com/entry/coronavirus-conspiracy-theories-social-media_n_5e83d701c5b6a1bb764f6d3b. Since this story was published, Buttar's channel on YouTube has reached 500.000 subscribers: https://www.youtube.com/user/drbuttar (last visited Feb 2, 2021).

[46] Political parties are responsible for 86% of ads, but only 48% of the spending. See Juan Carlos Medina Serrano et al., supra note 23 at 118.

[47] The testimony of a nurse from Arizona is compelling: "'We are feeling like our community — at least a good section of them — have turned on us,' he said. 'It's just amazing that nurses have gone from the number one trusted profession in the United States to now realizing that there's this subsection of the population that thinks we're trying to kill everyone.'" Jane Lytvynenko, *In 2020, Disinformation Broke The US*, BUZZFEED NEWS (Dec 6, 2020), https://www.buzzfeednews.com/article/janelytvynenko/disinformation-broke-us.

[48] The proposal is mainly directed at Facebook and Google, for the reasons already laid out in previous sections, though other large digital platforms with significant advertising revenue arguably should also be subjected to the same rules. "DPs are quickly controlling news distribution. Facebook is now the second largest news provider in terms of attention share in the United States. In the UK, Facebook is third, Google is fifth, and Twitter is tenth. By curating the news viewers receive, DPs have effectively appropriated the role that newspaper editors used to have in influencing readers' attention. This poses a concentration issue, as thousands of different viewpoints have now been replaced by a duopoly." Luigi Zingales & Filippo Maria Lancieri, *Stigler Committee on Digital Platforms: Policy Brief*, George J. Stigler Center for the Study of the

---

standards, including metadata regarding the URL, content, identity of the advertiser, amount spent, and all audience profile targeting options selected by the advertiser. The set of data about each ad and the format accessibility requirements in the proposal create obligations that go beyond current voluntary ad libraries. Such increased transparency might not be necessary in all cases, but this proposal targets a specific group of advertised content with a strong indication of abuse: content that has been fact-checked by accredited fact-checking organizations. The libraries should be designed with an eye towards ease of use, particularly by sophisticated users, including ensuring that the material is published in open and machine-readable formats.

The metadata about each ad should include, at a minimum, the URL, content, identity of the advertiser, amount spent, and all audience profile targeting options selected by the advertiser. The full repository should be available for download, and ideally there should also be an option to select data according to a particular timeframe. Access should not be curtailed by offering only the results of limited searches, or the visualizations presented in interactive dashboards. These features are useful to certain categories of readers, but they should be offered as an alternative to downloading the raw data. It is crucial for experienced users to be able to easily obtain the entirety of the relevant data they are looking for.

The most novel element of this proposal is the criterion to allow for the ad library to be cross-referenced against the results of fact-checks. In order to serve as a robust tool against misinformation, ad libraries should be sorted according to "assessments of the validity of claims made by public officials and institutions with an explicit attempt to identify whether a claim is factual".[49]

To be clear, the point is not that all promoted content should be subsequently fact-checked. That would be prohibitively costly and overbroad. Rather, platforms are to cross-reference their databases of promoted content, on one side, and databases of fact-checked content, on the other, such that it is possible to produce a new open dataset containing all promoted content that has been fact-checked. While there is some cost to sorting promoted content in this manner, this is a reasonable burden for the digital platforms to bear, given their own role in the problem.

Economy and the State (Sep 2019), https://www.chicagobooth.edu/-/media/research/stigler/pdfs/policy-brief---digital-platforms---stigler-center.pdf, at 5.
[49] Nathan Walter et al., *Fact-Checking: A Meta-Analysis of What Works and for Whom*, 37(3) POLITICAL COMMUNICATION 1 (2019), at 2.

The point of this proposal is less about increasing the chance that people exposed to false ads are also inoculated with a fact-check, than it is about providing increased public scrutiny over the misinformation ecosystem. Nonetheless, the tested effects of fact-checking misinformation do matter. Even before 2016, scientific research already pointed to fact-checking's ability to sway the opinion of the reader.[50] Politically-savvier people have an increased chance of being influenced by fact-checking results, though fact-checking initiatives can also resonate with less politically sophisticated readers.[51] Indeed, fact-checking results can "positively affect beliefs, irrespective of political ideology, preexisting positions, context (campaign vs. routine), and whether it refutes the entire false statement or just parts of a statement", though this is less effective during elections.[52]

Just as algorithmic moderation obfuscates what news stories are actually shown on recommended feeds and popularly read, frustrating accountability for social media platforms,[53] programmatic advertising hides where a company has placed an ad, which prevents public accountability against both the publisher and the advertiser.[54] Boosting the clarity and accessibility of ad libraries would not only convey useful information to the public at large, but also to the companies paying for the ads in the first place, giving them back awareness of where their ads appear, and thus some of the control that programmatic advertising has taken away.

Sleeping Giants, a global activism organization which works to pressure advertisers who support problematic content, currently carries out an effective strategy based on manually taking screenshots of ads displayed on particular websites. With these enhanced ad libraries, digital platforms would publish this material automatically, allowing campaigns against fact-checking misinformation to be effectively scaled. Civil society organizations, academia, and the press would

---

[50] Kim Fridkin, Patrick J. Kenney & Amanda Wintersieck, *Liar, Liar, Pants on Fire: How Fact-Checking Influences Citizens' Reactions to Negative Advertising*, 32(1) POLITICAL COMMUNICATION 127 (2015), at 139.
[51] Id at 145.
[52] Nathan Walter et al., supra note 50 at 17-18.
[53] Zeynep Tufekci, *Algorithmic Harms Beyond Facebook and Google: Emergent Challenges of Computational Agency*, 13 COLORADO TECHNOLOGY LAW REVIEW 203 (2015).
[54] Joshua A. Braun & Jessica L. Eklund, supra note 17 at 11. See also Adam J. Mills, Christine Pitt & Sarah Lord Ferguson, supra note 7 at 5: "Programmatic advertising is highly efficient, extremely cost-effective, virtually immediate, and designed almost completely around reaching the individual target consumer. To target in this manner, however, prioritizing consumers over con-text, advertisers must cede almost all control over advertisement placement to the intermediaries."

have access to much more relevant material to engage with the public discourse around misinformation. To be sure, "activism around programmatic advertising is necessarily complex and therefore challenging".[55] However, public perceptions towards the brands that inadvertently fund misinformation are a key piece of the puzzle, as Sleeping Giants' work shows.[56] Evidence demonstrates that a connection with fake news adversely impacts the public image of companies, by causing the reader to discredit the outlet. Lower levels of source credibility in turn negatively impact the reader's opinion of the advertised brand.[57] A "new consumer movement is rising" around this exposure dynamic,[58] providing traction to traditional regulatory tools such as transparency, in combination with market effects such as brand perception and corporate reputation management.

Implementation

This proposal is a variation of an initiative that Facebook, Google and Twitter have already taken on, which suggests that it should not be excessively costly or challenging to implement.[59] As a regulatory response, it does not deny consumers any service or feature, and as such does not stifle innovation.[60] Facebook, Google and Twitter each offer some of the components described above, but never all of them together. However, while ad libraries already exist in various forms, a comprehensive version, combined with international fact-checking databases, would be a major step forward in terms of supporting critical work by civil society and journalists in shedding light on how misinformation spreads. It should be clear,

---

[55] Joshua A. Braun, John D. Coakley & Emily West, supra note 11 at 70.

[56] "In this vein, researchers should go beyond merely investigating the direct effect of the ad and address the broader set of causal relationships involved in the formation of brand attitude and purchase intentions." Marco Visentin, Gabriele Pizzi & Marco Pichierri, *Fake News, Real Problems for Brands: The Impact of Content Truthfulness and Source Credibility on consumers' Behavioral Intentions toward the Advertised Brands*, 45 JOURNAL OF INTERACTIVE MARKETING 99 (2019), at 100.

[57] Id at 106.

[58] "(⋯) a new consumer movement is rising, and activists believe that where votes failed, wallets may prevail. This struggle is about much more than ads on Breitbart News — it's about using corporations as shields to protect vulnerable people from bullying and hate crimes." Pagan Kennedy, *How to Destroy the Business Model of Breitbart and Fake News*, THE NEW YORK TIMES (Jan 7, 2017), https://www.nytimes.com/2017/01/07/opinion/sunday/how-to-destroy-the-business-model-of-breitbart-and-fake-news.html.

[59] Facebook went so far as boasting the ad library to be one of its main measures against misinformation. See Sheryl Sandberg, *Hearing Before the United States Senate Select Committee on Intelligence*, UNITED STATES SENATE (Sep 5, 2018), https://www.intelligence.senate.gov/sites/default/files/documents/os-ssandberg-090518.pdf, at 5.

[60] Mark D. Fenwick, Wulf A. Kaal & Erik P.M. Vermeulen, *Regulation Tomorrow: What Happens When Technology Is Faster than the Law?* 6(3) AMERICAN UNIVERSITY BUSINESS LAW REVIEW 561 (2017), at 587.

---

however, that this proposal does not in any way entail information currently published in ad libraries to be removed if it does not relate to any fact-checking. This is about mandating the publication of additional data in a specific format, not about erasing any information already in the public realm.

Abby Wood and Ann Ravel have argued that the availability of open ad lists would enable citizens to know when they are targeted by "divide and conquer"-style exploitative misinformation schemes, perpetrated through programmatic advertisements, while also preventing ads from being used to squash counter speech.[61] Likewise, enhancing transparency in the advertising space would support projects like NewsGuard, which works "with the advertising industry to make apolitical, transparent trust ratings for news sites available to advertisers on large ad exchanges".[62] One of the key challenges to addressing exploitative misinformation is that we do not know what we do not know.[63] Ideally, empirical scholarship on platform practices should inform smart regulation. As it stands, the law often empowers companies to prosecute researchers for the retrieval and use of publicly available data on their digital platforms.[64]

The most technically challenging aspect of this proposal is that platforms would have to decide whether the vast the amounts of promoted content which they host have been fact-checked. While it is easy enough to create a list of URLs which have been specifically subject to fact-checks, carrying out an analysis as to whether this content appears elsewhere is more challenging. Carrying out this task with regard to videos or images adds yet another layer of complexity.

However, in considering the technical challenges underlying this problem, it is worth bearing in mind that digital platforms have years of experience in the automated identification of specific types of content. Implementing the proposal, either in compliance with legislation or on a voluntary basis, would require an

---

[61] Abby K. Wood & Ann M. Ravel, supra note 19 at 1259-1260.

[62] Matt Skibinski, *Misinformation won't stop unless we stop it* (Predictions for Journalism 2021, NiemanLab, 2020), https://www.niemanlab.org/2020/12/misinformation-wont-stop-unless-we-stop-it/

[63] "If we do not know whether social media has overall positive or negative effects on our polity, we have to blame the DPs themselves. All the data they generate is proprietary and they deny outside, independent researchers access to almost all of it. In doing so, they also prevent our societies from comprehending their true impacts." Luigi Zingales & Filippo Maria Lancieri, supra note 49 at 10.

[64] Mandating the publication of expanded ad libraries would solve, for the purposes of research about exploitative misinformation, the problem aptly spotlighted by Thomas E. Kadri, *Platforms as Blackacres*, 68 UCLA L. REV. 1 (forthcoming 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3742086.

organizing effort that is automatic, transparent, ex post and centralized.[65] Artificial intelligence is already a common tool in content moderation, and techniques such as the development of adversarial networks have greatly increased the accuracy of automated decisions.[66] Google, for instance, notoriously uses AI in its Content ID feature to make moderation decisions. While there are significant criticisms with how Content ID has been implemented to *remove user-generated content*,[67] these concerns do not apply in the same way with regards to automated decisions for ad libraries. While false positives generated under this proposal are concerning insofar as they could undermine the efficacy of the ad library as a tool for transparency, they do not raise the same freedom of expression concerns as the existing use of Content ID.

Another more complex element is deciding what repositories of fact-checks to use. If the platforms refused to implement this proposal voluntarily, devising a regulatory structure which defined the scope of an "acceptable" fact-check could be both technically challenging and, in some instances, engage difficult constitutional questions. Any regulatory requirement would have to be precise enough to enable enforcement, but the task of defining the scope of the repository would ideally be delegated to an independent agency. There are numerous fact-checking agencies operating in different countries,[68] under varying institutional governance systems, such as the "newsroom model" and the "NGO model", with varying levels of source transparency.[69]

One starting point could be to defer to the International Fact-Checking Network, and in particular to reference the signatories to their code of principles.[70] The list currently has 29 members, including in South Africa, France, Spain, the United States, Brazil, Serbia, Turkey, and Bosnia & Herzegovina. The Reporter's Lab at Duke University manages a separate database of over 300 active fact-checking

---

[65] James Grimmelmann, *The Virtues of Moderation*, 17(1) YALE JOURNAL OF LAW AND TECHNOLOGY 42 (2015).
[66] Cambridge Consultants, *Use of AI in Online Content Moderation*, OFCOM (2019), https://www.ofcom.org.uk/__data/assets/pdf_file/0028/157249/cambridge-consultants-ai-content-moderation.pdf, at 74.
[67] See Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016).
[68] Fact-checking agencies are more likely to be found in countries with "increased democratic governance and Internet accessibility". Michelle A Amazeen, *Journalistic interventions: The structural factors affecting the global emergence of fact-checking*, 21(1) JOURNALISM 95 (2020), at 103.
[69] See Edda Humprecht, *How Do They Debunk "Fake News"? A Cross- National Comparison of Transparency in Fact-checks*, 8(3) DIGITAL JOURNALISM 310 (2020).
[70] International Fact-Checking Network, *International Fact-Checking Network fact-checkers' code of principles*, https://www.poynter.org/ifcn-fact-checkers-code-of-principles/ (last visited Feb 3, 2021).

---

institutions worldwide that have met certain criteria, including transparency of sources, disclosure of affiliations, and an even-handed approach to checking statements from different political parties.[71] These lists could be a starting point for determining the parameters of the ad libraries.

V. Conclusion

Programmatic advertising has both removed the transparency of ad placement and restricted advertisers' choices in deciding who to support. As microtargeting selects based on a specific set of eyeballs instead of the location of ads, publisher reputation is no longer an element that directly influences which ad is shown where, with a detrimental impact on the commercial value of journalistic standards. This has drained resources from reputable news media, directing a continuous flow of opaque funds to hoax websites.

Transparency constitutes a structure-oriented measure for fighting misinformation, whereby regulation creates increased disclosure obligations that uncover the money trail of for-profit fake news. Instead of command-and-control regulatory responses that invite more platform censorship in order to fight abusive content while denying civil society a meaningful role,[72] this is a solution that looks to transparency obligations as a procedural guarantee.[73] In recent years, digital platforms have created repositories of political ads to try and remedy the problem. However, these initiatives are fundamentally flawed in that their visibility criteria are skewed towards election-related advertising, and they fail to comprehensively illuminate the funding which drives the production of misinformation.

Platforms should include information in their ad libraries about the promotion and financial subsidy of any content that has been fact-checked, regardless of whether these companies believe it to be "political". This is not a revolutionary measure that would entail major new costs, but rather a logical next step in a policy that these

---

[71] Bill Adair & Mark Stencel, *How We Identify Fact-Checkers*, REPORTER'S LAB (June 22, 2016), https://reporterslab.org/how-we-identify-fact-checkers/.

[72] Robert Gorwa, *The platform governance triangle: conceptualising the informal regulation of online content*, 8(2) INTERNET POLICY REVIEW 1 (2019), at 14.

[73] Barbora Bukovská, *The European Commission's Code of Conduct for Countering Illegal Hate Speech Online. An analysis of freedom of expression implications* (The Transatlantic Working Group Papers Series, May 7th, 2019), https://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/EC_Code_of_Conduct_TWG_Bukovska_May_2019.pdf, at 6.

digital platforms have already acknowledged as warranted and started to gradually implement.

In implementing this regulatory response, governments should recognize the central role played by fact-checking agencies in a healthy public sphere and avoid much harsher, complex and speech-restrictive measures that would target misinformation content for removal by granting digital platforms undue power and responsibilities over the shape of online speech.[74] Enhanced ad libraries are a natural response as they fix an obscurity problem with disclosure, realigning advertiser and publisher incentives without trending towards problematic forms of censorship.

---

[74] Ivar Hartmann, *A new framework for online content moderation*, 36 COMPUTER LAW & SECURITY REVIEW 1 (2020).

Source Triangulation Skills and the Future of Digital Inclusion: How Information Literacy Policy Can Address Misinformation and Disinformation Challenges

Jonathan A. Obar, PhD, York University

I. Introduction

As we navigate treacherous information environments finding answers to questions about politics, healthcare, nutrition, and beyond, the ability to distinguish fact from fallacy and even fake is vital. To help people develop this ability, policy responses to misinformation and disinformation challenges should promote equitable, distributive supports[1] to encourage source triangulation. The source triangulation concept[2] refers to information validation through access and assessment of multiple sources with the aim of finding credible commonality.[3] This is what Google nudges us to do when posting links to "more resources" adjacent to

---

[1] Instead of a "one-size-fits-all" Utilitarian approach to policy, Rawlsian/Senian distributional justice approaches to addressing societal inequities should be the focus. As Amit M. Schejter and Noam Tirosh explain, "In a Rawlsian solution the goal of the policy would be bringing the disconnected to connectivity before attending to any other betterment of the situation of the connected, or the interests of incumbent service providers. [⋯] A Senian interpretation would lead us to develop policies that ensure people actually communicate while keeping their social presence high in a rich media environment; moving from simple access to usage capabilities. Senian solutions will thus require teaching individuals to use the media and ensuring they can express themselves over them now that they can access them." Amit M. Schejter and Noam Tirosh, *"Seek the Meek, Seek the Just": Social Media and Social Justice*, 39 TELECOMMUNICATIONS POLICY 796 (2015). *See also* AMIT M. SCHEJTER and NOAM TIROSH, A JUSTICE-BASED APPROACH FOR NEW MEDIA POLICY: IN THE PATHS OF RIGHTEOUSNESS (2016).

[2] The concept is adapted here from the literature on methods of academic inquiry, and is often associated with academic or scientific research that assesses and compares data via different research methodologies and perspectives, *see* NORMAN K. DENZIN, THE RESEARCH ACT: A THEORETICAL INTRODUCTION TO SOCIOLOGICAL METHODS (2017); Michael Quinn Patton, *Enhancing the Quality and Credibility of Qualitative Analysis*, 34 HEALTH SERVICES RESEARCH 1189 (1999). The concept is also adapted from examples in the literature that discuss how to determine credibility from the evaluation of online sources, e.g. Marc Meola, *Chucking the Checklist: A Contextual Approach to Teaching Undergraduates Web-Site Evaluation,* 4 LIBRARIES AND THE ACADEMY 331 (2004); Miriam J. Metzger, Making Sense of Credibility on the Web: Models for Evaluating Online Information and Recommendations for Future Research, 58 JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY 2078 (2007). See also: Nicole A. Cooke, *Posttruth, Truthiness, and Alternative Facts: Information Behavior and Critical Information Consumption for a New Age,* 87 THE LIBRARY QUARTERLY 211 (2017).

[3] *See* Patton *supra* note 2.

YouTube videos, or what Twitter invites us to consider by adding links to alternate perspectives (see Figures 1 and 2). As an individual learns source triangulation, they develop one component of an information literacy education. Information literacy is a broader, multi-layered concept which includes various forms of knowledge and skill for distinguishing helpful from unhelpful information.[4] Source triangulation supports could complement platform-down content moderation efforts,[5] concurrently enhancing user agency[6] from the bottom-up.

Figure 1. Image of YouTube Video with Links to Supplementary Sources
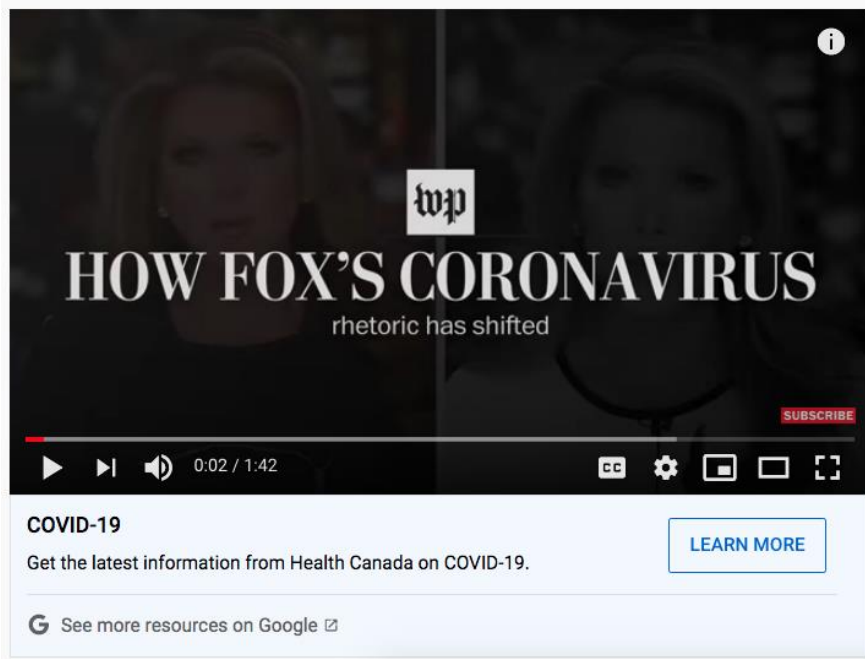
Note. Uploaded by the Washington Post on March 17, 2020,
https://www.youtube.com/watch?v=ifKbwDf51bA&t=1s

---

[4] *See* Michael B. Eisenberg, *Information Literacy: Essential Skills for the Information Age*, 28 DESIDOC JOURNAL OF LIBRARY & INFORMATION TECHNOLOGY 39

[5] *See* SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOWS OF SOCIAL MEDIA (2019); TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA (2018); Laura DeNardis and Andrea M. Hackl, *Internet Governance by Social Media Platforms*, 39 TELECOMMUNICATIONS POLICY 761 (2015).

[6] Presenting a nuanced definition of individual action, Couldry describes agency as more than "[···] acts (of clicking on this button, pressing 'like' to this post) but [···] the longer processes of action based on reflection, giving an account of what one has done, even more basically, making sense of the world *so as* to act within it." Nick Couldry, *A Necessary Disenchantment: Myth, Agency and Injustice in a Digital World*, 62 THE SOCIOLOGICAL REVIEW 880 (2014). Connections between information literacy and individual agency are present in the academic literature, e.g. James Elmborg, *Critical Information Literacy: Implications for Instructional Practice*, 32 THE JOURNAL OF ACADEMIC LIBRARIANSHIP 192 (2006); Cooke *supra* note 2.

Figure 2. Two Images from Twitter with Supplementary Sources

Note. The first tweet was posted to Twitter on November 4, 2020, the second on November 7, 2020.

Any new proposal connecting technology and education must acknowledge that U.S. policy tends to focus primarily on closing first-level digital divides (e.g. computer and internet access).[7] While funding must continue to address access, this proposal asserts that it is time to also address what the United Nations calls "meaningful universal connectivity."[8] As the Secretary-General of the International Telecommunications Union explains, "digital inclusion can only be meaningful and effective if and when Internet users feel empowered to use the technology."[9]

---

[7] Eszter Hargittai, The Digital Divide and What to do About it, in NEW ECONOMY HANDBOOK (2003).
[8] UNESCO. New Report on Global Broadband Access Underscores Urgent Need to Reach the Half of the World Still Unconnected (2019 September 23), https://en.unesco.org/news/new-report-global-broadband-access-underscores-urgent-need-reach-half-world-still-unconnected.
[9] *Id.*

While the context of these comments is second-level digital divides[10]outside of the United States, some American digital inclusion initiatives advance similar goals. For example, the National Digital Inclusion Alliance (NDIA) connects digital inclusion[11] and digital equity[12] with digital literacy.[13] Access and use of technologies are central to NDIA's work, suggesting digital inclusion "requires intentional strategies and investments to reduce and eliminate historical, institutional and structural barriers."[14] One attempt at addressing these barriers is The Digital Equity Act of 2019.[15] With direction from NDIA's definitions for digital inclusion and digital equity,[16] the Act proposes $1.25 billion in funding for two Federal programs facilitating grants through the National Telecommunications and Information

---

[10] The second-level digital divide definition refers to differences between individual ability in terms of internet use. *See* Eszter Hargittai, *Second-Level Digital Divide: Differences in People's Online Skills*, 7 FIRST MONDAY (2002).

[11] Digital inclusion definition "Digital Inclusion refers to the activities necessary to ensure that all individuals and communities, including the most disadvantaged, have access to and use of Information and Communication Technologies (ICTs). This includes 5 elements: 1) affordable, robust broadband internet service; 2) internet-enabled devices that meet the needs of the user; 3) access to digital literacy training; 4) quality technical support; and 5) applications and online content designed to enable and encourage self-sufficiency, participation and collaboration. Digital Inclusion must evolve as technology advances. Digital Inclusion requires intentional strategies and investments to reduce and eliminate historical, institutional and structural barriers to access and use technology." National Digital Inclusion Alliance. Definitions (n.d.), https://www.digitalinclusion.org/definitions/.

[12] Digital equity definition: "Digital Equity is a condition in which all individuals and communities have the information technology capacity needed for full participation in our society, democracy and economy. Digital Equity is necessary for civic and cultural participation, employment, lifelong learning, and access to essential services." *Id.*

[13] Digital literacy often refers to the technical skills for computer use, though newer conceptualizations include additions such as information literacy. For example, this is the NDIA definition of digital literacy: "NDIA recommends the American Library Association's definition of Digital Literacy via their Digital Literacy Taskforce: "*Digital Literacy is the ability to use information and communication technologies to find, evaluate, create, and communicate information, requiring both cognitive and technical skills.*" A Digitally Literate Person: Possesses the variety of skills – technical and cognitive – required to find, understand, evaluate, create, and communicate digital information in a wide variety of formats; Is able to use diverse technologies appropriately and effectively to retrieve information, interpret results, and judge the quality of that information; Understands the relationship between technology, life-long learning, personal privacy, and stewardship of information; Uses these skills and the appropriate technology to communicate and collaborate with peers, colleagues, family, and on occasion, the general public; and Uses these skills to actively participate in civic society and contribute to a vibrant, informed, and engaged community." *Id.*

[14] *Id.*

[15] *See* The Digital Equity Act of 2019, https://www.digitalequityact.org/.

[16] *See* notes 11 and 12.

Administration "to promote digital equity nationwide."[17] Indeed, coupling connectivity and use initiatives is essential as misinformation and disinformation challenges proliferate, and as ongoing research reveals that the state of individual ability to address these challenges is "bleak."[18]

This paper offers two policy recommendations:

1) U.S. federal government agencies addressing technology and education questions should develop an expansive vision of digital inclusion that includes the use of technologies, with a focus on information literacy initiatives in general, and source triangulation outcomes in particular; and

2) The work of the National Digital Inclusion Alliance should be complemented by funding efforts to realize information literacy skills via schools, libraries, and additional non/for-profit platform and infomediation initiatives to address misinformation and disinformation challenges.

These efforts should follow a distributive justice approach,[19] with the aim of achieving an equitable distribution of resources that acknowledges the challenges unique to marginalized and vulnerable communities.[20] In particular, we must address gender-based divides such as those UNESCO identifies,[21] struggles unique to individuals whose internet experiences are mobile-only or mobile-mostly,[22] and growing concerns about Big Data discrimination.[23] It is essential that

---

[17] As noted on www.digitalequityact.org: "One program would be carried out through state governments, with funding allocated by formula, and would incorporate state-by-state digital equity planning followed by implementation grants to qualifying programs. […] The other would be an annual national competitive grant program, run by the NTIA, to support digital equity projects undertaken by individual groups, coalitions, and/or communities of interest anywhere in the U.S." *Supra* note 15.

[18] Sam Wineburg, Sarah McGrew, Joel Breakstone, and Teresa Ortega. Evaluating Information: The Cornerstone of Civic Online Reasoning (2016), http://purl.stanford.edu/fv751yt5934 at 4.

[19] *See* Schejter and Tirosh, *supra* note 1.

[20] *See* Alice E. Marwick and danah boyd, *Understanding Privacy at the Margins*, 12 INTERNATIONAL JOURNAL OF COMMUNICATION 1157 (2018); SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018).

[21] *Supra* note 8; UNESCO. I'd Blush if I Could: Closing Gender Divides in Digital Skills Through Education (2019), https://en.unesco.org/Id-blush-if-I-could.

[22] Philip M. Napoli and Jonathan A. Obar, *The Emerging Mobile Internet Underclass: A Critique of Mobile Internet Access*, 30 THE INFORMATION SOCIETY 323 (2014).

[23] *See* Solon Barocas and Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIFORNIA LAW REVIEW 104 (2016); Marwick and boyd, *supra* note 20; Noble, *supra* note 20; RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); Jonathan A. Obar and Brenda McPhail, Preventing

---

misinformation and disinformation challenges be understood as including, for example, biased algorithms, data sets, and resulting eligibility determinations that can contribute to Big Data discrimination. Recognizing the extent and the variations is important if we are to stop the damage. We must address the threat mosaic of deteriorating information environments, especially for those most likely to be victimized. The consequences of inaction may be residence in online and offline worlds that are both dangerous and treacherous, where we cannot navigate, understand, or govern.

The next section briefly describes findings about the state of online information literacy in the United States. To begin to answer the information literacy challenge, a conceptualization of source triangulation is presented, and a set of policy proposals for future information literacy efforts in the United States follows. The conclusion suggests that these proposals might serve as a model for efforts internationally.

II. The Lack of Online Information Literacy in the United States

The state of online information literacy in the United States appears problematic. Assessments of the extent of the problem include studies from Stanford University's "Civic Online Reasoning"[24] project (a concept synonymous with online information literacy). In a 2016 report,[25] the group summarizes various assessments of thousands of middle-school, high school and university students at both poorly and well-resourced schools across a variety of U.S. states. A primary focus was the extent to which students could distinguish between helpful and unhelpful information online, check sources, as well as identify and compare evidence to test for credibility. Results suggest that "[o]verall, young people's ability to reason about the information on the Internet can be summed up in one word: bleak."[26] A follow-up study in 2019[27] of 3,446 students, "a national sample that matches the demographic profile of high school students in the United States,"[28] similarly suggests that despite policy efforts put in place since the first study, the results of

---

Big Data Discrimination in Canada: Addressing Design, Consent and Sovereignty Challenges. Data Governance in the Digital Age: Special Report (2018).

[24] *Supra* note 18; Stanford History Education Group, Civic Online Reasoning (February 24, 2021), https://cor.stanford.edu/.

[25] *Id.*

[26] *Id.* at 4.

[27] Joel Breakstone, Mark Smith, Sam Wineburg, Amie Rapaport, Jill Carle, Marshall Garland, and Anna Saavedra. Students' Civic Online Reasoning: A National Portrait (2019), https://purl.stanford.edu/gf151tb4868.

[28] *Id.* at 3.

the second study are "troubling." The authors summarize that "[n]early all students floundered [⋯] Education moves slowly. Technology doesn't. If we don't act with urgency, our students' ability to engage in civic life will be the casualty."[29]

PEW research suggests these concerns extend to older individuals as well. A recent report found that when classifying five factual statements that were "news-related" and five opinion statements, only 26 percent of more than 5,000 participants could correctly identify the factual statements, and 35 percent for the opinions.[30] More concerning is what PEW calls a "democratic deficit,"[31] suggesting the struggle "extends to a growing and troubling public distrust in each other's ability to make informed decisions about democratic leadership."[32] Beyond the American context, UNESCO emphasizes that "[t]oday billions of people have access to affordable devices and broadband networks, but do not have the requisite skills to take advantage of this technology to improve their lives."[33]

The online information literacy challenge is not only about the ability to read and understand content. Analyzing, critiquing and comparing content are initial steps, but more is necessary. For robust information literacy, we must also understand information environments where filter bubbles,[34] echo chambers,[35] bots,[36] trolls,[37]

---

[29] *Id.* at 3.

[30] Amy Mitchell, Jeffrey Gottfried, Michael Barthel, and Nami Sumida, Distinguishing Between Factual and Opinion Statements in the News (2018 June 18), https://www.journalism.org/2018/06/18/distinguishing-between-factual-and-opinion-statements-in-the-news/.

[31] Michael Dimock, How Americans View Trust, Facts, and Democracy Today (2020 February 19), https://www.pewtrusts.org/en/trust/archive/winter-2020/how-americans-view-trust-facts-and-democracy-today.

[32] *Id.*

[33] UNESCO *supra* note 21 at 17.

[34] *See* ELI PARISER, THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK (2011).

[35] *See* KATHLEEN HALL JAMIESON AND JOSEPH N. CAPPELLA. ECHO CHAMBER: RUSH LIMBAUGH AND THE CONSERVATIVE MEDIA ESTABLISHMENT (2008); Elizabeth Dubois, and Grant Blank. *The Echo Chamber is Overstated: The Moderating Effect of Political Interest and Diverse Media*, 21 INFORMATION, COMMUNICATION & SOCIETY 729 (2018).

[36] *See* Fenwick McKelvey and Elizabeth Dubois. Computational Propaganda in Canada: The Use of Political Bots (2017), http://oxis.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Canada.pdf.

[37] *See* KATHLEEN HALL JAMIESON, CYBERWAR: HOW RUSSIAN HACKERS AND TROLLS HELPED ELECT A PRESIDENT: WHAT WE DON'T, CAN'T, AND DO KNOW (2020).

and our own confirmation bias[38] introduce challenges. Beyond this is also the need to understand the policy, political economy, and infrastructure that contributes to the evolution of the internet.[39] Thus, while the policy recommendations herein suggest the need for source triangulation skills, to achieve a thorough online information literacy, multiple layers to the necessary education are vital if we are to achieve digital inclusion and "meaningful universal connectivity."[40]

III. Source Triangulation as a Learning Outcome

There seems to be comfort in the efficiency of accessing a single source of information we feel we can trust.[41] One problem with blindly trusting a single source however, is that its information may be questionable and require checking. If a problematic source provides material that is comforting, or reinforces prior beliefs, it can be challenging to break from it.[42]

Instead of learning outcomes that result from reliance on a single source, the ability to find, evaluate, and draw accurate information by "comparison and corroboration"[43] of multiple sources should be a primary focus of educational investment.[44] Source triangulation is one aspect of information literacy development which involves learning to validate sources by conducting research and comparing multiple sources to find a commonality that is credible.[45] The term "credible" is important, as credibility is context-specific. What is credible in a corporate context differs from what is credible in a critical studies context, which connects us to one of the oldest philosophical and political economic debates about

---

[38] Defined as: "(when people) view information consistent with their preexisting beliefs as more persuasive than dissonant information," David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. *The Science of Fake News,* 359 SCIENCE 1094 (2018) at 1095.

[39] *See* Leslie Regan Shade and Tamara Shepherd, *Viewing Youth and Mobile Privacy Through a Digital Policy Literacy Framework* 18 FIRST MONDAY (2013).

[40] *Supra* note 8.

[41] Accepting information as accurate based on feelings is associated with confirmation bias, Stephen Colbert's "truthiness," and even descriptions of a "posttruth era." Cooke *supra* note 2. For example, "It is now being said that we live in a posttruth era—an era in which audiences are more likely to believe information that appeals to emotions or existing personal beliefs, as opposed to seeking and readily accepting information regarded as factual or objective." Cooke *supra* note 2 at 212. For a definition of confirmation bias see note 38.

[42] Individuals applying critical theory to the study of the media have historically raised similar concerns. E.g. THEODOR W. ADORNO, INTRODUCTION TO THE SOCIOLOGY OF MUSIC (ASHTON, E. B., TRANS.) (1976).

[43] Meola *supra* note 2 at 332.

[44] *See* Metzger *supra* note 2.

[45] *See* Patton *supra* note 2.

---

manufactured consent,[46] mediums as messages,[47] "pictures in our heads,"[48] and shadows on cave walls.[49] A necessary caveat must clarify that truth will not necessarily emerge via commonality alone. Many important truths are with sources, some with access to the marketplace of ideas, some without, that are not understood or known broadly. Warnings about echo chambers are also important, which is why source triangulation skill is just one aspect of broader information literacy training necessary for navigating treacherous online information environments.

Despite the challenges with credible commonality, source triangulation is consistently viewed by information literacy experts and advocates as a primary strategy for helping users find the information necessary for democracy to function.[50] The triangulation concept draws from the practice of surveying land.[51] As Patton states:

> Knowing a single landmark only locates you somewhere along a line in a direction from the landmark, whereas with two landmarks you can *take bearings* in two directions and locate yourself at their intersection. The notion of triangulating also works metaphorically to call to mind the world's strongest geometric shape-the triangle.[52] (emphasis added)

Central to determining information-strength via triangulation is the challenge that single sources are often easily refutable by alternative explanations.[53] Critical analyses of multiple sources, and checking for consistencies across them, supports

---

[46] EDWARD S. HERMAN and NOAM CHOMSKY, MANUFACTURING CONSENT: THE POLITICAL ECONOMY OF THE MASS MEDIA (1988).

[47] MARSHALL MCLUHAN, UNDERSTANDING MEDIA: THE EXTENSIONS OF MAN (1964).

[48] WALTER LIPPMANN, PUBLIC OPINION (1922/1997) at 3.

[49] PLATO, THE REPUBLIC (380 BC).

[50] While information validation through assessment of multiple sources is a common strategy proposed, it is often presented via different concepts and frameworks. For example, the CRAAP test, Sarah Blakeslee, Evaluating Information - Applying the CRAAP Test (2010 September 17) https://library.csuchico.edu/sites/default/files/craap-test.pdf; "Civic Online Reasoning" *supra* note 18; Stanford History Education Group *supra* note 24; *see* International Federation of Library Associations and Institutions (IFLA), How to Spot Fake News (n.d.) https://commons.wikimedia.org/wiki/File:How_to_Spot_Fake_News.jpg, the IFLA notes that the infographic draws from Eugene Kiely and Lori Robertson, How to Spot Fake News, FactCheck.org, November 18, 2016, https://www.factcheck.org/2016/11/how-to-spot-fake-news/; Cooke *supra* note 2. See also Meola *supra* note 2.

[51] *See* Patton *supra* note 2.

[52] *Id*. at 1192.

[53] *Id*.

---

individuals as they attempt to determine credible commonality.[54] It is also important to identify facts and perspectives that may be missing from the consideration of a single source. The intention is both to help people identify "information problems"[55] while also learning to "take charge of the information allowed into their minds and decision-making schemas."[56]

One place to begin when teaching source triangulation is with Stanford University's "Civic Online Reasoning" project.[57] In addition to the curriculum available on the project platform,[58] published studies identify skills similar to source triangulation. For example, upon arriving at a piece of information online, the research recommends first to "take bearings"[59] (note the language consistent with Patton above[60]) or the process of taking time to determine a plan for evaluating the information available. Rushing to trust a single source is discouraged. A similar skill is "click restraint,"[61] relevant to evaluating search results or social media feeds, where individuals should begin to assess information before clicking, considering, for example, differences associated with the curation of results.[62] These two skills emphasize that information literacy efforts should help individuals learn to evaluate information sources as well as the platforms that curate access to those sources. These initial steps prepare the individual to then attempt a third, which is similar to source triangulation - "lateral reading."[63] As the authors note: "When reading laterally, one leaves a website and opens new tabs along the browser's horizontal axis, drawing on the resources of the Internet to learn more about a site and its claims."[64] Remaining on the site and only evaluating its content to determine credibility would be "reading vertically,"[65] and introduces the risks associated with

---

[54] Patton *supra* note 2; Meola *supra* note 2; Devon Greyson, *Information Triangulation: A Complex and Agentic Everyday Information Practice*, 69 JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY 869.
[55] Eisenberg, *supra* note 4 at 39.
[56] Cooke, *supra* note 2 at 216.
[57] Stanford History Education Group *supra* note 24.
[58] *Id.*
[59] Sam Wineburg, and Sarah McGrew, *Lateral Reading and the Nature of Expertise: Reading Less and Learning More When Evaluating Digital Information,* 121 TEACHERS COLLEGE RECORD 1 (2019) at 30.
[60] Patton *supra* note 2.
[61] *Supra* note 59 at 32.
[62] *See* Pariser *supra* note 34.
[63] *Supra* note 59 at 31.
[64] *Id.*
[65] *Id.* at 23.

trusting single sources. In a study of how fact-checkers engaged in lateral reading, "(f)act checkers, in short, learned most about a site by leaving it."[66]

This returns the discussion to the question of how to determine information credibility, especially when attempting to triangulate sources during lateral reading. Indeed, both the pedagogy and the learning outcomes for developing information literacy in the twenty-first century are areas in need of further research. At the same time, there is an abundance of resources (and related terms) available online about addressing misinformation and disinformation, and at different stages of development. Should we teach digital literacy? Media literacy? Metaliteracy? Critical information literacy? This requires educators to have information literacy skills just to navigate the resources for teaching information literacy. The need for further research and resource organization justifies the call for government leadership central to the policy recommendations that follow.

IV. Policy Recommendations

Information literacy and source triangulation skills are not primary goals of U.S. government funding mechanisms connecting technology and education. Funding efforts in this space often focus on closing first-level digital divides—differences in access to higher-quality hardware, software, and internet connections.[67] While access to technology is essential, particularly technology that gets users beyond mobile-only or mobile-mostly experiences,[68] considerable investment in a more "meaningful universal connectivity,"[69] with a specific focus on internet use is vital.

The Obama administration's digitalliteracy.gov project, now offline,[70] suggests government-funded online information literacy programs were at one time possible. The initiative was a collaboration between the National Telecommunications and Information Administration (Department of Commerce), the Department of Education, the Institute of Museum and Library Services, the Federal Communications Commission, and the Department of Housing and Urban

---

[66] *Id.* at 31.

[67] *Supra* note 7.

[68] *Supra* note 22.

[69] *Supra* note 8.

[70] The site was active from 2011-2016. Library of Congress, Web Archive: Digitalliteracy.gov https://www.loc.gov/item/lcwaN0014325/.

Development, among others.[71] While the focus of the Obama administration's efforts appear to have emphasized skills for the digital economy, such as professional development training, the few materials still available suggest programs prioritizing internet use were possible.

Policy responses should now revive and expand these efforts, funding research and implementation of a variety of supports to train and engage people in source triangulation. In particular, this paper recommends: 1) that U.S. federal government agencies already connecting technology and education should develop and promote a vision of digital inclusion that prioritizes internet use through information literacy initiatives such as source triangulation training, and 2) these agencies should allocate funding to schools, libraries, and additional for/non-profit efforts to deliver source triangulation and other information literacy learning outcomes. This should serve both to support as well as to complement the efforts of groups including the National Digital Inclusion Alliance.

The next section addresses how U.S. federal agencies connecting technology with educational opportunities might pursue these recommendations. A brief discussion follows of additional for/non-profit approaches involving platforms and infomediaries.

IV.i Vision and Funding from Federal Agencies for Information Literacy Supports

Libraries and centers for information and library science scholarship already lead efforts promoting online information literacy to address misinformation and disinformation challenges. To enhance and support these efforts in the United States, the federal agency known as the Institute of Museum and Library Services (IMLS) should prioritize source triangulation skills as a part of its vision for libraries across the country and fund efforts to develop and promote that vision. While the agency lists information literacy as a "twenty-first century skill," it is not one of the agency's "priority areas."[72] While the Library Services and Technology Act,[73] which governs the agency, appears to prioritize access, professional development and STEM development are IMLS priority areas, suggesting that

---

[71] Commerce.gov, Fact Sheet: Digital Literacy (2011), https://2010-2014.commerce.gov/news/fact-sheets/2011/05/13/fact-sheet-digital-literacy.html.

[72] Institute of Museum and Library Services, Priority Areas, https://www.imls.gov/our-work/priority-areas.

[73] Library Services and Technology Act, 20 USC 9101, https://www.govinfo.gov/content/pkg/STATUTE-110/pdf/STATUTE-110-Pg3009.pdf.

information literacy initiatives would be consistent with current practice. Beyond a clear vision prioritizing source triangulation, the funding programs the IMLS oversees, such as the Grants to States program which distributes hundreds of millions of dollars to libraries across the U.S., should support these learning outcomes. Such a shift might encourage all State Library Administrative Agencies to update priorities to include online information literacy initiatives.

Another federal agency that should prioritize and fund source triangulation efforts is the U.S. Department of Education's Office of Educational Technology. This agency "develops national educational technology policy and establishes the *vision* for how technology can be used to transform teaching and learning and how to make everywhere, all-the-time learning possible for early learners through K-12, higher education, and adult education."[74] (emphasis added) Source triangulation should be central to that vision. For example, while the 2017 National Education Technology Plan[75] makes one mention of information literacy in passing, source triangulation should be central to plan updates.

This agency also oversees massive funding efforts that focus primarily on closing first-level digital divides. For example, a recent initiative associated with the Coronavirus Aid, Relief, and Economic Security (CARES) Act, involves an allocation of $30.75 billion to support the emergency shift to online and blended forms of learning during the pandemic.[76] While funds should not be diverted from emergency efforts, it is important to emphasize that the Office of Educational Technology, and its various funding mechanisms, could be a place for more robust information literacy programs to emerge.

It is important to clarify that while state and local entities should have the freedom to independently pursue initiatives, ensuring flexibility within federal granting contexts, federal leadership remains vital. Instead of focusing on the specific details of programs or curricula, federal government agencies should prioritize the development of a clear and normative vision, as articulated in 20 U.S.C. 3425 - Office of Educational Technology:

---

[74] U.S. Department of Education, Office of Educational Technology, What We Do (2020) https://tech.ed.gov/what-we-do/.

[75] U.S. Department of Education, Office of Educational Technology, Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update (2020) https://tech.ed.gov/files/2017/01/NETP17.pdf.

[76] U.S. Department of Education, Office of Educational Technology, Funding Digital Learning (2021) https://tech.ed.gov/funding/.

The Director of the Office of Educational Technology [⋯] shall— (1) in support of the overall national technology policy and in consultation with other Federal departments or agencies which the Director determines appropriate, *provide leadership to the Nation in the use of technology* to promote achievement of the National Education Goals and to increase opportunities for all students to achieve State content and challenging State student performance standards.[77] (emphasis added)

The Office of Education Technology's website explains that the details of the National Education Technology Plan "align to the Activities to Support the Effective Use of Technology (Title IV A) of Every Student Succeeds Act as authorized by Congress in December 2015."[78] This section of the Act, emphasizing "21st Century Schools," states with regard to the allocation of funds:

ACTIVITIES TO SUPPORT THE EFFECTIVE USE OF TECHNOLOGY (a) Uses of Funds. --Subject to section 4106(f), each local educational agency, or consortium of such agencies, that receives an allocation under section 4015(a) shall use a portion of such funds to improve the use of technology to improve the academic achievement, academic growth, and digital literacy of all students.[79]

Returning to the list of agencies supportive of President Obama's digital literacy efforts,[80] there are a variety of additional agencies that might provide leadership and funding. One is the Federal Communications Commission (FCC). Via the Universal Service Fund (USF), and specifically, the E-Rate program, which the FCC oversees, eligible libraries and schools receive funds to create and enhance internet connections. While historically the view was that USF-programs should make closing first-level digital divides a priority,[81] there is evidence that FCC efforts to support information literacy might be also possible. In updates to its "Lifeline

---

[77] 20 U.S.C. 3425, Office of Educational Technology (1994), https://www.govinfo.gov/app/details/USCODE-1994-title20/USCODE-1994-title20-chap48-subchapII-sec3425.

[78] Department of Education, Office of Educational Technology, National Education Technology Plan, https://tech.ed.gov/netp/.

[79] Every Student Succeeds Act, 20 USC 6301, https://www.govinfo.gov/content/pkg/PLAW-114publ95/html/PLAW-114publ95.htm.

[80] Whereas the term digital literacy can refer only to technical skills for using computers, broader definitions can include information literacy skills.

[81] Heather E. Hudson, The Future of the E-Rate: US Universal Service Fund Support for Public Access and Social Services, in AND COMMUNICATIONS FOR ALL: A POLICY AGENDA FOR A NEW ADMINISTRATION (2009); Paul T. Jaeger, John Carlo Bertot, Kim M. Thompson, Sarah M. Katz & Elizabeth J. DeCoster, *The Intersection of Public Policy and Public Access: Digital Divides, Digital Literacy, Digital Inclusion, and Public Libraries*, 31 PUBLIC LIBRARY QUARTERLY 1 (2012).

Program for Low-Income Consumers," the FCC prioritized "digital inclusion" making it a "national priority."[82] The updates make specific reference to the efforts of the NDIA, and clarify that with the changes the FCC is starting "an ongoing campaign to build the Commission's digital literacy capacity."[83] The update tasked the Consumer and Governmental Affairs Bureau to develop a plan for achieving various digital inclusion outcomes. The resulting plan,[84] released in 2017, mentions digital literacy repeatedly, making clear that teaching individuals to use the internet must be central to efforts that aim to promote digital inclusion. This suggests that the FCC could make information literacy and source triangulation part of its priorities and funding efforts. It is worth considering how historical interpretations of the public interest, as a measure for determining access to the radio spectrum, for example, might be updated to establish new requirements for organizations operating within this model.[85]

To achieve source triangulation learning outcomes across the country there will need to be investments in the development of pedagogical approaches and accompanying software. Public-private partnerships may be one approach to pursue these efforts. As David L. Cohen, Senior Executive Vice President at Comcast, comments about the FCC's Lifeline updates:

> [W]e will not move the needle on broadband adoption in this country unless there is a robust commitment to digital literacy and relevancy training. [···] Closing the digital divide is a great challenge, but I'm convinced that, by working together, we can get more Americans connected to the amazing and transformative power of the Internet.[86]

It appears the focus on connectivity in federal funding mechanisms is slowly shifting towards programs for teaching people how to use the internet. Federal

---

[82] Benton Institute for Broadband & Society, FCC's Lifeline Reform Makes Digital Inclusion a National Priority (2016, May 10), https://www.benton.org/blog/lifeline-reform-makes-digital-inclusion-national-priority.
[83] Federal Communications Commission, Third Report and Order, Further Report and Order, and Order on Reconsideration, In the Matter of Lifeline and Link Up Reform and Modernization, Telecommunications Carriers Eligible for Universal Service Support, Connect America Fund (2016, March 31), https://docs.fcc.gov/public/attachments/FCC-16-38A1.pdf at 141.
[84] Consumer and Governmental Affairs Bureau, Federal Communications Commission, Strategies and Recommendations for Promoting Digital Inclusion (2017, January 11), https://docs.fcc.gov/public/attachments/DOC-342993A1.pdf.
[85] *See* PHILIP M. NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST: MEDIA REGULATION IN THE DISINFORMATION AGE (2019).
[86] David L. Cohen, Modernizing Lifeline Essential to Closing the Digital Divide (2016 March 8), https://corporate.comcast.com/comcast-voices/modernizing-lifeline-essential-to-closing-the-digital-divide.

agencies should help lead this important work and ensure the vision and funding to address misinformation and disinformation challenges.

IV.ii Information Literacy Supports Via Digital Platforms and Infomediaries

Another approach to supporting information literacy in general and source triangulation in particular is through supplementary platform efforts. As noted earlier, Google and Twitter are providing links to supplementary information next to YouTube videos and tweets.[87] By nudging individuals to "read beyond," and "consider the source,"[88] platforms are beginning to convey that users should learn not to trust single sources and instead consider multiple sources as a default information engagement strategy.

These efforts could expand to include more dynamic engagement and educational opportunities, similar to how experimental forms of dynamic online consent processes are being considered as part of initiatives to help users realize digital privacy protections.[89] Public-private partnerships will likely be necessary, as software, user-interface design, and embedded pedagogy are the subject of further research.

New strategies for achieving information literacy outcomes, as well as generating demand for skill development, should be considered. For example, with funding from the U.S. government, imagine if Jeff Bezos and Jack Dorsey organize a coalition of media companies to develop software and pedagogy for engaging users in source triangulation education. With Bezos leading the initiative, a subscription to *The Washington Post* could be a benefit of user participation. The program could develop a new online platform that would become central to K-12 information literacy curricula, if states chose to participate. One potential concern with this approach is the possibility of unfairly privileging dominant media players, such as

---

[87] David Gilbert, YouTube Thinks Wikipedia can Solve its Fake News Problem. Wikipedia Isn't so Sure, VICE NEWS (14 March 2018).

[88] This corresponds to advice provided by the International Federation of Library Associations and Institutions for how to identify fake news, IFLA *supra* note 50, drawing from the Eugene Kiely and Lori Robertson post, How to Spot Fake News, FactCheck.org (November 18, 2016), https://www.factcheck.org/2016/11/how-to-spot-fake-news/.

[89] *See* Office of the Privacy Commissioner of Canada, Guidelines for Obtaining Meaningful Consent (2018) https://www.priv.gc.ca/en/privacy-topics/collecting-personal-information/consent/gl_omc_201805/.

The Washington Post.[90] Ensuring participation in the coalition for smaller media companies, as well as those representing a diversity of ownership and content might help mitigate this.[91] This potential revitalization of the fourth estate,[92] might combine providing individuals with the information they require to be self-governing in a democracy (central to historical views of journalism[93]), with supports for developing information literacy.

Another possibility is the development and funding of infomediation services.[94] These entities might emulate similar services emerging in the online reputation management industry, helping individuals address the challenges of understanding and even controlling online information. As the management of misinformation and disinformation grows as a challenge relevant to individual as well as organizational reputations, various research, market, and non-market responses should be the focus of government funding. At this early stage, one goal might be to support infomediation services working to prevent organizations from making biased decisions based on flawed data. For example, consultants with expertise in source triangulation could assist human resource professionals as they attempt to navigate biased data and algorithms when evaluating job candidates.[95] These efforts might evolve into an array of private and non-profit infomediation services that could also support individuals as they encounter misinformation and disinformation challenges.[96] What is clear is that policy efforts must follow a distributive justice approach, acknowledging the unique challenges faced by those most likely to be harmed by the treacherous information environments online.

V. Conclusion

While this paper focuses on developing policy recommendations for implementation in a U.S. framework, the relevance of these proposals goes

---

[90] Pickard calls for "bold new (policy) models" for improving media systems, with a specific emphasis on public media, VICTOR PICKARD, AMERICA'S BATTLE FOR MEDIA DEMOCRACY: THE TRIUMPH OF CORPORATE LIBERTARIANISM AND THE FUTURE OF MEDIA REFORM (2015) at 231.

[91] C. EDWIN BAKER, MEDIA CONCENTRATION AND DEMOCRACY: WHY OWNERSHIP MATTERS (2006).

[92] *Supra* note 90.

[93] *See* BILL KOVACH and TOM ROSENSTIEL, THE ELEMENTS OF JOURNALISM: WHAT NEWSPEOPLE SHOULD KNOW AND THE PUBLIC SHOULD EXPECT (2014)

[94] Metzger was calling attention to a web-based form of this type of support in 2007. *See* Metzger *supra* note 2.

[95] *Supra* notes 20 and 23.

[96] *See* Jonathan A. Obar, *Big Data and The Phantom Public: Walter Lippmann and the Fallacy of Data Privacy Self-Management*, 2 BIG DATA & SOCIETY 1 (2015); Jonathan A. Obar, *Searching for Data Privacy Self-Management: Individual Data Control and Canada's Digital Strategy*, 44 CANADIAN JOURNAL OF COMMUNICATION 35 (2019).

beyond any single national context. In January 2017, the International Federation of Library Associations and Institutions (IFLA) posted an infographic entitled "How to Spot Fake News"[97] (see Figure 3) to the global resource, Wikimedia Commons.[98] The first four recommendations all connect to the source triangulation concept, suggesting individuals should read critically and review multiple sources before finding answers to questions. Even "check your biases"[99] asks individuals to consider whether less-rigorous information evaluation strategies, perhaps linked to the trust of individual sources or confirmation bias,[100] should be reconsidered.

Figure 3: "How to Spot Fake News" Infographic

Note. Uploaded to the Wikimedia Commons by the International Federation of Library Associations and Institutions, January 27, 2017, https://commons.wikimedia.org/wiki/File: How_to_Spot_Fake_News.jpg The file description notes "IFLA infographic based on FactCheck.org's 2016 article "How to Spot Fake News""

---

[97] IFLA *supra* note 51.

[98] Wikimedia Commons, File: How to Spot Fake News.jpg (January 27, 2017) https://commons.wikimedia.org/wiki/File:How_to_Spot_Fake_News.jpg.

[99] *Id.*

[100] *Supra* note 38. In the FactCheck.org post that the IFLA states is the basis for its infographic, it notes: "Check your biases. We know this is difficult. Confirmation bias leads people to put more stock in information that confirms their beliefs and discount information that doesn't." Kiely and Robertson, *supra* note 50.

Self-described as "the global voice of the library and information profession"[101] the IFLA claims to have at least 1,500 members in more than 150 countries.[102] A credible commonality check reveals that the IFLA's Facebook page is followed by more than 50,000 individuals.[103] The American Library Association notes the IFLA "wants to show the power of libraries in achieving the UN Sustainable Development Goals (SDGs) and has asked library associations around the world to join in a global effort."[104] These two additional pieces of information suggest the IFLA and its statements may be credible.

This brief exercise suggests the following. First, while the IFLA resource is a place to start, more research is necessary to determine what information literacy skills are important and how they might be taught. The "Civic Online Reasoning" curriculum[105] provides some helpful resources, but more are needed. Second, the presence of the IFLA and its commitment to information literacy education suggests that the policy proposals herein may extend internationally. As the United Nations works to promote "meaningful universal connectivity"[106] it is vital that international digital inclusion initiatives prioritize the vision and the funding to support individuals as they attempt to distinguish between fact, fallacy, and fake.

As technology changes, imagining new governance models can be difficult, especially when current (and ineffective) approaches have been in place for some time.[107] Answers to these questions require further research into how source triangulation may contribute to democratic discourses internationally. One thing is clear: we are unprepared to meet the "information problems"[108] present today and threatened by challenges soon to be realized in a quickly emerging tomorrow.[109]

---

[101] International Federation of Library Associations and Institutions, Homepage (2021) https://www.ifla.org/.
[102] International Federation of Library Associations and Institutions, About (2021) https://www.ifla.org/about.
[103] International Federation of Library Associations and Institutions, Facebook Page (2021) https://www.facebook.com/IFLA.org/.
[104] American Library Association, ALA Task Force on United Nations 2030 Sustainable Development Goals (April 20, 2020) http://www.ala.org/aboutala/ala-task-force-united-nations-2030-sustainable-development-goals.
[105] Stanford History Education Group *supra* note 24.
[106] *Supra* note 8.
[107] Jonathan A. Obar and Steven S. Wildman, *Social Media Definition and the Governance Challenge: An Introduction to the Special Issue*, 39 TELECOMMUNICATIONS POLICY 745 (2015).
[108] Eisenberg, *supra* note 4 at 39.
[109] This assertion is inspired by comments from former Canadian Privacy Commissioner Stoddart speaking about privacy law in Canada. The parallel context of data privacy challenges is important to consider. As the Commissioner states "It is increasingly clear that the law is not up to the task of meeting the challenges of today – and certainly not those of tomorrow." Office of the Privacy Commissioner of Canada. News release: New privacy challenges demand stronger protections for Canadians (2013) https://www.priv.gc.ca/en/opc-news/news-and-announcements/2013/nr-c_130523.