
ARTICLE

SOCIAL MEDIA SELF-REGULATION AND THE RISE OF VACCINE MISINFORMATION

ANA SANTOS RUTSCHMAN[†]

This essay examines the main characteristics and shortcomings of mainstream social media responses to vaccine misinformation and disinformation. Parts I and II contextualize the recent expansion of vaccine information and disinformation in the online environment. Part III provides a survey and taxonomy of ongoing responses to vaccine misinformation adopted by mainstream social media. It further notes the limitations of current self-regulatory modes and illustrates these limitations by presenting a short case study on Facebook—the largest social media vehicle for vaccine-specific misinformation, currently estimated to harbor approximately half of the social media accounts linked to vaccine misinformation. Part IV examines potential ways to improve stringency of ongoing modes of self-regulation of vaccine misinformation, as well as the creation of cooperative monitoring and mutual assistance networks dedicated to addressing issues specific to the field of vaccine misinformation.

INTRODUCTION.....	26
I. A PUBLIC HEALTH PROBLEM: VACCINE MISINFORMATION, TRUST AND HESITANCY	30
A. <i>A Public Health Paradox</i>	30
B. <i>Charting the Rise of Vaccine Misinformation</i>	34

[†] Assistant Professor of Law, Saint Louis University School of Law, Center for Health Law Studies, S.J.D., LL.M., Duke Law School. This essay has benefited from feedback received during presentations at the University of Pennsylvania Law School's *Journal of Law and Innovation*, the Health and Technology Policy 2020 Symposium at Gonzaga University School of Law, the *Inoculating Against Misinformation on the Internet* panel organized by the National Vaccine Advisory Committee in February 2020 in Washington DC, and a presentation to European Council's Working Party on Information. I am particularly grateful to Jennifer Rothman for serving as a commentator for this article during this Journal's 2021 symposium. I also thank Rachel Sachs and Anders Walker for comments on an earlier draft of the piece. I am also indebted to Cheryl Cooper, Kaena Kao and Hannah Schweissguth for research assistance.

II. VACCINE MISINFORMATION: THE ROLE OF SOCIAL MEDIA.....	37
A. <i>The Dual Role of Social Media in Vaccine-related Communications</i>	38
B. <i>The Automatization and Weaponization of Vaccine Misinformation Through Social Media</i>	40
III. ONGOING SOCIAL MEDIA RESPONSES TO VACCINE MISINFORMATION ...	43
A. <i>A Taxonomy of Mainstream Social Media Responses</i>	43
1. Suppression or “Zero Tolerance” Approaches.....	43
2. Limiting and Downgrading Approaches.....	45
3. Educational Approaches.....	45
B. <i>Shortcomings of Current Self-Regulatory Approaches</i>	46
1. Shortcomings of Social Media Self-Regulation.....	47
2. The Growth of Vaccine Misinformation on Social Media During the COVID-19 Pandemic: The Case of Facebook	51
IV. CURBING VACCINE MISINFORMATION AND DISINFORMATION: THE CASE FOR MORE STRINGENT SOFT-LAW APPROACHES	58
A. <i>A Starting Point: Expanding the Existing Regulatory and Institutional Infrastructure</i>	59
B. <i>Incorporating Tools to Address Vaccine-specific Misinformation and Disinformation</i>	65

INTRODUCTION

The world is everything that is the case.

The world is the totality of facts, not of things.

The world is determined by the facts, and by these being all the facts.

For the totality of facts determines both what is the case, and also all that is not the case.

The facts in logical space are the world.

The world divides into facts.

(...)

What is the case, the fact, is the existence of atomic facts.

– Ludwig Wittgenstein, *Tractatus Logico-Philosophicus* (1921)

The history of science, like the history of all human ideas, is a history of irresponsible dreams, of obstinacy, and of error. But science is one of the very few human activities — perhaps the only one — in which errors are systematically criticized and fairly often, in time, corrected.

This is why we can say that, in science, we often learn from our mistakes, and why we can speak clearly and sensibly about making progress there.

– Karl Popper, *Conjectures and Refutations: The Growth of Scientific Knowledge* (1963)

Vaccine hesitancy – the reluctance or refusal to vaccinate despite the availability of vaccines – threatens to reverse progress made in tackling vaccine-preventable diseases.

– World Health Organization, *Top Ten Threats to Global Health in 2019* (adding vaccine hesitancy to the list)

The circulation of inaccurate information among national and transnational communities has been documented throughout history.¹ A more recent development has been the use of online pathways to propagate inaccurate information on increasingly larger scales.² The popularization of social media, in particular, has significantly accelerated and amplified the spread of misinformation.³ This growth has left virtually no field untouched, from discourses on political and electoral themes to climate change, finance and pop culture.⁴

¹ See Robert Darnton, *The True History of Fake News*, N.Y. BOOK REV. (Feb. 13, 2017), <https://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/> (“[T]he equivalent of today’s poisonous, bite-size texts and tweets can be found in most periods of history, going back to the ancients.”).

² U.N. EDUC., SCI. & CULTURAL ORG., JOURNALISM, ‘FAKE NEWS’ & DISINFORMATION: HANDBOOK FOR JOURNALISM EDUCATION AND TRAINING 17 (2018) (“Increasingly, it is also possible to engineer audio and video in ways that go beyond legitimate news editing in order to make it appear that a particular individual said or did something in some place, and to pass this off as an authentic record . . .”).

³ See Michela Del Vicario et al., *The Spreading of Misinformation Online*, 113 PNAS 554, 554 (2016) (“[T]he World Wide Web (WWW) also allows for the rapid dissemination of unsubstantiated rumors and conspiracy theories that often elicit rapid, large, but naive social responses . . .”).

⁴ See generally CAILIN O’CONNOR & JAMES OWEN WEATHERALL, *THE MISINFORMATION AGE* 3 (2018) (“[A]bout six weeks before Donald Trump was elected president of the United States, a website calling itself ETF News (endingthefed.com) posted a story with the headline ‘Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement.’”); Claire Wardle, *Misinformation Has Created a New World Disorder*, SCI. AM. (2019), <https://www.scientificamerican.com/article/misinformation-has-created-a-new-world-disorder/> (“misleading and conspiratorial content did not begin with the 2016 presidential race, and it will not end after”); Shimon Kogan et al., *Social Media and Financial News Manipulation* 5 (Sept. 15, 2021), <https://ssrn.com/abstract=3237763> (“Misleading information can impact social, political, and economic relationships.”); Kathie M. d’I. Treen et al., *Online Misinformation About Climate Change*, WIREs CLIMATE CHANGE (Jun. 18, 2020), <https://onlinelibrary.wiley.com/doi/full/10.1002/wcc.665> (“Once in the public

In recent years, levels of hesitancy towards vaccines have been increasing in many areas of the globe, but especially among Western countries.⁵ The reasons for this increase are multi-factorial and are not solely attributable to the growing circulation of misinformation specifically focused on vaccine-related themes.⁶ However, the accelerated and virtually unencumbered dissemination of vaccine misinformation in the online environment—and particularly through social media—has profoundly reshaped this area.⁷ It has facilitated and accelerated the creation of like-minded communities; provided renewed visibility to vaccine-questioning and anti-vaccine discourses; reoriented many of the efforts of anti-vaccine activists towards online and social media channels; and attracted the attention of a broader category of players, who seek to increase demand for, and monetize the purchase of, “alternative” health goods, and which now populate social media with anti-vaccine or vaccine-questioning accounts.⁸

Vaccine misinformation has been exponentially exacerbated through use (and misuse) of the manifold avenues for the dissemination of content opened up by the popularization of social media.⁹ As seen in Part II, the online circulation of vaccine misinformation has now been linked to the growth of vaccine mistrust and hesitancy. Addressing the public health and technological paradox posed by insufficient uptake of available vaccines thus

domain, characteristics of online social networks . . . provide fertile ground for misinformation to spread” about climate change).

⁵ See Peter Hotez, *America and Europe’s New Normal: The Return of Vaccine-Preventable Diseases*, 85 PEDIATRIC RES. 912, 912 (2019) (“Heading into 2019 we are seeing a return of vaccine-preventable diseases in Europe and the United States, much of it because of well-organized, well-funded, and mostly unopposed antivaccine groups and lobbies . . .”).

⁶ See generally Daniel A. Salmon et al., *Vaccine Hesitancy: Causes, Consequences, and a Call to Action*, 49 AM. J. PREV. MED. S391, S391 (2015), <https://www.sciencedirect.com/science/article/abs/pii/S0749379715003141> (“There is a broad range of factors contributing to vaccine hesitancy, including the compulsory nature of vaccines, their coincidental temporal relationships to adverse health outcomes, unfamiliarity with vaccine-preventable diseases, and lack of trust in corporations and public health agencies.”).

⁷ See generally Andis Robeznieks, *Stopping the Scourge of Social Media Misinformation on Vaccines*, AM. MED. ASS’N (Mar. 15, 2019), <https://www.ama-assn.org/delivering-care/public-health/stopping-scurge-social-media-misinformation-vaccines> (“It is common that patient searches for information and products related to the word ‘vaccine’ yield top results pointing to harmfully inaccurate information about immunization safety. This place of prominence given to medical disinformation is deeply troubling to America’s physicians, especially amid alarming new reports regarding measles, tetanus and other vaccine-preventable conditions.”).

⁸ See *infra* notes 155-156 and accompanying text.

⁹ See, e.g., Robeznieks, *supra* note 7 (“The [American Medical Association] sent a letter to top executives at Amazon, Facebook, Google, Pinterest, Twitter and YouTube urging them to do even more to stem the ‘proliferation’ of ‘health-related misinformation’ that has helped vaccine-preventable diseases to reemerge.”).

requires considering how vaccine misinformation propagates online; how this propagation has been instrumentalized by actors with certain ideological or monetization purposes, or both; and the policy and legal options available to curb the spread of vaccine misinformation.

This essay begins exploring these topics by explaining how the intertwined phenomena of vaccine trust and vaccine hesitancy are presently shaped by the recent spur in the circulation of inaccurate content about vaccines in mainstream social media. Part III provides a survey and taxonomy of recent and ongoing responses to vaccine misinformation from mainstream social media and online social networks. It further notes the limitations of current self-regulatory modes and illustrates these limitations by presenting a short case study about Facebook—the largest social media vehicle for vaccine-specific misinformation, currently estimated to harbor approximately half of the social media accounts linked to vaccine misinformation. Part IV examines potential ways to improve stringency of ongoing modes of self-regulation of vaccine misinformation, as well as the creation of cooperative monitoring and mutual assistance networks dedicated to addressing issues specific to the field of vaccine misinformation.

A framing note: while multiple definitions of these phenomena populate the literature,¹⁰ the term “misinformation” has come to be broadly understood as the dissemination of false or misleading content.¹¹ A growing number of commentators treat the phenomenon of “disinformation” separately to reference situations in which such dissemination is done with the deliberate purpose of sowing doubt around a particular topic or increasing discord among holders of competing opinions or worldviews.¹² This essay uses the word “misinformation” as an umbrella term to denote the existence of inaccuracies in vaccine-related content, irrespective of intent. Throughout the essay, the word “disinformation” is reserved for cases in which there is an intent to deceive in the dissemination of inaccurate vaccine-related content; and in which it references a source that adopts the word as an umbrella term itself—as is the case of the European Union’s policy in this area.¹³

¹⁰ See, e.g., Emily K. Vraga & Leticia Bode, *Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation*, 37 POL. COMM. 136, 136 (2020) (“[D]efining ‘misinformation’ in a consistent and coherent way has been a challenge for the field.”).

¹¹ Gordon Pennycook et al., *Understanding and Reducing the Spread of Misinformation Online*, 592 NATURE 590, 590 (2021).

¹² Claire Wardle & Hossein Derakhshan, *Information Disorder: Toward an interdisciplinary Framework for Research and Policy Making*, Council Eur. Rep. DGI(2017)09, at 16 (Sept. 27, 2017) (“The term ‘fake news’ has also begun to be appropriated by certain politicians and commentators to describe news organizations whose coverage they find disagreeable.”).

¹³ See *infra*, Part IV.A.

However, instead of adopting the umbrella expression “anti-vaccine,” the essay distinguishes between instances of anti-vaccine discourses (statements that directly contradict current scientific consensus about the safety and efficacy of vaccines, or that promote vaccine refusal) and vaccine-questioning discourses (statements from individuals or organizations seeking more information about vaccines)¹⁴ to mirror the range of behaviors and motivations in this ideologically charged area.

I. A PUBLIC HEALTH PROBLEM: VACCINE MISINFORMATION, TRUST AND HESITANCY

A. *A Public Health Paradox*

Vaccines have long been regarded as one of the most cost-effective public health preparedness tools, playing an instrumental role in the prevention of outbreaks of infectious diseases,¹⁵ as well as in the response to ongoing outbreaks, as presently illustrated by the COVID-19 pandemic

Outside the context of highly disruptive—and often sudden—large-scale public health crises, robust administration of vaccines recommended by public health authorities not only improves public health outcomes, but also results in considerable savings to health systems, as well as in positive externalities for different economic sectors.¹⁶ Recent estimates from the World Health Organization indicate that vaccination prevents between two and three million deaths per year across the globe.¹⁷ A study analyzing vaccination practices in the U.S. found that vaccinating around four million infants helps prevent 42,000 early deaths and twenty million cases of illness.¹⁸ The study also calculated that avoidance of death and illness in this

¹⁴ Examples include individuals or organizations interrogating the safety of COVID-19 vaccines, which were developed according to a timeline perceived by the general public as exceptional.

¹⁵ Sachiko Ozawa et al., *Return On Investment From Childhood Immunization In Low-And Middle-Income Countries, 2011–20*, 35 HEALTH AFF. 199, 199 (2016).

¹⁶ Vanessa Rémy et al., *Vaccination: The Cornerstone of an Efficient Healthcare System*, 3 J. MRT. ACCESS HEALTH & POL'Y. 1, 1 (2015) (“[The] health effects [of vaccines] translate into positive economic results, as vaccination can provide significant savings by avoiding the direct and indirect costs associated with treating the disease and possible long-term disability. A recent US study estimated that every dollar spent on childhood vaccination could save US\$4 from a payer perspective and US\$10 from a societal perspective.”).

¹⁷ WORLD HEALTH ORG., *Immunization coverage* (2020), <https://www.who.int/en/news-room/fact-sheets/detail/immunization-coverage>.

¹⁸ Fangjun Zhou et al., *Economic Evaluation of the Routine Childhood Immunization Program in the United States*, 2009, 133 PEDIATRICS 577, 577 (2014).

context would translate into \$13.5 billion savings in net direct costs,¹⁹ as well as \$68.8 billion in net savings in societal costs.²⁰

While there are several vaccine-preventable diseases for which we currently lack commercially available vaccines,²¹ the number of vaccines targeting new diseases available to indicated populations has increased steadily from the mid-twentieth century onwards, leading to the eradication or near-eradication of many devastating diseases, as well as to significant reductions in the burden of many others.²²

In recent years, however, wavering trust in vaccines has been deemed one of the most significant contributing factors towards declining rates of vaccination, particularly across the Western world.²³ Problems of trust related to health technologies, and in particular vaccines, are not new. Different communities have historically challenged the public health value, safety and efficacy of vaccines, even when presented with the best available scientific evidence supporting the use of a given vaccine, and of vaccines in general, as further described in Part II.²⁴

Periods of heightened vaccine mistrust are linked to increased behavioral hesitancy affecting the timely administration of recommended vaccines to

¹⁹ *Id.* at 577. This category encompasses the costs of medical treatment, as well as of non-medical interventions, such as the costs associated with special education services necessary for children suffering from disabilities. *Id.*

²⁰ This category encompasses productivity losses and opportunity costs resulting from contracting a vaccine-preventable disease or caring for children suffering from a vaccine-preventable disease. *Id.*

²¹ See Stanley A. Plotkin et al., *Establishing a Global Vaccine-Development Fund*, 373 NEW ENGL. J. MED. 297, 297 (2015) (“[W]e still lack a vaccine [for Ebola] that has been shown to be safe and effective.”).

²² ARTHUR ALLEN, VACCINE: THE CONTROVERSIAL STORY OF MEDICINE’S GREATEST LIFESAVER 50 (2007); DAVID M. OSHINSKY, POLIO: AN AMERICAN STORY (2006); see also U.S. FOOD & DRUG ADMIN., *Vaccines Licensed for Use in the United States* (Dec. 1, 2021) <https://www.fda.gov/vaccines-blood-biologics/vaccines/vaccines-licensed-use-united-states> (listing currently licensed vaccines in the United States).

²³ See Walter A. Orenstein et al., *Assessing the State of Vaccine Confidence in the United States: Recommendations from the National Vaccine Advisory Committee*, 130 PUB. HEALTH REP. 573, 574-75 (2015) (“Reluctance, hesitation, concerns, or a lack of confidence has caused some parents to question or forgo recommended vaccines.”); Heather MacDougall & Laurence Monnais, *Vaccinating in the Age of Apathy: Measles Vaccination in Canada 1963–1998*, 190 CAN. J. MED. ASS’N E399, E399 (2018) (“[L]ack of uptake meant that vaccination was ‘a battleground where apathy is [a] greater enemy than disease.’); Heidi Larson et al., *State of Vaccine Confidence in the European Union in 2018*, 29 EUR. J. PUB. HEALTH 141, 141 (2019) (“We find that most age groups under 65 surveyed have less confidence in the safety and importance of both the MMR and seasonal influenza vaccines (and vaccines generally) than over 65’s.”).

²⁴ See generally JONATHAN M. BERMAN, ANTI-VAXXERS: HOW TO CHALLENGE A MISINFORMED MOVEMENT (2020).

indicated patients.²⁵ The World Health Organization currently defines “vaccine hesitancy” as the “reluctance or refusal to vaccinate despite the availability of vaccines.”²⁶ In 2019, the World Health Organization added vaccine hesitancy to the list of the top ten threats to global health.²⁷

We thus face a technological paradox with salient implications for public health and preparedness frameworks.²⁸ Notwithstanding the broad availability of health technologies that can prevent or lessen the burden of vaccine-preventable diseases, mistrust and hesitancy towards these technologies hamper their deployment as tools of public health.²⁹ As a result, vaccine-preventable diseases for which a vaccine is available are making a comeback. The 2019 outbreaks of measles across the United States, for example, have been directly linked to growing hesitancy towards childhood vaccines that have been recommended by public health authorities, commercialized, and administered for decades.³⁰

During the COVID-19 pandemic, public health authorities in the United States have had to contend with sub-optimal levels of public trust in newly developed vaccines, a phenomenon that continues to cast doubts on whether ongoing vaccine manufacturing and vaccination efforts will be enough to reach the critical mass required to achieve herd immunity within projected

²⁵ See Eve Dubé et al., *Vaccine Hesitancy, Vaccine Refusal and the Anti-Vaccine Movement: Influence, Impact and Implications*, 14 EXP. REV. VACCINES 99, 100 (2015) (“Vaccine-hesitant parents may refuse some vaccines, but agree to others; they may delay vaccines or accept them according to the recommended schedule, but feel unsure in doing so.”).

²⁶ WORLD HEALTH ORG./Rada Akbar, TEN THREATS TO GLOBAL HEALTH IN 2019, <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (last visited Dec. 19, 2021).

²⁷ *Id.*

²⁸ See CTRS. DISEASE CONTROL & PREVENTION, *Public Health Emergency Preparedness and Response Capabilities: National Standards for State, Local, Tribal, and Territorial Public Health* (Jan. 25, 2021, 3:15 PM), <https://www.cdc.gov/cpr/readiness/capabilities.htm> (“Community preparedness is the ability of communities to prepare for, withstand, and recover from public health incidents in both the short and long term.”).

²⁹ See, e.g., Saad B. Omer et al., *Vaccine Refusal, Mandatory Immunization, and the Risks of Vaccine-Preventable Diseases*, 361 NEW ENGL. J. MED. 1981, 1981 (2009) (“[T]he success of an immunization program depends on high rates of acceptance and coverage.”); *id.* (“[P]ublic concern about real or perceived adverse events associated with vaccines has increased. This heightened level of concern often results in an increase in the number of people refusing vaccines.”).

³⁰ See David A. Broniatowski et al., *Facebook Pages, the “Disneyland” Measles Outbreak, and Promotion of Vaccine Refusal as a Civil Right, 2009–2019*, 110 AM. J. PUB. HEALTH S312, S312 (2020) (“Activity in pages promoting vaccine choice as a civil liberty increased in January 2015, April 2016, and January 2019 The ‘Disneyland’ measles outbreak drew vaccine opposition into the political mainstream, followed by promotional campaigns conducted in pages framing vaccine refusal as a civil right.”).

timelines.³¹ This provides a stark contrast with the vaccine trust environment that characterized earlier vaccine races. The strong vaccine uptake that followed the development and approval of the first polio vaccines in the mid-1950s set in motion a series of public health initiatives that ultimately resulted in a 99% reduction of the incidence of the disease worldwide.³² By contrast, there were signs early on in the COVID-19 pandemic that a significant number of potential vaccine recipients were hesitant about receiving the vaccines, or planned to skip COVID-19 vaccination altogether.³³ For instance, less than three months after the World Health Organization declared COVID-19 a pandemic, only around 50% of Americans indicated that they were planning on receiving the vaccine, if one were to be made available during the pandemic.³⁴

The intertwined problems of vaccine trust and vaccine hesitancy at the root of the resurgence of vaccine-preventable diseases—or protracted duration of outbreaks of new diseases—cannot be attributed to a single factor.³⁵ Some of these factors have deep historical and philosophical origins,

³¹ See Alex Kacik, *Half of Americans are 'highly likely' to get COVID-19 Vaccinations*, MOD. HEALTH CARE (Aug. 28, 2020), <https://www.modernhealthcare.com/safety/half-americans-are-highly-likely-get-covid-19-vaccinations> (“Only around half of Americans are highly likely to get vaccinated for COVID-19, according to a new survey. . . . That reflects a sense of lack of safety around vaccines”); see also Lauran Neergaard & Hannah Fingerhut, *AP-NORC Poll: Only Half in US Want Shots as Vaccine Nears*, ASSOCIATED PRESS (Dec. 9, 2020), <https://apnews.com/article/ap-norc-poll-us-half-want-vaccine-shots-4d98dbfc0a64d60d52ac84c3065dac55> (“Many on the fence have safety concerns and want to watch how the initial rollout [of the COVID-19 vaccine] fares.”).

³² WORLD HEALTH ORG., *10 Facts on Polio Eradication* (Apr. 1, 2017), <https://www.who.int/features/factfiles/polio/en/>.

³³ Emily A. Harrison & Julia W. Wu, *Vaccine Confidence in the Time of COVID-19*, 22 EUR. J. EPIDEMIOLOGY 1, 1 (2020).

³⁴ Warren Cornwall, *Just 50% of Americans Plan to Get a COVID-19 Vaccine. Here's How to Win Over the Rest*, SCIENCE (June 30, 2020), <https://www.sciencemag.org/news/2020/06/just-50-americans-plan-get-covid-19-vaccine-here-s-how-win-over-rest>; see also Alex Hartlage, *Vaccine Hesitancy Post-Covid-19: Will Our Memory Fade or Last?*, VACCINES TODAY (July 1, 2020), <https://www.vaccinestoday.eu/stories/vaccine-hesitancy-post-covid-19-will-our-memory-fade-or-last/> (“[O]ver the last two decades, vaccine hesitancy has risen so substantially that the WHO now considers it a major threat to global health.”).

³⁵ See WORLD HEALTH ORG., *Report of the SAGE Working Group on Vaccine Hesitancy* 11 (Oct. 1, 2014), http://www.who.int/immunization/sage/meetings/2014/october/1_Report_WORKING_GROUP_vaccine_hesitancy_final.pdf (“[V]accine hesitancy is the behaviour that results from the decision-making process and reflects a constellation of factors that may influence the decision to accept some or all vaccines in accordance with the recommended schedule.”); *id.* at 11-12 (describing two models for understanding vaccine hesitancy, including one that focuses on complacency, confidence, and convenience); Daniel A. Salmon et al, *Vaccine Hesitancy: Causes, Consequences, and a Call to Action*, 33 VACCINE D66, D66 (2015) (“There is a broad range of factors contributing to vaccine

including heterogenous forms of resistance to government-endorsed interventions, transversal distrust of health technologies that require the insertion of extraneous substances into the human body, and general concerns with how biomedical research has historically been conducted.³⁶ In other cases, vaccine trust is compromised due to the emergence of more recent problems, techniques and even political or social agendas.³⁷ The topic of vaccine misinformation, particularly in the online context, falls under the latter category. While not new, vaccine misinformation has been exponentially exacerbated through use (and misuse) of the manifold avenues for the dissemination of content opened up by the popularization of social media. As seen in Part II, the online circulation of vaccine misinformation is now a leading cause of the growth of vaccine mistrust and hesitancy. Addressing the public health and technological paradox posed by insufficient uptake of available vaccines thus entails considering how vaccine misinformation propagates online; how this propagation has been instrumentalized by actors with ideological or monetization purposes, or both; and the policy and legal options available to curb the spread of vaccine misinformation. This essay now begins that exploration, starting with the historical and proximate roots of the growth of online vaccine misinformation.

B. *Charting the Rise of Vaccine Misinformation*

Topics surrounding the development and administration of vaccines have long been polarizing. For example, instances of popular opposition to vaccination recommended by public health authorities have been documented as early as in nineteenth-century England and America.³⁸ The expansion of

hesitancy, including the compulsory nature of vaccines, their coincidental temporal relationships to adverse health outcomes, unfamiliarity with vaccine-preventable diseases, and lack of trust in corporations and public health agencies.”); E. David G. McIntosh et al., *Vaccine Hesitancy and Refusal*, 175 EUR. PEDIATRIC ASS’N 248, 248 (2016) (describing vaccine hesitancy has a “multifactorial nature” that is driven primarily by opinions from the antivaccine movement).

³⁶ See, e.g., HIST. VACCINES, *Cultural Perspectives on Vaccination* (Jan. 10, 2018), <https://www.historyofvaccines.org/content/articles/cultural-perspectives-vaccination> (“Public opinions about vaccination include varied and deep-seated beliefs, a result of the tension between divergent cultural viewpoints and value systems. Several key cultural perspectives on vaccination stem from (1) individual rights and public health stances toward vaccination, (2) various religious standpoints and vaccine objections, and (3) suspicion and mistrust of vaccines among different U.S. and global cultures and communities.”). See generally HARRIET WASHINGTON, *MEDICAL APARTHEID* (2006).

³⁷ See *infra*, Part II.

³⁸ See Martin Kaufman, *The American Anti-Vaccinationists and Their Arguments*, 41 BULL. HIST. MED. 463, 464-66 (1967) (“In the 1850s, opposition to vaccination arose, largely from the irregular physicians, the advocates of unorthodox medical theories.”). See generally

vaccination mandates in the 1960s and 1970s,³⁹ while overwhelmingly contributing to the reduction of the incidence of vaccine-preventable diseases, was also used instrumentally to fuel doubts about vaccines and vaccination policies in many areas across North America.⁴⁰ And, in what remains perhaps the most well-known episode in the history of vaccine misinformation, in 1998 a study published by then-doctor Andrew Wakefield in *The Lancet*—one of the world’s leading peer-reviewed medical journals—fraudulently implied the existence of a causal link between the MMR vaccine (measles, mumps and rubella) and the development of autism in children.⁴¹ The study was immediately disputed by the medical and research communities and eventually retracted in 2010, the same year in which the General Medical Council in the United Kingdom declared that Wakefield had acted “dishonestly and irresponsibly” and found him guilty of over thirty charges of professional misconduct and banning him from practicing medicine.⁴² Nevertheless, the discredited study remains influential among communities questioning current approaches to vaccination and is often cited outside the scientific world in connection with claims contesting the safety of vaccines,⁴³ or linking vaccination to conspiracy theories involving the pursuit

BERMAN, *supra* note 24.

³⁹ In its seminal 1905 decision, the Supreme Court upheld the authority of public health authorities to mandate vaccination. *Jacobson v. Massachusetts*, 197 U.S. 11 (1905).

⁴⁰ MacDougall & Monnais, *supra* note 23, at E400 (“In Ontario, the Committee Against Compulsory Vaccination and the Association for Vaccine Damaged Children emerged in response to the [mandatory vaccination] legislation. . . . these groups claimed that doctors and public health nurses misled parents about infant vaccination requirements and called for parental choice.”).

⁴¹ See Fiona Godlee & Jane Smith, *Wakefield’s Article Linking MMR Vaccine and Autism Was Fraudulent*, 342 BRIT. MED. J. 64, 64 (2011) (“Over the following decade [after Wakefield’s article was published], epidemiological studies consistently found no evidence of a link between the MMR vaccine and autism. By the time the paper was finally retracted 12 years later . . . few people could deny that it was fatally flawed both scientifically and ethically.”).

⁴² See Clare Dyer, *Wakefield Was Dishonest and Irresponsible over MMR Research, Says GMC*, 340 BRIT. MED. J. c593, c593 (2010) (“Andrew Wakefield . . . has been found guilty of dishonesty and irresponsibility by the General Medical Council. The UK regulator held that Dr Wakefield abused his position, subjected children to intrusive procedures such as lumbar puncture and colonoscopy that were not clinically indicated, carried out research which flouted the conditions of ethics committee approval and brought the medical profession into disrepute.”); *A Timeline of the Wakefield Retraction*, 16 NAT. MED. 248, 248 (2010) (documenting the time from Wakefield’s initial publication to the retraction of said publication).

⁴³ See Clyde Haberman, *A Discredited Vaccine Study’s Continuing Impact on Public Health*, N.Y. TIMES (Feb. 1, 2015), <https://www.nytimes.com/2015/02/02/us/a-discredited-vaccine-study-s-continuing-impact-on-public-health.html> (“Nonetheless, despite [Dr. Wakefield] being held in disgrace, the vaccine-autism link has continued to be accepted on faith by some.”).

of hidden political or social agendas.⁴⁴

Although these examples are not exhaustive, they illustrate the idiosyncratic landscape against which vaccine-specific misinformation has historically proliferated. More recently, expressions of vaccine misinformation have gained new life—and reached new audiences—through online channels. For well over a decade, the internet has become one of the prime venues for the dissemination of content about vaccines and vaccination.⁴⁵ Online channels are used to convey both accurate and inaccurate information about vaccines. The same channels that carry messages from the WHO or the U.S. Centers for Disease Controls and Prevention can be—and are routinely—used by citizen-publishers posting or reposting scientifically inaccurate vaccine-related information.⁴⁶ Even more recently, they have also been used by individuals and automated programs purposefully circulating vaccine content with the aim of sowing or increasing discord—vaccine disinformation.⁴⁷

Social media have become the prime venues for the circulation of both misinformation and disinformation related to vaccines. As the second decade

⁴⁴ See Matthew Hornsey et al., *The Psychological Roots of Anti-Vaccination Attitudes: A 24-Nation Investigation*, 37 HEALTH PSYCHOL. 307, 308 (2018) (“For some, however, individual conspiracy beliefs are not held in isolation, but rather emerge from a unitary ‘conspiracist’ world-view People who feel this is the way the world works . . . might be motivated to believe conspiracies about science, with negative impacts on vaccination intentions. Indeed, there is evidence from an online American sample that people’s willingness to endorse conspiracies generally . . . are positively correlated with a range of ‘anti-science’ attitudes, including antivaccination attitudes.”).

⁴⁵ See, e.g., Anna Kata, *Anti-vaccine activists, Web 2.0, and the postmodern paradigm—an overview of tactics and tropes used online by the anti-vaccination movement*, 30 VACCINE 3778, 3778 (2012) (documenting this phenomenon with regard to anti-vaccine activity); David A. Broniatowski et al., *Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate*, 108 AM. J. PUB. HEALTH 1378, 1378 (2018) (documenting this phenomenon in the context of the spread of automated vaccine-related content in social media).

⁴⁶ See, e.g., Cornelia Betsch et al., *Opportunities and Challenges of Web 2.0 for Vaccination Decisions*, 30 VACCINE 3727, 3727 (2012) (“Websites that allow and promote interaction among users are an increasingly popular source of health information. Users of such so-called Web 2.0 applications (e.g. social media), while still in the minority, represent a growing proportion of online communicators, including vocal and active anti-vaccination groups as well as public health communicators.”); Kata, *supra* note 45, at 3779 (“Web 2.0 lets patients actively engage in their own care. While medical knowledge was previously bound to textbooks and journals, the Internet allows access to the ‘school of lay medicine’, shifting the locus of power from doctors as sole directors of a patient’s care to the patients themselves.”).

⁴⁷ Broniatowski et al., *supra* note 45, at 1288 (“Whereas bots that spread malware and unsolicited content disseminated antivaccine messages, Russian trolls promoted discord. Accounts masquerading as legitimate users create false equivalency, eroding public consensus on vaccination.”).

of the twenty-first century drew to a close, mainstream social media like Facebook, Instagram, Twitter, and YouTube had become the largest venues for the propagation of vaccine misinformation and disinformation. In 2018, a pivotal study on online disinformation reported escalating levels of activity specifically focused on the propagation of inaccurate vaccine-related content, labeling these recent developments as a form of “weaponized” of health communication.⁴⁸

The increased circulation of inaccurate information about vaccines in social media poses serious challenges to public health strategies designed to curb the emergence and spread of infectious diseases. It presents heightened hurdles when compared to previous embodiments of vaccine-specific misinformation, as social media enable spreaders of misinformation to reach wider audiences, as well as tapping into densely interconnected networks focused on discussions surrounding highly divisive non-vaccine topics, such as political and electoral themes, fake news and the role of mainstream media, and conspiracy theories on a variety of topics.⁴⁹

Part II describes the specificities of vaccine misinformation within the social media ecosystem. It is worth noting here that a growing body of vaccine-focused research finds that the propagation of inaccurate vaccine content through social media is creating significant problems for the implementation of vaccination campaigns and contributing to the erosion of overall levels of trust in vaccines.⁵⁰

II. VACCINE MISINFORMATION: THE ROLE OF SOCIAL MEDIA

We now turn to the specific role played by social media in the propagation of vaccine-related messages. Part A describes how social media platforms can function as conveyers of both accurate and inaccurate information about vaccines. Part B then focuses on mechanisms used to disproportionately amplify anti-vaccine or vaccine-questioning content shared through mainstream social media.

⁴⁸ *Id.*

⁴⁹ See Betsch, *supra* note 46, at 3728 (“Web 2.0 [is defined] as Internet applications that enable users to create and upload new content, comment on existing content, and share content with other users. . . . Social media, for example, provide opportunities to publicly express support for an issue and forward information to friends without great effort.”); Kata, *supra* note 45, at 3779 (noting a similar definition).

⁵⁰ See, e.g., Broniatowski et al., *supra* note 30, at S312 (“Tightly knit communities that collectively refuse to vaccinate lack herd immunity . . . and damage herd immunity for the broader population. In 2019, the US Centers for Disease Control and Prevention reported outbreaks of measles in several US states and worldwide, all of which struck communities with low vaccination rates Finally, some have raised concerns that the COVID-19 ‘infodemic’ could trigger vaccine refusal.”).

A. *The Dual Role of Social Media in Vaccine-related Communications*

Social media rose to prominence during the transition from the 2000s to the 2010s. The Pew Research Center started collecting data on social media usage in 2005.⁵¹ At that point, only 5% of adults in the United States used at least one of the then-largest social media platforms.⁵² In 2011, half of all Americans were using at least one of these platforms, and by 2019 that number was approaching three quarters of the U.S. population.⁵³ Among Americans aged between 18 and 29 years old, 90% used at least one mainstream social media—which for purposes of the Pew Research Center study comprised, as of 2019, Facebook, Instagram, LinkedIn, Twitter, Pinterest, Snapchat, YouTube, WhatsApp and Reddit.⁵⁴ The study further showed that usage of mainstream social media was fairly evenly distributed when race, gender and income are considered.⁵⁵ However, usage of a particular social media varied significantly within these and additional parameters, including formal levels of education.⁵⁶

The expansion of social media throughout the late 2000s and the 2010s has had an impact on health-related communications in the online environment. The use of social media has been shown to help the dissemination of public health research, inform policy debates and create venues for the formation of professional, patient, and activist communities.⁵⁷

In the case of vaccine-related communications, social media have been shown to play a dual role, functioning as fora and conduits for content promoting both accurate and inaccurate information about vaccines. On the one hand, they have become an important vehicle for efforts from public health-oriented institutions seeking to disseminate accurate vaccine information as part of educational campaigns, as well as campaigns designed to curb vaccine hesitancy.⁵⁸ On the other, recent studies have shown that

⁵¹ PEW RES. CTR., *Social Media Fact Sheet* (Jun. 12, 2019), <https://www.pewresearch.org/internet/fact-sheet/social-media/>.

⁵² *Id.*

⁵³ *Id.*

⁵⁴ *Id.*

⁵⁵ *Id.*

⁵⁶ *Id.*

⁵⁷ See Jessica Y. Breland, et al., *Social Media as a Tool to Increase the Impact of Public Health Research*, 107 AM. J. PUB. HEALTH 1890, 1891 (2017) (“Social media provides a way to share evidence for or against health policies with the public, policymakers, and other key stakeholders. . . . Social media can be used for numerous research purposes, including surveillance activities and behavioral interventions. . . . The increasing presence of academics, clinicians, industry professionals, public health departments, and health care systems on social media provides many opportunities for professional connection outside traditional settings.”).

⁵⁸ See Mark Dredze et al., *Understanding Vaccine Refusal: Why We Need Social Media*

social media function as amplifiers of both misinformation and disinformation on vaccine-related topics.⁵⁹

This amplification results in a minority of social media users holding views that are not supported by the scientific status quo, yielding a disproportionate amount of influence in the online environment by spreading anti-vaccine or vaccine-questioning content more efficiently—through highly connected networks—than individuals or institutions spreading accurate information about vaccines.⁶⁰ This phenomenon is particularly salient in the case of parents of young children—the latter being typically indicated to receive more vaccines than adult populations, thus making parents preferred targets of misinformation in this area. Researchers have long established that an overwhelming majority of parents hold favorable views on vaccination. A study conducted in 2018 in the United Kingdom found that 91% of parents regarded vaccines as “important.”⁶¹ At the same time, the study found that 41% of parents using social media reported encountering “negative messages” about vaccines or vaccination on a relatively regular basis.⁶² Among parents of very young children—defined as under five years old—that percentage climbed to 50%.⁶³

Disproportionate amplification of vaccine misinformation also exposed users who are not invested in parental vaccination debates to inaccurate content about vaccines. In a 2019 experiment, for example, a group of journalists analyzed the results of nearly 100 million individuals expressing vaccination views on Facebook.⁶⁴ The majority of views individuals expressed were either undecided on the question of vaccine efficacy or against vaccination.⁶⁵ As discussed in Part III.B.2, data related to vaccine misinformation practices drawn from social media sites is especially relevant—for example, as much as half the tweets about vaccination on

Now, 50 AM. J. PREV. MED. 550, 551 (2016) (“The messages observed and shared [on social media] can provide a real-time, detailed picture of public attitudes toward vaccination . . . because the Internet allows such rapid spread of anti-vaccine arguments, it is essential to harness the strength of the Internet to combat them.”).

⁵⁹ Betsch, *supra* note 46; Kata, *supra* note 45.

⁶⁰ See Neil F. Johnson et al., *The Online Competition Between Pro- and Anti-Vaccination Views*, 582 NATURE 230, 230 (2020) (“Although smaller in overall size, anti-vaccination clusters manage to become highly entangled with undecided clusters in the main online network, whereas pro-vaccination clusters are more peripheral.”).

⁶¹ ROYAL SOC’Y PUB. HEALTH, MOVING THE NEEDLE: PROMOTING VACCINATION UPTAKE ACROSS THE LIFE COURSE 3 (2019), <https://www.rsph.org.uk/static/uploaded/3b82db00-a7ef-494c-85451e78ce18a779.pdf>.

⁶² *Id.*

⁶³ *Id.*

⁶⁴ Johnson, *supra* note 60, at 230.

⁶⁵ *Id.* at 230-31.

Twitter are associated with anti-vaccine or vaccine-questioning beliefs.⁶⁶

In recent years, the spread of vaccine-related content in social media has also found new avenues through the use of automated programs that can spread pro- and anti-vaccine, vaccine-questioning content, or both at the same time, as I explain in the following section.

B. *The Automatization and Weaponization of Vaccine Misinformation Through Social Media*

The use of software to disseminate online content is very common, within and outside social media. Studies indicate that towards the later stages of the 2010s, automatically generated content became pervasive across the internet. The title of a piece published in *The Atlantic* in 2017—*The Internet Is Mostly Bots*—aptly captured the emerging online landscape.⁶⁷

A study surveying online activity throughout 2018 calculated that around 40% of online traffic was likely automated.⁶⁸ Another study calculated that programs spreading malicious content⁶⁹ accounted for 20.4% of automated traffic.⁷⁰ Although this constituted a 6.4% drop in traffic linked to malicious software when compared to the findings of a similar study looking at data from 2017, the levels of sophistication of these programs remained consistently high.⁷¹

The use of automated software combined with increasingly sophisticated techniques has enabled the propagation of vaccine-related content on increasingly larger scales. In 2015, the U.S. Defense Advanced Research Projects Agency (DARPA) conducted a four-week challenge—dubbed the Twitter Bot Detection Challenge—during which participants analyzed vaccine-related tweets and attempted to identify which ones were associated with certain types of activity.⁷² In particular, the challenge focused on

⁶⁶ Broniatowski, *supra* note 45, at 1378.

⁶⁷ Adrienne LaFrance, *The Internet Is Mostly Bots*, THE ATLANTIC (Jan. 31, 2017), <https://www.theatlantic.com/technology/archive/2017/01/bots-bots-bots/515043/>.

⁶⁸ Matthew Hughes, *Bots drove nearly 40% of internet traffic last year — and the naughty ones are getting smarter*, TNW (Apr. 18, 2019, 1:00 PM UTC), <https://thenextweb.com/security/2019/04/17/bots-drove-nearly-40-of-internet-traffic-last-year-and-the-naughty-ones-are-getting-smarter/>.

⁶⁹ The study defined malicious content as content spread by “bad bots.” GLOBAL DOTS, 2019 BAD BOT REPORT: THE BOT ARMS RACE CONTINUES 10 (2019), <https://www.globaldots.com/resources/blog/industry-report-bad-bot-landscape-2019-the-bot-arms-race-continues/>. These bots are automated programs that “scrape data from sites without permission in order to reuse it (e.g., pricing, inventory levels) and gain a competitive edge.” *Id.*

⁷⁰ *Id.*

⁷¹ *Id.*

⁷² V.S. Subrahmanian et al., *The DARPA Twitter Bot Challenge*, COMPUTER, June 13,

“influence bots,” which are “realistic, automated identities that illicitly shape discussions on social media sites like Twitter and Facebook, posing a risk to freedom of expression.”⁷³ While the study was largely procedurally oriented—seeking to examine strategies to detect, analyze, and categorize bots—it called attention to the growing sophistication of content automation strategies.⁷⁴

A seminal study published in 2018 analyzed the dissemination of vaccine-specific content through Twitter between July 2014 and September 2017.⁷⁵ The study contrasted automated programs (bots and content pollutants) with programs operated by humans but engaging in malicious behaviors (trolls).⁷⁶ In line with definitions of “bot” in both the technical and popular literatures, the study looked at bots as “accounts that automate content promotion” in the field of vaccines.⁷⁷ Content polluters were defined as “malicious accounts identified as promoting commercial content and malware” related to vaccines.⁷⁸ And trolls were characterized as accounts operating in the vaccine space and “exhibiting malicious behaviors yet operated by humans.”⁷⁹

The study found that accounts associated with sophisticated automated programs tweeted vaccine-related content at significantly higher rates than non-automated accounts.⁸⁰ Sophisticated programs are more complex and better at avoiding detection than average programs.⁸¹ The study also found that “content polluters” propagated the highest amount of anti-vaccine content, at a rate of 75% more than non-automated accounts.⁸²

One of the most salient findings of the study had to do with vaccine content disseminated through accounts located in areas of the world associated with the online promotion of misinformation and disinformation

2016, at 38, 39.

⁷³ *Id.* at 38. Other categories of bots categorized in the study but not the goal of the challenge were spambots, which spread content qualifying as spam, and paybots, which use different types of strategies to illicitly generate revenue for the account associated with that bot. *Id.*

⁷⁴ *Id.* at 44; see also Emerging Technology from the arXiv, *How DARPA Took On the Twitter Bot Menace with One Hand Behind Its Back*, MIT TECH. REV. (Jan. 28, 2016), <https://www.technologyreview.com/2016/01/28/163553/how-darpa-took-on-the-twitter-bot-menace-with-one-hand-behind-its-back/> (“The battle between bots and bot-hunters is one that is constantly evolving. With papers like this, the bot-hunters are revealing their hand in a way that allows bot-makers to design strategies to specifically defeat these algorithms. . . . Either way, this cat-and-mouse battle is set to continue.”).

⁷⁵ Broniatowski, *supra* note 45, at 1378.

⁷⁶ *Id.*

⁷⁷ *Id.*

⁷⁸ *Id.* at 1381.

⁷⁹ *Id.* at 1382.

⁸⁰ *Id.* at 1380.

⁸¹ *Id.* at 1382.

⁸² *Id.*

in other, more visible areas—namely Russia and the post-Soviet republics, which have been linked with political misinformation and disinformation.

Trolls traced backed to Russia were particularly active in spreading vaccine misinformation and disinformation.⁸³ They employed a two-pronged strategy, spreading both pro- and anti-vaccine content.⁸⁴ They did this as part of a broader strategy to tap into divisive topics among in the United States and increase discord.⁸⁵ Sophisticated Twitter bots traced to other locations also engaged in this type of activity.⁸⁶

An additional strategy employed by both bots and trolls included “flooding the discourse,” a technique that consists in increasing the circulation of content centered on a specific topic—in this case, through tweets about vaccines or vaccination—with the purpose of capturing social media traffic.⁸⁷ A related strategy consisted in the use of “astroturfing,”⁸⁸ a technique through which the origin of a message is made to appear as originating from grassroots organizations supporting a particular view.⁸⁹ In the case of vaccines, astroturfing on Twitter was linked to the propagation of anti-vaccine content, seemingly with the purpose of conveying the impression of the existence of grassroots debates about the efficacy of vaccines.⁹⁰ The study posited that trolls and semi-automated accounts (human-operated accounts occasionally also associated with bot activity) employed this strategy especially often, but recognized some limitations on and called for further study of the use of this particular technique in connection with Twitter vaccine discourse.⁹¹

The use of automated programs to spread vaccine-specific content has recently become intertwined with larger and more structured efforts to spread disinformation on multiple themes for geopolitical reasons. These efforts, emanating predominantly from Russia and certain post-Soviet actors, are meant to increase discord among Western communities by tapping into topics known to be polarizing.⁹² These efforts further seek to undermine the

⁸³ *Id.* at 1382.

⁸⁴ *Id.*

⁸⁵ *Id.*

⁸⁶ *Id.*

⁸⁷ *Id.* at 1380; *see also* Jeanette Sutton, *Health Communication Trolls and Bots Versus Public Health Agencies’ Trusted Voices*, 108 AM. J. PUB. HEALTH 1281, 1281 (2018) (contextualizing flooding techniques to propagate vaccine content through social media).

⁸⁸ Broniatowski, *supra* note 45, at 1382.

⁸⁹ *Id.* at 1382-84.

⁹⁰ *Id.* at 1382.

⁹¹ *Id.* at 1382, 1384.

⁹² *See, e.g.*, Jessica Glenza, *Coronavirus: US says Russia behind disinformation campaign*, GUARDIAN (Feb. 22, 2020), <https://www.theguardian.com/world/2020/feb/22/coronavirus-russia-disinformation-campaign-us-officials> (“Russia’s intent is to sow discord and undermine US institutions and alliances from within, including through covert and

credibility of public-sector institutions in the Western world, with a particular focus on the United States.⁹³ These strategies were employed even before COVID-19 was declared a pandemic, with social media accounts traced back to Russia disseminating disinformation across multiple health-related areas in the context of COVID-19 debates.⁹⁴

III. ONGOING SOCIAL MEDIA RESPONSES TO VACCINE MISINFORMATION

As more information on the specifics of vaccine misinformation has become available, social media have gradually taken steps to curb its spread. From 2019 onwards, mainstream social media have adopted a variety of approaches to deal with inaccurate vaccine-specific content made available by their users. As seen below, the onset of the COVID-19 pandemic in 2020 contributed to the adoption of seemingly increasingly stringent approaches to the moderation of vaccine-related content, described in Part A. However, as detailed in Part B, ongoing modes of self-regulation are highly heterogenous and present several practical problems.

A. *A Taxonomy of Mainstream Social Media Responses*

1. Suppression or “Zero Tolerance” Approaches

Suppression approaches, also known as “zero tolerance” policies, consist in the removal of content qualifying as inaccurate information. In the context of vaccine misinformation propagated through mainstream social media, this approach was pioneered by Pinterest. In February 2019, the company announced that it would block all vaccine-related search results.⁹⁵ The

coercive malign influence campaigns,’ said Philip Reeker, the acting assistant secretary of state for Europe and Eurasia.”). Iran has also been linked to the exploitation of health-related disinformation during the COVID-19 pandemic. *See, e.g.,* Ali Breland, *Russia Isn’t the Only Country Pushing Coronavirus Disinformation*, MOTHER JONES (Mar. 13, 2020), <https://www.motherjones.com/politics/2020/03/russia-iran-coronavirus-bioweapon/> (“Iranian media organizations linked to the government have also been taking advantage of fears over the virus to spread other conspiracies. . . . Iranians are pushing coronavirus conspiracies even more strongly than Russia.”).

⁹³ Glenza, *supra* note 92.

⁹⁴ *Id.*

⁹⁵ *See, e.g.,* Taylor Telford, *Pinterest is blocking search results about vaccines to protect users from misinformation*, WASH. POST (Feb. 21, 2019), <https://www.washingtonpost.com/business/2019/02/21/pinterest-is-blocking-all-vaccine-related-searches-all-or-nothing-approach-policing-health-misinformation/> (“As social media companies wrestle with how to police dangerous health misinformation on their platforms, Pinterest has taken an extreme approach: blocking search results related to vaccinations, whether the results are medically

decision was prompted by data revealing that most Pinterest searches about vaccines yielded results containing information that contradicted current scientific standards.⁹⁶

In the case of Pinterest, this approach still allowed for users to pin vaccine-related content to their personal board, but that content is not made available through the search function, remaining confined to personal pages.⁹⁷

Importantly, this was designed as a temporary measure.⁹⁸ Eventually Pinterest paired this approach with interventive steps, consisting of the display of vaccine content originating from legitimate sources—including the World Health Organization, the Centers for Disease Control and Prevention, and the American Academy of Pediatrics.⁹⁹

Also in early 2019, Pinterest blocked accounts linked to groups or individuals propagating vaccine misinformation and disinformation.¹⁰⁰ This approach resulted in the blocking of content promoted on Pinterest by the National Vaccine Information Center, an organization originally named Dissatisfied Parents Together and had no connection to federal agencies or public health authorities, which was launched in the 1980s and has since become one of the leading promoters of vaccine misinformation in the United States.¹⁰¹ Pinterest also blocked Larry Cook, a prominent figure in the anti-vaccine movement, particularly known for his use of social media channels to spread vaccine misinformation.¹⁰²

As seen below, while Pinterest took a zero-tolerance approach to the moderation of vaccine content even before the COVID-19 pandemic, other social media took less restrictive approaches. This allowed anti-vaccine discourses to continue circulating within large swaths of the mainstream social media space and, with the onset of the COVID-19 pandemic, the reach of anti-vaccine or vaccine-questioning content increased substantially. Larry

accurate or not.”).

⁹⁶ *Id.*

⁹⁷ *Id.*

⁹⁸ *Id.*

⁹⁹ Erin Brodwin, *How Pinterest beat back vaccine misinformation — and what Facebook could learn from its approach*, STAT (Sept. 21, 2020), <https://www.statnews.com/2020/09/21/pinterest-facebook-vaccine-misinformation/>. For an illustration of how Pinterest shares only vaccine content produced by reputable institutions in the public health space, see Pinterest, *Immunization* (last accessed Apr. 25, 2021), <https://www.pinterest.com/thephf/immunization/> (listing vaccine-related “pins” from institutions including the World Health Organization and UNICEF).

¹⁰⁰ Julia Carrie Wong, *Anti-vaxx propaganda has gone viral on Facebook. Pinterest has a cure*, GUARDIAN (Feb. 21, 2019), <https://www.theguardian.com/technology/2019/feb/20/pinterest-anti-vaxx-propaganda-search-facebook>.

¹⁰¹ *Id.*

¹⁰² *Id.*

Cook, for instance, maintained a Facebook group with close to 200,000 members until Facebook deleted it in late 2020.¹⁰³

The increased circulation of vaccine misinformation during the COVID-19 pandemic, which this Essay describes in greater detail in Part III.B.2, eventually prompted several mainstream social media to adopt suppression measures. In mid-October 2020, YouTube implemented a policy of removal of videos sharing vaccine misinformation.¹⁰⁴

2. Limiting and Downgrading Approaches

A different type of moderating approach consists of allowing anti-vaccine or vaccine-questioning discourses to circulate within a given social media (or group of related social media), but use techniques that will make this type of content less prominent to users.

Shortly after Pinterest implemented its initial suppression approach, Facebook announced that it would continue to allow anti-vaccine content to be shared—both on Facebook and on Facebook-owned Instagram—but that it would limit its reach by not allowing anti-vaccine content to be promoted through recommendations and ads.¹⁰⁵ Additionally, the company announced that it would downgrade vaccine misinformation showing up as search results.¹⁰⁶

As is the case with suppression approaches, limiting and downgrading approaches can also be paired with other interventions, such as the display of credible information on vaccines originating from reputable institutions.

3. Educational Approaches

Educational approaches consist of the active promotion of accurate content about vaccines and vaccination, typically sourced from credible scientific or public health-oriented organizations. For example, a search

¹⁰³ Aatif Sulleyman, *Facebook Bans One of the Anti-Vaccine Movement's Biggest Groups for Violating QAnon Rules*, NEWSWEEK (Nov. 18, 2020, 12:36 PM EST), <https://www.newsweek.com/facebook-bans-anti-vaccine-group-violating-qanon-rules-154840>.

¹⁰⁴ See Robert Hart, *YouTube Bans Covid-19 Vaccine Misinformation*, FORBES (Oct. 14, 2020, 12:10 PM EDT), <https://www.forbes.com/sites/roberthart/2020/10/14/youtube-bans-covid-19-vaccine-misinformation/?sh=8e0f4d868a9b> (“YouTube is cracking down on content peddling Covid-19 vaccine falsehoods, including banning anything that suggests a vaccine would kill people, cause infertility or be used to implant microchips into people getting vaccinated. The video platform has pledged to delete videos containing misleading claims about Covid-19 vaccines, it said in a blog post.”).

¹⁰⁵ Louise Matsakis, *Facebook Will Crack Down on Anti-Vaccine Content*, WIRED (Mar. 7, 2019, 5:29 PM), <https://www.wired.com/story/facebook-anti-vaccine-crack-down/>.

¹⁰⁶ *Id.*

performed on Facebook for the word “vaccine” in early February 2021 directed the user to content from prominent organizations in the following order: first, the U.S. Centers for Disease Control; second, the American Academy of Pediatrics; third, the American Cancer Society; fourth, the World Health Organization; and fifth, UNICEF.¹⁰⁷ A similar search performed on Twitter led to the appearance of a large informational panel noting that “[t]o make sure you get the best information on vaccinations, resources are available from the US Department of Health & Human Services,” and providing a link for, and the Twitter handle of, vaccines.gov, a website maintained by the Department of Health and Human Services.¹⁰⁸

Educational approaches were adopted broadly during the COVID-19 pandemic by players in the online environment, both within and outside social media. For instance, in December 2020, as the first COVID-19 vaccines were being authorized across the world, Google began displaying informational panels about each type of COVID-19 vaccine as part of search results related to vaccines.¹⁰⁹ This replicated a strategy previously employed by Google to counter general misinformation about the pandemic by showing informational panels with content provided by credible organizations on the SARS-CoV-2 virus and its spread, public health measures taken and recommended by public health authorities, as well as the location of COVID-19 testing centers.¹¹⁰

B. *Shortcomings of Current Self-Regulatory Approaches*

Overall, the steps taken by mainstream social media in response to the growing problem of vaccine misinformation have increased the amount and visibility of accurate information about vaccines in the online environment. Yet, as the example of moderation of vaccine content by Facebook presented in this section shows, social media responses have largely left the sources of

¹⁰⁷ The search was performed in English and from the United States.

¹⁰⁸ The search was similarly performed in English and from the United States. *See* U.S. DEP’T HEALTH & HUM. SERVS., *Vaccine Safety* (last accessed May 31, 2021), <https://www.vaccines.gov/basics/safety> (describing vaccine safety and providing answers to common questions about vaccine side effects).

¹⁰⁹ *See, e.g.*, Jon Porter, *Google search panels launch to counter vaccine misinformation*, THE VERGE (Dec. 10, 2020), <https://www.theverge.com/2020/12/10/22167185/google-vaccine-information-search-results-youtube-information-panels> (“Google is launching new vaccine information panels in its search results to counter misinformation and educate people about what’s available in their location, the company announced today. . . . The panels include information on each individual vaccine, and aim to address vaccine misinformation, and give guidance about how, when, and where people can receive the shots. Google has used similar information panels to share information about COVID-19 generally, as well as the locations of testing centers.”).

¹¹⁰ *Id.*

vaccine-specific misinformation untouched. Part III.B first articulates the general shortcomings of current modes of self-regulation and then presents a case study that illustrates the shortcomings in the response to vaccine misinformation adopted by the social media that is home to the largest amount of vaccine misinformation, Facebook.

1. Shortcomings of Social Media Self-Regulation

Social media sites have adopted policies to combat vaccine misinformation that are, on balance, somewhat similar. The dominant approach, particularly during the COVID-19 pandemic, has been to favor educational strategies, in some cases coupled with downgrading or other limiting policies.¹¹¹ Attempts to remove postings containing information flagged as vaccine misinformation have been rarer, with Pinterest's suppression approach applying to vaccine content in general remaining exceptional.¹¹²

The availability of accurate information about vaccines on social media is insufficient to counter the growing detrimental effects of online vaccine misinformation.¹¹³ Dissemination of accurate information alone does not guarantee that the content will be consumed or critically assimilated, and its impact on effecting behavioral changes in social media users pre-disposed to consume vaccine misinformation has been minimal.¹¹⁴

Large actors promoting accurate vaccine information through social media predominantly resort to vaccine literacy campaigns, which consist of sharing information from credible sources in response to searches for vaccine- or vaccination-related topics. However, as illustrated below in the case of Facebook, studies have suggested that these campaigns have a modest impact and are likely ineffectual in countering the disproportionate escalation of anti-vaccine and vaccine-questioning discourses in social media.¹¹⁵

Pages or accounts on mainstream social media promoting pro-vaccine views tend to have significantly larger followings than anti-vaccine or

¹¹¹ See e.g., Maryke S. Steffens et al., *Using Social Media for Vaccination Promotion: Practices and Challenges*, 6 DIGITAL HEALTH 1, 7 (2020) (surveying social media responses to vaccine misinformation).

¹¹² *Id.*

¹¹³ See Terry Connolly & Jochen Reb, *Toward Interactive, Internet-Based Decision Aid for Vaccination Decisions: Better Information Alone is Not Enough*, 30 VACCINE 3813, 3813 (2012) (“[I]nformation alone is not sufficient. Even with access to complete and well-validated information about vaccine- and disease-related outcomes and probabilities, the decision maker needs a procedure or mechanism to translate the information into an action recommendation—that is, a way to use the information effectively to make a decision.”).

¹¹⁴ *Id.* at 3813, 3816-17.

¹¹⁵ See *infra* Part III.B.2.

vaccine-questioning pages and accounts.¹¹⁶ For instance, the Facebook page for the Centers for Disease Control and Prevention (CDC) had around two million followers in early 2020. At the same time, a large Facebook page promoting anti-vaccine views would draw around 40,000 followers.¹¹⁷ Yet, anti-vaccine or vaccine-questioning pages or accounts tend to be dedicated to either just vaccine content or a narrow segment of health-related subjects. By contrast, pages or accounts associated with the CDC or the WHO are much more general purpose. As seen below, pages and accounts spreading negative views about vaccines were much more efficacious at spreading their content and reaching other users than pages conveying general-purpose, albeit pro-vaccine, information.¹¹⁸

Additionally, there is a significant asymmetry between the production of accurate and inaccurate vaccine-related content, particularly when the former is conceived as part of a set of educational materials.¹¹⁹ Content qualifying as misinformation or disinformation can be generated fairly quickly and inexpensively, as it does not rely on the collection, treatment and analysis of supporting data, such as studies or surveys. The creation of countervailing educational content, on the other hand, typically requires more resources, coordination of multiple data entry points and the time lag associated with obtaining those data and treating them for educational purposes. This can be especially pronounced in the case of emerging diseases like COVID-19, about which relatively little is known at first within the scientific community, whereas actors in the misinformation and disinformation arenas can quickly repackage existing content (e.g., unsubstantiated claims about measles vaccines producing a specific detrimental effect) and link it to quickly expanding debates in social media about new pathogens or public health problems (e.g., by making unsubstantiated claims about COVID-19 vaccines producing the same detrimental effect(s) claimed previously in connection with measles vaccines).¹²⁰

Another problem with vaccine literacy campaigns promoted by social media is that they are currently structured against the backdrop of the heterogenous and largely permissive approaches to vaccine content moderation. The adoption of less stringent modes of responses to the growth of social media-based vaccine misinformation creates a porous ecosystem in which this type of content can continue to circulate with relative ease. This

¹¹⁶ Jonhson, *supra* note 60, at 230.

¹¹⁷ *Id.*

¹¹⁸ *Id.* at 231.

¹¹⁹ See e.g., Renée DiResta, *Virus Experts Aren't Getting the Message Out*, ATLANTIC (May 6, 2020), <https://www.theatlantic.com/ideas/archive/2020/05/health-experts-dont-understand-how-information-moves/611218/> ("If the authorities can't satisfy the public's desire to know more, others will fill the void with misinformation.").

¹²⁰ *Id.*

porosity, allied with the time lag in responses between different social media, allows for the migration of inaccurate vaccine content. Consider the case of Larry Cook, one of the leading figures in the online anti-vaccination movement. Mr. Cook has long been known for orchestrating social media campaigns urging followers to both question the safety of vaccines and refrain from receiving recommended vaccines.¹²¹ Pinterest blocked Mr. Cook in February 2019.¹²² In response, he focused his activity on much larger platforms, Facebook and Twitter.¹²³ It was not until mid-November 2020, when Mr. Cook's role in spreading QAnon conspiracy theories (in addition to vaccine misinformation) became the subject of more publicized discussions, that these two platforms also blocked Mr. Cook.¹²⁴ This example highlights some of the costs of reliance on heterogenous modes of self-regulation in a borderless environment—namely how the lack of concerted efforts leaves significant swaths of online territory uncovered by the response of a single player; and the related time-gap problem, which in the example surveyed here covered both the months leading to the COVID-19 pandemic, as well as the entire period of the pandemic during which the first vaccines were being developed, tested and assessed by regulatory authorities in the United States and abroad.

Additionally, some social media platforms might not be motivated to respond to vaccine misinformation for a variety of reasons, including the fact that the dissemination of vaccine misinformation has become increasingly linked with possibilities of content monetization, as noted above. The case study on Facebook presented in the following section further illustrates how social media platforms themselves—and not just originators of misinformation—often make money if vaccine misinformation is consumed by large groups of users.¹²⁵

Finally, in addition to problems inherent to lack of uniformity, stringency, and motivation to intervene, current social media responses are also

¹²¹ See Sulleyman, *supra* note 103 (“Larry Cook, a prominent figure in anti-vaccine circles and the creator of the ‘Stop Mandatory Vaccination’ group on Facebook, has had his Facebook and Twitter accounts suspended. . . . In February, it emerged that the mother of a four-year-old boy who died of flu had opted not to use the Tamiflu medication that had been prescribed by a doctor, and was discussing natural ‘remedies’ with members of the Stop Mandatory Vaccination group instead.”); Wong, *supra* note 100 (“The policy change cleared the way for Pinterest to deploy a number of technological approaches to combating anti-vaxx propaganda. The company has banned boards by a number of prominent anti-vaccine propagandists, including the National Vaccine Information Center and Larry Cook, who runs the website and Facebook group ‘StopMandatory Vaccination.’”).

¹²² Wong, *supra* note 100.

¹²³ *Id.*; Sulleyman, *supra* note 103.

¹²⁴ Wong, *supra* note 100.

¹²⁵ See *infra* Part III.B.2.

hampered by technical limitations.¹²⁶ These limitations further accentuate concerns that overall approaches to the spread of vaccine misinformation across social media platforms may be lacking. Examples of these limitations are manifold. For instance, Facebook uses a machine-learning algorithm to detect both “hate speech and disinformation.”¹²⁷ Most content flagged by the algorithm as falling under either of these categories is then screened by a human, although in cases in which the algorithm determines that there is a high probability that the content amounts to hate speech or misinformation, the program removes it automatically.¹²⁸ The algorithm is very good at performing its screening function in areas where it has been trained on data for extended periods. By late 2019 it was able to detect 88.8% of all hate speech on Facebook, for example.¹²⁹ However, the emergence of events—and by extension, data—materially different from the ones on which the algorithm has been trained drastically reduces the accuracy and speed of automated screening. The time it takes to train the algorithm on emerging data needed to contextualize the screening process renders it much less effective for addressing large-scale events with a quick onset accompanied by an exponential increase in the proliferation of new forms of problematic content. This was the case of the COVID-19 pandemic, during which Facebook had to rely primarily on human-based screening of content potentially qualifying as COVID-related misinformation.¹³⁰

A study published in October 2020, as the first COVID-19 vaccines were poised to receive emergency authorization in the United States and Europe, showed that YouTube’s newly adopted policy of removing videos propagating COVID-19 misinformation routinely failed to capture a significant amount of anti-vaccine videos in Portuguese that were being shared on YouTube.¹³¹ Portuguese is the seventh most-spoken language in

¹²⁶ See *infra* notes 127-130 and accompanying text.

¹²⁷ Karen Hao, *Facebook’s AI is still largely baffled by covid misinformation*, MIT TECH. REV. (May 12, 2020), <https://www.technologyreview.com/2020/05/12/1001633/ai-is-still-largely-baffled-by-covid-misinformation/>.

¹²⁸ *Id.*; see also Guy Rosen, *Community Standards Enforcement Report, November 2019 Edition*, FACEBOOK (Nov. 13, 2019), <https://about.fb.com/news/2019/11/community-standards-enforcement-report-nov-2019/> (“[W]e’ve used these [AI] systems to proactively detect potential hate speech violations and send them to our content review teams since people can better assess context where AI cannot.”).

¹²⁹ Hao, *supra* note 127.

¹³⁰ *Id.*; see also Roshan Sumbaly et al., *Using AI to detect COVID-19 misinformation and exploitative content*, FACEBOOK (May 12, 2020), <https://ai.facebook.com/blog/using-ai-to-detect-covid-19-misinformation-and-exploitative-content> (“Since the pandemic began, we’ve used our current AI systems and deployed new ones to take COVID-19-related material our fact-checking partners have flagged as misinformation and then detect copies when someone tries to share them.”).

¹³¹ See Dayane Fumiyo Tokojima Machado et al., *Natural Stings: Selling Distrust About*

the world, with around 220 million native speakers.¹³²

All these shortcomings combine to form a social media ecosystem in which vaccine misinformation continues to travel with relative ease, even as more social media platforms revise their vaccine-specific misinformation policies—many of them driven by pressure stemming from the growth of misinformation during the COVID-19 pandemic.

2. The Growth of Vaccine Misinformation on Social Media During the COVID-19 Pandemic: The Case of Facebook

The propagation of misinformation about health-related topics grew significantly in recent years and gained new momentum with the COVID-19 pandemic, especially in the online environment.¹³³ As the race to develop COVID-19 vaccines became central to national and global responses to the pandemic, misinformation focused specifically on vaccines and vaccination proliferated especially quickly.¹³⁴

A study published by the Center to Counter Digital Hate in July 2020 examined over four hundred accounts associated with the largest social media platforms – Facebook, Facebook-owned Instagram, YouTube, and Twitter – looking for increases in the following of anti-vaccine or vaccine-questioning content.¹³⁵ Among the accounts surveyed, the largest 197 (measured by number of followers) had added over 8 million followers since 2019.¹³⁶ While

Vaccines on Brazilian YouTube, FRONTIERS COMM. (Oct. 26, 2020), <https://www.frontiersin.org/articles/10.3389/fcomm.2020.577941/full> (“Despite YouTube’s statements, M&D about vaccines continue to be disseminated in videos in Portuguese, reaching a large audience. We found 52 videos containing M&D about vaccines.”); *see also* Fernanda Ferreira, *Antivaccine videos slip through YouTube’s advertising policies, new study finds*, SCIENCE (Nov. 2, 2020), <https://www.sciencemag.org/news/2020/11/antivaccine-videos-slip-through-youtube-s-advertising-policies-new-study-finds> (“[A] new study in Brazil suggests YouTube’s moderation policies miss many anti-vaccination videos in Portuguese.”).

¹³² James Lane, *The 10 Most Spoken Languages in the World*, BABEL (Sept. 6, 2019), <https://www.babbel.com/en/magazine/the-10-most-spoken-languages-in-the-world>.

¹³³ Editorial Board, *Coronavirus Misinformation Needs Engagement*, 581 NATURE 355, 355 (2020).

¹³⁴ *See, e.g.*, Jason Murdock, *Anti-Vax Posts Against Future COVID-19 Vaccine Steadily Increasing on Social Media, Researchers Warn*, NEWSWEEK (Aug. 14, 2020), <https://www.newsweek.com/coronavirus-antivax-misinformation-rising-social-media-pinterest-research-1525073> (“Misinformation about a vaccine linked to the ongoing COVID-19 pandemic is ‘steadily rising’ on social media, experts say. . . . One of the things we are seeing on social media now is a steady increase of anti-future COVID-19 vaccine posts.”).

¹³⁵ CTR. COUNTERING DIGITAL HATE, *The Anti-Vaxx Industry: How Big Tech Powers and Profits from Vaccine Misinformation*, July 2020, at 5 [hereinafter *The Anti-Vaxx Industry*], https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_6910f8ab94a241cfa088953dd5e60968.pdf.

¹³⁶ *Id.*

the period surveyed in the study partly pre-dates the onset of the COVID-19 pandemic, it illustrates a trend that continued to unfold in later months of the pandemic.¹³⁷

Another 2020 study, published in *Nature*, provided data on the dynamics of vaccine misinformation disseminated through Facebook.¹³⁸ The takeaways from this study are especially valuable because Facebook, at over 2.6 billion users,¹³⁹ is both the most widely used social media platform and the largest social media venue for the propagation of vaccine misinformation.¹⁴⁰

The *Nature* study looked for Facebook users who had shared views about vaccines, irrespective of their leanings on the topic. It identified 100 million relevant accounts for which there was evidence that vaccine content was being sharing or consumed.¹⁴¹ Users active in the vaccine space on Facebook were located in multiple countries and operated in a variety of languages.¹⁴² Despite the ideological differences between users sharing or consuming pro- and anti-vaccine content, the study showed that Facebook users with an interest in vaccine content were becoming increasingly more connected.¹⁴³

¹³⁷ See, e.g., Alexandra S. Levine, *Misinformation About the vaccine could be worse than disinformation about the elections*, POLITICO (Dec. 21, 2020), <https://www.politico.com/news/2020/12/21/social-media-vaccine-misinformation-449770> (“Health experts have been warning about the need to prepare for vaccine misinformation since the early days of the pandemic.”); Lois Beckett, *Misinformation ‘superspreaders’: Covid vaccine falsehoods still thriving on Facebook and Instagram*, GUARDIAN (Jan. 6, 2021), <https://www.theguardian.com/world/2021/jan/06/facebook-instagram-urged-fight-deluge-anti-covid-vaccine-falsehoods> (“Researchers say big Facebook accounts still condemn vaccines while anti-vaxxers banned from Facebook have fled to Instagram. . . . Conspiracy theories and misinformation about the coronavirus vaccine are still spreading on Facebook and Instagram, more than a month after Facebook pledged it would take them down.”); Shannon Bond, *‘The Perfect Storm’: How Vaccine Misinformation Spread To The Mainstream*, NPR (Dec. 10, 2020), <https://www.npr.org/2020/12/10/944408988/the-perfect-storm-how-coronavirus-spread-vaccine-misinformation-to-the-mainstream> (noting the unabated circulation of vaccine misinformation in late 2020 and early 2021).

¹³⁸ Johnson, *supra* note 60, at 230.

¹³⁹ Facebook had over 2.6 billion users at the time of the *Nature* study. Andrew Hutchinson, *Facebook Closes in on New Milestone of 3 Billion Total Users Across its Platforms*, SOC. MEDIA TODAY (Apr. 29, 2020), <https://www.socialmediatoday.com/news/facebook-closes-in-on-new-milestone-of-3-billion-total-users-across-its-pla/577048/>. By the end of 2020, that number had increased to 2.8 billion. STATISTA, *Number of monthly active Facebook users worldwide as of 4th quarter 2020* (Feb. 2, 2021), <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>.

¹⁴⁰ See *The Anti-Vaxx Industry*, *supra* note 135 (“It is also clear that the anti-vaxx movement is most popular on Facebook.”).

¹⁴¹ Johnson, *supra* note 60, at 230.

¹⁴² *Id.*

¹⁴³ *Id.*

Mimicking what happens outside the context of social media, the majority of Facebook users active in the sharing or consumption of vaccine content have favorable views about vaccines.¹⁴⁴ Nonetheless, users questioning vaccines and sharing or consuming anti-vaccine content formed more clusters. A cluster was defined as a Facebook page and associated followers. The study found that vaccine-questioning and anti-vaccine clusters were twice as numerous as pro-vaccine clusters.

The same study also revealed that anti-vaccine clusters on Facebook were much better at reaching Facebook users with seemingly undecided views on vaccines or vaccination than pro-vaccine ones.¹⁴⁵ The study also showed that undecided users, estimated to amount to 50 million, are now considerably more active than what was previously known, conducting searches for vaccine related-content and engaging with anti-vaccine or vaccine-questioning clusters.¹⁴⁶ Strategies used by clusters disseminating anti-vaccine or vaccine-questioning content had become “robust and resilient,” thus significantly limiting the persuasive effect of techniques employed by pro-vaccine Facebook clusters—which consisted primarily of sharing vaccine information from credible institutions and sources.¹⁴⁷

During the first months of the COVID-19 pandemic, Facebook added 850,000 users following anti-vaccine or vaccine-questioning pages.¹⁴⁸ This bolstered the number of Facebook users following this type of content to close to 30 million – which corresponds to nearly half of all followers of vaccine misinformation across the most popular social media platform.¹⁴⁹

Before and throughout the pandemic, Facebook also remained the prime social media venue for the monetization of anti-vaccine or vaccine-questioning content through implementation of “marketing funnel” strategies.¹⁵⁰ These strategies consist of attracting followers through an initial channel, which then redirects users interested in further related content to venues outside the social media platform, where they are often invited to

¹⁴⁴ See *supra* note 61 and accompanying text.

¹⁴⁵ Johnson, *supra* note 60, at 230-31.

¹⁴⁶ *Id.* at 230 (“[A]nti-vaccination clusters dominate the main network patch in which they are heavily entangled with a very large presence of undecided clusters.”).

¹⁴⁷ *Id.*; see also Ana Santos Rutschman, *Facebook’s Latest Attempt to Address Vaccine Misinformation—And Why It’s Not Enough*, HEALTH AFF. BLOG (Nov. 5, 2020), <https://www.healthaffairs.org/doi/10.1377/hblog20201029.23107/full/> (“Facebook’s newest set of vaccine-specific measures. . . leaves the problem of the circulation of vaccine misinformation—the dissemination of inaccurate content—largely untouched and does virtually nothing to remove the well-established sources of vaccine misinformation within the Facebook network.”).

¹⁴⁸ *The Anti-Vaxx Industry*, *supra* note 135, at 14.

¹⁴⁹ *Id.*

¹⁵⁰ *Id.*

purchase goods or services.¹⁵¹ In the context of vaccine misinformation on Facebook, marketing funnels target users or followers of anti-vaccine pages by inviting them to sign up for email newsletters or similar information vehicles, through which additional vaccine- or health-related material is made available free of charge.¹⁵² This material often consists of videos hosted on other social media or personal webpages.¹⁵³ The final step in these funneling strategies occurs when this free content is then linked to offers to purchase additional content or goods. In the case of vaccines, these goods are often associated with products perceived as “alternative medicine,” “natural” (as opposed to vaccines, which in anti-vaccine discourse are portrayed as chemical-dense and thus harmful) or similar signifiers.¹⁵⁴ In some cases, consumption of these products can be problematic: studies show that Facebook pages enabling marketing funneling strategies in this area often link to goods like dietary supplements, a type of product over which the FDA has extremely tenuous oversight.¹⁵⁵ Some of the dietary supplements known to be offered for purchase at the end of marketing funneling strategies initiating from anti-vaccine and vaccine-questioning pages on Facebook have been the subject of warnings issued by public-health oriented agencies, such as the Food and Drug Administration. For instance, a dietary supplement sometimes promoted in this context is colloidal silver (silver particles in a liquid solution), for which the NIH maintains a warning, cautioning consumers that “[s]cientific evidence doesn’t support the use of colloidal silver dietary supplements for any disease or condition” and that “[i]n fact,

¹⁵¹ See, e.g., Anatoli Colicev et al., *Modeling the Relationship Between Firm and User Generated Content and the Stages of the Marketing Funnel*, 39 INT’L J. RES. MKT. 100, 100 (2018).

¹⁵² *The Anti-Vaxx Industry*, *supra* note 135, at 14.

¹⁵³ *Id.* at 15.

¹⁵⁴ See, e.g., Timothy Caulfield et al., *Injecting Doubt: Responding to the Naturopathic Anti-Vaccination Rhetoric*, 4 J. LAW & BIOSCIENCES 229, 235 (2017) (citing examples of language from naturopathic websites discussing vaccinations).

¹⁵⁵ See 21 U.S.C. § 321(ff) (2018) (defining “dietary supplement”); Roseann B. Termini & Vincent A. Sannuti, *A Look Back at the DSHEA — Over 25 Years Later: The Dangers of a Reactionary Approach to Dietary Supplement Regulation*, 22 QUINNIPIAC HEALTH L.J. 171, 176 (2019) (“Although many consumers are unaware, dietary supplements are not regulated in the same manner as conventional drug products.”); Joanna K. Sax, *Dietary Supplements Are Not All Safe and Not All Food: How the Low Cost of Dietary Supplements Preys on the Consumer*, 41 AM. J.L. & MED. 374, 374 (2015) (“Dietary supplements are regulated as food, even though the safety and efficacy of some supplements are unknown. These products are often promoted as ‘natural.’ This leads many consumers to fail to question the supplements’ safety, and some consumers even equate ‘natural’ with safe. But, ‘natural’ does not mean safe.”); Richard E. Nowak, *DSHEA’S Failure: Why a Proactive Approach to Dietary Supplement Regulation Is Needed to Effectively Protect Consumers*, 2010 U. ILL. L. REV. 1045, 1045 (collectively highlighting the shortcomings of the current regulatory regime for dietary supplements).

colloidal silver can be dangerous to your health.”¹⁵⁶

In addition to strategies that enable potential monetization of anti-vaccine or vaccine-questioning content in mediated ways, social media also provide a pathway for the direct monetization of this type of content, and social media platforms themselves can benefit from this pathway. The study conducted by the Center to Counter Digital Hate used the average revenue per person rate used by Facebook to determine how much the company could make directly by allowing anti-vaccine and vaccine-questioning content to be shared on its pages and groups.¹⁵⁷ By looking at both Facebook and Facebook-owned Instagram, the study estimated this direct revenue stream to be worth around US \$23 million.¹⁵⁸

While the COVID-19 pandemic presented heightened challenges, it also replicated some of the phenomena observed during social media responses to vaccine misinformation during preceding, smaller-scale public health crises.¹⁵⁹ The response from Facebook is also telling from this perspective, as pre-COVID-19 analyses of social media approaches to the spread of vaccine misinformation should have been regarded as informative about the shortcomings of ongoing forms of self-regulation.

As separate outbreaks of the same vaccine-preventable diseases (measles) occurred throughout the United States in 2019, several mainstream social media began altering their policies regarding the moderation of vaccine-specific content.¹⁶⁰ While Pinterest adopted a suppression approach, later paired with educational measures,¹⁶¹ Facebook and other mainstream social media elected to continue allowing users to post and share content qualifying

¹⁵⁶ U.S. NAT’L INST. HEALTH, *Colloidal Silver* (Apr. 2017), <https://www.nccih.nih.gov/health/colloidal-silver>.

¹⁵⁷ *The Anti-Vaxx Industry*, *supra* note 135, at 31.

¹⁵⁸ *Id.*

¹⁵⁹ See Broniatowski, *supra* note 30 (“The ‘Disneyland’ measles outbreak drew vaccine opposition into the political mainstream, followed by promotional campaigns conducted in pages framing vaccine refusal as a civil right. Political mobilization in state-focused pages followed in 2019.”).

¹⁶⁰ See Rachel Becker, *Facebook outlines plans to curb anti-vax conspiracy theories*, THE VERGE (Mar. 7, 2019), <https://www.theverge.com/2019/3/7/18255107/facebook-anti-vaccine-misinformation-measles-outbreaks-group-page-recommendations-removal> (“Facebook announced today that it will remove groups and pages that share anti-vaccine misinformation from its recommendations. The company also pledged that it will stop allowing advertisers to target people who Facebook’s advertising algorithm identifies as interested in ‘vaccine controversies.’”); Julia Belluz, *Facebook, Pinterest, and YouTube are cracking down on fake vaccine news*, VOX (Sept. 5, 2020), <https://www.vox.com/2019/3/1/18244384/measles-outbreak-vaccine-washington> (“Facebook is changing how users see vaccine content — part of an ongoing effort by social platforms to correct, block, and minimize vaccine misinformation amid a record measles outbreak.”).

¹⁶¹ See *supra* notes 95-99 and accompanying text.

as vaccine misinformation.¹⁶² Data show that, under a policy that left most of the sources of vaccine misinformation available on Facebook untouched, anti-vaccine and vaccine-questioning clusters grew exponentially more than other clusters.¹⁶³ The most active anti-vaccine or vaccine-questioning clusters grew by more than 300%.¹⁶⁴ By contrast, not a single pro-vaccine cluster exceeded a 100% increase in growth, with most growth rates for pro-vaccine clusters averaging less than 50%.¹⁶⁵ No similarly comprehensive data are yet available for other mainstream social media, although emerging studies indicate that social media that do not attempt to restrict anti-vaccine and vaccine-questioning content, opting primarily for educational approaches, are unlikely to curb the spread of vaccine misinformation.¹⁶⁶

The experience with the moderation of vaccine content before the pandemic also indicated that, even when companies announce policies that seemingly restrict the posting or sharing of vaccine misinformation, the actual effects of these policies can be minimal.¹⁶⁷ Once again, the case of Facebook is illuminating. In October 2020, the company announced that it would ratchet up its vaccine misinformation policy as part of an effort primarily designed to promote accurate information on vaccines in collaboration with public health agencies and other organizations.¹⁶⁸ As flu season was approaching, the new policy was built around the publicization of information about flu vaccines, touching indirectly on the possibility of COVID-19 vaccines becoming available shortly.¹⁶⁹ Per the new policy, Facebook would start promoting content about seasonal flu vaccines provided by credible organization, as well as providing users with sharable vaccination reminders.¹⁷⁰ At the same time, the company announced that the

¹⁶² Broniatowski, *supra* note 30, at S312; Belluz, *supra* note 160.

¹⁶³ Johnson, *supra* note 60, at 231.

¹⁶⁴ *Id.*

¹⁶⁵ *Id.*

¹⁶⁶ *Id.*; see also Beckett, *supra* note 137 (“The social network says it has limited the reach of some prominent anti-vaxx Facebook pages, and that few people are seeing some of the latest coronavirus misinformation. But misinformation experts say the platform’s actions amount to far too little, too late.”).

¹⁶⁷ See Johnson, *supra* note 60, at 230 (“Social media companies are struggling to control online health dis- and misinformation, for example, during the COVID-19 pandemic in 2020.”); Rutschman, *supra* note 147 (“Facebook’s newest set of vaccine-specific measures . . . leaves the problem of the circulation of vaccine misinformation—the dissemination of inaccurate content—largely untouched and does virtually nothing to remove the well-established sources of vaccine misinformation within the Facebook network.”).

¹⁶⁸ Kang-Xing Jin et al., *Supporting Public Health Experts’ Vaccine Efforts*, FACEBOOK (Oct. 13, 2020), <https://about.fb.com/news/2020/10/supporting-public-health-experts-vaccine-efforts/>.

¹⁶⁹ *Id.*

¹⁷⁰ *Id.*

policy would also include the rejection of ads explicitly “discouraging people from getting vaccinated.”¹⁷¹

Although the new policy concerning ads appears to fall under—and was described as—a suppression-like approach, in practice it was extremely limited.¹⁷² The policy created a very broad exception to the prohibition on anti-vaccine ads, by continuing to allow ads as long as they can be construed as advocating “for or against legislation or government policies around vaccines – including a COVID-19 vaccine.”¹⁷³ Facebook coupled the exceptional approach for these ads with a screening mechanism it deploys to moderate certain categories of speech, including political speech: the company screens the ad before making it available, and adds a label to the ad identifying who paid for it.¹⁷⁴ Similarly, the revised policy did not cover private pages sharing vaccine-related content on Facebook, whether pro- or anti-vaccine, effectively leaving the major Facebook-specific fora for the propagation of vaccine misinformation untouched.¹⁷⁵

Although the findings summarized in this section are based on data and studies focused on Facebook—reflecting a particular interest in the company, given its disproportionately large footprint in enabling online vaccine discourses—they are not restricted to Facebook.¹⁷⁶ They illustrate the norm, with suppression approaches like the one adopted by Pinterest being the exception. The current social media environment is therefore characterized by self-regulatory approaches to the moderation of vaccine-related content that do not remove or meaningfully isolate the sources of misinformation. As further discussed below, several commentators have begun to call for more uniformly stringent approaches to vaccine-specific misinformation circulating across social media, a problem to which the Essay now turns.

¹⁷¹ *Id.*

¹⁷² Rutschman, *supra* note 147.

¹⁷³ *Id.*

¹⁷⁴ *See id.* (describing the policy for vaccine-related ads as adopted in October 2020); *see also* FACEBOOK, *Get Authorized to Run Ads About Social Issues, Elections or Politics*, (Mar. 3, 2021), https://www.facebook.com/business/help/208949576550051?id=288762101909005&recommended_by=253606115253606 (displaying the most recent embodiment of the overall policy for ads relating to political and electoral content, as well as certain types of social debates).

¹⁷⁵ Johnson, *supra* note 60, at 230.

¹⁷⁶ Claire Wardle & Eric Singerman, *Too Little, Too Late: Social Media Companies’ Failure to Tackle Vaccine Misinformation Poses a Real Threat*, 372 BRIT. MED. J. 1, 1 (2021); *see also* Belluz, *supra* note 160 (“Facebook and Twitter and Google (which owns YouTube) have stated that they will take more action against false and misleading information about COVID-19 vaccines. This is undeniably positive, but these policy updates will not cover many types of posts that have the potential to lead to vaccine hesitancy.”).

IV. CURBING VACCINE MISINFORMATION AND DISINFORMATION: THE CASE FOR MORE STRINGENT SOFT-LAW APPROACHES

The Essay now considers potential ways of adding stringency to ongoing modes of self-regulation of vaccine misinformation and explores pathways to build cooperative monitoring and mutual assistance networks in this area. From a prescriptive perspective – and despite the limitations of ongoing modes of self-regulation, as described in Part III – the Essay focuses on improving self-regulation frameworks instead of exploring other types of regulation of vaccine content disseminated through social media for several reasons.

First, the regulation of speech about vaccines, if performed outside the context of self-regulatory approaches, raises significant legal issues. Because social media platforms are private actors, self-regulation as practiced in recent years largely escapes First Amendment scrutiny, which places the bulk of its limitations on the regulation of speech by government actors.¹⁷⁷ Proposals shifting the focus from self-regulation to other modes of regulating vaccine-related speech on social media would potentially run into First Amendment hurdles. Seeking to add stringency to the status quo by shifting away from self-regulation would potentially amount to imposing restrictions on speech based on its content. Content-based speech restrictions have long been understood as presumptively invalid¹⁷⁸ and, in light of both historical patterns and recent caselaw,¹⁷⁹ it seems that, at a minimum, proposals to

¹⁷⁷ U.S. CONST. amend. I. As Dawn Carla Nunziato has recently explained, the ways in which social media has responded to both political and health-related misinformation have been consistent with both substantive and procedural First Amendment safeguards applicable to private actors, as well as safeguards established elsewhere (e.g., by providing prior notice before restricting speech, social media has complied with due process requirements). See Dawn Carla Nunziato, *Misinformation Mayhem: Social Media Platforms' Efforts to Combat Medical and Political Misinformation*, 19 FIRST AMEND. L. REV. 32, 35–36 (2020) (“[A]pproaches undertaken by the major social media platforms are generally consistent with First Amendment values, both the substantive values in terms of what constitutes protected and unprotected speech, and the procedural values, in terms of process accorded to users whose speech is restricted or otherwise subject to action by the platforms.”).

¹⁷⁸ See *Simon & Schuster, Inc. v. Members of N.Y. State Crime Victims Bd.*, 502 U.S. 105, 115 (1991) (“A statute is presumptively inconsistent with the First Amendment if it imposes a financial burden on speakers because of the content of their speech.”); *Ashcroft v. Am. Civ. Liberties Union*, 542 U.S. 656, 660 (2004) (“[T]he Constitution demands that content-based restrictions on speech be presumed invalid.”).

¹⁷⁹ See *R.A.V. v. St. Paul*, 505 U.S. 377, 377 (1992) (“A few limited categories of speech . . . may be regulated because of their constitutionally proscribable content.”); *N.Y. Times Co. v. Sullivan*, 376 U.S. 254, 264 (1964) (“We hold that the rule of law applied by the Alabama courts is constitutionally deficient for failure to provide safeguards for freedom of speech and of the press that are required by the First and Fourteenth Amendments.”);

impose restrictions on vaccine-related speech on social media are likely to face close (and potentially harsh) constitutional scrutiny – a process that is both protracted in time and socially polarizing. The Essay is agnostic on the merits, legality or even desirability of regulating vaccine-related speech through instruments only available outside the context of self-regulation – it merely takes an approach that favors expediency in the face of the growing footprint left by vaccine misinformation circulating on social media, particularly against the backdrop of a pandemic during which levels of vaccine hesitancy have remained at problematic levels.

Therefore, the Essay focuses on measures that can be adopted in the short-term to improve existing, albeit imperfect, self-regulatory modes of addressing the problems posed by the propagation of vaccine misinformation through social media. Part IV.A looks at recent steps taken by the European Union in response to the proliferation of online misinformation—surveying the frameworks adopted before the COVID-19 pandemic, as well as changes made or proposed in direct response to the growth of misinformation during the pandemic. Although largely reliant on a soft law approach—through the adoption of non-binding instruments¹⁸⁰—European regulators succeeded in creating the first large-scale framework for the regulation of misinformation propagated through social media, as well as in attracting most mainstream social media as signatories.¹⁸¹ While the framework, which was designed to address misinformation in general, has had limited success in curbing the spread of vaccine misinformation, Part IV. B argues that the basic features of this framework can and should be adapted to respond to specific challenges posed by vaccine misinformation. The Essay concludes by suggesting several starting points in the creation of such a tailored framework.

A. *A Starting Point: Expanding the Existing Regulatory and Institutional Infrastructure*

The steps taken by the European Union to curb the growth of misinformation and disinformation during the COVID-19 pandemic built on pre-existing soft law frameworks targeting misinformation and

Miller v. California, 413 U. S. 15, 24 (1973) (“[W]e now confine the permissible scope of such regulation to works which depict or describe sexual conduct.”); United States v. Alvarez, 567 U.S. 709, 719 (2012) (“[F]alsity alone may not suffice to bring the speech outside the First Amendment.”).

¹⁸⁰ See, e.g., DANIEL BODANSKY, LEGALLY BINDING VERSUS NON-LEGALLY BINDING INSTRUMENTS, TOWARDS A WORKABLE AND EFFECTIVE CLIMATE REGIME 155 (Scott Barrett et al. eds., 2015) (“[C]onfident assertions, one way or the other, on the degree to which the legally binding nature of the Paris agreement does or does not matter seem unwarranted.”).

¹⁸¹ See *infra* note 213 and accompanying text (noting the absence of WhatsApp and TikTok).

disinformation at large. In 2015, following the reporting of online disinformation campaigns originating in Russia, the European Council established an ad hoc taskforce—the East Strategic Communication Task Force¹⁸²—and directed different groups to gather information, develop best practices and collaborate with European Union institutions, international organizations and other stakeholders across areas impacted by misinformation and disinformation.¹⁸³ Even though the first steps taken by the European Union were driven in large part by concerns and disinformation in connection with upcoming elections,¹⁸⁴ the resulting action plan, published in 2018, provided a general-purpose framework to address disinformation problems affecting both the European Union as a whole and individual member-states.¹⁸⁵ Specifically, the *Action Plan Against Disinformation* focused on disinformation spread through online channels and led to the development and adoption of the Code of Practice on Disinformation, the creation of the European Union Observatory Against Disinformation and of several educational initiatives on media and online literacy.¹⁸⁶

The Code of Practice on Disinformation has been hailed as marking the “first time worldwide that industry has agreed, on a voluntary basis, to self-regulatory standards to fight disinformation.”¹⁸⁷ It provides definitional clarity about what constitutes disinformation in the online environment; enunciates commitments adopted by industry signatories (which included several, although not all, of the mainstream social media platforms); and includes an annex describing a set of best practices to be adopted by industry, linking each recommended practice to a policy adopted by a specific social media platform, search engine or other online actor.¹⁸⁸

¹⁸² See EUR. EXTERNAL ACTION SERVICE, *Questions and Answers about the East StratCom Task Force* (Dec. 5, 2018), https://eeas.europa.eu/headquarters/headquarters-homepage/2116/-questions-and-answers-about-the-east-stratcom-task-force_en (“The Task Force was set up to address Russia’s ongoing disinformation campaigns.”).

¹⁸³ Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Action Plan against Disinformation*, EUR. COMM’N 1 (Dec. 5, 2018), <https://ec.europa.eu/digital-single-market/en/news/action-plan-against-disinformation>.

¹⁸⁴ *Id.* at 1-2.

¹⁸⁵ *Id.* at 1.

¹⁸⁶ EUR. UNION DISINFORMATION LAB, *Action Plan Against Disinformation*, <https://www.disinfo.eu/resources/eu-actions/other-eu-initiatives/>.

¹⁸⁷ EUR. COMM’N, *Code of Practice on Disinformation* [hereinafter *Code of Practice on Disinformation*], <https://ec.europa.eu/digital-single-market/en/code-practice-disinformation> (last updated May 28, 2021).

¹⁸⁸ *Id.*; see also EUR. COMM’N, *Code of Practice on Disinformation*, Annex II [hereinafter *Code of Practice Annex II*], https://ec.europa.eu/information_society/newsroom/image/document/2018-29/annex-to-msf-cop-on-disinformation-13-07-99F63CFE-A8CE-39BF687C68BFC0668569_53544.pdf.

The Code defines disinformation as “verifiably false or misleading information” meeting two cumulative criteria: first, it refers to content “created, presented and disseminated for economic gain or to intentionally deceive the public”; and second, such content “may cause public harm.”¹⁸⁹ Importantly, the Code defines public harm in the context of the spread of online disinformation as “threats to democratic political and policymaking processes as well as public goods such as the protection of EU citizens’ health, the environment or security.”¹⁹⁰ The inclusion of harms to public or individual health is especially relevant for debates on how to best address the specific problems caused by vaccine- and health-related misinformation.

Still at the definitional level, the Code also expressly recognizes categories that fall outside the purview of ongoing efforts undertaken by the European Union to curb the spread of disinformation in the online environment. It states that the operational concept of disinformation adopted by signatories of the Code “does not include misleading advertising, reporting errors, satire and parody, or clearly identified partisan news and commentary, and is without prejudice to binding legal obligations, self-regulatory advertising codes, and standards regarding misleading advertising.”¹⁹¹

The requirements set forth by the Code range from obligations relative to the placement of advertisements (such as allowing for third-party verification of content and using verification tools)¹⁹² to the prioritization of the promotion of accurate information to the detriment of misinformation.¹⁹³ To monitor actual adoption of these obligations, the Code required signatories to provide annual updates on their policies and practices,¹⁹⁴ and established a one-year assessment period.¹⁹⁵ Annex II of the Code currently complements this framework through a fourteen-page document providing language that develops each of the commitments articulated in the Code into discrete best practices.¹⁹⁶ The Annex further provides a link next to each best practice directing readers to actual policies adopted by signatories and currently in place.¹⁹⁷

The self-regulation framework laid out by the Code was complemented by the establishment of the European Union Observatory Against

¹⁸⁹ *Code of Practice on Disinformation*, *supra* note 187, Preamble.

¹⁹⁰ *Id.*

¹⁹¹ *Id.*

¹⁹² *Id.* at I, II.A.

¹⁹³ *Id.* at II.D.

¹⁹⁴ *Id.* at II.E., III.

¹⁹⁵ *Id.* at III.

¹⁹⁶ *Code of Practice Annex II*, *supra* note 188.

¹⁹⁷ *Id.*

Disinformation.¹⁹⁸ The Observatory collects information on disinformation, publishes reports, organizes educational events (such as webinars), publishes informational materials (such as toolkits on topics like data intelligence¹⁹⁹) and provides free fact checking technology to fact checkers, media organizations, researchers, social media and policy makers interested in collaborating with the Observatory.²⁰⁰

And finally, the educational portion of the European Union's strategy to curb disinformation was further developed through a 2018 amendment to the 2010 Media Directive.²⁰¹ While directives are binding legal instruments under European law²⁰² and hence escape the roster of soft law mechanisms surveyed in this section, it is worth noting here that it showcases how the European Union embedded a part (albeit small) of its programmatic response to online disinformation into a hard law instrument.

Drawing on this pre-existing framework, the European Union noted that the growth of online misinformation and disinformation during the COVID-19 pandemic required the adoption of more stringent measures to monitor, report and curb the spread of misinformation—particularly in cases in which the promotion of inaccurate content could lead to harms to individual or public health.²⁰³ In June 2020, the European Commission issued a

¹⁹⁸ EUR. UNION OBSERVATORY AGAINST DISINFORMATION, <https://www.disinfobservatory.org> (last visited June 1, 2021).

¹⁹⁹ *Id.*; see also SOCIAL OBSERVATORY FOR DISINFORMATION AND SOCIAL MEDIA ANALYSIS, *D3.3 Data Intelligence Toolkit Description*, 5 (Oct. 30, 2019), <https://www.disinfobservatory.org/wp-content/uploads/2020/05/D3.3-Data-Intelligence-toolkit-description.pdf> (“In this document, we focus on the algorithmic and technological engine of the platform: a data-driven social disinformation toolbox named DisInfoNet that supports the users of the SOMA platform in collecting and processing social media data with the goal of tracking popular and controversial contents, understanding the dynamics of (fake) news dissemination in social media, and identifying the origin and the broadcasters of false information.”).

²⁰⁰ SOCIAL OBSERVATORY FOR DISINFORMATION AND SOCIAL MEDIA ANALYSIS, *supra* note 199, at 5.

²⁰¹ Directive (EU) 2018/1808 of the European Parliament and of the Council (Nov. 14, 2018), amending Directive 2010/13/EU on the Coordination of Certain Provisions Laid Down by Law, Regulation or Administrative Action in Member States Concerning the Provision of Audiovisual Media Services (Audiovisual Media Services Directive) in View of Changing Market Realities, <https://eur-lex.europa.eu/eli/dir/2018/1808/oj>.

²⁰² See e.g., EUR. UNION, *Regulations, Directives and Other Acts*, https://europa.eu/european-union/law/legal-acts_en (last visited June 1, 2021) (“A ‘directive’ is a legislative act that sets out a goal that all EU countries must achieve.”).

²⁰³ Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Tackling COVID-19 Disinformation - Getting the Facts Right*, EUR. COMM’N 1 (Jun. 10, 2020) [hereinafter *Tackling COVID-19 Disinformation - Getting the Facts Right*], <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020JC0008>.

Communication about disinformation in the context of the pandemic.²⁰⁴ The Communication framed the proliferation of misinformation and disinformation as an “infodemic,”²⁰⁵ a term used to highlight “the perils of misinformation phenomena during the management of disease outbreaks.”²⁰⁶ It then called for a better calibration of response to misinformation and disinformation, and it urged stakeholders to distinguish “between illegal content, as defined by law, and content that is harmful but not illegal.”²⁰⁷ Noting that efforts should focus on cases in which there is “intention to deceive or cause public harm, or to make economic gain,” the Communication then called for “all parts of society” to be involved in a more robust response to the infodemic.²⁰⁸ The specific measures proposed by the European Commission included the development of better monitoring and reporting channels both within the European Union (between member-states, as well as between member-states and European regulators)²⁰⁹ and between the European Union and regional or international players.²¹⁰ At the regional level, the European Union singled out the Western Balkans and Africa as its “immediate neighbourhood,” and gave priority status to collaborations with countries in these areas.²¹¹

The Communication also requested more transparency from social media platforms in the implementation of their misinformation and disinformation policies, and argued that the monitoring and reporting mechanisms established in the Code of Practice on Disinformation should apply to non-signatories.²¹² This proposal specifically reacted to the absence of prominent social media from the self-regulation framework created by the Code, with the Commission singling out WhatsApp and TikTok.²¹³ However, the

²⁰⁴ *Id.*

²⁰⁵ *Id.* The use of the concept of “infodemic” matches conceptual approaches adopted by multiple international institutions. See WORLD HEALTH ORG., *Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation* (Sept. 23, 2020), <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> (providing a statement from several international organizations, including several institutions affiliated with the United Nations).

²⁰⁶ Matteo Cinelli et al., *The COVID-19 social media infodemic*, 10 NATURE SCI. REP. 1, 1 (Oct. 6, 2020) <https://www.nature.com/articles/s41598-020-73510-5>; see also J. Zarocostas, *How to Fight an Infodemic*, 395 LANCET 676, 676 (2020) (noting the proliferation of the word “infodemic” to describe the spread of misinformation).

²⁰⁷ *Tackling COVID-19 Disinformation - Getting the Facts Right*, *supra* note 203, at 3.

²⁰⁸ *Id.* at 3-4.

²⁰⁹ *Id.* at 5-6.

²¹⁰ *Id.* at 5-7.

²¹¹ *Id.* at 6-7.

²¹² *Id.* at 8.

²¹³ *Id.*

Communication was silent on how a non-binding framework could be extended to non-signatories or, alternatively, how non-signatories could be persuaded to self-regulate under the European framework.

Additional proposed measures included the allocation of more resources towards fact-checking²¹⁴ and educational campaigns.²¹⁵ Moreover, noting the exponential growth of health-related misinformation and disinformation during the COVID-19 pandemic, the Communication underscored how some aspects of disinformation can be dealt with through established bodies of law, particularly when monetization of health-related disinformation occurs through the offer to sell products for which there is very little to no regulatory oversight:

Manipulation, deceptive marketing techniques, fraud, and scams exploit fears in order to sell unnecessary, ineffective and potentially dangerous products under false health claims, or to lure consumers into buying products at exorbitant prices. Whilst this kind of content may contain disinformation, if it infringes the consumers' *acquis* [common rights and obligations binding all E.U. member states] it is illegal content and requires a different response, under consumers' protection laws and by competent authorities.²¹⁶

In response to this problem, the Commission proposed increasing funding for the Consumer Protection Cooperation Network, a network of national authorities responsible for the enforcement of consumer protection laws.²¹⁷ The Network was established in 2007 to allow national authorities to share best practices and to function as a “mutual assistance mechanism.”²¹⁸ European regulations make the provision of assistance mandatory through timely compliance with information requests and requests to assist with enforcement measures, as well as the obligation to alert the European Commission and other member states when a breach occurs.²¹⁹ Moreover,

²¹⁴ *Id.* at 9-10.

²¹⁵ *Id.* at 13-14.

²¹⁶ *Id.* at 14.

²¹⁷ *Id.* at 15; see also EUR. COMM'N, *Consumer Protection Cooperation Network*, https://ec.europa.eu/info/live-work-travel-eu/consumer-rights-and-complaints/enforcement-consumer-protection/consumer-protection-cooperation-network_en (last visited June 13, 2021) (“Thanks to the updated Consumer Protection Cooperation Regulation (EU) 2017/2394, national authorities now have stronger powers to detect irregularities and take speedy action against rogue traders. If national consumer authorities work together at EU level, they save taxpayers' money.”).

²¹⁸ EUR. COMM'N, *Single Market Scoreboard: Consumer Protection Cooperation (CPC): Reporting period: 01/2018 – 12/2018*, at 2 [hereinafter *Single Market Scoreboard*], https://ec.europa.eu/internal_market/scoreboard/_docs/2018/cpc/2018-scoreboard-cpc_en.pdf (last visited Dec. 19, 2021).

²¹⁹ Regulation (EU) 2017/2394 of the European Parliament and of the Council on cooperation Between National Authorities Responsible for the Enforcement of Consumer

Network engages in area-specific website monitoring, periodically conducting “sweeps,” systematic screenings of websites in a given sector “offering fixed/mobile phone, internet, audio and video streaming services.”²²⁰ The first sweep, conducted in 2017, monitored 207 websites and found 163 violations of European consumer law.²²¹ For the following sweeps, the Network partnered with authorities in countries outside the European Union (Iceland and Norway) to expand the reach of its monitoring activities, and increased the number of websites being monitored (560 in 2018 and 481 in 2019).²²²

Most recently, in late January 2021, the European Union called for greater efforts to nudge social media to begin de-monetizing disinformation.²²³ The European Commission is currently pondering the adoption of specific measures to achieve this goal, including potential additions to the Digital Services Act,²²⁴ a set of reforms—currently in draft form—to the European laws regulating the activity of online platforms (a category that includes, but is not limited to, social media platforms).²²⁵

B. *Incorporating Tools to Address Vaccine-specific Misinformation and Disinformation*

The combined reliance on soft law and social media voluntarism as the primary way of curbing the escalation of misinformation circulating online has so far failed to meaningfully address the problems posed by the propagation of vaccine-specific content—and especially inaccurate vaccine content, as described in Part III. Yet, ongoing soft law frameworks applicable to misinformation in general provide several important clues on how to develop monitoring and response mechanisms tailored to the characteristics

Protection Laws and Repealing Regulation, 2017 O.J. (L 345) 1-26, <https://eur-lex.europa.eu/eli/reg/2017/2394/oj>.

²²⁰ *Single Market Scoreboard*, *supra* note 218, at 3-5.

²²¹ *Id.* at 4.

²²² *Id.*; see also EUR. COMM’N, *Single Market Scoreboard: Consumer Protection Cooperation (CPC): Reporting period: 01/2019 – 12/2019*, https://ec.europa.eu/internal_market/scoreboard/performance_by_governance_tool/consumer_protection_cooperation_network/index_en.htm (last visited June 1, 2021) (“[T]he CPC authorities in 27 countries (25 EU countries, Norway and Iceland) screened 481 websites selling clothing and footwear, furniture, household items and appliances.”).

²²³ Samuel Stolton, *EU Commission Presses Platforms to De-Monetise Disinformation*, EURACTIV (Jan. 28, 2021), <https://www.euractiv.com/section/digital/news/eu-commission-presses-platforms-to-de-monetise-disinformation>.

²²⁴ Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC, ECOM/2020/825, Eur. Comm’n, <https://eur-lex.europa.eu/legal-content/en/TXT/?qid=1608117147218&uri=COM%3A2020%3A825%3AFIN> (last visited June 1, 2021).

²²⁵ *Id.*

of, and problems posed by, vaccine-specific misinformation. The Essay sets forth three arguments in support of creating a tailored system through soft law and self-regulation as opposed to other types of interventions. It then explains how this system could be developed to target vaccine misinformation.

First, creating a framework that specifically targets vaccine misinformation propagated through social media is necessary in light of the idiosyncrasies of the vaccine misinformation ecosystem, as social media communities active in vaccine debates operate in highly specialized ways vis-à-vis communities engaged in debates focused on other themes. Additionally, a tailored framework is necessary because of the specific challenges that consumption of vaccine misinformation may pose to individual and public health. Exposure to vaccine misinformation on social media has been linked with the growth of skepticism towards vaccines,²²⁶ the increase in vaccination delays²²⁷ and a decrease in vaccination rates.²²⁸ Research has also suggested that repeated exposure to vaccine misinformation on social media is likely to increase vaccine hesitancy,²²⁹ which in turn may lead individuals to skip or unnecessarily delay receiving a vaccine, or having their children vaccinated.²³⁰ The failure of current approaches in curbing the expansion of vaccine misinformation through loosely designed self-regulatory frameworks, allied to the health-related problems connected with consumption of vaccine misinformation circulating on social media, suggests that additional steps are necessary—and that future interventions should take

²²⁶ Broniatowski et al., *supra* note 30, at S312.

²²⁷ *Id.*; see also Daniel Jolley & Karen M. Douglas, *The effects of Anti-Vaccine Conspiracy Theories on Vaccination Intentions*, PLOS ONE, Feb. 2014, at 2 (describing how misinformation decreases overall vaccination rates, as shown in two studies). See generally Dubé, *supra* note 25.

²²⁸ Broniatowski et al., *supra* note 30, at S312.

²²⁹ *Id.*; see also Gordon Pennycook et al., *Prior Exposure Increases Perceived Accuracy of Fake News*, 147 J. EXP. PSYCH. GEN. 1865, 1865 (2018) (surveying the problem of repeated exposure to misinformation on a general level).

²³⁰ See Alessandro Facciola et al., *Vaccine Hesitancy: An Overview on Parents' Opinions About Vaccination and Possible Reasons of Vaccine Refusal*, 8 J. PUB. HEALTH RES. 13, 17 (2019) (“The argument of the vaccine hesitancy is particularly important for public health because, due to the parents’ refusal to vaccinate their children, we have assisted in last years to the recruitment of VPDs in many countries. This refusal is due especially to a fear about potential side effects of vaccines. However, many studies have shown that vaccines are sure and efficacies.”). But see Ángel V. Jiménez et al., *No Evidence That Omission and Confirmation Biases Affect the Perception and Recall of Vaccine-Related Information*, PLOS ONE, Mar. 2020, at 1 (“Contrary to the hypotheses, perceptions of severity and the recall of symptoms/side effects were not associated with experimental condition, failing to support the omission bias, nor did they interact with attitudes towards vaccines, failing to support the confirmation bias. This cast doubt on the possibility that the spread of anti-vaccination claims can be explained by these particular universal cognitive biases.”).

into greater account the features of vaccine-specific misinformation that differentiate it from other types of misinformation.

Second, and in spite of the failures noted above,²³¹ building a framework tailored to vaccine misinformation is best accomplished through continued reliance on soft law as a mechanism to achieve more stringency in self-regulation, rather than through interventions entailing the use of hard law mechanisms. A shift towards the adoption of binding legal mechanisms can provide a homogenous solution—by requiring that all social media implement the same measures; impose a stringent approach to the moderation of vaccine-related content—such as default suppression approaches that extend to private groups or pages within social media, coupled with the suppression of monetization channels;²³² and create an environment that facilitates enforcement actions—by establishing actionable behaviors and corresponding consequences. Yet, such a shift towards hard law would likely be met with several political economy hurdles and may even be counterproductive in light of the dynamics of contemporary anti-vaccine and vaccine-questioning discourses. On the political economy side of things, passing legislation at the domestic level is a time-consuming and politically fraught endeavor. Moreover, legislation imposing restrictions on content-based speech is bound to face protracted scrutiny and, depending on the specific embodiments of these restrictions, be of dubious legality—for instance, under long-held First Amendment jurisprudence in the United States, which has traditionally approached content-based restrictions as presumptively invalid²³³ and created very few permissible categorical exceptions.²³⁴ If hard law approaches were instead (or complementary) adopted at the transnational level, time and buy-in problems would also likely occur, as the negotiation of treaties and other multi-lateral agreements is notoriously lengthy, both procedurally and due to the need to harmonize competing geopolitical interests in areas often adjacent to the one being regulated.

In addition to political economy problems, singling out vaccine-related speech as exceptional may be counterproductive by increasing animosity towards government interventions in the ideologically fraught area of vaccines.²³⁵ As such, it can fuel the instrumentalization of anti-vaccine and

²³¹ See *supra* Part III.A (summarizing the shortcomings of current self-regulatory approaches to vaccine misinformation).

²³² See *infra* notes 150-156 and accompanying text.

²³³ *Simon & Schuster, Inc. v. Members of N.Y. State Crime Victims Bd.*, 502 U.S. 105, 115 (1991); *Ashcroft v. Am. Civ. Liberties Union*, 542 U.S. 656, 660 (2004).

²³⁴ *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942); *Brandenburg v. Ohio*, 395 U.S. 444 (1969); *R.A.V. v. St. Paul*, 505 U.S. 377 (1992); *United States v. Alvarez*, 567 U.S. 709 (2012).

²³⁵ See, e.g., Lucas B. Stolle et al., *Fact vs Fallacy: The Anti-Vaccine Discussion*

vaccine-questioning discourses—in particular, the strands of these discourses that border on conspiracy theories according to which the government, or representatives thereof, are portrayed as promoting semi-hidden vaccination agendas in multiple ways, including through the suppression of dissenting voices in vaccine debates.²³⁶ This is a problem that the use of soft law cannot fully address—although, by interposing social media and other entities (such as monitoring organizations) in the regulation of content related to vaccines, it may lessen perceived links between efforts to curb vaccine-specific misinformation and the role of the government in vaccine policy.

Against this backdrop, the third argument in favor of bolstering current frameworks for social media self-regulation of vaccine misinformation relies on the fact that there is already an incipient regulatory framework in place that can be further tailored to specific needs in this area. The approach taken by the European Union—which resulted in the current, albeit insufficient, self-regulation framework adopted by social media—provides some clues on how to create a tailored and potentially more stringent regime targeting vaccine misinformation. To begin with, a framework akin to the Code of Practice on Disinformation can and should be developed for vaccine misinformation: a “Code of Practice on Vaccine Misinformation,” providing conceptual clarity on what constitute both vaccine misinformation and vaccine disinformation; articulating commitments to be adopted by industry; and providing “best practice” language that can be incorporated into social media vaccine-specific policies.²³⁷

While creating a code does not translate into automatic industry adoption of the proposed framework, it can have an important nudging effect, particularly in light of current concerns about, and goodwill towards, vaccines and vaccination campaigns in the context of a global pandemic. Bargaining with social media to persuade them to adopt more stringent forms of moderation of vaccine content (i.e., defaulting to suppression approaches that may include blocking private groups spreading vaccine-specific misinformation) disrupts fewer established interests than wider approaches. Moreover, the COVID-19 pandemic has not only reawakened public perceptions about the public health value of vaccines. It has also called attention to the fact that scientists have long called attention to the likely increasing frequency at which pandemics or large-scale epidemics are likely

Reloaded, 37 ADV. THER. 4481, 4484 (2020) (noting that “government institutions” are linked to sentiment of hesitancy).

²³⁶ Jon D. Lee, *The Utter Familiarity of Even the Strangest Vaccine Conspiracy Theories*, THE ATLANTIC (Jan. 11, 2021), <https://www.theatlantic.com/ideas/archive/2021/01/familiarity-strangest-vaccine-conspiracy-theories/617572/>.

²³⁷ See *supra* note 188 and accompanying text.

to occur in the near future.²³⁸ Because the development and deployment of vaccines is predicted to play an important role in the response to these events, self-regulating actors may be more willing to adopt more stringent frameworks in this area as opposed to in connection with misinformation in general.

In addition to a more stringent regulatory framework, there is also a need for greater institutional monitoring dedicated to vaccine-specific misinformation circulating in the social media environment. Creation of an observatory-like structure in the United States, potentially modeled after the Observatory Against Disinformation established by the European Union and focused exclusively on vaccine misinformation,²³⁹ would contribute to the collection and analytical treatment of data on the idiosyncrasies of this subtype of misinformation. A vaccine misinformation observatory can also play important informative and educational roles, similar to the ones currently played by the Observatory Against Disinformation in its general-purpose approach to disinformation. Lastly, a dedicated observatory can also provide a basic technological infrastructure, in the form of screening tools that can be used by fact-checkers and other monitoring agents—and even social media themselves—operating in the vaccine misinformation ecosystem.²⁴⁰

Still at the institutional level, part of the European response to the surge of activity qualifying as disinformation during the COVID-19 pandemic hinged on broadening its cooperation both within member-states and externally.²⁴¹ A tailored response to vaccine misinformation could benefit from the formation of monitoring networks with mutual assistance protocols (or even obligations, as in the case of the E.U. Consumer Protection Cooperation Network) dedicated to the monitoring of, and data sharing on, vaccine misinformation.²⁴² These networks can be established at the national or transnational level (or both). In addition to focusing only geographical “immediate neighborhoods” as the European Union did, some countries may consider the formation of “language neighborhoods.” For instance, the United States might benefit from forming monitoring networks with other English-speaking countries. Conversely, given the shortcomings of current screening practices when misinformation circulates in languages other than English,²⁴³ non-English speaking countries may stand to gain from cooperating with countries where large segments of the population speak the

²³⁸ GAVI, *5 Reasons Why Pandemics Like COVID-19 Are Becoming More Likely* (Apr. 28, 2020), <https://www.gavi.org/vaccineswork/5-reasons-why-pandemics-like-covid-19-are-becoming-more-likely>.

²³⁹ See *supra* note 198 and accompanying text.

²⁴⁰ See *supra* note 200 and accompanying text.

²⁴¹ See *supra* notes 217-222 and accompanying text.

²⁴² See *supra* notes 209-211 and accompanying text.

²⁴³ See *supra* notes 131-132 and accompanying text.

same language, or proximate languages.

The steps outlined above—already adopted to some extent in Europe, albeit without a specific focus on vaccine misinformation, and with less stringent self-regulatory standards than the ones proposed here—do not require the commitment of extensive infrastructure or monetary commitments. At a time when the programmatic role of vaccination in popular discourse has changed dramatically, the adoption of more stringent self-regulatory frameworks likely faces fewer political economy constraints than ever before. Moreover, the development of stronger monitoring and cooperation frameworks designed to address this particular sub-set of misinformation aligns with public health and consumer protection goals acknowledged—at least nominally—by policymakers and regulators in several countries, including the United States.

The steps proposed here are also admittedly limited. On the one hand, they do not move the needle on the need for voluntary industry commitment if the regulation of vaccine misinformation is to be harmonized across social media—and especially when such harmonization entails the adoption of more stringent paradigms. On the other hand, they still require complementary interventions in other areas, such as the development of more accurate algorithms and other screening tools.²⁴⁴ Yet, on balance, they can contribute to the formation of better monitoring and response frameworks attuned to the specific problems posed by vaccine misinformation; drive the improvement of current best practices; and, at a minimum, exert pressure on mainstream social media for the adoption of more homogenous or stringent self-regulatory frameworks—or, ideally, both.

²⁴⁴ See *supra* notes 127-130 and accompanying text.