

2022

Private Censorship, Disinformation and the First Amendment: Rethinking Online Platforms Regulation in the Era of a Global Pandemic

Tzu- Chiang Huang
University of Michigan Law School

Follow this and additional works at: <https://repository.law.umich.edu/mtlr>



Part of the [Communications Law Commons](#), [First Amendment Commons](#), and the [Mass Communication Commons](#)

Recommended Citation

Tzu- Chiang Huang, *Private Censorship, Disinformation and the First Amendment: Rethinking Online Platforms Regulation in the Era of a Global Pandemic*, 29 MICH. TECH. L. REV. 137 (2022).
Available at: <https://repository.law.umich.edu/mtlr/vol29/iss1/5>

This Note is brought to you for free and open access by the Journals at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Michigan Technology Law Review by an authorized editor of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

PRIVATE CENSORSHIP, DISINFORMATION AND THE FIRST AMENDMENT: RETHINKING ONLINE PLATFORMS REGULATION IN THE ERA OF A GLOBAL PANDEMIC

*Tzu- Chiang Huang**

ABSTRACT

The proliferation of online disinformation and the rise of private censorship are paradigmatic examples of the challenges to traditional First Amendment jurisprudence in an algorithmic society. The limitations of traditional First Amendment jurisprudence are amplified by the impact of the COVID-19 pandemic in two ways. On the one hand, in the wake of the pandemic, we have entered an “infodemic” era where the volume of disinformation, as well as the harm it causes have reached unprecedented levels. For example, health disinformation has contributed to vaccine hesitancy. On the other hand, even though the proliferation of online disinformation seems to suggest that it is desirable to enforce content moderation more rigorously, the pandemic has also revealed the importance of access to online information, raising concerns about censorship imposed by private platforms on social media users. Furthermore, the high degree of opacity and unpredictability of content moderation pose great danger to users’ First Amendment right. In light of the above consideration, this Note proposes a legal framework that would curtail online disinformation while ensuring users’ right to accessing online platforms. To achieve this goal, this Note argues that the First Amendment should be interpreted as not merely a negative right but also a positive right. That is, the traditional *laissez-faire* First Amendment jurisprudence, which considers public actors as the sole threat to freedom of speech and neglects the power asymmetry between private platforms and their users, should be rejected. The underlying principle of the positive approach is to design a regulatory regime that is least restrictive and fosters accountability and transparency in content moderation by introducing pro-

* L.L.M. 2022, University of Michigan Law School. I would like to thank Professor Ming-Hsin Lin for his guidance and encouragement—this Note grew out of my master dissertation under his supervision at National Taiwan University. Thank you to Josiah Himmelman and Sandra Sulzer for your helpful comments and patience throughout the publication process. Thank you to the editors of Michigan Technology Law Review for your insightful feedback and thoughtful editing that make this Note possible. All remaining errors are my own.

cedural requirements. In this regard, the recently introduced Digital Services Act in the European Union—which represents a paradigmatic shift from interpreting freedom of speech as a “negative right” (i.e., protecting users from government interferences) to a “positive right” (i.e., ensuring the government provides users with sufficient procedural safeguards to check against private platforms)—could provide some important lessons for the U.S. to reconstruct its online platforms regulation in the era of an algorithmic society.

TABLE OF CONTENTS

INTRODUCTION	138
I. THE CAUSE: BIAS AND ALGORITHMS	143
A. <i>Rational Audience and Falsehood</i>	144
B. <i>Algorithms, Echo Chambers and Bots</i>	146
II. THE HARM: DIMINISHED TRUST AND VACCINE HESITANCY	150
III. THE ANTIDOTE: TRANSPARENCY AND ACCOUNTABILITY IN ONLINE CONTENT MODERATION	153
A. <i>The Peril of Private Content Moderation</i>	154
B. <i>Reconstructing the First Amendment Jurisprudence in the Algorithmic Society</i>	157
C. <i>Towards a More Accountable and Democratic Platform</i>	159
CONCLUSION	162

INTRODUCTION

On May 18, 2022, Lee C. Bollinger, President of Columbia University and a preeminent First Amendment scholar, remarked the advent of “the Age of Disinformation” in his Commencement address at Columbia University.¹ The Age of Disinformation is characterized by the modern phenomenon of systematic campaigns of disinformation designed to undermine the democratic political process and erode “the very idea of deep knowledge and expertise itself” by soliciting fear and anger, such as denials of election integrity and the effectiveness of vaccines.² What makes the issue even more concerning is that propagandists exploit the technological advancement of communication tools, social media in particular, to augment the effectiveness of disinformation campaigns.³ In light of an over-abundance of freely expressed but deeply misguided speech that poses significant threats to democracies, Bollinger reflected on the laissez-faire First Amendment jurisprudence holding that the purpose of free speech is best served when it

1. Lee C. Bollinger, President, Columbia Univ., 2022 Commencement Address (May 18, 2022) (transcript available at <https://president.columbia.edu/content/2022-commencement-address>).

2. *Id.*

3. *Id.*

is free from government regulation—an approach that has dominated online platforms governance for more than two decades⁴— and asked: “Does this basic premise, does all of this still hold true?”⁵

The proliferation of online disinformation and the rise of private censorship are two fundamental challenges to the long-standing laissez-faire First Amendment jurisprudence in an algorithmic society. In the wake of the 2016 U.S. presidential election and the Brexit referendum, the study of “disinformation”⁶ (or “fake news”⁷) has become prevalent worldwide. It is worth noting that while the existence of disinformation is anything but new,⁸ what has made the problem of disinformation particularly problematic in recent years is the development of social media.⁹ Social media has made publishing and receiving information easier for the general public. However, the democratization of information has also made it easier than ever for hostile

4. See generally Danielle Keats Citron & Mary Anne Franks, *The Internet as a Speech Machine and Other Myths Confounding Section 230 Reform*, 2020 U. CHI. LEGAL F. 45 (2020).

5. Bollinger, *supra* note 1.

6. In recent years, most legal scholars and commentators have substituted the term “fake news” with “disinformation” due to its imprecision and its implication to undermine trust in mainstream media. See Lili Levi, *Real “Fake News” and Fake “Fake News”*, 16 FIRST AMEND. L. REV. 232, 233–34 (2018). Disinformation is generally defined as information containing falsehood that aims to mislead the public. See generally Ana Santos Rutschman, *The COVID-19 Vaccine Race: Intellectual Property, Collaboration(s), Nationalism and Misinformation*, 64 WASH. U. J.L. & POL’Y 167, 199 (2021) (defining disinformation as “inaccurate information is circulated with the specific aim of sowing doubt or increasing disagreements between people or institutions with different viewpoints”); Ruairi Harrison, *Tackling Disinformation in Times of Crisis: The European Commission’s Response to the Covid-19 Infodemic and the Feasibility of a Consumer-centric Solution*, 17 UTRECHT L. REV. 18, 22 (2021) (defining disinformation as “verifiably false or misleading information’ which is created for economic gain or to deceive the public, and may cause public harm.”).

7. The term “fake news” is ambiguous in nature, while the precise definition differs slightly amongst scholars, most of whom agree that fake news contains content that is intentionally and misleadingly false. See generally Edson C. Tandoc, Jr., Zheng Wei Lim & Richard Ling, *Defining “Fake News”*, 6 DIGIT. JOURNALISM 137, 141–47 (identifying six categories of fake news including news satire, news parody, news fabrication, photo manipulation, advertising and public relation and propaganda); Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSP.S 211, 213 (2017) (defining fake news as “news articles that are intentionally and verifiably false, and could mislead readers”); Ari Ezra Waldman, *The Marketplace of Fake News*, 20 U. PA. J. CONST. L. 845, 849 (2018) (defining fake news as “misinformation designed to mislead readers by looking like and coming across as traditional media”); Axel Gelfert, *Fake News: A Definition*, 38 INFORMAL LOGIC 85–86, 108 (2018) (defining fake news as “the deliberate presentation of (typically) false or misleading claim as news, where the claims are misleading by design.”).

8. See generally CALIN O’CONNOR & JAMES OWEN WEATHERALL, *THE MISINFORMATION AGE* 9–11 (2018).

9. In light of the significant role of social media in the recent phenomenon of disinformation, some commentators include using online communicational tools as the medium into the definition of disinformation or fake news. See, e.g., David O. Klein & Joshua R. Wueller, *Fake News: A Legal Perspective*, J. INTERNET. L. 6, 6 (2017) (defining fake news as “the online publication of intentionally or knowingly false statements of facts”).

actors to spread disinformation for political or economic gain.¹⁰ This is because social media algorithms allow posters to micro-target their audience and prevent them from receiving debunking information, while also substantially minimizing the cost of disseminating disinformation.¹¹

Many legal scholars have considered whether and how to reconstruct First Amendment jurisprudence in an algorithmic society. Some adhere to the traditional *laissez-faire* approach and assert that if an authority is allowed to police online speech, it would enable the authority to act as the “arbiter of truth,” which is a gross infringement of freedom of speech.¹² For example, the German model of Network Enforcement Act (“NetzDG”) has been subject to harsh criticism as it imposes censorship and strict speech law on social media platforms, resulting in overcorrection of otherwise lawful speech.¹³ Under this traditional *laissez-faire* approach, the First Amendment is considered to be a “negative right” that discourages government intervention and prioritizes self-regulation by social media platforms.¹⁴ In contrast, others have proposed a new First Amendment paradigm in response to the modern information ecosystem.¹⁵ Professor Balkin famously argued that the traditional “dualist model”¹⁶ of speech regulation has been replaced by a pluralist “triangle” model involving multiple players: governments, social media platforms, and speakers.¹⁷ Under this model, the government is no longer the only potential source of censorship—private

10. See generally Andrew M. Guess & Benjamin A. Lyons, *Misinformation, Disinformation, and Online Propaganda*, in *SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM* 10 (Nathaniel Persily & Joshua A. Tucker eds., 2020); Liza Potts & Stephanie Mahnke, *Subverting the Platform Flexibility of Twitter to Spread Misinformation*, in *PLATFORMS, PROTESTS, AND THE CHALLENGE OF NETWORKED DEMOCRACY* 157, 157–72 (John Jones & Michael Trice eds., 2020).

11. See *infra* notes 60–75 and accompanying text.

12. See Clay Calvert, Stephanie McNeff, Austin Vining & Sebastian Zarate, *Fake News and the First Amendment: Reconciling a Disconnect Between Theory and Doctrine*, 86 U. CIN. L. REV. 99, 136–38 (2018).

13. See Rebecca Zipursky, *Nuts About NETZ: The Network Enforcement Act and Freedom of Expression*, 42 FORDHAM INT’L L.J. 1325, 1328, 1359–63 (2019); Flavia Durach, Alina Bârgoanu & Cătălina Nastasiu, *Tackling Disinformation: EU Regulation of the Digital Space*, 20 ROMANIAN J. EUR. AFF.S 5, 12 (2020).

14. See generally Daniela C. Manzi, *Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News*, 87 FORDHAM L. REV. 2623, 2633–41 (2019).

15. See, e.g., Tim Wu, *Is the First Amendment Obsolete?*, 117 MICH. L. REV. 547, 554 (2018) (arguing that speech is no longer scarce, but it is listeners’ attention that is highly limited); PHILIP M. NAPOLI, *SOCIAL MEDIA AND THE PUBLIC INTEREST: MEDIA REGULATION IN THE DISINFORMATION AGE* 191–93 (2019) [hereinafter NAPOLI, *SOCIAL MEDIA AND THE PUBLIC INTEREST*] (arguing that freedom of speech should be understood as “collective rights” instead of “individual rights”).

16. Jack M. Balkin, *Free Speech Is a Triangle*, 118 COLUM. L. REV. 2012, 2013–14 (2018) [hereinafter Balkin, *Free Speech Is a Triangle*] (referring to a dualist system where a nation-state is the only actor regulating speech).

17. *Id.* at 2014–15.

platforms are as well.¹⁸ In addition, some legal scholars espouse transparency requirements instead of censorship. One example of this regulatory regime is the Honest Ads Act bill, which imposes disclosure requirements on online political advertisements.¹⁹

Though many legal frameworks have been proposed to combat online disinformation and ensure users' right to online information, disinformation remains rampant.²⁰ Users also remain subject to privatized content moderation which is based on unaccountable decision-making and opaque procedures.²¹ The challenges to free speech in an algorithmic society are further intensified by the impact of the COVID-19 pandemic in two ways.²² First, in the wake of the COVID-19 pandemic, we have entered an era of "infodemic"²³ where the volume of disinformation as well as the harm it causes have reached unprecedented levels. In 2020, the World Health Organization flagged the "uncontrolled dissemination of misinformation" as one of the most urgent health challenges of the next decade.²⁴ The two most problematic types of disinformation are political disinformation and health disinformation.²⁵ The two types may overlap where COVID-19 health disinformation is utilized to promote a political agenda, such as the spread of anti-vaccine sentiment, which has contributed significantly to growing vaccine hesitancy.²⁶

18. *Id.* at 201–25. See also Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2298 (2014); Ashutosh Bhagwat, *Free Speech Categories in the Digital Age*, in FREE SPEECH IN THE DIGITAL AGE 88, 91–94 (Susan J. Brison & Katharine Gelber eds., 2019).

19. Dawn Carla Nunziato, *The Marketplace of Ideas Online*, 94 NOTRE DAME L. REV. 1519, 1554–56 (2019).

20. See Blake Hounshell, Sheera Frenkel, Tiffany Hsu & Stuart A. Thompson, *A Journey Into the Misinformation Fever Swamps*, N.Y. TIMES (Aug. 26, 2022), <https://www.nytimes.com/2022/08/26/us/politics/misinformation-social-media.html>.

21. See Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149, 1197 (2018) [hereinafter Balkin, *Free Speech in the Algorithmic Society*]; see also Eric Goldman, *Content Moderation Remedies*, 28 MICH. TECH. L. REV. 1, 8–12 (2021).

22. See Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution to Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2496–98 (2020); see also GIOVANNI DE GREGORIO, *DIGITAL CONSTITUTIONALISM IN EUROPE: REFRAMING RIGHTS AND POWERS IN THE ALGORITHMIC SOCIETY* 182 (Mark Dawson et al. eds. 2022).

23. *Infodemic*, MACMILLAN DICTIONARY, <https://www.macmillandictionary.com/buzzword/entries/infodemic.html> (last visited May 6, 2020).

24. *Urgent Health Challenges for the Next Decade*, WORLD HEALTH ORG., <https://www.who.int/news-room/photo-story/photo-story-detail/urgent-health-challenges-for-the-next-decade> (Jan. 13, 2020) (noting that "[p]ublic health is compromised by the uncontrolled dissemination of misinformation in social media, as well as through an erosion of trust in public institutions.").

25. Harrison, *supra* note 6, at 20.

26. *Id.*

Second, even though the proliferation of online disinformation suggests that it is desirable to enforce content moderation more rigorously,²⁷ the pandemic has also revealed the importance of access to online information, raising concerns about censorship imposed by private platforms on social media users. Since the COVID-19 outbreak, most countries have imposed strict social-distancing policies and severely restricted face-to-face contact, forcing the public to rely even more heavily on online communication than ever before, thereby reinforcing the already enormous influence of social media platforms.²⁸ The decisions of Facebook and Twitter to block the account of former president Donald Trump in the aftermath of the capital riot is one noted example of social media platforms' absolute power over online information. After the ban, Trump's political influence diminished significantly due to the lack of a channel to convey his messages.²⁹ In times of a pandemic, accessing online platforms has become a prerequisite for individuals to participate in public discourse. However, the process of content moderation often lacks transparency and is enforced under platforms' sole discretion, thus preventing users from understanding how content is moderated in the digital environment.³⁰

This Note proposes a regulatory framework which would curtail online disinformation and ensure users' right to access online platforms. Part I discusses how social media algorithms and their enforcement of content moderation contribute to the proliferation of online disinformation, as well as its implications to the "more speech" doctrine enshrined in traditional First Amendment jurisprudence. Part II underlines the severity of the issue by analyzing how online disinformation frustrates democratic values by facilitating distrust in public institutions, which in turn engenders grave public health concerns such as increased vaccine hesitancy. Part III discusses how the traditional laissez-faire First Amendment paradigm cannot adequately address the contemporary issues of free speech and provides arguments in favor of a "positive right" approach. The underlying principle of the positive approach, rather than compelling platforms to carry or remove certain content, is to design a regulatory regime that is least restrictive and fosters accountability and transparency in content moderation by introducing procedural requirements. In this regard, the recently introduced Digital Services Act in the European Union— which represents a paradigmatic shift from interpreting freedom of speech as a "negative right" to a "positive right"—

27. See, e.g., Mike Isaac, *Facebook Says It Plans to Remove Posts with False Vaccine Claims*, N.Y. TIMES (Feb. 8, 2021), <https://www.nytimes.com/2021/02/08/technology/facebook-vaccine-misinformation.html>.

28. Klonick, *supra* note 22, at 2497.

29. See Dylan Byers, *How Facebook and Twitter Decided to Take Down Trump's Accounts*, NBC NEWS (Jan. 14, 2021), <https://www.nbcnews.com/tech/tech-news/how-facebook-twitter-decided-take-down-trump-s-accounts-n1254317>.

30. GREGORIO, *supra* note 22, at 207.

could provide valuable guidance for ways the U.S. could reconstruct its online platforms regulation in the era of an algorithmic society.

I. THE CAUSE: BIAS AND ALGORITHMS

A central theme of the First Amendment—commonly known as the “marketplace of ideas”—is that more speech is the most effective remedy against the dissemination and consumption of falsehood.³¹ The “marketplace of ideas” theory was first introduced into First Amendment jurisprudence by Justice Holmes in his dissent in *Abrams v. United States*,³² in which Justice Holmes argued that “the ultimate good desired is better reached by free trade in ideas.”³³ The notion was further elaborated on when Justice Brandeis proposed the counterspeech doctrine in *Whitney v. California*:³⁴ “If there be time to expose through discussion the falsehood and fallacies, to avert the evil by the process of education, the remedy to be applied is more speech, not enforced silence.”³⁵ More recently, the Supreme Court ruled that falsehood is constitutionally protected because “[o]ur constitutional tradition stands against the idea that we need Oceania’s Ministry of Truth.”³⁶ Therefore, falsehood should be overcome by counterspeech instead of censorship.³⁷ It is evidence that the “marketplace of ideas” theory, which endorses a laissez-faire approach, has become the dominant theory in First Amendment jurisprudence.³⁸ Under this laissez-faire approach, direct government regulation of speech is minimized in favor of an open and competitive speech environment in which truth would eventually prevail.

31. See generally ROBERT C. POST, *DEMOCRACY, EXPERTISE, AND ACADEMIC FREEDOM: A FIRST AMENDMENT JURISPRUDENCE FOR THE MODERN STATE* 5-18 (2012) (discussing the First Amendment and “marketplace of ideas”).

32. *Abrams v. United States*, 250 U.S. 616, 624–31 (1919) (Holmes, J., dissenting).

33. *Id.* at 630.

34. *Whitney v. California*, 274 U.S. 357, 377 (1927) (Brandeis, J., concurring).

35. *Id.*

36. *United States v. Alvarez*, 567 U.S. 709, 723 (2012) (plurality opinion).

37. See *id.* at 726. For a more detailed discussion of whether and how false speech is protected by the First Amendment, see Erwin Chemerinsky, *False Speech and the First Amendment*, 71 OKLA. L. REV. 1, 5–10 (2018).

38. See LEE C. BOLLINGER, *THE TOLERANT SOCIETY* 18 (1986) (“[T]he *Abrams* dissent of Holmes stands as one of the central organizing pronouncements for our contemporary vision of free speech”); Vincent Blasi, *The First Amendment and the Ideal of Civic Courage: The Brandeis Opinion in Whitney v. California*, 29 WM. & MARY L. REV. 653, 668 (1988) (“It is [. . .] arguably the most important essay ever written, on or off the bench, on the meaning of the first amendment”); Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1455 (2013) (“The best-known conception, and that most commonly invoked by the Supreme Court, is the marketplace of ideas.”). *But see* Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 48 (1984) (raising objections to this theory); Paul H. Brietzke, *How and Why the Marketplace of Ideas Fails*, 31 VAL. U. L. REV. 951, 952 (1997).

The “marketplace of ideas” is based on the following premises: (1) information is scarce, and therefore governmental regulation is the primary threat to freedom of speech;³⁹ (2) both truth and false information are equally accessible to the audience;⁴⁰ (3) after being exposed to both kinds of information, the audience is capable of distinguishing truth from falsity,⁴¹ and (4) most importantly, the audience has a preference for truthful information over falsity.⁴² However, these presumptions are fundamentally flawed. The following section first discusses how the assumption of a “rational audience” is challenged by behavioral economics and cognitive psychology studies,⁴³ and more critically — how social media rewrites the rules of mass communication and facilitates an information ecosystem promoting disinformation that poses a significant threat on both political process and public health issues.

A. Rational Audience and Falsehood

Individuals are influenced by filters, biases, and heuristics in the process of acquiring, selecting, and digesting information, making it impossible for most people to make decisions rationally.⁴⁴ That is, the choice of where to devote our limited attention in face of mass information will significantly affect how we think and act. In particular, people prefer information that confirms their original assumptions.⁴⁵ Studies have also illustrated that people tend to select information in a way that almost guarantees the confirmation of their pre-determined position, a phenomenon generally referred to as “confirmation bias.”⁴⁶ Therefore, the effectiveness of “more speech” is lim-

39. Wu, *supra* note 15; see also Vincent Blasi, *Holmes and the Marketplace of Ideas*, 2004 SUP. CT. REV. 1, 4 (2005).

40. See Vincent Blasi, *Reading Holmes through the Lens of Schauer: The Abrams Dissent*, 72 NOTRE DAME L. REV. 1343, 1357 (1997).

41. Ingber, *supra* note 38, at 15; Lyriisa Barnett Lidsky, *Nobody’s Fools: The Rational Audience as First Amendment Ideal*, 2010 U. ILL. L. REV. 799, 801 (2010); Philip M. Napoli, *The Marketplace of Ideas Metaphor in Communications Regulation*, 49 J. COMM’N 151, 153 (1999).

42. Philip M. Napoli, *What If More Speech Is No Longer the Solution: First Amendment Theory Meets Fake News and the Filter Bubble*, 70 FED. COMM. L.J. 55, 61 (2018) [hereinafter Napoli, *First Amendment Theory Meets Fake News*]; Alvin I. Goldman & James C. Cox, *Speech, Truth, and the Free Market for Ideas*, 2 LEGAL THEORY 1, 2 (1996).

43. See, e.g., R. Kelly Garrett & Natalie Jomini Stroud, *Partisan Paths to Exposure Diversity: Differences in Pro- and Counterattitudinal News Consumption*, 64 J. COMM’N 680, 693–94 (2014); Alessandro Bessi, Fabio Petroni, Michaela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli & Walter Quattrociocchi, *Homophily and Polarization in the Age of Misinformation*, 225 EUR. PHYSICAL J. SPECIAL TOPICS 2047 (2016).

44. Derek E. Bambauer, *Shopping Badly: Cognitive Biases, Communications, and the Fallacy of the Marketplace of Ideas*, 77 U. COLO. L. REV. 649, 673–74 (2006).

45. *Id.* at 678.

46. See generally Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCH. 175, 175–220 (1998).

ited as people are prone to selectively accept and emphasize information that reinforces their beliefs.⁴⁷ Furthermore, even if they are able to distinguish truth from falsehood, they may still choose to believe a falsehood that confirms to their assumption instead of truth.⁴⁸

In addition, when individuals interact in groups, they are even more likely to be influenced by bias and to choose falsehood over truth.⁴⁹ “Informational cascades” refers to the effect of information pooling in a group, which causes individuals to be swayed by the opinions of others in the group and to change their original opinions.⁵⁰ When individuals are not sufficiently sure about the issue, they tend to look for “information signals” in the group to know the group’s attitude toward the issue.⁵¹ Psychologists have also proposed the concept of “conformity bias,” which suggests that people are generally reluctant to contradict the views of the group majority.⁵²

Furthermore, “reputation cascades” refers to the phenomenon that people may prefer to agree with others, even though they know that others are not right, in order to avoid contradicting the opinions of others and causing embarrassment or affecting their reputation.⁵³ The more closed the community, the more pronounced the effect of a reputation cascade is, and the more likely it is that the minority who disagrees with the majority of the community will tend to self-censor and choose to remain silent to avoid possible social sanctions.⁵⁴ From this standpoint, what is critical to people’s acceptance of information is not quality and truthfulness, but whether the information is espoused by others.

47. See Edward Glaeser & Cass R. Sunstein, *Does More Speech Correct Falsehoods?*, 43 J. LEGAL STUD. 65, 68–69 (2014) (noting that the presentation of correct information could backfire and have a polarizing effect as it reinforces people’s commitment to their inaccurate beliefs).

48. See Bambauer, *supra* note 44, at 678–79.

49. See CASS R. SUNSTEIN, CONFORMITY: THE POWER OF SOCIAL INFLUENCES 35–77 (2019) [hereinafter SUNSTEIN, CONFORMITY]; see also MICHELLE BADDELEY, COPYCATS & CONTRARIANS: WHY WE FOLLOW OTHERS . . . AND WHEN WE DON’T 22–28 (2018); Cass R. Sunstein, *Deliberative Trouble? Why Groups Go to Extremes*, 110 YALE L.J. 71, 77–84 (2000).

50. CASS R. SUNSTEIN, ON RUMORS: HOW FALSEHOOD SPREAD, WHY WE BELIEVE THEM, WHAT CAN BE DONE, 19 (2009).

51. See *id.* at 23.

52. O’CONNOR & WEATHERALL, *supra* note 8, at 81–82.

53. SUNSTEIN, CONFORMITY, *supra* note 49, at 68–70.

54. *Id.* at 41–42.

B. Algorithms, Echo Chambers and Bots

As noted, while disinformation is hardly novel, the diminished gatekeeping function⁵⁵ of mass media and the digitization of information has transformed its conduct and impact. Traditionally, mass media organizations selected the content they considered newsworthy through their internal editorial process, which curtailed the dissemination of false news relative to legitimate news.⁵⁶ However, with the advent of social media, the gatekeeping function of mass media has been dramatically reduced.⁵⁷ As social media has become the dominant communicational tool, it might be tempting to conclude that social media contributes to the democratization of information and users' autonomy as the distribution of information is no longer monopolized by mass media. Nonetheless, what users are exposed to on social media is largely controlled by algorithm.⁵⁸ Whereas traditional mass media filter information based on the social significance of the message, social media algorithms determine what information is presented to the users on the basis of personal preference,⁵⁹ which then facilitates the dissemination of disinformation.

Since social media features personalized content, propagandists with political and economic motives to disseminate disinformation are now able to micro-target and manipulate their intended recipients—typically those who are the most vulnerable to disinformation—more efficiently and less costly than ever before.⁶⁰ Social media algorithms collect information and online footprints of their users (e.g., likes, shares) to create individual user profiles.⁶¹ Algorithms then analyze user preferences to personalize their interface, which allows platforms to maximize users' online attention, thus meeting the interests of companies interested in advertising their products and services online.⁶² Propagandists exploit this personalization feature to examine the users' online activities to understand their preferences, and then

55. The term “gatekeeping” refers to the decision-making process of “how certain events are being treated more newsworthy than others and how institutions or influential individuals determine which information passes to the receivers.”

Engin Bozdag, *Bias in algorithmic filtering and personalization*, 15 ETHICS & INFO. TECH. 209, 209 (2013).

56. Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 71.

57. *See id.* at 71–74; Bozdag, *supra* note 55, at 209; Jessica Pepp, Eliot Michaelson & Rachel Katharine Sterken, *What's New About Fake News?*, 16 J. ETHICS & SOC. PHIL. 67, 80–81 (2019).

58. *See* Bozdag, *supra* note 55, at 210–13.

59. Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 74–76.

60. *See, e.g.*, Sophie L. Vériter et al., *Tackling COVID-19 Disinformation: Internal and External Challenges for the European Union*, 15 HAGUE J. DIPL. 569, 573 (2020).

61. *See generally* Stephan Winter, Ewa Maslowska & Anne L. Vos, *The Effects of Trait-Based Personalization in Social Media Advertising*, 114 COMPUT.S HUM. BEHAV. 1, 2–3 (2021).

62. GREGORIO, *supra* note 22, at 173.

deliver personalized messages to them accordingly.⁶³ This model is known as “micro-targeting,” a technique applied by Cambridge Analytica—the data analysis company hired by the Trump campaign—and the Russian-run Facebook accounts during the 2016 U.S. presidential election.⁶⁴ Empirical research demonstrated that much of the disinformation targeted swing state voters during the 2016 election, where key voters were the most susceptible to disinformation.⁶⁵

The personalization feature also fragments online space into isolated groups and contributes to increased group polarization,⁶⁶ which allows online disinformation campaigns to weaponize the user communities and subsequently results in the proliferation of disinformation.⁶⁷ The personalized information on social media reflects an individual’s values and facilitates the gathering and communication of like-minded people—a phenomenon known as the “filter bubble”⁶⁸ or “echo chamber.”⁶⁹ The echo chamber effect often breeds group polarization as individuals are surrounded and supported by like-minded people, thus reinforcing their confidence.⁷⁰ Given that the “echo chamber” effect is inextricably linked to partisanship, and that partisanship is highly correlated with the production and acceptance of disinformation, the “echo chamber” effect tends to promote the circulation of disinformation and to substantially discourage the dissemination of debunking information—a phenomenon termed the “spiral of partisanship.”⁷¹ Researchers found that many social media platforms consist of many polarized, non-interactive groups. In addition, most users prefer to search for and share posts that support their beliefs, and they hardly interact with debunk-

63. Yochai Benkler, Robert Faris & Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, 269-70 (2018).

64. See generally BRITTANY KAISER, *TARGETED: THE CAMBRIDGE ANALYTICA WHISTLEBLOWER’S INSIDE STORY OF HOW BIG DATA, TRUMP, AND FACEBOOK BROKE DEMOCRACY AND HOW IT CAN HAPPEN AGAIN* (2019); KATHLEEN HALL JAMIESON, *CYBER WAR: HOW RUSSIAN HACKERS AND TROLLS HELPED ELECTED A PRESIDENT: WHAT WE DON’T, CAN’T AND DO KNOW* (2018).

65. Natasha Bertrand, *Twitter Users Spreading Fake News Targeted Swing States in the Run-up to Election Day*, BUS. INSIDER INDIA (Sept. 28, 2017, 23:00 IST), <https://www.businessinsider.in/twitter-users-spreading-fake-news-targeted-swing-states-in-the-run-up-to-election-day/articleshow/60874965.cms>.

66. W. Lance Bennett & Shanto Iyengar, *A New Era of Minimal Effects? The Changing Foundations of Political Communication*, 58 J. COMM. 707, 720 (2008).

67. Harrison, *supra* note 6, at 18.

68. ELI PARISER, *THE FILTER BUBBLE: HOW THE NEW PERSONALIZED WEB IS CHANGING WHAT WE READ AND HOW WE THINK*, 9 (2011).

69. CASS R. SUNSTEIN, *#REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA*, 116 (2017).

70. CASS R. SUNSTEIN, *GOING TO EXTREMES: HOW LIKE MINDS UNITE AND DIVIDE*, 22-30 (2011).

71. Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 78.

ing information.⁷² In other words, the personalization of information and the fragmentation of the online information ecology have allowed biased and manipulative messages to flourish.

Therefore, although social media should theoretically facilitate both the distribution of legitimate news and disinformation, social media tends to disproportionately favor the spread of disinformation.⁷³ Researchers have found that large proportions of the content about vaccines on popular social media platforms consist of scientifically incorrect information promoting anti-vaccination sentiment, indicating that users are more likely to be exposed to vaccine disinformation.⁷⁴ An analysis of Google search results also suggested that parents concerned about COVID-19 vaccination safety and thus searching for information about vaccination risks would encounter 3.6 times more vaccine disinformation per website than parents who use neutral terms (i.e., neither related to risks nor benefits).⁷⁵

In addition, the use of bots⁷⁶ has further distorted the online information ecosystem and facilitates the spread of disinformation.⁷⁷ Bots use fake accounts to generate large amounts of tweets and reposts to shape false public opinion. Bots are activated to tweet or retweet specific content, or to reply automatically when a particular keyword or hashtag is detected.⁷⁸ Research

72. Itai Himelboim, Stephen McCreery & Marc Smith, *Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter*, 18 J. COMPUT.-MEDIATED COMMUN. 154, 154 (2013); see also Eytan Bakshy, Solomon Messing & Lada A. Adamic, *Exposure to Ideologically Diverse News and Opinion on Facebook*, 348 SCI. 1130, 1131 (2015) (finding that the spectrum of political information to which users are exposed has been significantly limited without the knowledge of the viewer). But see Elizabeth Dubois & Grant Blank, *The Echo Chamber is Overstated: The Moderating Effect of Political Interest and Diverse Media*, 21 INFO. COMMUN. & SOC'Y 729, 740–741 (2018) (arguing that the “high-choice media environment” enables the audience to be exposed to diverse political views).

73. Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 SCI. 1146, 1146 (2018).

74. See, e.g., Beth L. Hoffman, Elizabeth M. Felter, Kar-Hai Chu, Ariel Shensa, Chad Hermann, Todd Wolynn, Daria Williams & Brian A. Primack, *It's Not All About Autism: The Emerging Landscape of Anti-Vaccination Sentiment on Facebook*, 37 VACCINE 2216 (2019); Naomi Smith & Tim Graham, *Mapping the Anti-Vaccination Movement on Facebook*, 22 INFO. COMMUN. SOC'Y 1310 (2017).

75. Jeanette B. Ruiz & Robert A. Bell, *Understanding Vaccination Resistance: Vaccine Search Term Selection Bias and the Valence of Retrieved Information*, 32 VACCINE 5776, 5779 (2014).

76. Philip N. Howard, Samuel Woolley & Ryan Calo, *Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration*, 15 J. INFO. TECH. & POL. 81, 82 (2018) (referring to “bots” as “an executable software that automates the interaction between a user and content or other users”).

77. Helen Norton, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441, 471–72 (2019) [hereinafter Norton, *Powerful Speakers*].

78. Patrick Kulp, *Mysterious Bots Flood the FCC with Fake Anti-Net Neutrality Comments*, MASHABLE (May 10, 2017), <http://mashable.com/2017/05/10/bots-net-neutrality-comments-fcc/#a8jzO1YQu5qE>.

has shown that Russian bots and troll farms have pushed anti-vaccination messages on a large scale on Western social media.⁷⁹ Disinformation claiming that COVID-19 was an American-developed weapon or caused by 5G cellular networks has also been pushed by Russian bot networks.⁸⁰ By automatically producing mass information in a short period of time, machine speech effectively marginalizes human speech and causes users to misjudge public opinion via information cascades.⁸¹

The changes in information ecosystems and advertising mechanisms have challenged the major assumptions of the “marketplace of ideas” theory.⁸² With the gradual decline of mass media and the emergence of social media, the criterion of the trustworthiness of information has gradually shifted from the reputation of media organizations to individuals who share the same ideology.⁸³ It is more difficult than before to access debunking information or counterarguments because of the echo chamber effect facilitated by social media which aims to provide personalized content catered to users’ political preferences.⁸⁴ In the context of the echo chamber, when users are exposed to “more speech,” they only come across more information—including disinformation—that conforms their original beliefs, as opposed to meeting Justice Holmes’ vision where more speech would facilitate public debate and cure the evil of falsehood.⁸⁵

Therefore, the basic assumption of a marketplace of ideas established in the broadcasting era—that balanced reports or opinions on the same platform and at the same time are effective in reaching the audience so that listeners can make careful judgments after receiving contrary information—has been seriously challenged in the social media era.⁸⁶ The evolution of communication technology has accelerated the dissemination of disinformation, further challenging Justice Holmes’ presumption of sufficient time for review by the audience. While the production of legitimate news can be

79. David A. Broniatowski et al., *Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate*, 108 AM. J. PUB. HEALTH 1378, 1383 (2018).

80. Amy Mackinnon, *Russia Knows Just Who to Blame for the Coronavirus: America*, FOREIGN POL’Y (Feb. 14, 2020, 10:28 AM), <https://foreignpolicy.com/2020/02/14/russia-blame-america-coronavirus-conspiracy-theories-disinformation>; James Meese, Jordan Frith & Rowan Wilken, COVID-19, 5G Conspiracies and Infrastructural Futures, 177 Media International Australia 30, 40 (2020).

81. See Tauhid Zaman, *Even a Few Bots Can Shift Public Opinion in Big Ways*, CONVERSATION (Nov. 5, 2018, 6:41 AM EST), <https://theconversation.com/even-a-few-bots-can-shift-public-opinion-in-big-ways-104377>.

82. See NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST, *supra* note 15, at 90–106; Nunziato, *supra* note 19, at 1526–27; Wu, *supra* note 15, at 554–58; Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 387, 390–94 (2020).

83. Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 71.

84. *Id.* at 78–79.

85. See GREGORIO, *supra* note 22, at 172–73.

86. See Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 68–77.

time-consuming, the production of disinformation is far quicker.⁸⁷ Furthermore, bots have brought the speed and scale of information delivery to a level unmatched by human users.⁸⁸

II. THE HARM: DIMINISHED TRUST AND VACCINE HESITANCY

Disinformation presents a challenge to freedom of speech because it undermines the integrity of the democratic political process. First, disinformation weakens voter competence because voters who mistake disinformation for legitimate information become ill-informed.⁸⁹ The assumption is that providing voters with information could increase voter competence.⁹⁰ However, if the information is deceptive or manipulative, voters are left worse off than if they had not received the disinformation in the first place as the purpose of a disinformation campaign is to influence the voters' attitude on a particular issue.⁹¹ This negative effect is magnified by algorithms designed to cater to users' preferences.⁹² Disinformation on social media is delivered in the same way as commercial advertising—micro-targeting specific groups of readers for precise delivery based on their political preferences and prejudices—making those users most susceptible to disinformation also the ones most likely to receive it.⁹³

Second, disinformation contributes to diminished trust in legitimate news producers and institutions.⁹⁴ Given that the “echo chamber” effect is highly linked to partisanship, and that partisanship is highly correlated with the production and acceptance of disinformation, online disinformation further fortifies the echo chamber effect.⁹⁵ This dynamic tends to promote the circulation of disinformation and increases group polarization, thus making users more skeptical of legitimate news producers or other institutions.⁹⁶

87. See NAPOLI, *SOCIAL MEDIA AND THE PUBLIC INTEREST*, *supra* note 15, at 92–93.

88. Alessandro Bessi & Emilio Ferrara, *Social Bots Distort the 2016 U.S. Presidential Election Online Discussion*, 21 *FIRST MONDAY* (2016), <https://firstmonday.org/article/view/7090/5653>.

89. Allcott & Gentzkow, *supra* note 7, at 219; Abby K. Wood & Ann M. Ravel, *Fool Me Once: Regulating “Fake News” and Other Online Advertising*, 91 *S. CAL. L. REV.* 1223, 1234 (2018).

90. Wood & Ravel, *supra* note 89, at 1235.

91. *Id.* at 1235–36; see also Nathaniel Persily, *Can Democracy Survive the Internet?*, 28 *J. DEMOCRACY* 63, 71 (2017).

92. See Helen Norton, *Robotic Speakers and Human Listeners*, 41 *SEATTLE U. L. REV.* 1145, 1147 (2018).

93. See *supra* notes 59–70 and accompanying text.

94. Allcott & Gentzkow, *supra* note 7, at 219; see also Veriter et al., *supra* note 60, at 571.

95. Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 78.

96. See generally Pablo Barberá, *Social Media, Echo Chambers, and Political Polarization*, in *SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD, PROSPECTS FOR REFORM* 10, 10–27 (Nathaniel Persily & Joshua A. Tucker eds., 2020); Samuel C. Rhodes,

This negative impact may be reinforced by supply-side responses. In other words, a reduced demand for unbiased reporting will reduce the incentives to invest in accurate and truthful reporting.⁹⁷

Third, disinformation contributes to increased extremism and radicalization by splintering the electorate.⁹⁸ Since the cost of disseminating information on social media is extremely low, it becomes possible to engage in political mobilization by using bots and trolls to tear voters apart.⁹⁹ The micro-targeting technique makes the more extreme voters more easily targeted for political mobilization, while a vast moderate center is left out of the discussion of issues surrounding the election.¹⁰⁰ The personalized feature also enables extremist groups to quickly identify each other and to spread fear, hate speech, or disinformation.¹⁰¹ Moreover, the customization of contents also leads to polarized politics.¹⁰²

In the context of the COVID-19 pandemic, the unprecedented “infodemic” not only constitutes a serious threat to political process but also poses grave danger to public health. Conspiracy theories often prosper in times of crisis, as the anxious public struggles to make sense of a highly uncertain situation.¹⁰³ Such conspiracy theories include the prevention of COVID-19 (e.g., claiming that 5G technology suppresses the immune system, thus making people more vulnerable to the coronavirus)¹⁰⁴ and the effect of COVID-19 vaccines (e.g., claiming that vaccines genetically manipulate the population or contain microchips that interact with 5G networks).¹⁰⁵ Research has found a statistically significant negative relationship between belief in COVID-19 conspiracy theories and compliance with public health guidance.¹⁰⁶ The stronger the belief in COVID-19 conspiracy thinking, the

Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation, 39 POL. COMM. 'N 1 (2022).

97. Allcott & Gentzkow, *supra* note 7, at 219.

98. See Wood & Ravel, *supra* note 89, at 1236.

99. See Samuel C. Woolley, Bots and Computational Propaganda: Automation for Communication and Control, in *SOCIAL MEDIA AND DEMOCRACY: THE STATE OF THE FIELD AND PROSPECTS FOR REFORM* 89, 89-90 (Nathaniel Persily & Joshua A. Tucker eds., 2020).

100. Wood & Ravel, *supra* note 89, at 1236.

101. See Richard L. Hasen, *Cheap Speech and What It Has Done (to American Democracy)*, 16 FIRST AMEND L. REV. 200, 215 (2017).

102. See generally *id.*

103. Veriter et al., *supra* note 60, at 573.

104. Wassim Ahmed, Joseph Downing, Marc Tutters & Peter Knight, *Four Experts Investigate How the 5G Coronavirus Conspiracy Theory Began*, THE CONVERSATION (June 11, 2020), <https://theconversation.com/four-experts-investigate-how-the-5g-coronavirus-conspiracy-theory-began-139137>.

105. Daniel Freeman et al., *Coronavirus Conspiracy Beliefs, Mistrust, and Compliance with Government Guidelines in England*, 52 PSYCH. MED. 251, 257 (2020).

106. Daniel Allington & Nayana Dhavan, *The Relationship between Conspiracy Beliefs and Compliance with Public Health Guidance with Regard to COVID-19* (Apr. 8, 2020).

less adherence to government guidelines such as taking antibody tests or abiding by mask mandates.¹⁰⁷

The proliferation of disinformation about COVID-19 vaccines also contributes to the rapid growth of vaccine hesitancy. Vaccine hesitancy refers to the phenomenon of “delay in acceptance or refusal of vaccination despite availability of vaccination services.”¹⁰⁸ Prior to the COVID-19 outbreak, the World Health Organization had already identified vaccine hesitancy as one of the top ten threats to global health.¹⁰⁹ A new pandemic outbreak in the era of social media and disinformation only further enables the spread vaccine hesitancy.¹¹⁰ While there are many factors that might contribute to vaccine hesitancy,¹¹¹ many empirical studies have concluded that exposure to disinformation about COVID-19 vaccines is central to such reluctance.¹¹² Studies found that large proportions of the content about vaccines on major social media platforms are anti-vaccination.¹¹³ The data demonstrated a significant relationship between organizations on social media and public doubts of vaccine safety as well as a substantial relationship between foreign disinformation campaigns and declining vaccination coverage.¹¹⁴ Another study observed high levels of disinformation exposure amongst participants, and

107. Freeman et al., *supra* note 105.

108. Noni E. MacDonald et al., *Vaccine Hesitancy: Definition, Scope and Determinants*, 33 *VACCINE* 4161, 4163 (2015).

109. World Health Organization, *Ten Threats to Global Health in 2019*, WORLD HEALTH ORGANIZATION, <https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019> (last visited Sept. 24, 2022).

110. *See generally* Ariel Fridman, Rachel Gershon & Ayelet Gneezy, *COVID-19 and Vaccine Hesitancy: A Longitudinal Study*, *PLOS ONE* (2021) (finding a decrease in intentions of getting a COVID-19 vaccine).

111. *See generally* Stacy Wood & Kevin Schulman, *Beyond Politics—Promoting COVID-19 Vaccination in the United States*, 384 *N. ENGL. J. MED.* e23 (2021) (finding that the politicization of the COVID-19 pandemic and the concerns over the expedited development of the vaccine lead to vaccine hesitancy); Jagdish Khubchandani et al., *COVID-19 Vaccination Hesitancy in the United States: A Rapid National Assessment*, 46 *J. CMTY. HEALTH* 270 (2021) (finding that education, employment and income are predictors of vaccine hesitancy); Timothy Callaghan et al., *Correlates and Disparities of Intention to Vaccinate Against COVID-19*, 272 *SOCIAL SCI. & MED.* 1 (2021) (finding that “African Americans, women, and conservatives” are the three groups with the highest rate of vaccine hesitancy).

112. *See, e.g.*, Francesco Pierri, Brea L. Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer & John Bryden, *Online Misinformation is Linked to Early COVID-19 Vaccination Hesitancy and Refusal*, 12 *SCI. REP.* 1, 1 (2022); Renee Farett & Sean D. Young, *Online Misinformation and Vaccine Hesitancy*, 11 *TRANSLATIONAL BEHAV. MED.* 2194 (2021).

113. Steven Lloyd Wilson & Charles Wiysonge, *Social Media and Vaccine Hesitancy*, 5 *BMJ GLOB. HEALTH* 1, 2 (2020).

114. *Id.* at 5.

that the more disinformation they were exposed to, the less likely they were vaccinated.¹¹⁵

The prevalence of vaccine hesitancy and online vaccine disinformation has broader implications—it not only engenders serious health issues by prolonging the pandemic, but it also promotes distrust in democratic institutions by sowing doubts about health authorities. It is well-established that vaccination is the most effective method to prevent COVID-19 and that widespread vaccination is necessary to achieve herd immunity.¹¹⁶ However, despite the widespread availability of COVID-19 vaccines, vaccine hesitancy driven by disinformation continues to impede the attainment of herd immunity.¹¹⁷ Furthermore, relying on scientific expertise to make policy choices has been recognized by the Supreme Court as one of the foundations of a democratic society.¹¹⁸ This foundation is undermined by online disinformation which leads the public to disobey public health guidance. From this perspective, disinformation poses a threat to democracy by corrupting the integrity of the political process and undermining the democratic values which shape public policies, including in the health and science sectors.¹¹⁹

III. THE ANTIDOTE: TRANSPARENCY AND ACCOUNTABILITY IN ONLINE CONTENT MODERATION

The discussions in Part I and Part II illustrate the first challenge to the First Amendment brought by the pandemic in an algorithmic society: the proliferation of online information. Social media platforms' responses to combating disinformation lead us to the second challenge: the rise of private censorship. In response to growing public pressure, social media platforms have more rigorously enforced content moderation. It is clear that social media platforms have taken a more interventionist stance in the fight against health disinformation than they have in the past with regard to political dis-

115. Stephen R. Neely, Christina Eldredge, Robin Ersing & Christa Remington, *Vaccine Hesitancy and Exposure to Misinformation: A Survey Analysis*, 37 J. GEN. INTERNAL MED. 179, 179 (2022).

116. Coronavirus disease (COVID-19): Herd Immunity, Lockdowns and COVID-19, WHO (Dec. 31, 2020), <https://www.who.int/news-room/questions-and-answers/item/herd-immunity-lockdowns-and-covid-19>

117. Neely, *supra* note 115, at 179.

118. See, e.g., *FCC v. Fox Television Stations, Inc.*, 556 U.S. 502, 552 (2009) (Breyer, J., dissenting) (“[When reviewing an agency’s decision], a reviewing court should take into account whether the agency . . . offered an explanation for its decision that runs counter to the evidence before the agency, or is so implausible that it could not be ascribed to a difference in view or the product of agency expertise.”); *Dep’t of Com. v. New York*, 139 S.Ct. 2551, 2585 (2019) (Breyer, J., concurring in part and dissenting in part).

119. Veriter et al., *supra* note 60, at 572.

information.¹²⁰ Facebook and YouTube, for example, have been actively removing anti-vaccination campaigns from their sites.¹²¹

At first glance, self-regulation by social media platforms seems to be more consistent with the First Amendment tradition that favors a free marketplace over government intervention. However, the First Amendment tradition might actually stand in the way of states introducing necessary reforms to address the issues raised by an algorithmic society.¹²² The following section discusses the problems associated with private content moderation and proposes a new First Amendment paradigm that could both curtail the spread of disinformation and safeguard platform users' First Amendment rights.

A. *The Peril of Private Content Moderation*

Transnational online platforms, such as Facebook and Twitter, play an increasingly critical role in facilitating the flow of information online and have revolutionized the way people communicate. In *Packingham v. North Carolina*,¹²³ the Supreme Court noted that social media platforms are “one of the most important places to exchange views,”¹²⁴ as they offer “relatively unlimited, low-cost capacity for communication of all kinds.”¹²⁵ In today’s society, using Google or Facebook is almost a necessity for entering public discourse.¹²⁶ Online platforms facilitate citizens’ participation in online public debate. However, that also means that our practical ability to share and receive information is subject to the control of private platforms.¹²⁷

Despite several social media platforms attempting to build a global community promoting freedom of speech, they often find themselves in need of policing content to protect their own business interests.¹²⁸ Twitter, most notably, though labelling itself as “the free speech wing of the free speech party,”¹²⁹ has gradually become less tolerant towards disinformation

120. Harrison, *supra* note 6, at 25.

121. See Flora Carmichael & Charlie Haynes, *Facebook Removes Anti-vax Influencer Campaign*, BBC (Aug. 10, 2021), <https://www.bbc.com/news/blogs-trending-58167339>; see also Gerrit De Vynck, *YouTube is Banning Prominent Anti-vaccine Activists and Blocking All Anti-vaccine Content*, WASH. POST (Sep. 21, 2021), <https://www.washingtonpost.com/technology/2021/09/29/youtube-ban-joseph-mercola/>.

122. See Hasen, *supra* note 101, at 216.

123. *Packingham v. North Carolina*, 137 S.Ct. 1730 (2017).

124. *Id.* at 1732.

125. *Id.* (quoting *Reno v. ACLU*, 521 U.S. 844 (1997)).

126. GREGORIO, *supra* note 22, at 169.

127. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1153.

128. Balkin, *Free Speech is a Triangle*, *supra* note 16, at 2023.

129. Josh Halliday, *Twitter’s Tony Wang: ‘We are the Free Speech Wing of the Free Speech Party’*, GUARDIAN (Mar. 22, 2012), <https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech>.

in response to public outcry.¹³⁰ In this regard, Professor Balkin aptly analogized social media platforms to “nineteenth century enlightened despots.”¹³¹ On the one hand, social media platforms often champion a set of enlightened values, such as espousing First Amendment rights; on the other hand, they still have discretion over whether to change or enforce those values as users are not involved in the decision-making process.¹³²

One might argue that social media platforms’ enforcement of content moderation appears preferable to government intervention. This assumption is based on the “marketplace of ideas” doctrine that deems the First Amendment to be a “negative right” discouraging government intervention and prioritizes self-regulation by private parties.¹³³ In the algorithmic society, however, this assumption is fundamentally flawed. It neglects to account for the power asymmetry between social media platforms and their users. Further, the high degree of opacity and unpredictability of content moderation threatens rather than safeguards users’ right to freedom of speech.¹³⁴

The State Action Doctrine limits the obligations imposed by the First Amendment to public actors, thereby enabling private parties, such as social media platforms, to exclusively control the rules for content moderation without any First Amendment concerns.¹³⁵ Social media platforms, despite

130. See Mike Wending, *Twitter Steps Up Ukraine Misinformation Fight*, BBC (May 19, 2022), <https://www.bbc.co.uk/news/business-61512261>; see also Stuart A. Thompson, *To Fight Election Falsehoods, Social Media Companies Ready a Familiar Playbook*, N.Y. TIMES (Aug. 23, 2022), <https://www.nytimes.com/2022/08/23/technology/midterms-misinformation-tiktok-facebook.html>.

131. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1201.

132. See *id.*

133. See Napoli, *First Amendment Theory Meets Fake News*, *supra* note 42, at 60-66.

134. One might argue that power asymmetry is inherent in all areas of publishing (for example, only a limited number of people have access to mass media); therefore, power asymmetry itself does not warrant a shift to a new interpretation of First Amendment. However, as Professor Balkin argued, the fundamental difference between mass media and social media makes the power asymmetry between platforms and their users particularly problematic:

Curatorial due process made little sense for twentieth-century mass media because twentieth-century mass media largely published their own content or the content of a relatively small number of people and businesses. Most people had no access to mass-media publication, and few people expected an explanation or a right to be heard either before or after the *New York Times* rejected their proposed letters to the editor. Twenty-first-century media companies, by contrast, primarily publish content by the general public. Digital curators exist to facilitate mass cultural participation, and their end users expect and depend on the fact that curators will help them in this process. Therefore, curators need to provide assurances that when they block or limit participation, they are not being overbroad or arbitrary.

Balkin, *Free Speech is a Triangle*, *supra* note 16, at 2045.

135. For more detailed discussions on whether social media platforms should be deemed as public entities under the state action doctrine, thus being subject to First Amendment constraints, see Ryan Kraski, *Combating Fake News in Social Media: U.S. and German Legal Approaches*, 91 ST. JOHN’S L. REV. 923, 928 (2017); see also Jonathan Peters, *The Sovereigns*

being the “new governors”¹³⁶ dictating what users can say online, are protected by the First Amendment rather than restricted by it. Therefore, the enforcement of community norms is not subject to due process requirements and it often lacks notice or transparency.¹³⁷ Even though most platforms provide their users with community guidelines, the guidelines do not necessarily reflect the reality of content moderation.¹³⁸ As Professor Gregorio has noted, the differences between publicly available community guidelines and private internal policies render the moderation process “more as an authoritarian determination than a democratic expression.”¹³⁹

As private actors, platforms may be more interested in promoting their self-interest rather than safeguarding public values when exercising their content moderation power. The logic of digital capitalism is characterized by the “attention economy” that seeks to capture users’ scarce attention.¹⁴⁰ Social media companies’ primary goals are increasing the interactions between users, and the quantity of content they post on social media, so that they can sell the users’ scarce attention to advertisers.¹⁴¹ To achieve these ends, social media platforms feed users with personalized content that conforms to their preferences, which fragments online space into isolated groups and contributes to increased polarization, thereby facilitating the dissemination of disinformation.¹⁴² As a result, rather than creating a common space for democratic deliberation and curtailing online disinformation, the fragmentation of the online information ecosystem have led to the flourishing of biased and manipulative information.

A challenge further raised by the algorithmic society stems from the fact that a limited number of private platforms possess the absolute power to set the boundary of freedom of speech on a global scale. The market for online platforms is characterized by a high degree of concentration in the contemporary digital environment.¹⁴³ The internet has fundamentally transformed the nature of public discourse. It is no longer public spaces such as parks and streets where people exchange ideas; instead, public discourse now takes place on a limited number of privately owned platforms whose

of Cyberspace and State Action: The First Amendment’s Application - Or Lack Thereof - To Third-Party Platforms, 32 BERKELEY TECH. L.J. 989, 1022-23 (2017).

136. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1663 (2018).

137. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1197.

138. Klonick, *supra* note 136, at 1630.

139. GREGORIO, *supra* note 22, at 185. *See also* Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1197 (arguing that platforms may “behave like absolutist monarchs, who claim to exercise power benevolently, but who make arbitrary exceptions and judgments in governing online speech.”).

140. *See generally* TIM WU, *THE ATTENTION MERCHANTS: THE EPIC SCRAMBLE TO GET INSIDE OUR HEADS* (2016).

141. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1192.

142. *See supra* notes 60–75 and accompanying text.

143. GREGORIO, *supra* note 22, at 170.

owners “have no constitutional obligations to grant nondiscriminatory access to their platforms, and who regularly exercise authority to control the content of expression on their platforms.”¹⁴⁴

In sum, the dominant role of social media platforms in shaping public opinion has made private content moderation both necessary and problematic. As noted above, the impact of the COVID-19 pandemic has demonstrated the paradoxical nature of this development.¹⁴⁵ On the one hand, the proliferation of online disinformation underscores that social media platforms have failed to stop, and may have contributed to, the spread of disinformation with their content moderation policies. On the other hand, the pandemic has also revealed the importance of access to online information as people become more reliant on digital platforms to exercise their First Amendment right, raising concern over censorship of social media users imposed by private platforms. Therefore, the question of how to safeguard individual rights to online speech in an era of a pandemic is of paramount importance.

B. *Reconstructing the First Amendment Jurisprudence in the Algorithmic Society*

Since content moderation plays such a crucial role in shaping the First Amendment right in an algorithmic society, it is worth contemplating how to avoid free speech being subject to private interests rather than democratic values. As Professor Balkin noted, “[t]he problems of free speech in any era are shaped by the communications technology available for people to use and by the ways that people actually use that technology.”¹⁴⁶ If the problems are derived from the characteristics of communications technology, then those characteristics should also be considered when devising the solution. The proliferation of online disinformation and the opaque nature of private censorship that characterize the contemporary online speech environment forced the question of whether the traditional laissez-faire approach of the First Amendment, which discourages any kind of governmental intervention, remains justified.

In an algorithmic society, despite that the state remains to pose a threat to freedom of speech, it is crucial that the state “serve as a necessary counterweight to developing technologies of private control.”¹⁴⁷ Therefore, it is worth proposing an interpretive shift that reads the First Amendment not merely as a negative right but also a positive one. The notion of positive right acknowledges that the freedom of speech requires “support and ena-

144. Bhagwat, *supra* note 18, at 93.

145. See *supra* notes 22–30 and accompanying text.

146. Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1151.

147. *Id.* at 1152.

blement”¹⁴⁸ from the states to facilitate robust public debate where “diversity of expression flourish[es].”¹⁴⁹ The duty to protect the First Amendment right involves an obligation for governments to promote this right and to provide for an environment where it can be meaningfully exercised without being unduly impeded by private parties.¹⁵⁰

The arguing against the laissez-faire approach to online platform regulation is not synonymous with advocating for more rigorous content-based regulations imposed by either the government or social media platforms. The positive dimension of the First Amendment imposes obligations on the state to introduce regulation promoting a pluralistic speech environment. However, the negative dimension of the First Amendment restricts the state from introducing regulation that is unnecessarily intrusive.¹⁵¹ Therefore, the underlying principle is to design a regulatory regime that introduce requirements protecting users from being subject to content moderation that is based on unaccountable decision-making and deliberately opaque procedures, instead of compelling platforms to carry or remove certain content.¹⁵²

While First Amendment jurisprudence overwhelmingly considers free speech to be a negative right, arguments for a shift to an interpretation that incorporates the positive dimension are actually anything but new.¹⁵³ Professor Emerson has argued that the First Amendment requires the government to affirmatively enable facilities for communication because the First Amendment protects audience interests in the form of “a right to hear.”¹⁵⁴ In more recent accounts, Professor Norton has argued for a “listener-centered” approach to the First Amendment which permits the government to regulate the speech of comparatively knowledgeable or powerful speakers.¹⁵⁵ Professor Balkin has argued that fiduciary relationships involving asymmetries of power, information, and transparency, justifies government intervention.¹⁵⁶ Similarly, Professor Napoli has argued for a “collectivist approach,” which interprets the First Amendment as prioritizing the welfare of the collective citizenry over that of the individual speaker, which grants government the

148. Andrew T. Kenyon, *Positive Free Speech: A Democratic Freedom*, in THE OXFORD HANDBOOK OF FREEDOM OF SPEECH (Adrienne Stone & Frederick Schauer eds. 2021) 1, 4.

149. *Id.*

150. Gregorio, *supra* note 22, at 206.

151. Kenyon, *supra* note 148; *see also* ANDREW T. KENYON, DEMOCRACY OF EXPRESSION: POSITIVE FREE SPEECH AND LAW 1 (2021) (arguing that the “state has obligations to act in support of communicative freedom as well as there being limits on state restriction of speech; both positive and negative dimensions of free speech need protection for democracy to have substantial communicative legitimacy”) [hereinafter KENYON, DEMOCRACY OF EXPRESSION].

152. *See* GREGORIO, *supra* note 22, at 159; Balkin, *Free Speech in the Algorithmic Society*, *supra* note 21, at 1198; Citron & Franks, *supra* note 4, at 66.

153. *See generally* KENYON, DEMOCRACY OF EXPRESSION, *supra* note 151, at 118–20.

154. THOMAS EMERSON, THE SYSTEM OF FREEDOM OF EXPRESSION 649 (1970).

155. Norton, *Powerful Speakers*, *supra* note 77, at 441–42.

156. Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1185–86 (2016).

authority to introduce regulation that enhances the free speech environment.¹⁵⁷

The point here is that power asymmetry between private platforms and their users, which characterizes the nature of an algorithmic society, may provide the most compelling case for this interpretive shift to eventually take place.¹⁵⁸ Grave threats to fundamental rights, such as “radically unequal economic power”¹⁵⁹ between private parties, can trigger the positive obligation of states to regulate private activities and protect fundamental rights.¹⁶⁰ As Professor Bakan has argued, the notion of free speech “requires a measure of social equality.”¹⁶¹ That is, freedom of speech is realized only when everyone is guaranteed “the capacity – the resources, abilities, and opportunities – to exercise freedom and avoid coercion by others.”¹⁶² Therefore, the power asymmetry between private platforms and their users justifies invoking the “positive” dimension of the First Amendment.

C. Towards a More Accountable and Democratic Platform

Despite social media platforms fundamentally transforming the way we communicate, there has not been any significant change in the regulatory framework for online platforms in the last twenty years. The traditional *laissez-faire* First Amendment jurisprudence, incorporating the “more speech” doctrine, has discouraged the government from introducing regulations.¹⁶³ After all, online platforms are viewed as a facilitator of democracy rather than a threat to public discourse.¹⁶⁴ The Communications Decency Act 230 (CDA 230) regime, which is one of the very few laws governing online platforms for more than two decades, epitomizes the *laissez-faire* approach long favored by the First Amendment tradition.¹⁶⁵ Though the legislative intent of CDA 230 is to provide legal shield for “Good Samaritan” filtering of harmful content, the courts held that in order to preserve the internet’s role

157. NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST, *supra* note 15, at 191–92.

158. *Id.* at 192. See also Martha Minow, Saving the News: Why the Constitution Calls for Government Action to Preserve Freedom of Speech 100 (2021) (arguing that the challenges to free speech in an algorithmic society warrant “a hard look at the development of First Amendment doctrine and an interpretative guide to make sure it does not produce unnecessary barriers to the government actions needed to revitalize the news business.”)

159. Jack M. Balkin, *Some Realism about Pluralism: Legal Realist Approaches to the First Amendment*, 1990 DUKE L.J. 375, 379 (1990).

160. GREGORIO, *supra* note 22, at 203.

161. JOEL BAKAN, JUST WORDS: CONSTITUTIONAL RIGHTS AND SOCIAL WRONGS 10 (1997).

162. *Id.*

163. *See supra* notes 31–38 and accompanying text.

164. This notion of Internet optimism is a view long held by the U.S. government. See generally Jack Goldsmith, The Failure of Internet Freedom, in *The Perilous Public Square: Structural Threats to Free Expression Today* 241, 241–48 (David E. Pozen ed., 2020).

165. *See generally* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans Sec. 230 Immunity*, 86 FORDHAM L. REV. 401 (2017).

in “myriad avenues for intellectual activity,”¹⁶⁶ CDA 230 must be interpreted broadly to immunize social media platforms from liability for third-party content posted on their platforms.¹⁶⁷ In the wake of the 2016 presidential election, the Senate proposed the Honest Ads Act that would filled the regulatory gap of online political advertisements if passed.¹⁶⁸ However, it is only a first step towards a comprehensive regulatory regime, as it only imposes transparency requirements on political advertisements and fails to consider the opacity of content moderation imposed by private platforms.

Under the *laissez-faire* approach to the interpretation of the First Amendment, the concept of editorial discretion safeguards the right of private companies to regulate speech disseminated through their platforms.¹⁶⁹ One recent example of this application is the Supreme Court’s decision to uphold a district court’s ruling blocking a Texas social media law from taking effect in *NetChoice, LLC v. Paxton*.¹⁷⁰ The contentious Texas law was prompted by conservative complaints aiming to free social media users from “Silicon Valley censorship,”¹⁷¹ which prohibited platforms from censoring users based on their viewpoints.¹⁷² In the district court’s ruling, the court explicitly recognized that “[s]ocial media platforms have a First Amendment right to moderate content disseminated on their platforms.”¹⁷³ Hence, the court concluded that the Texas law violated the First Amendment because it interfered with the platforms’ editorial discretion by compelling them to disseminate third-party content.¹⁷⁴

From an international perspective, many non-binding agreements have been introduced to enhance the transparency of content moderation. The European Commission’s Code of Practice on Disinformation introduced a requirement that clear policies be put in place regarding the identity and misuse of automated bots on platforms.¹⁷⁵ The Manila Principles on Intermediary Liability, which focus on the proceduralization of content moderation, prescribe that platform companies adopt “the least restrictive technical means” of content moderation.¹⁷⁶ The Manila Principles introduce procedural safeguards such as requiring platforms to establish clear and

166. Barnes v. Yahoo!, Inc., 570 F.3d 1096, 1099 (9th Cir. 2009).

167. Citron & Franks, *supra* note 4, at 50.

168. Tim Lau, The Honest Ads Act Explained, Brennan Center for Justice (Jan. 17, 2020), <https://www.brennancenter.org/our-work/research-reports/honest-ads-act-explained>.

169. Bhagwat, *supra* note 18, at 91.

170. NetChoice, LLC v. Paxton, 142 S.Ct. 1715 (2022).

171. Senator Bryan Hughes (@SenBryanHughes), Twitter (Mar. 5, 2021, 11:48 PM), <https://twitter.com/SenBryanHughes/status/1368061021609463812>.

172. Tex. Civ. Prac. & Rem. Code § 143A.002.

173. NetChoice, LLC v. Paxton, 573 F.Supp.3d 1092, 1106 (W.D. Tex. 2021).

174. *Id.* at 1109.

175. Harrison, *supra* note 6, at 22.

176. Manila Principles on Intermediary Liability, § 4(b) (Mar. 24, 2015), <https://manilaprinciples.org>.

public notice of the content-regulation policies¹⁷⁷ and to provide explanation for content removal.¹⁷⁸ Similarly, the Santa Clara Principles on Transparency and Accountability in Content Moderation prescribe due process requirements such as that platform companies provide notice to users whose content is removed and establish an appeal process available to users.¹⁷⁹

However, despite the good intentions of these proposals, this soft law approach has produced very limited progress. Though the flexibility encourages platforms to sign up without the fears of over-regulation, the voluntary nature of the approach establishes a regulatory asymmetry between signatories and non-signatories.¹⁸⁰ This leads to the possibility of disinformation remaining rampant at non-signatory platforms, which undermines many of the improvements made by the signatories.¹⁸¹ Furthermore, there is no effective instrument to evaluate the signatories' compliance or sanction them for breaches. The lack of any binding force in this system leaves online platforms free to decide whether to participate with these rules while maintaining their internal community standards.¹⁸²

In contrast to the US, where the *laissez-faire* approach fails to account for the reality of the dominant social media corporations, the EU has responded to the challenges of an algorithmic society with its rise of “digital constitutionalism.”¹⁸³ The primary concern is the lack of transparency and accountability, or redress mechanisms limiting platform power in the field of content moderation.¹⁸⁴ Recognizing that threats to constitutional rights also come from transnational private actors “whose freedoms are increasingly turning into forms of unaccountable power,”¹⁸⁵ freedom of speech is no longer considered merely a negative right but also a positive right under the framework of European digital constitutionalism.¹⁸⁶

Regarding the EU's approach, the recently introduced Digital Services Act symbolizes a paradigmatic shift from interpreting freedom of speech as merely a “negative right” to a “positive right.”¹⁸⁷ The Digital Services Act limits platform power by prescribing substantial obligations and procedural safeguards mandating platforms to disclose information, assess the risk for

177. *Id.* at § 6(c).

178. *Id.* at § 3(e).

179. The Santa Clara Principles on Transparency and Accountability in Content Moderation, Art. 3, <https://santaclaraprinciples.org/>.

180. Harrison, *supra* note 6, at 23.

181. *Id.*

182. GREGORIO, *supra* note 22, at 188.

183. *See generally id.* at 38–79.

184. *Id.* at 105, 119.

185. *Id.* at 65.

186. *Id.* at 205.

187. *Id.* at 211-14.

fundamental rights and provide redress mechanisms.¹⁸⁸ Without introducing content-based regulation, it requires online platforms to abide by procedural safeguards. For example, Article 12 requires platforms to disclose information regarding any policies, procedures, measures and tools used for the purpose of content moderation.¹⁸⁹ Article 14 prescribes that compliant procedures shall be “easy to access” and “user friendly.”¹⁹⁰ Article 15 stipulates that platforms must provide users with “a clear and specific statement of reasons” for their decisions on content moderation.¹⁹¹ Furthermore, to make this system more effective and to incentivize online platforms to comply, the Digital Services Act imposes fines on those violating the procedural requirement¹⁹² while maintaining the exemption of liability for online platforms that moderate content responsibly.¹⁹³ The procedure-centric approach, while allowing platforms to remove content they deem detrimental to the community, makes platforms more accountable for their decision-making, thereby mitigating the power asymmetry between platforms and their users.

CONCLUSION

Over the last twenty years, the internet, and social media in particular, has become the primary venue for the public to exercise their First Amendment right. The democratization of information and the ability to access online content have played a critical role in promoting free speech on a global scale. However, recent developments suggest that instead of facilitating a robust and uninhibited space for democratic deliberation, social media and algorithms create the “homophily of networks”¹⁹⁴ that engenders polarization and fragmentation.¹⁹⁵ The impact of the COVID-19 pandemic demonstrates two major challenges to the First Amendment right in an algorithmic society: how the excessive amount of online disinformation frustrates democracy, and how the online speech environment is governed by unaccountable private actors imposing opaque standards for content moderation. There has never been a more urgent time to address the long-standing question of how to overcome the deleterious influence of exposure to online disinformation while safeguarding users’ right to access online information.

188. 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, COM/2020/825 final (Dec. 12, 2020) (to be codified at 2020/0361(COD)).

189. *Id.* art. 12.

190. *Id.* art. 14.

191. *Id.* art. 15.

192. *Id.* art. 42.

193. *Id.* art. 5.

194. Miller McPherson, Lynn Smith-Lovin & James M. Cook, *Birds of a Feather: Homophily in Social Networks*, 27 ANN. REV. SOC. 415, 415 (2001).

195. *See supra* notes 60-72 and accompanying text.

This leads us back to the question that Professor Bollinger posed: does the basic premise of the “more speech” doctrine still hold true in the algorithmic society?¹⁹⁶ This Note argues that it does not. The First Amendment should be interpreted as both a negative and a positive right. By arguing against the laissez-faire approach to online platform regulation, this Note does not support more rigorous content-based regulation. As private censorship is inevitable in the era of social media, the solution is not to compel private platforms to either carry or remove certain content. Instead, the positive obligations of public actors should restrain platform powers to define the protection of freedom of speech online, thereby balancing the power asymmetry between platforms and their users. That is, the underlying principle of the positive approach is to design a regulatory regime that is least restrictive, and fosters accountability and transparency by protecting users from being subject to content moderation that is based on unaccountable decision-making and deliberately opaque procedures. In this regard, the recently introduced Digital Services Act in the EU, which emphasizes the transparency and procedural safeguards in content moderation rather than compelling platforms to carry or remove certain content, may offer some important lessons for the US to reconstruct its online platforms regulation in the era of an algorithmic society.

196. Bollinger, *supra* note 1.

