Regulating Algorithmic Disinformation

Haochen Sun*

ABSTRACT

Disinformation is endemic in the digital age, seriously harming the public interest in democracy, health care, and national security. Increasingly, disinformation is created and disseminated by social media algorithms. Algorithmic disinformation, a new phenomenon, thus looms large in contemporary society. Recommendation algorithms are driving the spread of disinformation on social media networks, and generative algorithms are creating deepfakes, both at unprecedented levels. The regulation of algorithmic disinformation is therefore one of today's thorniest legal problems.

Against this backdrop, this Article proposes a novel approach to regulating algorithmic disinformation effectively. It first explores why transparency, intelligibility, and accountability should be adopted as the three major principles of the legal regulation of algorithmic disinformation. Because of its market-based technology development and regulation policy, the United States has yet to adopt any laws regulating algorithmic disinformation, let alone these three principles. The Article then examines legislative reforms in France and China, where the three principles have been translated into legal rules requiring social media companies to disclose their disinformation-related algorithms, render them intelligible to users, and assume legal responsibility for curbing the spread of disinformation on their platforms.

Based on a critical discussion of the major problems with these legal rules, the Article puts forward a multi-stakeholder approach to better implement the three principles. It argues that the United States should take the lead in creating and piloting an algorithmic disinformation review system. This new system would empower the administrative oversight of algorithmic disinformation and promote the dynamic engagement of social media users and experts in policing algorithms that generate and disseminate disinformation. The ADRS would thus promote the transparency and intelligibility of algorithms and hold social media platforms accountable for curbing disinformation.

* Professor of Law, University of Hong Kong Faculty of Law; Short-Term International Visiting Professor of Law, Columbia Law School (Spring 2023).

TABLE OF CONTENTS

INTRODUCTION

Disinformation is a defining problem of the digital age. Its amplification on social media reflects "humanity's worst impulses."[1] Disinformation erodes democracy.[2] Research suggests that the disinformation swirling on Facebook swayed the outcomes of both the Brexit referendum and 2016 U.S. presidential election.[3] It also harms public health. The "infodemic"[4] accompanying the COVID-19 pandemic has radiated throughout the Internet, spreading fake news and conspiracy theories that have caused mental distress, bred mistrust in health authorities, and undermined pandemic relief measures.[5] Disinformation threatens other public interests as well, ranging from national security[6] to racial equality.[7]

Disinformation is not new, but it is now created and spread to an unprecedented extent through social media algorithms.[8] Owing to the amplification power of their algorithms, social media platforms can spread disinformation faster than the truth,[9] disrupting their users' sense of reality.[10] A recent study has revealed that it is the algorithms applied by social media platforms that play a larger role in spreading

1. Ella Lee, *'Humanity's Worst Impulses': Obama Says Online Disinformation Puts Democracy at Risk*, USA TODAY (Apr. 22, 2022), https://www.usatoday.com/story/news/politics/2022/04/22/obama-online-disinformation-democracy-at-risk/7408070001 [https://perma.cc/X462-B9PX] [https://web.archive.org/web/https://www.usatoday.com/story/news/politics/2022/04/22/obama-online-disinformation-democracy-at-risk/7408070001].

2. *See generally* YOCHAI BENKLER ET AL., NETWORK PROPAGANDA: DISINFORMATION, MANIPULATION, AND RADICALIZATION IN AMERICAN POLITICS (2018); PHILIP N. HOWARD, LIE MACHINES: HOW TO SAVE DEMOCRACY FROM TROLL ARMIES, DECEITFUL ROBOTS, JUNK NEWS OPERATIONS, AND POLITICAL OPERATIVES (2020); JOSHUA A. TUCKER ET AL., SOCIAL MEDIA, POLITICAL POLARIZATION, AND POLITICAL DISINFORMATION: A REVIEW OF THE SCIENTIFIC LITERATURE (2018).

3. *See* BENKLER ET AL., *supra* note 2, at 28.

4. *See Infodemic*, WORLD HEALTH ORGANIZATION, https://www.who.int/health-topics/infodemic#tab=tab_1 [https://perma.cc/7C2A-NQTM] [https://web.archive.org/web/20230220233320/https://www.who.int/health-topics/infodemic].

5. *See* Michael A. Gisondi et al., *A Deadly Infodemic: Social Media and the Power of COVID-19 Misinformation*, 24 J. MED. INTERNET RES. e35552 (2022).

6. *See* Steven Lee Myers & Eileen Sullivan, *Disinformation Has Become Another Untouchable Problem in Washington*, N.Y. TIMES (July 6, 2022), https://www.nytimes.com/2022/07/06/business/disinformation-board-dc.html [https://perma.cc/EX8P-5AN5] [https://web.archive.org/web/20230220233451/https://www.nytimes.com/2022/07/06/business/disinformation-board-dc.html] ("The Department of Homeland Security added the threat of false information to its periodic national terrorism advisory bulletins for the first time in February."); Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753 (2019).

7. *See generally* Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023 (2017); Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811 (2020); Thomas B. Nachbar, *Algorithmic Fairness, Algorithmic Discrimination*, 48 FLA. ST. U. L. REV. 509 (2021).

8. *See* NOAH GIANSIRACUSA, HOW ALGORITHMS CREATE AND PREVENT FAKE NEWS xi (2021) ("There has also been an increasing awareness of the role played by data-driven algorithms in the creation, dissemination, and detection/moderation of fake news.").

9. *See* Soroush Vosoughi, Deb Roy & Sinan Aral, *The Spread of True and False News Online*, 359 Science 1146 (Mar. 9, 2018).

10. *See* Daniela C. Manzi, *Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News*, 87 FORDHAM L. REV. 2623, 2628 (2019).

disinformation.[11] Against this backdrop, a judge has gone so far as to criticize disinformation-laden social media platforms such as Facebook as "a tool for evil."[12] The United Nations High Commissioner for Human Rights has denounced disinformation as "a symptom of global diseases," lamenting that legal regulation is still "insufficient."[13]

The regulation of disinformation, particularly that spread by algorithms, is one of today's thorniest legal problems.[14] Social media companies develop and utilize their algorithms as black boxes, rarely disclosing any information about their internal functioning and external ill effects.[15] Thus, regulators and the public alike lack access to the informatics of how social media platforms use algorithms.[16] The "trade secret" status of algorithms has made it exceedingly difficult to legally regulate the disinformation they create and/or disseminate.

Against this backdrop, this Article puts forward the concept of "algorithmic disinformation" to enhance comprehension of the severe harms brought about by algorithmically amplified disinformation on social media networks. Differing from the conventional wisdom,[17] this concept breaks disinformation into two categories: disinformation created and/or disseminated through algorithms and that created/disseminated without the use of algorithms. The Article demonstrates how recommendation algorithms have been utilized to spread disinformation via social

11. Pamela Madrid, *USC Study Reveals the Key Reason Why Fake News Spreads on Social Media* (Jan. 17, 2023), https://news.usc.edu/204782/usc-study-reveals-the-key-reason-why-fake-news-spreads-on-social-media [https://perma.cc/NSN8-LJED] [https://web.archive.org/web/20230512032943/https://news.usc.edu/204782/usc-study-reveals-the-key-reason-why-fake-news-spreads-on-social-media].

12. *Facebook is a "Tool for Evil", Says Judge as Mother Trolled Over Fake Claims She Tried To Kill a Baby Is Found Dead*, THE TELEGRAPH (Feb. 7, 2017), https://www.telegraph.co.uk/news/2017/02/07/facebook-tool-evil-says-judge-mother-trolled-fake-claims-tried [https://perma.cc/U87T-66CA] [http://web.archive.org/web/20230224173625/https://www.telegraph.co.uk/news/2017/02/07/facebook-tool-evil-says-judge-mother-trolled-fake-claims-tried].

13. *Rise of Disinformation A Symptom of 'Global Diseases' Undermining Public Trust: Bachelet*, UN NEWS (Jun. 28, 2022), https://news.un.org/en/story/2022/06/1121572 [https://perma.cc/X393-68D7] [http://web.archive.org/web/20230224173928/https://news.un.org/en/story/2022/06/1121572].

14. *See* Mark Verstraete et al., *Identifying and Countering Fake News*, 73 HASTINGS L.J. 821, 858 (2022) ("Fake news presents a complex regulatory challenge in the increasingly democratized and intermediated online information ecosystem."); Abby K. Wood & Ann M. Ravel, *Fool Me Once: Regulating "Fake News" and Other Online Advertising*, 91 S. CAL. L. REV. 1223, 1225 (2018) ("Untraceable online political advertising undermines key democratic values, and the problem is exacerbated by disinformation.").

15. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 191 (2015) (arguing that the black box society is unjust because "[d]ata is becoming staggering in its breadth and depth, yet often the information most important to us is out of our reach, available only to insiders").

16. *See infra* Part I.A.2.

17. Conventional wisdom divides the term "fake news" into three categories and then compares their differences: misinformation (false or misleading information disseminated without harmful intent); disinformation (false information disseminated with harmful intent); and mal-information (genuine information disseminated to cause harm). *See, e.g.*, Claire Wardle & Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*, COUNCIL EUR. (Sept. 27, 2017), https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c [https://perma.cc/2F8C-JFVX] [http://web.archive.org/web/20230120110652/https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c].

media networks,[18] and artificial intelligence ("AI")-powered generative algorithms used to create deepfake videos and convincingly journalistic fake news articles.[19]

The Article then suggests three legal principles for achieving the effective legal regulation of algorithmic disinformation: transparency, intelligibility, and accountability. It argues that to be effective such legal regulation must require social media companies to make their disinformation-related algorithms public, render them appropriately intelligible to users, and assume legal responsibility for curbing disinformation on their platforms.[20] Because of its market-based technology development and regulation policy, the United States has yet to adopt any laws regulating algorithmic disinformation, let alone these three legal principles.[21] In contrast, France and China recently passed laws aimed at regulating algorithmic information.[22] The Article evaluates the extent to which these laws have translated the principles of transparency, intelligibility, and accountability into legal rules and discusses the major problems with the principles' implementation.[23]

Based upon this discussion, the Article then proposes a new approach to regulating algorithmic disinformation. It recommends the launch of a multi-stakeholder governance mechanism that engages a suitable administrative agency with social media users and experts to examine the algorithmic treatment and policing of disinformation on social media platforms. In addition to Meta's Oversight Board, which deals primarily with content moderation,[24] the new mechanism would become another legal innovation in the regulation of algorithmic disinformation.

To carry out the proposed multi-stakeholder approach, the article proposes that the U.S. should take the lead in establishing an algorithmic disinformation review system ("ADRS") as a pilot program. Under the ADRS, the Federal Communications Commission ("FCC") would take charge of creating review panels comprising social media users and experts who volunteer to serve. Every two years, each panel would meet with two social media companies selected by the FCC and review their efforts to render their disinformation-related algorithms transparent and intelligible, as well as the effectiveness of their measures to reduce disinformation spread through their

18. *See infra* Part I.A.1.

19. *See infra* Part I.A.2.

20. *See infra* Part I.B; MARTHA MINOW, SAVING THE NEWS: WHY THE CONSTITUTION CALLS FOR GOVERNMENT ACTION TO PRESERVE FREEDOM OF SPEECH 108–09 (2021) (discussing social media platforms' responsibilities); HAOCHEN SUN, TECHNOLOGY AND THE PUBLIC INTEREST 121–55 (2022) (proposing three major responsibilities that should be imposed on technology companies).

21. *See infra* Part II.A.

22. *See infra* Part II.B–C.

23. *See infra* Part III.A.1.

24. *See* Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution To Adjudicate Online Free Expression*, 129 YALE L.J. 2418, 2425 (2020) ("Following his spring 2018 statement about a 'Supreme Court'-like structure, and in response to longstanding and increasingly vocal criticism demanding user accountability, Zuckerberg announced in November 2018 that Facebook would create an 'Independent Governance and Oversight' committee by the close of 2019 to advise on content policy and listen to user appeals on content decisions."); Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1603 (2018) (arguing that "platforms should be thought of as operating as the New Governors of online speech").

algorithms. Once the panels had issued their review reports and had them approved by the FCC, the social media company concerned would be required to take prompt action in accordance with the recommendations therein. Should the company fail to comply, the FCC would impose penalties.[25]

This Article makes three contributions to the scholarship on the legal regulation of disinformation in the digital age. First, it presents new legal principles as theoretical bases for developing legal rules regulating algorithmic disinformation. Having detected the growing danger of information pollution facilitated by algorithms, some scholars have called for the introduction of legal reforms to regulate algorithmic disinformation.[26] However, they have not suggested concrete regulatory principles for governing the legal treatment of such disinformation. This Article fills that void by articulating transparency, intelligibility, and accountability as the three major principles of a new legal mechanism for combating algorithmic disinformation.

Second, the Article constitutes the first attempt to conduct a comprehensive, systematic review of the legal regulatory approaches that various countries have adopted to tackle algorithmic disinformation. Scholars have highlighted the lack of any laws regulating online disinformation in the U.S.[27] and examined the anti-disinformation laws adopted in other jurisdictions such as Germany[28] and Singapore.[29] However, these laws do not entail legal rules that directly target algorithmic treatment and the policing of disinformation. The aforementioned laws recently passed in France and China steer legislative efforts toward the direct legal regulation of the algorithmic amplification of disinformation. The Article discusses the legislative purposes and major rules of these new laws and reveals their inadequacies in rendering disinformation-related algorithms sufficiently transparent and intelligible.

Third, the Article puts forward a new legal governance model aimed at effectively curbing algorithmic disinformation. Drawing upon fiduciary doctrine,[30] a number of

25. *See infra* Part III.B.

26. *See, e.g.*, PHILIP M. NAPOLI, SOCIAL MEDIA AND THE PUBLIC INTEREST: MEDIA REGULATION IN THE DISINFORMATION AGE 188–93 (2019) (calling for legal and regulatory reforms aimed at curbing disinformation in the public interest).

27. *See* Michael P. Goodyear, *Is There No Way To the Truth? Copyright Liability as a Model for Restricting Fake News*, 34 HARV. J.L. & TECH. 279, 282 (2020) ("Fake news is undoubtedly a problem in the United States, yet there are few legal constraints to stop it.").

28. *See* Ryan Kraski, *Combating Fake News in Social Media: U.S. and German Legal Approaches*, 91 ST. JOHN'S L. REV. 923, 923–24 (2017); Brittany Finnegan, *The Cost of Free Speech: Combating Fake News or Upholding the First Amendment?*, 75 U. MIAMI L. REV. 572, 579–605 (2021).

29. *See* Rebecca K. Helm & Hitoshi Nasu, *Regulatory Responses To 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation*, 21 HUM. RTS. L. REV. 302, 316–17 (2021) ("Singapore has adopted a more interventionist approach with information correction. Under the Protection from Online Falsehoods and Manipulation Act 2019, Singaporean authorities can issue a correction direction to require a person who has made a false statement, or the internet intermediary service provider, to make a correction notice in the specified form and manner.").

30. *See* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1209 (2016) ("An information fiduciary is a person or business who, because of their relationship with another, has taken on special duties with respect to the information they obtain in the course of the relationship.").

eminent scholars have suggested that social media companies can be trusted to police disinformation in good faith because they can be deemed trustees of the large swathes of user data they collect.[31] In this Article, I contend that it is the algorithms they use that are a main contributory force to the worsening disinformation chaos. Hence, my multi-stakeholder approach advocates for regulators to target the internal black-box informatics of such algorithms by exposing them to the light of day (the transparency principle) and to minimize their external ill effects by enabling users to understand them from the inside out (the intelligibility principle). My approach suggests the establishment of the ADRS to oversee algorithmic disinformation and to empower users' dynamic engagement in policing algorithms that generate and disseminate disinformation.[32]

The remainder of the Article proceeds as follows. Putting forward the new concept of algorithmic disinformation, Part I suggests transparency, intelligibility, and accountability as three principles for the legal regulation of such disinformation and discusses the challenges to bringing them to fruition. Part II then examines the regulatory approaches that the United States, France, and China have adopted to cope with algorithmic disinformation, with a focus on the laws recently passed in the latter two countries. Based on a critical evaluation of these approaches, Part III proposes a multi-stakeholder approach as an alternative means of achieving the effective legal regulation of such disinformation. It also presents a pathway toward establishing the ADRS in the United States and elsewhere.

I. ALGORITHMIC DISINFORMATION

In the age of social media and AI, companies are utilizing algorithms in their products, services, processes, and decision-making to an increasing extent. For example, platforms such as Facebook, YouTube, and Twitter use recommendation algorithms to target users with specialized content, while credit rating agencies use credit score algorithms to predict consumers' likelihood of a loan default.[33]

Algorithms also have tremendous power to create and disseminate disinformation. In this part, I define and discuss the new phenomenon of "algorithmic disinformation,"

31. *See* Philip M. Napoli & Fabienne Graf, *Social Media Platforms as Public Trustees: An Approach To the Disinformation Problem*, *in* ARTIFICIAL INTELLIGENCE AND THE MEDIA 108 (Taina Pihlajarinne & Anette Alén-Savikko eds., 2022) ("In the approach outlined here, we propose treating the massive aggregations of user data that serve as the economic foundation of these platforms as a *public resource*. Within the context of the public trust framework, this means treating aggregate user data as the *trust property* which effectively triggers the classification of the digital platforms as public trustees."); 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, 1162 (2018).

32. *See* Niva Elkin-Koren et al., *Social Media as Contractual Networks: A Bottom Up Check on Content Moderation*, 107 IOWA L. REV. 987, 994 (2022) (arguing that it is very important to "enable users to restrain platforms' discretion and safeguard their private interests").

33. *See* PASQUALE, *supra* note 15, at 4–6; SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 63–97 (2019); *see generally* DAVID SUMPTER, OUTNUMBERED: FROM FACEBOOK AND GOOGLE TO FAKE NEWS AND FILTER-BUBBLES—THE ALGORITHMS THAT CONTROL OUR LIVES (2018).

focusing in particular on the role played by recommendation algorithms and generative algorithms. I further consider how principles can be formed to govern the legal regulation of algorithmic disinformation.

A. TYPES OF DISINFORMATION

1. Recommendation Algorithms

Social media companies use recommendation algorithms to sort through vast amounts of content and present it to platform users in a manageable, appealing way that optimizes attention.[34] This automated service promotes news or other content based on users' membership of a particular demographic, their prior history of engagement on the platform, and/or the conduct of their family members or circle of friends.[35]

Recommendation algorithms can be divided into three major categories according to the filtering method employed. First, they can use collaborative filtering, which identifies the preferences of a large group of users and recommends content based upon the "underlying intuition . . . that if users A and B have similar taste in a product, then A and B are likely to have similar taste in other products as well."[36] Second, recommendation algorithms can employ content-based filtering, which focuses more specifically on the preferences and history of the individual user being targeted, recommending content similar to that in which the user has previously demonstrated interest.[37] Third, recommendation algorithms can employ hybrid systems. These use elements of both collaborative and content-based filtering, either by providing inputs to multiple algorithms in parallel and combining the results or by providing inputs to a single algorithm before passing on the output to further systems in sequence.[38]

Recommendation algorithms can easily spread and amplify fake news. One way in which they do so is by prioritizing the promotion of content that is aligned with a social media company's platform policy over content that users might be most interested in. For instance, in one of its earlier forms, Facebook's algorithm was designed to recommend content that received a large number of likes and clicks. Once it was discovered that this design had led to a rise in "clickbait," the algorithm target was changed to content that users were spending the most time consuming.[39] Then, after

34. TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION, AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 1–21 (2018).

35. Botambu Collins et al., *Trends in Combatting Fake News on Social Media—A Survey*, 5 J. INFO. & TELECOMM. 247, 250 (2021).

36. Vatsal, *Recommender Systems Explained*, TOWARDS DATA SCI. (July 12, 2021), https://towardsdatascience.com/recommendation-systems-explained-a42fc60591ed [https://perma.cc/V4J4-45KT] [https://web.archive.org/web/20230303221117/https://towardsdatascience.com/recommendation-systems-explained-a42fc60591ed?gi=755e28c89511].

37. *Id.*

38. *Id.*

39. Will Oremus et al., *How Facebook Shapes Your Feed*, WASH. POST (Oct. 26, 2021), https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works [https://perma.cc/

realizing that users were consuming content passively and increasingly taking their more active forms of communication to other platforms, Facebook again redesigned its algorithm to target "meaningful social interactions" by amplifying posts that attracted a large number of comments and replies. In reality, such posts were often those that had offended or angered users.[40] Facebook has reportedly conducted multiple studies indicating that the types of content most likely to promote engagement in the form of comments and replies can be considered harmful, such as politically divisive speech and misinformation.[41] Nevertheless, Facebook has continued to use this algorithm, even pushing pages with "engaging" content onto the feeds of users who do not follow those pages.[42] This dynamic has been exploited by troll farms that create fake news stories specifically designed to generate engagement so that the algorithm amplifies the content, in turn generating more clicks and thus ad revenue.[43]

Powered by technologies such as deep learning, recommendation algorithms have grown increasingly complex, with potentially negative consequences. Following a recent major policy change, YouTube shifted focus away from video clicks toward watch time.[44] Taken at face value, the new policy appears to benefit users by recommending videos that others are watching at length. Then, in 2015, YouTube developed a new algorithm incorporating deep learning technology to narrow down the vast pool of potential recommendations to a few hundred based on a user's "watched video history, keyword search history . . . the geographic region the user is logged in from, the type of device they are using, and the user's age and gender if they have provided that information," ranked according to both user-specific predictors and "a few hundred video-specific predictors, including details on the user's previous interactions with the channel the video is from."[45] The algorithm demotes a video each time it is recommended to a user but not clicked on.[46] The technology developed further in 2018 when YouTube introduced a deep reinforcement learning model designed to predict how long a user might spend watching the next recommended video, with the aim of hooking viewers into watching a succession of videos.[47] The algorithm operates through a reward function based on "something like the total amount of watch time each user spends in a sequence of up next recommendations

67D2-NG2P] [https://web.archive.org/web/20230303221141/https://www.washingtonpost.com/technology/interactive/2021/how-facebook-algorithm-works].

40. *Id.*

41. Karen Hao, *Troll Farms Reached 140 Million Americans a Month on Facebook Before 2020 Election, Internal Report Shows*, MIT TECH. REV. (Sept. 16, 2021), https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election [https://perma.cc/J2QX-763T] [https://web.archive.org/web/20230303192427/https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election].

42. *Id.*

43. *Id.*

44. *See* NOAH GIANSIRACUSA, HOW ALGORITHMS CREATE AND PREVENT FAKE NEWS 71 (2021).

45. *Id.* at 73.

46. *Id.*

47. *Id.* at 75.

before leaving the site."[48] This function encourages recommendations that might not initially generate much watch time but expose the user to a whole new topic or category of content, potentially leading them to remain active for longer than if they had consumed only familiar content.[49]

These advanced algorithmic designs have made fake news an endemic problem on YouTube. For instance, after a rise in the popularity of far-right politicians in Brazil, a 2019 Harvard study conducted for *The New York Times* found that following the chain of top recommendations from a video on a popular channel often led to videos on right-wing, conspiracy-filled channels on YouTube, including that of former President Jair Bolsonaro.[50] In a similar manner to Facebook, this effect is credited to the YouTube algorithm's emphasis on watch time, as people are drawn "in to content . . . [through] fear, doubt, and anger . . . the same emotions that right-wing extremists and conspiracy theorists have relied on for years."[51] In 2016, an engineer who had previously worked on YouTube's recommendation algorithm team designed a computer program to track where the platform would take a user from "seed" videos discovered after making what the engineer considered to be common or important searches relating to the U.S. presidential election.[52] His findings suggested that "divisive, sensational, and conspiratorial" content was systematically amplified by the platform, with, for example, the search "Who is Michelle Obama?" leading mostly to videos falsely claiming that she is a man.[53]

Falsehoods diffuse "significantly farther, faster, deeper, and more broadly than the truth in all categories of information" on social media.[54] According to a study looking at diffusion through retweets and the independent tweeting of true, false, and mixed rumors, the veracity of which were agreed upon by six independent fact-checking organizations,[55] the truth took about six times as long to reach 1500 people as falsehoods.[56] The study did not look specifically at the role of recommendation algorithms in structuring the presentation of content. Instead, it explained the difference in speed as a consequence of human behavioral tendencies, as false news evokes emotion and is more novel, and novel information is more likely to be retweeted.[57] Nevertheless, commentators have noted the importance of engagement to

48. *Id.*

49. *Id.*

50. *Id.* at 77; *see, e.g.*, Max Fisher & Amanda Taub, *How YouTube Radicalized Brazil*, N.Y. TIMES (Apr. 11, 2019), https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html [https://perma.cc/SU7V-RURX] [https://web.archive.org/web/20230303221357/https://www.nytimes.com/2019/08/11/world/americas/youtube-brazil.html].

51. *See* GIANSIRACUSA, *supra* note 44, at 79.

52. *Id.* at 86.

53. *Id.*

54. Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCIENCE 1146, 1147 (2018).

55. *Id.* at 1146 ("Here we investigate the differential diffusion of true, false, and mixed (partially true, partially false) news stories using a comprehensive data set of all of the fact-checked rumor cascades that spread on Twitter from its inception in 2006 to 2017. The data include ~126,000 rumor cascades spread by ~3 million people more than 4.5 million times.").

56. *Id.* at 1148.

57. *Id.* at 1149.

recommendation algorithms, suggesting that the aforementioned study demonstrates the dangers of such an approach.[58] If false information is more likely to generate emotional reactions and retweets, then it is possible that Twitter's algorithm has played a role in recommending tweets containing false rumors or will do so in the future.

Recommendation algorithms can also amplify disinformation to influence user beliefs by creating echo chambers and filter bubbles. Echo chambers refer to homogenous information environments resulting from users' own choices to follow like-minded individuals on social media, whereas filter bubbles are similar information environments created surreptitiously by a platform's automated efforts to understand individual user preferences and reflect them in its recommendations.[59] Recommendation algorithms can distort the character and form of our social capital, that is, our connections to others and, most relevantly here, "the level of trust (and trustworthiness) and the informal rules and common understandings that facilitate communication."[60] Filter bubbles trick us "into thinking our social and political 'bubbles' are representative."[61] Furthermore, filter bubbles can enable confirmation bias, as "human beings are wired to trust familiar sources that confirm their existing world view."[62] Social media platforms that amplify disinformation can be particularly problematic, as studies have shown that people are more inclined to accept information when they engage with it in the passive environment created by social media.[63]

2. Generative Algorithms

As the preceding section demonstrates, recommendation algorithms spread and amplify algorithmic disinformation. Algorithms can also generate disinformation in the form of deepfake news. I refer to algorithms of this type in general as generative algorithms.

58. Robinson Meyer, *The Grim Conclusions of the Largest-Ever Study of Fake News*, ATLANTIC (Mar. 8, 2018), https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104 [https://perma.cc/VAR2-WV27] [https://web.archive.org/web/20230303221531/https://www.theatlantic.com/technology/archive/2018/03/largest-study-ever-fake-news-mit-twitter/555104] ("Tromble, the political-science professor, said that the findings would likely apply to Facebook, too. 'Earlier this year, Facebook announced that it would restructure its News Feed to favor 'meaningful interaction,' she told me. 'It became clear that they would gauge 'meaningful interaction' based on the number of comments and replies to comments a post receives. But, as this study shows, that only further incentivizes creating posts full of disinformation and other content likely to garner strong emotional reactions,' she added.").

59. Samuel C. Rhodes, *Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals To Be Less Critical of Political Misinformation*, 39 POL. COMMC'N 1, 4–5 (2022).

60. *See* ELISABETH COSTA & DAVID HALPERN, THE BEHAVIOURAL SCIENCE OF ONLINE HARM AND MANIPULATION AND WHAT TO DO ABOUT IT 24 (2019), https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf [https://perma.cc/N4HT-QUBY] [https://web.archive.org/web/20230216160800/https://www.bi.team/wp-content/uploads/2019/04/BIT_The-behavioural-science-of-online-harm-and-manipulation-and-what-to-do-about-it_Single.pdf].

61. *Id.*

62. *See* Rhodes, *supra* note 59, at 6.

63. *Id.*

a. Visual Deepfakes

Deepfakes are a type of media in which a person's likeness in an existing image or video is replaced with that of someone else using technologies such as deep learning algorithms.[64] Toward the end of 2017, for example, an anonymous Reddit user posted a deepfake video on the platform[65] that he or she had created using deep learning algorithms to replace the face of an actress in a pornographic video with the face of the actress Gal Gadot. The video went viral, encouraging the user to create similar videos through algorithmic face-swapping.[66]

There are two main steps involved in producing deepfakes. First, deep learning algorithms are used to digitally draw artificial faces. The algorithm is fed numerous photographs of people with a box manually drawn around each face, and the algorithm eventually learns how to draw the boxes on its own. Second, a face-mapping algorithm is applied. Called an autoencoder, this algorithm can learn whatever it needs to describe a face with the aid of data. It might, for example, note the locations of different facial features, quantify their shape, and represent different hairstyles and colors numerically.[67]

Visual deepfakes are intended to spread disinformation.[68] In early November 2020, amid the contentions of voter fraud in the U.S. presidential election, a video was edited to make it appear that Joe Biden was admitting to voter fraud when in fact he was discussing his campaign's efforts to prevent it. The video was viewed more than 17 million times on social media platforms.[69] Another video aired by a British television station in December 2020 featured a deepfake Queen Elizabeth delivering her Christmas message while dancing.[70] Other examples include deepfake videos depicting

64. Riana Pfefferkorn, *The Threat Posed By Deepfakes To Marginalized Communities*, BROOKINGS (Apr. 21, 2021), https://www.brookings.edu/techstream/the-threat-posed-by-deepfakes-to-marginalized-communities [https://perma.cc/B8MT-KFVF] [https://web.archive.org/web/20230303222039/https://www.brookings.edu/techstream/the-threat-posed-by-deepfakes-to-marginalized-communities].

65. GIANSIRACUSA, *supra* note 44, at 46 ("This event marked the ominous beginning of a dark saga in the history of artificial intelligence that continues to unfold today.").

66. *Id.* at 46.

67. GIANSIRACUSA, *supra* note 44, at 47.

68. Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J.L. & TECH. 1, 9–13 (2020) (describing how deepfakes create "war on reality" by confirming preconceived notions and causing individuals to question authenticity of any audiovisual record); Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1762–63 (2019); *Deepfake Queen to Deliver Channel 4 Christmas Message*, BBC (Dec. 23, 2020), https://www.bbc.com/news/technology-55424730 [https://perma.cc/BQ37-K6XS] [https://web.archive.org/web/20230224234227/https://www.bbc.com/news/technology-55424730].

69. Sheera Frenkel, *Deceptively Edited Video of Biden Proliferates on Social Media*, N.Y. TIMES (Nov. 2, 2020), https://www.nytimes.com/2020/11/02/technology/biden-video-edited.html [https://perma.cc/97NR-AJDS] [https://web.archive.org/web/20230224234526/https://www.nytimes.com/2020/11/02/technology/biden-video-edited.html].

70. Zamira Rahim, *'Deepfake' Queen Delivers Alternative Christmas Speech, In Warning About Misinformation*, CNN (Dec. 25, 2020), https://www.cnn.com/2020/12/25/uk/deepfake-queen-speech-christmas-intl-gbr/index.html [https://perma.cc/6NTN-MZ9T] [https://web.archive.org/web/https://edition.cnn.com/2020/12/25/uk/deepfake-queen-speech-christmas-intl-gbr/index.html].

"an Israeli soldier committing an atrocity against a Palestinian child, a European Commission official offering to end agricultural subsidies on the eve of an important trade negotiation, and a Rohingya leader advocating violence against the security forces in Myanmar."[71]

The number of visual deepfakes has been growing exponentially on social media, with the number doubling roughly every six months.[72] As of June 2020, approximately 50,000 deepfake videos had been identified.[73] A range of applications equipped with sophisticated algorithms, including DFaker, faceswap, faceswap-GAN, FakeApp, and DeepFaceLab, have been made available to the public, enabling the quick and easy production of deepfake video and audio.

Deepfake videos can cause enormous harm. Given the right timing and a convincing script, they can spark violence in cities, bolster insurgent narratives of alleged atrocities, and even exacerbate the political divisions within a society. They can also present opportunities for sabotage, say, in the context of sensitive international trade deal negotiations conducted via digital means.[74]

b. Textual Deepfakes

AI algorithms can be used to create entirely false or misleading news articles or to detect them. GROVER, for example, is an AI system that employs algorithms to create fake news articles but can also be used to identify articles written by machines. In a recent study, digital technology experts found GROVER to have churned out news articles falsely asserting that vaccines are linked to autism spectrum disorder written in the distinctive style of specific news outlets. According to the study, readers found these articles to be more convincing than those written by human beings.[75]

Writing in the style of *The New York Times* science section, for example, GROVER generated a headline, an author name, and the opening of an article crediting the discovery of the aforementioned link to scientists from the University of California-San Diego and to those working for the federal government. The article began as follows:

> Those who have been vaccinated against measles have a more than 5-fold higher chance of developing autism, researchers at the University of California San Diego School of Medicine and the

71. Robert Chesney & Danielle K. Citron, *Disinformation on Steroids: The Threat of Deep Fakes*, COUNS. ON FOREIGN RELATIONS (Oct. 16, 2018), https://www.cfr.org/report/deep-fake-disinformation-steroids [https://perma.cc/PXZ8-JF6C] [https://web.archive.org/web/https://www.cfr.org/report/deep-fake-disinformation-steroids].

72. GIANSIRACUSA, *supra* note 44, at 49.

73. *Id.*

74. Chesney & Citron, *supra* note 71.

75. Dan Robitzski, *New AI Generates Horrifyingly Plausible Fake News*, FUTURISM (May 31, 2019), https://futurism.com/ai-generates-fake-news [https://perma.cc/A8N7-DQ6S] [https://web.archive.org/web/https://futurism.com/ai-generates-fake-news].

> Centers for Disease Control and Prevention report today in the Journal of Epidemiology and Community Health.[76]

The study also demonstrated GROVER's ability to refine its output over time to better match specific publications.[77] The researchers supplied the system with a headline claiming that vaccines cause autism and instructed it to write a full article in the style of *Wired*, which it duly did. Another example cited by the researchers was an article that GROVER "wrote" in the style of *The Washington Post* claiming that Donald Trump had been impeached based on new evidence from the Mueller Report:

> WASHINGTON—The House voted to impeach President Donald Trump Wednesday after releasing hundreds of pages of text messages that point to clear evidence of obstruction of justice and communication with the head of the Trump Organization about a potential business deal in Russia. The 220-197 vote came after weeks of debate over whether new evidence released by special counsel Robert Mueller's office signaled sufficient grounds for Trump's removal from office. The president personally denounced the move, announcing his intent to veto the resolution and accusing Democrats of plotting to remove him from office through a "con job."[78]

Although deepfake video and audio may be used in the future to create sensational stories that come to dominate the news, perhaps to distract from a scandal, the potential of "textfakes" is arguably even more sinister. These fraudulent texts, when distributed among and disguised as real posts on Twitter, Facebook, and/or Reddit, could be used to manufacture opinion in "sophisticated, extensive influence campaigns." If produced in sufficiently large numbers by more powerful generative AI systems such as ChatGPT, textfakes have the potential to warp the social communication ecosystem, with algorithmically generated content receiving algorithmically generated responses that, in turn, feed into recommendation algorithms.[79]

76. *Id.*

77. *Id.*

78. *Id.*

79. Renee Diresta, *AI-Generated Text Is the Scariest Deepfake of All*, WIRED (July 31, 2020), https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all [https://perma.cc/WL6C-6R9N] [https://web.archive.org/web/https://www.wired.com/story/ai-generated-text-is-the-scariest-deepfake-of-all]; Luke Hurst, *Rapid Growth of "News" Sites Using AI Tools like ChatGPT is Driving the Spread of Misinformation*, EURONEWS (May 2, 2023), https://www.euronews.com/next/2023/05/02/rapid-growth-of-news-sites-using-ai-tools-like-chatgpt-is-driving-the-spread-of-misinforma [https://perma.cc/ZX6U-69AP] [https://web.archive.org/web/20230512033630/https://www.euronews.com/next/2023/05/02/rapid-growth-of-news-sites-using-ai-tools-like-chatgpt-is-driving-the-spread-of-misinforma].

B. MAJOR CHALLENGES TO EFFECTIVE LEGAL REGULATION

Given that technology companies develop and apply recommendation algorithms and generative algorithms, they can and should be legally required to regulate algorithmic disinformation. They are the entities best placed to do so because they can detect and remove disinformation. In this section, I put forward three major legal principles for governing the effective regulation of algorithmic disinformation by technology companies and also explore potential challenges to the implementation of these principles.

1. Transparency

Appropriate disclosure of information on social media algorithms that shape the experience of platform users is an essential step in neutralizing algorithmic disinformation. By allowing the public to monitor and scrutinize algorithms' functioning and potential for harm, a transparency requirement would likely push technology companies to develop and apply algorithms in a manner less conducive to the spread of disinformation.[80]

Algorithms will constitute a critical component of virtually every business in the future because almost all of the business insights and decisions of tomorrow will be data driven.[81]In reality, however, there are commercial barriers to achieving algorithmic transparency. The value of algorithms means that technology companies routinely develop and operate them in a "black box" manner. In 2002, for example, Google's search algorithm earned the company half a billion dollars, while the latest version earns it that much in just three days.[82] Technology companies defend the confidentiality of their algorithms on the grounds that they need to protect such commercially valuable systems and maintain security to protect both their own gatekeeping role and the privacy of users.[83] Although it is widely known that the Google search algorithm influences web positioning, no one outside Google knows all of the parameters that influence the ranking of a web page or the percentage weighting of those parameters.

80. *See* Nicholas Diakopoulos & Michael Koliska, *Algorithmic Transparency in the News Media*, 5 DIGIT. JOURNALISM 809 (2017) ("The growing use of difficult-to-parse algorithmic systems in the production of news, from algorithmic curation to automated writing and news bots, problematizes the normative turn toward transparency as a key tenet of journalism ethics. Pragmatic guidelines that facilitate algorithmic transparency are needed.").

81. Salim Ismail, *Why Algorithms Are The Future Of Business Success*, GROWTH INST., https://blog.growthinstitute.com/exo/algorithms [https://perma.cc/4JZN-YY3W] [https://web.archive.org/web/https://blog.growthinstitute.com/exo/algorithms].

82. Megan Graham & Jennifer Elias, *How Google's $150 Billion Advertising Business Works*, CNBC (May 18, 2021), https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html [https://perma.cc/9KK6-BKAZ] [https://web.archive.org/web/https://www.cnbc.com/2021/05/18/how-does-google-make-money-advertising-business-breakdown-.html].

83. *See* Paddy Leerssen, *The Soap Box as a Black Box: Regulating Transparency in Social Media Recommender Systems*, EUROPEAN J.L. & TECH., Sept. 30, 2020, at 13.

Given Google's governance of the main activities on the web, understanding this part of its operation would confer a valuable competitive advantage.

In the case of Twitter's recommendation algorithm, little is known with respect to the exact science behind it, which made Elon Musk's stated intention to disclose the secret formula to the public a major talking point before his purchase of the platform.[84] Both ends of the political spectrum had expressed concern about the algorithm's potential for manipulation. For instance, conservative media outlets are still echoing Musk's claims about Twitter's capacity for manipulation and sharing with audiences his instructions for configuring timelines to display tweets in chronological order rather than based on recommendations.[85] Although some studies have asserted that right-leaning political content is more likely to be amplified by Twitter's algorithm,[86] concerns about the spread of disinformation on Twitter exist across the political divide.

2. Intelligibility

Even if technology companies publicize their algorithms, people still need help understanding how these algorithms operate to amplify disinformation or prevent its dissemination. Publicizing algorithms in and of itself may be insufficient to allow users to learn much of great value about the operations of platforms. For instance, whereas the microblogging platform Mastodon, which was launched in response to concerns over Twitter's dominance, has posted its code on the software repository GitHub, users remain in the dark about the business structures and processes involved in Twitter's development.[87] Therefore, making algorithms appropriately intelligible is another essential step in regulating algorithmic disinformation.

However, making algorithms intelligible is exceedingly difficult for several reasons. First, there are technological difficulties. The decisions of algorithms are guided by

84. Reed Albergotti, *Elon Musk Wants Twitter's Algorithm To Be Public. It's Not That Simple*, WASH. POST (Apr. 16, 2022), https://www.washingtonpost.com/technology/2022/04/16/elon-musk-twitter-algorithm [https://perma.cc/88UW-QGKW] [https://web.archive.org/web/https://www.washingtonpost.com/technology/2022/04/16/elon-musk-twitter-algorithm].

85. *See, e.g.*, Anders Hagstrom, *Elon Musk Says Twitter 'Manipulating' Users, Issues a How-To on Fixing Your Feed*, FOX BUS. (May 15, 2022), https://www.foxbusiness.com/politics/elon-musk-twitter-algorithm-advice-manipulated [https://perma.cc/E5ZA-49TA] [https://web.archive.org/web/https://www.foxbusiness.com/politics/elon-musk-twitter-algorithm-advice-manipulated]; Brian Freeman, *Musk Tells Users To Switch Off Manipulative Twitter Algorithm*, NEWSMAX (May 15, 2022), https://www.newsmax.com/newsfront/elon-musk-twitter-news-feed-algorithm/2022/05/15/id/1069981 [Perma.cc link is unavailable.] [https://web.archive.org/web/https://www.newsmax.com/newsfront/elon-musk-twitter-news-feed-algorithm/2022/05/15/id/1069981].

86. Shoaib Jameel, *Twitter's Algorithm Favours the Political Right, a Recent Study Finds*, CONVERSATION (Jan. 31, 2022), https://theconversation.com/twitters-algorithm-favours-the-political-right-a-recent-study-finds-175154 [https://perma.cc/6VM4-VY7Z] [https://web.archive.org/web/https://theconversation.com/twitters-algorithm-favours-the-political-right-a-recent-study-finds-175154].

87. Chris Stokel-Walker, *The Problem With Elon Musk's Plan To Open-Source the Twitter Algorithm*, MIT TECH. REV. (Apr. 27, 2022), https://www.technologyreview.com/2022/04/27/1051472/the-problems-with-elon-musks-plan-to-open-source-the-twitter-algorithm [https://perma.cc/XQV3-SZBW] [https://web.archive.org/web/https://www.technologyreview.com/2022/04/27/1051472/the-problems-with-elon-musks-plan-to-open-source-the-twitter-algorithm].

complex machine learning processes that are inscrutable to outside observers.[88] Social media recommendation algorithms include an array of inputs and mechanisms such as moderation and filtering, the promotion of paid content, and user profiles.[89] Second, the importance of user behavior in steering the operation of recommendation algorithms, such as user likes, comments, ratings, and clicks, means that the systems are not fully pre-determined or controlled by their platforms.[90] This complexity means that efforts such as Facebook's "Why am I seeing this?" have been criticized for failing to adequately explain algorithm functioning.[91]

3. Accountability

Another major challenge for the effective legal regulation of algorithmic disinformation is how to make technology companies accountable.[92] As social media are increasingly becoming people's main source of news, this accountability issue looms large in the face of the social and political controversies, disputes, conflicts, and even catastrophes caused by disinformation. For example, with respect to the 2021 U.S. Capitol attack, a leading commentator has lamented:

> [Social media companies] have not only allowed [politicians] to lie and sow division for years, their business models have exploited our biases and weaknesses and abetted the growth of conspiracy-touting hate groups and outrage machines. They have done this without bearing any responsibility for how their products and business decisions affect our democracy; in this case, including allowing an insurrection to be planned and promoted on their platforms.[93]

Indeed, social media companies' failure to prevent the spread of algorithmic disinformation necessitates urgent consideration of how and why they should be held accountable for such failure. Two major accountability issues arise. The first issue concerns the basis on which a social media company should be held legally accountable for amplifying a specific piece of disinformation. The enormous amount of disinformation circulating on any given social media platform makes such a case-by-

88. Leerssen, *supra* note 83, at 3.

89. Matthew Gooding, *Elon Musk's Plan for an Open-Source Algorithm Won't Solve Twitter's Problems*, TECH MONITOR (Apr. 26, 2022), https://techmonitor.ai/technology/ai-and-automation/open-source-twitter-algorithm-elon-musk [https://perma.cc/8VGX-P25G] [https://web.archive.org/web/https://techmonitor.ai/technology/ai-and-automation/open-source-twitter-algorithm-elon-musk].

90. *See* Leerssen, *supra* note 83, at 4–5.

91. *Id.* at 13.

92. *See generally* HAOCHEN SUN, TECHNOLOGY AND THE PUBLIC INTEREST 121-55 (2022); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 60 (2019).

93. Yaël Eisenstat, *How To Hold Social Media Accountable for Undermining Democracy*, HARV. BUS. REV. (Jan. 11, 2021), https://hbr.org/2021/01/how-to-hold-social-media-accountable-for-undermining-democracy [https://perma.cc/7ZFA-XGTE] [https://web.archive.org/web/20230217145518/https://hbr.org/2021/01/how-to-hold-social-media-accountable-for-undermining-democracy].

case approach difficult to sustain. As algorithms are the major vehicle for creating and disseminating disinformation on social media, it would make more sense to hold social media companies accountable for the algorithms they develop or apply. Should such an approach be adopted, a follow-up issue would be how to ascertain whether their algorithms have actually led to disinformation. This issue is closely intertwined with the transparency and intelligibility principles discussed above because both principles are necessary for determining how the algorithms concerned have contributed to the creation and spread of disinformation.

The second issue relates to which party should be held accountable if a case involves both a social media platform and its user(s). Policy-wise, it makes sense to assert that organizations using algorithms must be held accountable for the decisions made by those algorithms.[94] In reality, however, it is very difficult to ascertain how and why they should be deemed accountable. Take recommendation algorithms as an example. They themselves are susceptible to manipulation. One risk comes from bots: software-controlled social media accounts designed to emulate human activities but at a much higher volume of output.[95] Research into bots has found their capabilities to include searching Twitter for keywords or phrases and automatically retweeting posts containing them; automatically following users who follow a particular account or make posts containing certain keywords or phrases; automatically replying to tweets meeting certain criteria; and searching Google to find news articles meeting certain criteria and linking them in automatic replies to other users.[96] Given recommendation algorithms' focus on engagement, bots have the potential to harness these systems to generate synthetic virality by automating these forms of engagement.[97] For instance, one empirical study of Twitter found that, after being particularly active in amplifying content in the very early spreading moments before an article goes viral and targeting "influential users through replies and mentions," bots are able to amplify low-credibility content "to the point that it is statistically indistinguishable" from fact-checked

94. *See* Nicholas Diakopoulos, *Algorithmic Accountability: Journalistic Investigation of Computational Power Structures*, 3 DIGIT. JOURNALISM 398 (2015); Megan Rose Dickey, *Algorithmic Accountability*, TECHCRUNCH (Apr. 30, 2017), https://techcrunch.com/2017/04/30/algorithmic-accountability [https://perma.cc/HP9T-FSHN] [https://web.archive.org/web/20230216225641/https://techcrunch.com/2017/04/30/algorithmic-accountability].

95. Emilio Ferrara, *Bots, Elections, and Social Media: A Brief Overview*, *in* DISINFORMATION, MISINFORMATION, AND FAKE NEWS IN SOCIAL MEDIA 95–96 (Kai Shu et al. eds., 2020).

96. Alessandro Bessi & Emilio Ferrara, *Social Bots Distort the 2016 U.S. Presidential Election Online Discussion*, FIRST MONDAY (Nov. 7, 2016), https://firstmonday.org/ojs/index.php/fm/article/view/7090/5653 [https://perma.cc/6DTQ-J8Z6] [https://web.archive.org/web/20230216225736/https://firstmonday.org/ojs/index.php/fm/article/view/7090/5653].

97. Chengcheng Shao et al., *The Spread of Low-credibility Content by Social Bots*, 9 NATURE COMMC'NS 1 (2018); Lisa-Maria Neudert, *Future Elections May Be Swayed by Intelligent, Weaponized Chatbots*, MIT TECH. REV. (Aug. 22, 2018), https://www.technologyreview.com/s/611832/future-elections-may-be-swayed-by-intelligent-weaponized-chatbots [https://perma.cc/QM3N-N2M2] [https://web.archive.org/web/20230216225746/https://arxiv.org/abs/1707.07592].

articles.[98] Against this backdrop, it is quite difficult to determine whether a platform such as Twitter or the bot account user should be held accountable for the disinformation concerned.

II. CURRENT REGULATORY APPROACHES

How can governments effectively curb algorithmic disinformation by responding to the challenges of transparency, intelligibility, and accountability presented above? In this part, I show that governments, in general, have developed three regulatory approaches. The United States has adopted a market-based approach, allowing technology companies to regulate algorithmic disinformation by themselves. In contrast, France has enacted a law requiring technology companies to report their efforts to combat disinformation and publicize their algorithms. As this part shows, this French law represents a modest legislative approach to regulating algorithmic disinformation. China, in contrast, has adopted a stringent legislative approach that requires technology companies to not only publicize their algorithms but also to make them relatively intelligible.

A. UNITED STATES' MARKET-BASED APPROACH

In the United States, Congress has enacted no specific law to regulate the creation and spread of algorithmic disinformation despite its enormous potential to disrupt American society, and nor has any federal agency adopted administrative regulations or strategic measures to counter algorithmic disinformation.[99] Instead, the United States relies on self-regulation by social media platforms. We can call the U.S. approach a market-based approach to regulating disinformation because it allows market actors to regulate in line with their own cost and benefit calculations.

This market-based approach is attributable to three legal and policy aspects of the platform economy in the United States. First, shareholder value theory dominates the corporate policy-making process and has to date prevented social media companies from taking on additional responsibilities to proactively curb disinformation.[100] According to shareholder value theory, a company's sole social responsibility is to make

98. *Id.* at 10. A further analysis found that bots were employed to distort online discussion in the runup to the 2016 U.S. presidential election and 2018 midterms, and to push the "MacronLeaks" disinformation campaign in the runup to the 2017 French presidential election. *See* Ferrara, *supra* note 95, at 109–10.

99. Philip M. Napoli & Fabienne Graf, *Social Media Platforms as Public Trustees: An Approach To The Disinformation Problem, in* ARTIFICIAL INTELLIGENCE AND THE MEDIA 94 (Taina Pihlajarinne & Anette Alén-Savikko eds., 2022) ("Yet, from a regulation and policy standpoint, the federal government in the USA has done virtually nothing to confront the social media disinformation problem. None of the many pieces of social media-related legislation that are at various stages of consideration within Congress address the disinformation problem in any direct way. Nor have any of the regulatory agencies with potential jurisdiction in this space (such as the Federal Trade Commission (FTC), the Federal Communications Commission (FCC), and the Federal Election Commission (FEC)) introduced substantive regulatory interventions.").

100. SUN, *supra* note 92, at 105–106.

as much profit as possible for its shareholders. The theory's main champion, Milton Friedman, stated in *Capitalism and Freedom* that

> [t]he view has been gaining widespread acceptance that corporate officials and labor leaders have a "social responsibility" that goes beyond serving the interest of their stockholders or their members. This view shows a fundamental misconception of the character and nature of a free economy. In such an economy, there is one and only one social responsibility of business—to use its resources and engage in activities designed to increase its profits so long as it stays within the rules of the game, which is to say, engages in open and free competition, without deception or fraud.[101]

As a result of shareholder value theory, U.S. corporate law treats companies as profit-maximizing institutions with virtually no social responsibilities.[102] With the rise of the platform economy, the use of recommendation algorithms to provide advertising services has become the most significant source of revenue for most social media companies. In 2017 alone, $200 billion was spent on advertising in the United States. An estimated 437 billion hours of ad-supported content was consumed in the United States in 2016.[103] Social media platforms have become key targets for investment in advertising, largely because their algorithms can keep users engaged, thereby increasing the amount of attention available for sale.[104] Platforms utilize recommendation algorithms to determine the order in which content is presented or suggested to users on the platform interface. To keep users engaged, Facebook, for example, draws upon an estimated twenty-five times the computational power of that used by IBM's Deep Blue supercomputer in the operation of News Feed's presentation of content posted by friends, shared by paying advertisers, and suggested by the platform's recommendation algorithms.[105] Shareholder value theory has encouraged social media companies to use recommendation algorithms to maximize shareholder interests through advertising

101. *Id.* at 106.

102. *See, e.g.*, Francis v. United Jersey Bank, 432 A.2d 814, 824 (N.J. 1981); Revlon, Inc. v. MacAndrews & Forbes Holdings, Inc., 506 A.2d 173, 179 (Del. 1986); Thomas A. Smith, *The Efficient Norm for Corporate Law: A Neotraditional Interpretation of Fiduciary Duty*, 98 MICH. L. REV. 214, 214 (1999); Gregory S. Crespi, *Rethinking Corporate Fiduciary Duties: The Inefficiency of the Shareholder Primacy Norm*, 55 SMU L. REV. 141, 141 (2002).

103. Tim Wu, *Blind Spot: The Attention Economy and the Law*, 82 ANTITRUST L.J. 771, 784 (2019).

104. COSTA & HALPERN, *supra* note 60, at 20 ("[L]ikes or retweets capture attention and prompt frequent rechecking, bottomless pages keep users scrolling, and swipes and streaks make browsing feel like a game, removing breaks that might otherwise prompt a natural end to the task and our attention to shift elsewhere.").

105. *Id.* YouTube initially sought to increase video views through its recommendation algorithm, but, in 2012, it made the decision to instead focus on increasing watch time in order to maximize user attention and increase potential advertising revenue. In 2018, YouTube's Chief Product Officer indicated that its system was having the desired effect by announcing that seventy percent of the total time users spent watching videos came from recommendations. *See* GIANSIRACUSA, *supra* note 44, at 70–71.

profits. Curbing disinformation is not considered a major responsibility for these companies.[106]

Second, Section 230 of the Communications Decency Act[107] does not obligate social media platforms to monitor and remove disinformation. Rather, it prevents them from being treated as the publishers of user-generated content, thereby protecting them from any legal liability for such content.[108] In other words, social media companies enjoy broad immunity against civil lawsuits over online content on their networks or platforms unless they materially contributed to the creation of unlawful content.[109] Therefore, Section 230 allows platforms to evade responsibility for amplifying or recommending disinformation created or disseminated by users.[110]

Third, the First Amendment to the U.S. Constitution may present legal challenges to the active regulation of algorithms. The First Amendment guarantees freedom of speech by prohibiting Congress from enacting any law restricting the right of the press or individuals to speak freely.[111] Under First Amendment jurisprudence, computer code[112] and search engine results produced by algorithms are protected speech.[113] Underlying the ranking and recommendation algorithms that amplify disinformation on online platforms is computer code. In addition, such algorithms perform in a similar way to those producing search results. The regulation of ranking and recommendation algorithms would therefore be subject to a judicial review that might well find it to be in violation of the First Amendment.[114]

106. *See* SUN, *supra* note 92, at 106.

107. Telecommunications Act of 1996, Pub. L. No. 104–104, §§ 501–61, 110 Stat. 56, 133–43 (codified as amended in scattered sections of 18 and 47 U.S.C.).

108. Zeran v. Am. Online, Inc., 129 F.3d 327, 328 (4th Cir. 1997) (holding that Section 230 granted AOL an affirmative defense of immunity and that AOL was not liable for defamatory statements posted on the platform).

109. Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L.J. 639, 655–56 (2014).

110. Danielle K. Citron & Robert Chesney, *Disinformation on Steroids: The Threat of Deep Fakes*, CYBER BRIEF (2018), [https://scholarship.law.bu.edu/shorter_works/30] [https://perma.cc/V9VF-J4DF] [https://web.archive.org/web/20230309142235/https://scholarship.law.bu.edu/shorter_works/30] ("Social media platforms have long been insulated from liability for distributing harmful content. Section 230 of the Communications Decency Act of 1996 broadly immunizes online service providers in relation to harms caused by user-generated content, with only a few exceptions.").

111. U.S. CONST. amend. I.

112. *See, e.g.*, Universal City Studios, Inc. v. Corley, 273 F.3d 429 (2d Cir. 2001); Green v. U.S. Dep't of Just., 392 F. Supp. 3d 68 (D.D.C. 2019); Bernstein v. U.S. Dep't of State, 922 F. Supp. 1426 (N.D. Cal. 1996).

113. *See, e.g.*, E-ventures Worldwide, LLC v. Google, Inc., 188 F. Supp. 3d 1265, 1273–75. (M.D. Fla. 2016) ("Google's PageRanks are pure opinions of the website's relevancy to a user's search query, incapable of being proven true or false."); Zhang v. Baidu.com Inc., 10 F. Supp. 3d 433, 437–39 (S.D.N.Y. 2014); Kinderstart.com, LLC v. Google, Inc., No. C 06-2057 JF (RS), 2007 WL 831806, at *13–15 (N.D. Cal. Mar. 16, 2007) (finding Google not liable for alleged First Amendment violation because it was not a government actor); Langdon v. Google, Inc., 474 F. Supp. 2d 622, 629–30 (D. Del. 2007); Search King, Inc. v. Google Tech., Inc., No. CIV-02-1457-M, 2003 WL 21464568, at *2–3 (W.D. Okla. May 27, 2003).

114. Alan K. Chen, *Free Speech, Rational Deliberation, and Some Truths About Lies*, 62 WM. & MARY L. REV. 357, 361 (2020) ("One reason for this may be that the First Amendment's Free Speech Clause likely represents a significant barrier to such efforts. State regulation of fake news dissemination would be inherently content-based, and therefore suspect under current doctrine, particularly since the Supreme Court has rejected the proposition that lies are categorically exempt from First Amendment protection.").

Any potential disinformation regulation in the United States would have to be reconciled with the "marketplace of ideas" doctrine underpinning the constitutional protection of free speech.[115] This doctrine holds that the best test of truth is for an idea to be accepted in the marketplace through free competition with other ideas.[116] There is a strong belief in the U.S. that the government must remain neutral in the exchange of ideas, avoiding discrimination based on differing viewpoints.[117] The U.S. Supreme Court has therefore ruled that free speech protection extends to false speech. For example, in *United States v. Alvarez*,[118] the Stolen Valor Act criminalizing falsehoods about the receipt of military honors was struck down owing in part to concerns that it would create a dangerous precedent for the overly broad regulation of ideas and allow the state to become the arbiter of truth.

Because of the market-based approach that prevails in the United States, social media companies are not legally bound to actively tackle disinformation and enjoy nearly complete latitude in deciding what kinds of disinformation to remove from their platforms. Owing to the proliferation of disinformation in recent years, however, social media companies have come under mounting pressure from the public to tackle disinformation and have become more proactive in this regard. Following the Cambridge Analytica scandal, for example, Facebook enhanced the transparency of its takedown procedures, publishing internal enforcement guidelines stating that disinformation would not be removed but simply presented lower down in news feeds to reduce the economic incentives for disinformation.[119] Facebook also established an appeal body overseen by an independent board to review the most challenging content decisions, committed itself to being fully bound by the board's decisions, and suggested that the board could set policy moving forwards.[120] The establishment of an independent external appellate body responsible for the private transnational Internet adjudication of online speech is a major example of self-regulation in the disinformation context.[121]

YouTube, too, has adopted several measures to address the issue of fake news on its platform. First, it reviews problematic content identified by machine learning and users. Second, machine learning systems are used to elevate information from authoritative sources. Third, fact check information panels are used to display third-party fact-checked articles above search results for relevant queries to give more context

115. Clay Calvert et al., *Fake News and the First Amendment: Reconciling a Disconnect Between Theory and Doctrine*, 86 U. CIN. L. REV. 99, 99–100 (2018); 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, 58–59, 97–98 (2018); Ari Ezra Waldman, *The Marketplace of Fake News*, 20 U. PA. J. CONST. L. 845, 847–48 (2018).

116. Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

117. Irini Katsirea, *"Fake News": Reconsidering the Value of Untruthful Expression in the Face of Regulatory Uncertainty*, 10 J. MEDIA L., 159, 184 (2018).

118. United States v. Alvarez, 567 U.S. 709 (2012).

119. Amélie Heldt, *Let's Meet Halfway: Sharing New Responsibilities in a Digital Age*, 9 J. INFO. POL. 355, 357 (2019).

120. *Id.* at 356.

121. *See generally* Kate Klonick, *The Facebook Oversight Board: Creating an Independent Institution To Adjudicate Online Free Expression*, *supra* note 24.

for videos across YouTube.[122] Twitter, meanwhile, advocates a community-driven approach to content moderation, allowing users to report misinformation and to provide informative context for tweets in the form of notes.[123]

B. FRANCE'S MODERATE LEGISLATIVE REGULATION

1. Background and Purpose

In contrast to the United States, France has abandoned the market-based approach and adopted a law that actively regulates technology companies' dealings with disinformation. Drawing upon general principles gleaned from global commentary, French commentators have expressed serious concerns about algorithmic disinformation. For instance, they have examined the role of algorithms in creating echo chambers and filter bubbles that accentuate confirmation bias and reinforce prejudices or preconceptions by exposing users to content they are already inclined to agree with and consume.[124] They have also paid attention to the possibility that these algorithms might be manipulated by bots or click workers, such as those based in Thailand and Macedonia, that generate synthetic virality.[125] Concerning algorithmic disinformation's negative impacts on political polarization, they have suggested that time away from social media or greater exposure to opposing views leads to a slight reduction in such polarization.[126]

In response, French regulators have focused on how disinformation spreads and influences political discourse for two main reasons. The first is the potential impact of disinformation on major elections in France. A study into narratives of the "fake news" present in French media found that the term first emerged about a week after the 2016 U.S. presidential election, with articles discussing the role that Facebook had played in the victory of Donald Trump.[127] During the 2017 presidential election in France, clickbait and false information posted by satirical news sites were co-opted and spread by right-wing websites and user profiles.[128] In one case, *Le Gorafi*, a satirical news site likened to *The Onion*, shared an article claiming that presidential candidate Emmanuel Macron had expressed disdain for the disadvantaged and wiped his hands after coming

122. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, *supra* note 24, at 1621.

123. *Id.* at 1621.

124. Patrick Troude-Chastenet, *Fake News et Post-Vérité. De l'Extension de la Propagande au Royaume-Uni, aux États-Unis et en France* [*Fake News and Post-Truth. Of the Extension of Propaganda in the United Kingdom, the United States and France*], 96 QUADERNI 87, 91–92 (2018).

125. *Id.*

126. Rémy Demichelis, *Médias et Sphères de Justice: Réduire la Prédominance des Plateformes Internet de Recommandation* [*Media and Spheres of Justice: Reducing the Dominance of Internet Recommendation Platforms*], 2 LES CAHIERS DU JOURNALISME 67, 74 (2021); *see also* Hunt Allcott et al., *The Welfare Effects of Social Media*, 110 AM. ECON. REV. 629 (2020) (the study referenced by Demichelis).

127. Angeliki Monnier, *Narratives of the Fake News Debate in France*, 5 IAFOR J. ARTS & HUMANITIES 3, 7 (2018).

128. *See* Troude-Chastenet, *supra* note 124, at 88.

into contact with the poor.[129] In advance of the second round of voting, this article was shared more than 600,000 times.[130] Like their counterparts in the United States, French politicians brought disinformation into the mainstream discourse by referencing fringe Internet conspiracies and reports. For instance, Macron's 2017 opponent Marine Le Pen regularly referenced false reports originating from 4chan that Macron was benefitting from offshore bank accounts in the Bahamas.[131] Macron faced further false accusations that he was financially backed by Saudi Arabia, wanted to introduce Sharia Law in the French territory of Mayotte, and so forth.[132]

The second reason that French regulators have focused on how disinformation spreads and influences political discourse is the interference in French politics by Russia and Russian news organizations. Prior to announcing the proposed anti-disinformation legislation, Macron criticized *Russia Today* and *Sputnik*, stating that they were "organs of influence" responsible for spreading "deceitful propaganda,"[133] and he also aired such views during a joint news conference with Russian President Vladimir Putin. The French authorities have even gone so far as to create difficulties for Russian outlets attempting to secure accreditation to cover French government events.[134] It has also been suggested that Russian-backed Twitter accounts were part of the alleged manipulation efforts.[135]

Against this backdrop, the French legislature in 2018 proposed a bill to prevent online disinformation from disturbing the electoral process.[136] The belief that disinformation was already having serious consequences for the functioning of democracy was reflected in the initial bill's stated intention to address the current electoral climate, with reference made to both the 2016 U.S. presidential election and U.K. referendum on EU membership.[137] In its advisory opinion, the Council of State was receptive to the bill's provisions designed to target online platforms, noting such platforms' capacity to amplify false information and create echo chambers. Although

129. *Id.*

130. *Id.*

131. James McAuley, *France Weighs a Law To Rein in 'Fake News,' Raising Fears for Freedom of Speech*, WASH. POST (Jan. 10, 2018, 3:58 PM), https://www.washingtonpost.com/world/europe/france-weighs-a-law-to-rein-in-fake-news-raising-fears-for-freedom-of-speech/2018/01/10/78256962-f558-11e7-9af7-a50bc3300042_story.html [https://perma.cc/NVD9-CJJB] [https://web.archive.org/web/20230213043401/https://www.washingtonpost.com/world/europe/france-weighs-a-law-to-rein-in-fake-news-raising-fears-for-freedom-of-speech/2018/01/10/78256962-f558-11e7-9af7-a50bc3300042_story.html].

132. *See* Troude-Chastenet, *supra* note 124, at 93.

133. Jan van der Made, *Russian Outlets Sparked Macron's Fake News Law Plan, Analysts*, RFI (Jan. 4, 2018, 5:15 PM), https://www.rfi.fr/en/europe/20180104-france-fake-news-law-macron-russia-angry-deny-sputnik-rt [https://perma.cc/F8C4-Q3ZC] [https://web.archive.org/web/20230213043516/https://www.rfi.fr/en/europe/20180104-france-fake-news-law-macron-russia-angry-deny-sputnik-rt].

134. *See* McAuley, *supra* note 131.

135. *See* van der Made, *supra* note 133.

136. Irène Couzigou, *The French Legislation Against Digital Information Manipulation in Electoral Campaigns: A Scope Limited by Freedom of Expression*, 20 ELECTION L.J. 98, 103 (2021).

137. Ciara Nugent, *France Is Voting on a Law Banning Fake News. Here's How It Could Work*, TIME (Jun. 7, 2018, 1:09 PM), https://time.com/5304611/france-fake-news-law-macron [https://perma.cc/D5NK-JQVQ] [https://web.archive.org/web/20230213043646/https://time.com/5304611/france-fake-news-law-macron].

the Council challenged some aspects of the proposals, noting in particular that the sharing of false information is not necessarily unlawful, it was supportive of the requirement for platforms to publicize the resources they devote to combatting the spread of misinformation.[138]

However, some critics in the French Senate argued that platform regulation should be carried out at the European Union rather than national level.[139] Representatives in the National Assembly raised an instrumental criticism of the bill, charging that the platform-focused components of the proposed law needed to target more specifically the role played by algorithms in the dissemination of fake news.[140] Following this criticism, on July 8, 2018 the National Assembly adopted Amendment No. 136, which requires Internet platforms that employ recommendation algorithms to publish statistics on their operation.[141] This led to the final adoption of Law No. 2018-1202 of December 22, 2018 relating to the fight against the manipulation of information (known as the Manipulation of Information Law).

2. Nature and Scope

The Manipulation of Information Law establishes France's new approach to regulating disinformation through amendments and additions to existing legislation and several standalone provisions. Before outlining the duties established for online platforms under Title III of the law, I will first discuss the broader regime within which they exist.

Title IV of the Manipulation of Information Law introduces several changes to the French Code on Education, requiring greater consideration of the Internet and online communication in schools, with a particular focus on fostering a critical attitude toward the reliability of online information.[142] Title II introduces provisions designed to curb the broadcast of misinformation by foreign states, with Article 5 revising Article 33-1

138. *Avis Consultatif: Lutte Contre les Fausses Informations* [*Advisory Opinion: Fight Against False Information*], CONSEIL D'ÉTAT (May 4, 2018), https://www.conseil-etat.fr/avis-consultatifs/derniers-avis-rendus/a-l-assemblee-nationale-et-au-senat/lutte-contre-les-fausses-informations [https://perma.cc/8DUD-KXKW] [https://web.archive.org/web/20230512165621/https://www.conseil-etat.fr/avis-consultatifs/derniers-avis-rendus/a-l-assemblee-nationale-et-au-senat/lutte-contre-les-fausses-informations].

139. Djazia Tiourtite, *What Does the Future Hold for French Anti Fake News Laws?*, MEDIAWRITES (Jan. 3, 2019), https://mediawrites.law/what-does-the-future-hold-for-french-anti-fake-news-laws [https://perma.cc/LL9A-DVAZ] [https://web.archive.org/web/20230403192002/https://mediawrites.law/what-does-the-future-hold-for-french-anti-fake-news-laws].

140. Paula Forteza, *La Proposition de Loi sur la Lutte Contre les Fake News* [*The Proposed Law on the Fight Against Fake News*], FORTEZA (Jul. 13, 2018), https://forteza.fr/2018/07/13/la-proposition-de-loi-sur-la-lutte-contre-les-fake-news [https://perma.cc/99ZL-X7E9] [https://web.archive.org/web/20230213043825/https://forteza.fr/2018/07/13/la-proposition-de-loi-sur-la-lutte-contre-les-fake-news].

141. *Amendement n°136* [*Amendment No. 136*], ASSEMBLÉE NATIONALE, https://www.assemblee-nationale.fr/dyn/15/amendements/0990/AN/136 [https://perma.cc/69U9-49N5] [https://web.archive.org/web/20230213043918/https://www.assemblee-nationale.fr/dyn/15/amendements/0990/AN/136].

142. *See* Rachael Craufurd Smith, *Fake News, French Law and Democratic Legitimacy: Lessons for the United Kingdom?*, 11 J. MEDIA L. 52, 63 (2019).

of Law 86-1067 on Freedom of Communication to empower the French broadcasting authorities to refuse to enter into broadcast agreements that pose a grave risk to certain important interests.[143] Moreover, Article 6 of the Manipulation of Information Law introduces Article 33-1-1 of the Freedom of Communication Law, empowering the authorities to suspend distribution by any electronic means of a television or radio service owned or controlled by a foreign state during electoral periods.[144] Articles 33-1 and 33-1-1 initially specified the Superior Audiovisual Council (CSA) as the authority responsible for overseeing Title II. However, following the CSA's merger with the Higher Authority for the Dissemination of Works and the Protection of Rights on the Internet, it was clarified in a 2021 amendment that responsibility would pass to the new Regulatory Authority for Audiovisual and Digital Communication (ARCOM).[145]

Some of the most substantial provisions, however, are introduced in Title I of the Manipulation of Information Law. First, Article 1 inserts the new Article L163-2 into the French Electoral Code, establishing that during "the three months preceding the first day of the month of general elections and until the date of the ballot," when inaccurate or misleading allegations likely to affect the integrity of an election result are disseminated in a "deliberate, artificial or automated and massive way through an online public communication service," an application can be made for interim measures to stop the dissemination.[146] The judge in chambers is required to rule within forty-eight hours of the referral, and, in the case of an appeal, the court is similarly expected to rule within forty-eight hours.[147]

Second, Article 1 of the Manipulation of Information Law adds Article L163-1 to the Electoral Code. This provision holds that, during the same three-month period, online platform operators with users exceeding a particular threshold are required to (i) provide transparent information regarding the identity of the person or company paying for the promotion of content relating to a debate of general interest; (ii) provide users with transparent information on the use of their data in this context; and (iii) publicize the amount of renumeration received in cases where the amount received is over a particular threshold.[148] Article D102-1 of the Electoral Code clarifies that the threshold for platforms is 5 million unique users each month and that that for renumeration is 100 euros excluding tax. When platforms publicize this information,

143. *Id.* at 58 ("These include the dignity, liberty and property of others; maintenance of a diverse flow of thoughts and opinions in society; protection of children; safeguarding of public order; and protection of the fundamental interests of the nation, including the proper functioning of the nation's institutions. The transmission of false information can threaten a number of these interests, notably the last. Refusal of a convention is also possible where this would lead to a breach of existing law and, as noted above, a number of French laws explicitly tackle the publication of false news.").

144. *Id.* at 59.

145. LOI n° 2021-1382 du 25 octobre 2021 relative à la régulation et à la protection de l'accès aux œuvres culturelles à l'ère numérique [Law 2021-1382 of October 25, 2021 relating to the Regulation and Protection of Access to Cultural Works in the Digital Age], JOURNAL OFFICIEL DE LA RÉPUBLIQUE FRANÇAISE [J.O.] [OFFICIAL GAZETTE OF FRANCE], Oct. 26, 2021.

146. Code électoral [C. électoral] [Electoral Code] art. L163-2 (Fr.).

147. *Id.*

148. Code électoral [C. électoral] [Electoral Code] art. L163-1 (Fr.).

they are required to do so electronically in an open format and to regularly update the information during the three-month election period.[149] As user profiles containing personal data can constitute one of the many inputs considered by recommendation algorithms,[150] the Article L163-1 Electoral Code requirement that platforms publicize how they have employed user data in the promotion of content during election periods relates to algorithms. However, algorithms are addressed more directly and specifically within the duties of cooperation established by Title III of the new law.

The primary Title III duties established in Article 11 of the Manipulation of Information Law are that online platform operators must "implement measures to combat the dissemination of false information likely to disturb public order or alter the sincerity" of ballots and set up an accessible system for users to report such information. The article does not state that these duties are limited to the three months preceding an election in order to encourage platforms to engage more generally with misinformation beyond that period.[151] Article 12 inserts a new provision into the Freedom of Communication Law stating that these duties will be overseen by the CSA, which will do so by issuing recommendations and publishing periodic reviews.[152] As platforms can only be named and shamed, this provision suggests that the duties are not constraining.[153] Although the latest version of the law no longer specifies the form of oversight, it clarifies that ARCOM is the agency responsible for regulating relevant matters.[154] The broad duty to combat the dissemination of false information provides platforms with a wide margin of discretion as to the measures they can take.[155] However, Article 11 of the Manipulation of Information Law specifies several potential actions, including promoting reliable content, fighting accounts that massively propagate false information, informing users about the origin of content, and implementing measures relating to the transparency of their recommendation algorithms.[156]

Article 14 of the Manipulation of Information Law provides more information on the transparency requirement concerning platforms' algorithms. It targets operators falling under Article L163-1 of the Electoral Code (those exceeding 5 million unique users each month, as described above) that use algorithms for the recommendation,

149. *Id.*

150. Engin Bozdag, *Bias in Algorithmic Filtering and Personalization*, 15 ETHICS INFO. TECH. 209, 213 (2013).

151. *See* Smith, *supra* note 142, at 62 (2019).

152. LOI n° 86-1067 du 30 septembre 1986 relative à la liberté de communication (Loi Léotard) [Law 86-1067 of September 30, 1986 on Freedom of Communication (Léotard Law)], art. 17-2 (Version in force from December 24, 2018, to December 24, 2020) (Fr.).

153. Marine Guillaume, *Combatting the Manipulation of Information—a French Case*, 16 HYBRID COE STRATEGIC ANALYSIS 1, 5 (2019).

154. *See* LOI n° 86-1067 du 30 septembre 1986 relative à la liberté de communication (Loi Léotard) [Law 86-1067 of September 30, 1986 on Freedom of Communication (Léotard Law)], art. 17-2 (Version in force from October 27, 2021) (Fr.).

155. *See* Irène Couzigou, *The French Legislation Against Digital Information Manipulation in Electoral Campaigns: A Scope Limited by Freedom of Expression*, 20 ELECTION L.J. 98, 110 (2021).

156. *Id.*

classification, or referencing of information content relating to debates of general interest.[157] The aim of this high threshold is to target only platforms with a large audience, and thus only those capable of influencing opinion through their recommendation algorithms.[158] Article 14 of the Manipulation of Information Law further clarifies that the statistics published should include 1) the share of direct access made to content without recourse to recommendation and 2) the share of indirect access attributable to either the platform's internal search engine algorithm or recommendation algorithm. Finally, Article 14 also states that the statistics should be published online and made accessible to all using a free and open format.

3. Implementation

Implementation of the Manipulation of Information Law has met with a number of difficulties. Famously, the French government saw its own social media voter registration campaign blocked by Twitter, as the platform claimed that it did not know how to effectively comply with the requirement that it publish information concerning content sponsorship in this context and was thus choosing to avoid the problem altogether.[159] The French government denied that the law had backfired, criticizing Twitter's conduct and arguing that the register to vote message should not be considered a political campaign.[160]

Before responsibility was transferred from the CSA to ARCOM, the CSA published two annual reports on the efforts made by platforms to fulfil their obligations. The gist of the 2020 report was that platforms' efforts in 2019 could be greatly improved upon, with the CSA raising criticisms concerning the limits of the information and statistics provided to it, for example on content moderation.[161] With respect to algorithm transparency, the CSA complained that the information provided by platforms differed very little from the information already available on their websites and expressed concern over the lack of clarity on platforms' efforts to make their algorithms intelligible and understandable.[162] For their part, platforms have proved resistant to

157. LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information [LAW No. 2018-1202 of December 22, 2018 Relating to the Fight Against the Manipulation of Information], art. 14 (Fr.).

158. *See* Couzigou, *supra* note 155, at 109.

159. *Twitter Blocks French Government with Its Own Fake News Law*, BBC NEWS (Apr. 3, 2019), https://www.bbc.com/news/world-europe-47800418 [https://perma.cc/7S4A-2D64] [https://web.archive.org/web/20230226193125/https://www.bbc.com/news/world-europe-47800418].

160. *Id.*

161. Alexandre Piquard, *Loi «Fake News»: le CSA Veut Davantage de Transparence des Réseaux Sociaux* [*"Fake News" Law: The CSA Wants More Transparency of Social Networks*], LE MONDE (Jul. 30, 2020), https://www.lemonde.fr/economie/article/2020/07/30/loi-fake-news-le-csa-veut-davantage-de-transparence-des-reseaux-sociaux_6047715_3234.html [https://perma.cc/WHK3-GWL6] [https://web.archive.org/web/20230303165532/https://www.lemonde.fr/economie/article/2020/07/30/loi-fake-news-le-csa-veut-davantage-de-transparence-des-reseaux-sociaux_6047715_3234.html].

162. *Summary—Combatting the Dissemination of False Information on Online Platforms: An Evaluation of the Application and Effectiveness of Measures Implemented by Operators in 2019 (English Version)*, CONSEIL SUPÉRIEUR L'AUDIOVISUAL 1, 7 (2020), https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-

providing further information in this respect, citing concerns over business secrets and the risk of sharing information that could assist their competitors.[163] As commentators have noted, the CSA was not granted sanctioning powers under the Manipulation of Information Law, meaning that its only available recourse was to put pressure on platforms through the publication of its reports.[164]

The 2020 report therefore proceeded to outline several recommendations on how platforms could improve their compliance. First, it asked platforms to provide more information concerning the intelligibility and accountability of their algorithms, inviting them to provide the CSA with the principles of their recommendation approaches and lists of the criteria and data used by their algorithms.[165] Second, the CSA encouraged platforms to provide greater transparency to users concerning the operation of their algorithms, noting that the density and complexity of the information provided thus far was not conducive to understanding.[166] Third, the CSA recommended that platforms provide users with greater transparency concerning how they could adjust algorithm settings.[167] Finally, the CSA noted an analysis of platform practices that appeared to suggest increased reliance on algorithmic curation, inviting platforms to provide more information "on the respective roles and extent of human and algorithmic curation, in both the detection and processing of false information."[168]

In its subsequent 2021 report, the CSA noted that eleven platforms, including Facebook, Twitter, Google (Google Search and YouTube), and Verizon Media (Yahoo Search), had made a declaration concerning their overall compliance efforts.[169] However, it stated that the quality of the information provided was inconsistent, with Verizon Media singled out for its "particularly poor declaration."[170] With respect to algorithm transparency in particular, the CSA noted that there had been a "considerable increase in the amount of information provided by certain operators," but stressed that in some cases a lack of transparency remained concerning the operations and consequences of recommendation algorithms.[171] Moreover, the CSA requested

les-infox-le-CSA-publie-son-premier-bilan [https://perma.cc/9WFM-DNFY] [https://web.archive.org/web/20230303170007/https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-les-infox-le-CSA-publie-son-premier-bilan].

163. *See* Piquard, *supra* note 161.

164. *Id.*

165. *See* CONSEIL SUPÉRIEUR L'AUDIOVISUAL, *supra* note 162, at 8.

166. *Id.*

167. *Id.*

168. *Id.*

169. *Summary—Combatting the Dissemination of False Information on Online Platforms: An Evaluation of the Application and Effectiveness of Measures Implemented by Operators in 2020 (English Version)*, CONSEIL SUPÉRIEUR L'AUDIOVISUAL 1, 3 (2021), https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-la-manipulation-de-l-information-le-CSA-publie-le-bilan-des-mesures-mises-en-oeuvre-par-les-plateformes-en-ligne-en-2020 [https://perma.cc/V42X-YSQ8] [https://web.archive.org/web/20230303170710/https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-la-manipulation-de-l-information-le-CSA-publie-le-bilan-des-mesures-mises-en-oeuvre-par-les-plateformes-en-ligne-en-2020].

170. *Id.*

171. *Id.* at 5–6.

additional information regarding ethical issues implicated by algorithms, such as the efforts taken to counter bias.[172] It concluded by emphasizing the impact of recommendation algorithms on the manipulation of information and stressing that, despite considerable improvements, more information (confidential if necessary) should be supplied in order to counter these effects.[173]

The 2021 annual report identified two major areas for improvement. First, it suggested that platforms should provide users with features enabling them to understand the effects of algorithms "in a personalized and context-driven way if possible."[174] Second, it recommended that platforms provide information concerning their fight against algorithmic bias, including "dedicated resources, tools, subsequent modifications, [and] results."[175] It remains to be seen whether ARCOM will be satisfied with the steps taken by platforms in this respect when it publishes its first annual report on the efforts made in 2021.

C. CHINA'S STRINGENT LEGISLATIVE REGULATION

1. Background and Purpose

For several years now, China has been implementing a crackdown on online platforms.[176] In a 2016 speech, President Xi Jinping outlined the government's intention to further the development of China's information industry, albeit with an increased emphasis on making the Internet "better" for the people of China.[177] Since 2016, China has implemented several measures aimed at achieving this goal. For instance, the Provisions on the Ecological Governance of Network Information Content came into effect in 2020 with the aim of creating a good network ecology, protecting the rights and interests of citizens, and safeguarding national security and public interests.[178] The Chinese government has also penalized online platforms for such conduct as false advertising and monopolistic behavior, accusing the major companies of mistreating their users.[179]

172. *Id.* at 6.

173. *Id.*

174. *Id.* at 8.

175. *Id.*

176. *See generally*, Angela Huyue Zhang, *Agility Over Stability: China's Great Reversal in Regulating the Platform Economy*, 63 HARV. INT. LAW J. 457 (2022).

177. Xi Jinping, *Zai Wangluo Anquan He Xinxi Hua Gongzuo Zuotan Hui Shang De Jianghua* [*Speech at the Symposium on Network Security and Informatization*], GUOJIA HULIANWANG XINXI BANGONGSHI [CYBERSPACE ADMIN. OF CHINA] (Apr. 25, 2016), http://www.cac.gov.cn/2016-04/25/c_1118731366.htm [https://perma.cc/XA6U-2H5H] [https://web.archive.org/web/20230303171425/http://www.cac.gov.cn/2016-04/25/c_1118731366.htm].

178. Wangluo Xinxi Neirong Shengtai Zhili Guiding [Provisions on the Ecological Governance of Network Information Content] (adopted at the executive meeting of the Cyberspace Administration of China Dec. 15, 2019, effective Mar. 1, 2020), art. 1.

179. Brenda Goh & Josh Horwitz, *Factbox: How China's Regulatory Crackdown Has Reshaped Its Tech, Property Sectors*, REUTERS (Apr. 30, 2022), https://www.reuters.com/world/china/education-bitcoin-chinas-season-regulatory-crackdown-2021-07-27 [https://perma.cc/AYC9-58V2] [https://web.archive.org/web/

A recent addition to these efforts is the adoption of the Provisions on the Administration of Algorithm Recommendations for Internet Information Services (Recommendation Algorithm Provisions),[180] a new law designed to regulate the algorithms applied by technology companies. Starting in around 2018, and influenced by the broader international discourse, Chinese commentators began discussing the potential harms of recommendation algorithms.[181] For instance, they have raised concerns about the risk of "information cocoons" filled with vile and vulgar content, the promotion of exaggerated advertisements, and the dissemination of "extremely emotional" articles.[182] Commentators have acknowledged the risks of algorithmic disinformation, claiming that recommendation algorithms are making it difficult to distinguish between true and false information and presenting users with "inferior information" containing insufficient depth and confusing value orientations.[183] They have also argued that recommendation algorithms are imbued with the values of the platforms on which they operate, contending that the "traffic is king" approach needs correction and replacement with more positive and healthy guiding values.[184]

The issues that China faces are not merely theoretical. Like the U.S. and France, it has had to grapple with instances of disinformation spreading on Internet platforms. For instance, the authorities have cracked down on "self-media accounts," which post sensational or fabricated stories with clickbait headlines to attract advertisers.[185] One particularly prolific account, which earned a reported USD 112,000 in ad revenue, was shut down after attracting controversy for sharing a story about a young man from a disadvantaged background who had graduated from a top university and then failed to find employment before dying of cancer, a story that was discovered to be false.[186] In November 2018, the Cyberspace Administration of China claimed to have shut down more than 9,800 similar accounts.[187] Tencent, the owner and operator of WeChat, was forced to launch a fake news debunking initiative after the widespread circulation of

20230226205648/https://www.reuters.com/world/china/education-bitcoin-chinas-season-regulatory-crackdown-2021-07-27].

180. Hualianwang Xinxi Fuwu Suanfa Tuijian Guanli Guiding [Provisions on the Administration of Algorithm Recommendations for Internet Information Services] (adopted at the 20th office meeting of the State Internet Information Office Nov. 16, 2021, effective Mar. 1, 2022) [hereinafter *Recommendation Algorithm Provisions*].

181. Gilad Abiri & Xinyu Huang, *The People's (Republic) Algorithms*, 12 NOTRE DAME J. INT. & COMP. L. 16, 22 (2022).

182. *Wangluo Shidai Yinggai Ruhe Guifan* [*How To Regulate Algorithm in the Era of Internet*], RENMIN WANG [PEOPLE. CN] (July 4, 2018), http://it.people.com.cn/n1/2018/0704/c1009-30124289.html [https://perma.cc/L4A7-SFJF] [https://web.archive.org/web/20230403192420/http://it.people.com.cn/n1/2018/0704/c1009-30124289.html].

183. *Id.*

184. *Id.*

185. Celia Chen, *China's 'Social Media Queen' Mimeng Closes WeChat Account With 13 Million Followers After Outrage Over Fake Story*, S. CHINA MORNING POST (Feb. 22, 2019), https://www.scmp.com/tech/apps-social/article/2187225/chinas-social-media-queen-closes-wechat-account-13-million [https://perma.cc/PT7Q-9RUE] [https://web.archive.org/web/20230226205344/https://www.scmp.com/tech/apps-social/article/2187225/chinas-social-media-queen-closes-wechat-account-13-million].

186. *Id.*

187. *Id.*

disinformation, including a story claiming that onions could kill the flu virus that was shared more than 400,000 times.[188] Most notably, in early 2020, following the initial outbreak of COVID-19, Chinese Internet platforms faced the rapid proliferation of rumors that large numbers of patients in the United States were committing suicide and that a Russian government official had confirmed that the virus was synthetically created.[189]

Commentators in China have also highlighted a more specific concern about the spread of information through recommendation algorithms. Building on the previously stated concern that recommendation algorithms promote news with a "confusing value orientation," the focus has been not only on preventing the spread of disinformation but also on promoting correct messages. For instance, commentators have expressed concern that algorithms overwhelm and undermine the influence and cohesion of the top-down dissemination of mainstream ideological values through official newspapers and media.[190] Although commentators in the United States have expressed similar concerns about the marketplace of ideas potentially being overwhelmed or bastardized by the proliferation of disinformation,[191] their counterparts in China have clearly gone much further because there is a strongly held belief in the United States that the government should remain value-neutral in relation to the marketplace.[192] Emphasizing the promotion of a mainstream value orientation would likely raise objections overseas, as international commentators have accused China of itself engaging in disinformation campaigns. For instance, an Atlantic Council report claims that China's state-run media are involved in spreading disinformation and propaganda.[193]

188. Iris Deng, *Tencent's Fake News Debunkers Reached Nearly 300 Million WeChat Users Last Year*, S. CHINA MORNING POST (Jan. 22, 2019), https://www.scmp.com/tech/apps-social/article/2183124/tencents-fake-news-debunkers-reached-nearly-300-million-wechat [https://perma.cc/CGV2-FB62] [https://web.archive.org/web/20230226205107/https://www.scmp.com/tech/apps-social/article/2183124/tencents-fake-news-debunkers-reached-nearly-300-million-wechat].

189. Grace Ye, *The Chinese Fact Checkers Taking on the Covid Infodemic—a Rumour at a Time*, S. CHINA MORNING POST (June 26, 2022), https://www.scmp.com/news/china/politics/article/3183102/chinese-fact-checkers-taking-covid-indofemic-rumour-time [https://perma.cc/HP7T-AAH8] [https://web.archive.org/web/20230309153257/https://www.scmp.com/news/china/politics/article/3183102/chinese-fact-checkers-taking-covid-indofemic-rumour-time].

190. Dongde Hou & Liping Zhang, *Suanfa Tuijian Yishi Xingtai Fengxian De Falu Fangfan* [*Legal Prevention of the Ideological Risk of Algorithmic Recommendation*] (2021) 321 CHONGQING SHEHUI KEXUE [CHONGQING SOC. SCIENCES] 77, 80 (2021); *see also* Bao Wu, *Suanfa Tuijian Shidai Zhuliu Yishi Xingtai Chunabo Mianlin de Tiaozhan Ji Qi Yingdui* [*Challenges and Countermeasures of Mainstream Ideology Propagation in the Era of Algorithmic Recommendation*], 37 ZHONGGUO SHIYOU DAXUE XUEBAO (SHEHUI KEXUE BAN) [J. OF CHINA UNIV. OF PETROLEUM (SOC. SCIENCE EDITION)] (2021), http://dx.doi.org/10.13216/j.cnki.upcjess.2021.04.0014.

191. *See, e.g.*, Irini Katsirea, *"Fake News": Reconsidering the Value of Untruthful Expression in the Face of Regulatory Uncertainty*, 10 J. MEDIA L. 159, 184 (2018).

192. *See, e.g.*, Ari Ezra Waldman, *The Marketplace of Fake News*, 20 J. CONST. L. 845, 866–869 (2018).

193. DEXTER ROBERTS, ATLANTIC COUNCIL, CHINA'S DISINFORMATION STRATEGY—ITS DIMENSIONS AND FUTURE, (Dec. 2020), https://www.atlanticcouncil.org/wp-content/uploads/2020/12/CHINA-ASI-Report-FINAL-1.pdf [https://perma.cc/S385-P388] [https://web.archive.org/web/20230225235132/https://www.atlanticcouncil.org/wp-content/uploads/2020/12/CHINA-ASI-Report-FINAL-1.pdf].

In response to concerns about recommendation algorithms, China published an opinion-seeking draft of the Recommendation Algorithm Provisions in August 2021.[194] The draft provisions were quickly reviewed and approved at the twentieth meeting of the State Internet Information Office in November 2021, passed by the Cyberspace Administration of China, and then came into force in March 2022. The aim of the Recommendation Algorithm Provisions are to increase the transparency of algorithm functioning, grant Internet users more control over the data employed in recommendation decision-making, and ensure that recommendation systems do not produce negative social and moral consequences.[195] Article 1 of the Provisions clarifies the latter goal as including the promotion of pure socialist values and the safeguarding of national security and social public interests.[196] One commentator has claimed that the Provisions have broken new regulatory ground and that China has pioneered a new approach that empowers people over algorithms.[197]

2. Nature and Scope

Part 1 of the Recommendation Algorithm Provisions offers a brief introduction to the purpose, scope, and definitions of the regulatory regime thereunder. Article 2 subjects the Internet platforms that use algorithms for the generation and synthesis, personalized push, sorting and selection, and filtering of content to the various requirements set out in the Provisions.[198] It also targets platforms using algorithms to schedule decision-making,[199] including assigning such work as the delivery of food.[200] The platforms defined by Article 2 are required to abide by laws and regulations, respect social morality and ethics, and follow principles of fairness and transparency and are encouraged to strengthen industry standards and self-discipline.[201]

Part 2 of the Recommendation Algorithm Provisions establishes specific duties for platforms providing algorithm services, while Part 3 introduces provisions designed for the protection of users. As not all of the harms caused by recommendation algorithms are related to the spread of disinformation, the other articles of the Provisions are not relevant to this article. For instance, Article 15 holds that platforms shall not use

194. *See* Abiri & Huang, *supra* note 181, at 2.

195. Arendse Huld, *China's Sweeping Recommendation Algorithms in Effect From March 1*, CHINA BRIEFING (Jan. 6, 2022), https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022 [https://perma.cc/6KV6-E8C9] [https://web.archive.org/web/20230303183848/https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022]; *see also* Abiri & Huang, *supra* note 181, at 25–26.

196. *See Recommendation Algorithm Provisions*, *supra* note 180, at art. 1.

197. *Id.*; Alberto Romero, *China Has Pioneered a Law To Empower People Over Algorithms*, ONEZERO (Mar. 9, 2022), https://onezero.medium.com/china-has-pioneered-a-law-to-empower-people-over-algorithms-70b29ba6285f [https://perma.cc/8C4E-5RBS] [https://web.archive.org/web/20230309154954/https://onezero.medium.com/china-has-pioneered-a-law-to-empower-people-over-algorithms-70b29ba6285f].

198. *See Recommendation Algorithm Provisions*, *supra* note 180, at art. 2.

199. *Id.*

200. *See* Abiri & Huang, *supra* note 181, at 25.

201. *See Recommendation Algorithm Provisions*, *supra* note 180, at art. 4–5.

algorithms for the purpose of restricting other platforms or hindering competition.[202] Moreover, Articles 18 and 19 introduce specific cautions for platforms that provide services to minors or the elderly. Finally, given that gig-work platforms fall within the scope of the Recommendation Algorithm Provisions, Article 20 introduces a provision protecting the legitimate rights and interests of gig-workers.

Several articles of the Provisions do not specifically address disinformation, but they could still have an impact on its reduction given the major role recommendation algorithms play in the spread of disinformation. For instance, Article 7 establishes the responsibilities of platforms, including the establishment and improvement of algorithm mechanism review and scientific and technological ethics review and the formulation and publication of norms relating to the operation of their recommendation services. Article 8 establishes further responsibilities, including the regular assessment of algorithm mechanisms, data, and outcomes to ensure that they are not contrary to public order or positive customs. By encouraging self-assessment and ethical standards, the Provisions may help platforms to detect and prevent the unintended amplification of fake news.

More notably, Article 9(1) requires the strengthening of content management through the creation and operation of databases used for the identification of illegal or harmful information, also noting that "synthetic information" must be marked as such before dissemination can continue. Article 9(2) clarifies that when illegal information is identified, transmission must be stopped entirely, records must be made, and measures such as deletion should be taken to prevent its spread, while harmful information is to be dealt with as required under the Provisions on the Ecological Governance of Network Information Content.[203] These provisions indicate what constitutes illegal and harmful information.[204] Illegal information includes, but is not limited to, content that jeopardizes national security, subverts state power, undermines national unity, damages the reputation or interests of the state, or spreads rumors that disturb the economic and social order.[205] Examples of harmful information include misleading clickbait, content hyping gossip or scandals, and content promoting indecency.[206] Although other jurisdictions may disagree with these definitions, platforms in China are at least given clear guidance on the types of content they must avoid recommending and develop strategies to manage, as well as on how they should respond when they discover such content being recommended on their platforms.

Article 10 further states that platforms should not use illegal or harmful information as keywords for user interests or user tags for the basis of recommending content. Finally, Article 12 states that platforms should implement strategies to avoid creating harmful consequences for users, such as content de-duplication and optimization of the transparency and comprehensibility of search, ranking, and selection criteria. Although

202. *Id.* at art. 15.
203. *See* Provisions on the Ecological Governance of Network Information Content, *supra* note 178.
204. *Id.* at art. 6–7.
205. *Id.* at art. 6.
206. *Id.* at art. 7.

none of these provisions is specifically designed to combat disinformation, they all provide specific technical guidance that can help platforms to mitigate the risks of recommendation algorithms, including the dissemination of disinformation. Moreover, Article 14 prohibits platforms that provide recommendation services from using algorithms to falsely register users, generate false likes, shares, and comments, or manipulate topic lists or search rankings. Although this article appears to address some of the recommendation manipulation problems associated with bots, because its wording refers specifically to "algorithmic recommendation service providers," it is unclear how it will address the use of bots by entities other than platforms themselves.[207]

The most substantial provisions targeting recommendation algorithms with respect to disinformation include transparency requirements and requirements to offer users genuine choice. Article 28(2), for example, introduces a provision requiring platforms to preserve network records and cooperate with relevant government departments when conducting security assessment, supervision, and inspection work in accordance with the law. Article 16 establishes a transparency obligation, stating that users must be told clearly what algorithmic services are being provided and requiring platforms to publish "the basic principles, purposes, and main operating mechanisms of algorithmic recommendation services in an appropriate manner."[208] Given the black box nature of recommendation algorithms and the probable difficulty of forming clear explanations from complex computer science, compliance with this particular provision may prove challenging. Nevertheless, if Chinese platforms are able to do so, lawmakers elsewhere may be encouraged to introduce similar requirements.

Moreover, the Recommendation Algorithm Provisions follow up the transparency obligation with opportunities for users to exercise choice in response to the information communicated to them. Article 11 provides that platforms have an obligation to establish mechanisms for manual interventions and autonomous user choice. Article 17 goes further by clarifying that users must be provided with the option not to have their personal characteristics targeted or to switch off recommendations entirely, as well as the option to choose or delete user tags related to their personal characteristics. Article 17 has drawn significant interest from commentators worldwide, with some noting the user autonomy it could provide and proposing it as something worth implementing or at least considering elsewhere.[209] However, some have questioned the potentially substantial investment that would be required to develop a working opt-out system capable of translating user profile inputs "into a form understandable to users and then adjust the AI profile in a way that fits the choices made

207. *See* Abrams v. United States, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

208. *See Recommendation Algorithm Provisions*, *supra* note 180, at art. 16.

209. *See, e.g.*, Romero, *supra* note 197; Jonathan Keane, *China and Europe Are Leading the Push To Regulate A.I.—One of Them Could Set the Global Playbook*, CNBC (May 26, 2022), https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html [https://perma.cc/MYP9-5QXB] [https://web.archive.org/web/20230303185317/https://www.cnbc.com/2022/05/26/china-and-europe-are-leading-the-push-to-regulate-ai.html].

by the user."[210] While identifying and presenting users with the tags that make up their recommendation profile may prove challenging, the option to turn down recommendation services entirely is certainly less so. It remains to be seen whether Article 17 will price out smaller platforms but, by mid-March 2022, many of the major Chinese platforms had already implemented an opt-out system.[211] Another concern is that Article 17 might undermine platforms such as Douyin, which have built their business models and entire user experience around a highly personalized recommendation service.[212] However, this concern should be lessened by the fact that users will initially be presented with the ordinary recommendation model, only opting out if they object to it. Moreover, users unhappy with the new service will always have the option to opt back into the old model.

Part 4 of the Recommendation Algorithm Provisions establishes two liability standards for the various principles established in the regime. First, Article 31 establishes that violation of the various provisions, including Articles 7, 8, 9(1), 10, 14, 16, and 17, will result in the issue of warnings, reports containing criticism, and orders to rectify the conduct in question within a specified period of time. Second, under grave circumstances or when a platform refuses to comply with a rectification order, a provisional suspension of service may be ordered and a fine of between RMB 10,000 and RMB 100,000 issued.[213] In contrast to France's algorithm regime, China's appears to be backed by real consequences for non-compliance. The provisions contained within Article 31 focus primarily on the technical standards expected of recommendation systems. The fact that an opportunity to rectify violations is included is important for platforms employing recommendation systems, because recommendation algorithms continuously respond to new data and are therefore "in constant metamorphosis" and "hard to pin down."[214] As platforms cannot be aware of the content their algorithms are recommending at all times, they need some leeway in identifying and correcting harmful recommendation practices.

For those provisions that specifically target disinformation or require platforms to uphold certain mainstream values, including Articles 6, 9(2), 11, and 13, it is established in Article 32 that violations will be dealt with by relevant government departments under relevant rules or regulations. Such broad wording leaves it unclear how each individual provision will be dealt with by regulators, whether fines will be issued, and whether platforms will be given any opportunities to rectify mistakes. However, the wording of the Ecological Governance of Network Information Content Provisions indicates that where "illegal information" is involved platforms may be required to issue warnings and demand rectification, restrict functions, suspend updates, or terminate

210. *See* Abiri & Huang, *supra* note 181, at 29.

211. Claudia Vernotti, *Digital Policy Experts Weigh in on China's New Algorithm Regulation*, TECHNODE (Apr. 5, 2022), https://technode.com/2022/04/05/digital-policy-experts-weigh-in-on-chinas-new-algorithm-regulation [https://perma.cc/U7BN-K847] [https://web.archive.org/web/20230303185649/https://technode.com/2022/04/05/digital-policy-experts-weigh-in-on-chinas-new-algorithm-regulation].

212. *See* Abiri & Huang, *supra* note 181, at 29.

213. *See Recommendation Algorithm Provisions*, *supra* note 180, at art. 31.

214. *See* Abiri & Huang, *supra* note 181, at 30.

accounts in a timely manner.[215] Such wording suggests an approach that focuses on the removal of content and users rather than on affording platforms the leeway to identify and correct harmful recommendation practices.

III. A MULTI-STAKEHOLDER APPROACH

In this part, I first assess the merits and demerits of the three approaches described in Part II. I argue that the United States should scrap the market-based approach to regulating algorithmic disinformation because it has failed completely to accommodate the transparency, intelligibility, and accountability principles. With respect to the French and Chinese legislative approaches, both have advantages and disadvantages in meeting these principles.

I then put forward a multi-stakeholder approach as a legal and policy strategy for effectively regulating algorithmic disinformation. Drawing on the merits of the French and Chinese laws, this approach would dynamically engage social media users, experts, and governmental agencies, such as the U.S. Federal Communications Commission ("FCC"), in curbing the creation and dissemination of algorithmic disinformation.

A. ASSESSING MAJOR PROBLEMS WITH THE THREE APPROACHES

1. Death of the Market-Based Approach?

As demonstrated in Part II, the market-based approach adopted by the United States leaves social media platforms with the power to tackle algorithmic disinformation. Neither Congress nor any federal administrative agency has adopted regulatory rules or measures aimed at countering algorithmic disinformation and constraining social media companies' considerable discretion in the self-regulation of such disinformation.[216] Accordingly, there is no legal requirement for social medial platforms to publicize the recommendation and generative algorithms that are so instrumental in spreading disinformation. Absent such a transparency requirement, platforms need take no action to make the operation of their algorithms intelligible to affected parties or the public.[217]

As a result, the U.S. market-based approach has rendered algorithms black boxes, with none of the transparency, intelligibility, or accountability principles needed to alleviate algorithmic disinformation. Without being subject to these principles, social media companies remain far from adequately responsive in blocking and removing

215. *See* Provisions on the Ecological Governance of Network Information Content, *supra* note 178, at art. 34.

216. Philip M. Napoli & Fabienne Graf, *Social Media Platforms as Public Trustees: An Approach To the Disinformation Problem*, *in* ARTIFICIAL INTELLIGENCE AND THE MEDIA 94 (Taina Pihlajarinne & Anette Alén-Savikko eds., 2022) ("Yet, from a regulation and policy standpoint, the federal government in the USA has done virtually nothing to confront the social media disinformation problem.").

217. *See* Fernando Nuñez, *Disinformation Legislation and Freedom of Expression*, 10 U.C. IRVINE L. REV. 783 (2020).

disinformation.[218] In fact, as we have seen, disinformation is instead being amplified by social media algorithms. According to a Mozilla Foundation report, seventy-one percent of all videos containing disinformation, violent content, hate speech, and/or scams reported as "regrettable" were actively recommended by social media algorithms.[219] Viewers were forty percent more likely to regret watching a recommended video than one they had searched for.[220] It has also been revealed that Facebook's algorithm enabled mass-scale foreign propaganda campaigns during the 2020 U.S. presidential election, and Eastern European troll firms run Facebook's most popular pages for Christian and Black American content, reaching 140 million U.S. users per month.[221]

A key reason for social media companies' failure to curb algorithmic disinformation is that they place their commercial interests above the interests of the public. Facebook is a textbook example. In 2021, whistle-blower Frances Haugen testified before the Senate Committee on Commerce, Science and Transportation on the dangers of Facebook's algorithms.[222] According to Haugen, Facebook knowingly promotes harmful content that amplifies "division, extremism, and polarization."[223] She accused Facebook of consistently putting its profits above users' health and safety, and urged Congress to resolve the Facebook-created crisis.[224] Facebook's own internal research

218. Paul Barrett, *Social Media Can Be an "Arbiter of the Truth" After All*, POLITICO (Apr. 14, 2020), https://www.politico.com/news/agenda/2020/04/14/social-media-coronavirus-184438 [https://perma.cc/9CTF-MJLZ] [https://web.archive.org/web/20230303190405/https://www.politico.com/news/agenda/2020/04/14/social-media-coronavirus-184438]; Ellen P. Goodman & Karen Kornbluh, *Social Media Platforms Need To Flatten the Curve of Dangerous Misinformation*, SLATE (Aug. 21, 2020), https://slate.com/technology/2020/08/facebook-twitter-youtube-misinformation-virality-speed-bump.html [https://perma.cc/9A82-MUQJ] [https://web.archive.org/web/20230303190756/https://slate.com/technology/2020/08/facebook-twitter-youtube-misinformation-virality-speed-bump.html].

219. MOZILLA FOUNDATION, YOUTUBE REGRETS: A CROWDSOURCED INVESTIGATION INTO YOUTUBE'S RECOMMENDATION ALGORITHM 3 (2021), https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf [https://perma.cc/6AHU-S4DR] [https://web.archive.org/web/https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf].

220. *Id.*

221. Karen Hao, *Troll Farms Reached 140 Million Americans a Month on Facebook Before 2020 Election, Internal Report Shows*, MIT TECH. REV. (Sept. 16, 2021), https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election [https://perma.cc/739M-ZYKV] [https://web.archive.org/web/20230303192427/https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-report-us-2020-election].

222. Karen Hao, *The Facebook Whistleblower Says Its Algorithms Are Dangerous. Here's Why.*, MIT TECH. REV. (Oct. 5, 2021), https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms [https://perma.cc/GR8K-ZW5L] [https://web.archive.org/web/20230303192748/https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms].

223. Statement of Frances Haugen: Hearing Before the Subcomm. on Consumer Protection, Product Saftey, and Data Security of the Comm. on Commerce, Science and Transportation, 117th Cong. (Oct. 4, 2021), https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49 [https://perma.cc/5QF7-FU24] [https://web.archive.org/web/20230512043939/https://www.commerce.senate.gov/services/files/FC8A558E-824E-4914-BEDB-3A7B1190BD49].

224. *Id.*

has also revealed that "misinformation, toxicity, and violent content are inordinately prevalent" in material reshared by users and promoted by the company's algorithms.[225]

In response to this dire situation, scholars and policymakers alike are increasingly rejecting the market-based approach and calling for proactive governmental interventions:

> [T]ackling the disinformation problem on social media remains purely within the completely voluntary efforts undertaken by the platforms themselves. Given their disappointing track record thus far, many have questioned whether the platforms are sufficiently incentivized to perform as well as they could; which raises the question of whether some sort of government oversight could provide further incentive.[226]

Legislators have taken action to fill the gigantic loopholes arising from market-based policy, proposing new laws to regulate algorithms in the United States.[227] One set of legislative proposals is intended to deal with the black box nature of recommendation algorithms, with bipartisan efforts in the Senate and the House of Representatives to introduce the Filter Bubble Transparency Act.[228] If passed, the act would require online platforms such as Facebook and Google to notify their users that they use algorithms to determine the order or manner in which information is delivered and provide users with an option to use the platforms without the operation of algorithms,[229] thereby giving users the right to opt out of content selection by personal data-driven algorithms.[230]

Other legislative proposals relate to disclosure. The Algorithmic Justice and Online Platform Transparency Act[231] proposes requiring online platforms to disclose to users the types of algorithmic processes they employ and the categories of personal information they collect to power their algorithms. Similarly, the Algorithmic Fairness

225. Jeff Horwitz, *The Facebook Whistleblower, Frances Haugen, Says She Wants To Fix the Company, Not Harm It*, WALL ST. J. (Oct. 3, 2021), https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-to-fix-the-company-not-harm-it-11633304122 [https://perma.cc/53SW-EC82] [https://web.archive.org/web/20230309161932/https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-to-fix-the-company-not-harm-it-11633304122].

226. Philip M. Napoli & Fabienne Graf, *Social Media Platforms as Public Trustees: An Approach To The Disinformation Problem*, *in* ARTIFICIAL INTELLIGENCE AND THE MEDIA 100 (Taina Pihlajarinne & Anette Alén-Savikko eds., 2022).

227. Tony Samp et al., *US Congress Tries To Decode Algorithms*, DLA PIPER (Jan. 27, 2022), https://www.dlapiper.com/en/insights/publications/ai-outlook/2022/us-congress-tries-to-decode-algorithms [https://perma.cc/YH3T-XKA4] [https://web.archive.org/web/20230309162311/https://www.dlapiper.com/en/insights/publications/ai-outlook/2022/us-congress-tries-to-decode-algorithms].

228. Filter Bubble Transparency Act, S. 2024, 117th Cong. (2021).

229. *Id.*

230. *Id.*

231. Algorithmic Justice and Online Platform Transparency Act, H.R. 3611, 117th Cong. (2021); S. 1896, 117th Cong. (2021).

Act[232] would make it mandatory for platforms to notify users who have been the subject of an algorithmic eligibility determination and provide them with the data employed to make the determination and an opportunity to correct the data.

While these legislative initiatives may have sounded the death knell for the market-based approach, none has yet been passed by Congress despite a sufficient lapse of time. Worse still, these emerging regulations have not incentivized social media companies to alter the way in which they self-regulate algorithmic disinformation.

2. Inadequacies of the French and Chinese Approaches

As the preceding part demonstrates, France and China have acted as forerunners in legislating to regulate algorithmic disinformation. Both countries have introduced legal rules that obligate social media companies to meet the algorithmic transparency requirement. Although France's Manipulation of Information Law does not have explicit rules holding social media companies accountable for failing to meet the transparency requirement, the French authority ARCOM can still make recommendations on how such companies can improve their practices to meet it.[233] China's Recommendation Algorithm Provisions, in contrast, contain a range of penalties that render social media companies accountable for their failure to meet various requirements under the provisions.[234]

Both laws are pathbreaking but have major problems. First, despite explicitly championing transparency, the French Manipulation of Information Law has not adopted measures to promote the intelligibility of the algorithms that create and/or spread disinformation. Information on the operation of algorithms offers no meaningful value to users if they cannot do anything with the information.[235] Platform users tend to be time- and resource-poor and to lack the expertise necessary to interpret and act upon information provided through a platform's duty of transparency.[236] Article 11 of the French Manipulation of Information Law simply creates a duty for platforms to combat disinformation dissemination through measures that may include algorithmic transparency. It is essential to explore whether the French regime provides sufficient guidance on what should be done with the transparency required. Otherwise, this approach to combatting disinformation ultimately seems to do little more than place a burden on users to "to seek out information about a system, interpret it, and determine its significance, only then to find out they have little power to change things anyway."[237]

Other provisions of the Manipulation of Information Law itself require social media companies to make their recommendation and generative algorithms intelligible to

232. Algorithmic Fairness Act of 2020, S. 5052, 116th Cong. (2020).

233. *See supra* Part II.B.

234. *See supra* Part II.C.

235. Lillian Edwards & Michael Veale, *Slave To the Algorithm? Why a 'Right to Explanation' Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 67 (2017).

236. *Id.*

237. *Id.*

users. As outlined in the preceding part, the original wording of the CSA's oversight responsibilities included the issuing of reports and recommendations in response to the information that platforms provide.[238] In its first annual report, for instance, the CSA encouraged platforms to provide users with greater transparency on how to personalize or adjust algorithm settings.[239] Commentators have argued that choice architecture of this nature, that is, architecture allowing users to opt in or out of algorithm inputs, or opt out of recommendations altogether, can counter some of the negative effects of algorithmic amplification.[240] However, as the CSA's recommendations are only advisory, the regime itself clearly does little to ensure algorithmic transparency.

Second, both the French and Chinese laws have given rise to concerns about freedom of expression. The Manipulation of Information Law has faced substantial criticism concerning its provisions addressing elections and the broadcast of foreign media, especially with respect to their impacts on free speech. However, owing to the law's prioritization of transparency or intelligibility rules and minimal focus on the potential harms of algorithms, less commentary has been directed toward the adequacy of its provisions for addressing the latter.

The French law's impact on free speech is likely to be limited. However, the Chinese Recommendation Algorithm Provisions may seriously curb free speech activities on social media platforms by requiring platforms to control the diversity of opinions. Article 6, for example, states that platforms must uphold mainstream value orientations, vigorously disseminate positive energy, use algorithms in the direction of good, and not use recommendation services to engage in activities that harm national security or the public interest. Article 11 requires platforms to vigorously present information that conforms to mainstream values in key areas of their services, such as on their home pages, in hot search terms, and in selected topics. Finally, Article 13 states that platforms providing news information services must be licensed in accordance with the law, must standardize their news collation services, must not generate or synthesize fake news information, and must not "disseminate news information not published by work units in the State-determined scope." This provision is not entirely new, as, under Article 11(2) of the 2019 Provisions on the Administration of Online Audio and Video Information Services,[241] platforms are already prohibited from using AI to create, publish, or disseminate disinformation in audio or video form.

B. CREATING A MULTI-STAKEHOLDER APPROACH

How can we create a regulatory regime that effectively curbs algorithmic disinformation? As I demonstrate in Section A, scholars, policymakers, and legislators

238. *See, e.g.*, Léotard Law, *supra* note 154.

239. *See supra* Part II.B.

240. Jennifer Cobbe & Jatinder Singh, *Regulating Recommending: Motivations, Considerations, and Principles*, 10 EUROPEAN J.L. & TECH., Dec. 30, 2019, at 18.

241. Wangluo Yin Shipin Xinxi Fuwu Guanli Guiding [Provisions on the Administration of Online Audio and Video Information Services] (Promulgated by the State Internet Information Office, Ministry of Culture and Tourism, and State Administration of Radio and Television, Nov. 18, 2019).

have repudiated the market-based approach on the grounds that self-regulation by social media platforms is problematic and damaging to society.[242] Legislative intervention and administrative oversight, which both France and China have initiated, are needed to effectively regulate algorithmic disinformation. However, the approaches adopted by France and China have major limitations, as noted.

Against this backdrop, I propose that the United States should take the lead by adopting a multi-stakeholder approach to legislating and administering the legal regulation of algorithmic disinformation in the public interest. This approach, as I will demonstrate, would be better equipped than current approaches to promote the principles of transparency, intelligibility, and accountability through the more dynamic engagement of social media users and experts and more robust oversight by the relevant administrative agencies.

In a nutshell, the multi-stakeholder approach would engage panels of social media users and experts in reviewing the transparency and intelligibility of the social media company algorithms involved in the creation and dissemination of disinformation every two years. After each review exercise, each panel would make recommendations on how the social media company concerned should improve or rectify its algorithms. Governmental agencies such as the FCC would operate the panel review system and impose penalties on social media companies that failed to meet the review requirements or follow approved recommendations.

1. The Algorithmic Disinformation Review System (ADRS)

a. Formation of the ADRS

The FCC should take the initiative in establishing an algorithmic disinformation review system ("ADRS") comprising panels of social media users and experts. The ADRS is to be tasked with conducting a biennial review of the transparency and intelligibility of a social media company's disinformation-related algorithms. Each review panel is to consist of seven social media users and three experts who will write recommendations on improvements based on the review results.

To this end, the FCC may issue a call for participation, inviting members of the public to volunteer to serve on the review panels. Any adult U.S. citizen who is an active social media user would be eligible to apply. Applicants would be asked to supply information on their social media experience, including the duration of their use of specific accounts, and personal particulars such as age, gender, and ethnicity. Provided that applicants can prove their social media knowledge in the application and subsequent interview, the FCC should make selection decisions to ensure that the panel membership is as diverse as possible in terms of gender, race, sexual orientation, geographic location, education, age, disability, wealth, and political views.[243] The seven

242. *See supra* Part III.A.

243. Meta applies a similar diversity standard to select experts to serve on its Oversight Board. *See* Oversight Board, *Ensuring Respect for Free Expression, Through Independent Judgement*, https://www.

applicants selected for each review panel would be asked to serve on the condition that they agree to participate in reviews once every two years at a place and time designated by the FCC and serve for four consecutive years without remuneration.

Each review panel is also to be staffed by three experts: a journalist, a computer scientist, and a legal professional. Applicants would be required to demonstrate their journalistic, computer science, or legal expertise that is most relevant to the regulation of algorithmic disinformation. Again, the FCC should choose applicants with the aim of promoting diversity, with the selected experts asked to sign an agreement with terms similar to those for the social media users. After participating in the panel review process along with the chosen social media users, the experts would write a report containing recommendations on how the social media company under review should improve the transparency and intelligibility of its disinformation-related algorithms.

I envision each panel being asked to review the operation of the disinformation-related algorithms of several social media companies to allow a comparison of the practices and standards adopted to deal with disinformation and further understanding of how algorithms are developed and applied to deal with disinformation. I suggest that the FCC should consider subjecting social media companies with more than 30 million users in the United States (approximately ten percent of the U.S. population) to this biennial review exercise. This benchmark would cover the most used social networking platforms in the United States, namely, Facebook, Instagram, TikTok, Twitter, Pinterest, LinkedIn, and Reddit,[244] as well as smaller social networking platforms such as Nextdoor[245] and video-sharing and messaging planforms such as YouTube and WhatsApp.[246]

b. The Scope of the Review Process

For each review, the social media companies subject to review would have to prepare materials about the algorithms they have developed and apply in relation to disinformation, submit them to the FCC ahead of the review period, and supply officials

oversightboard.com [https://perma.cc/9535-MKNW] [https://web.archive.org/web/20230303203454/https://www.oversightboard.com] ("When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds.").

244. *See US Social Media Statistics 2022*, THE GLOBAL STATS., https://www.theglobalstatistics.com/united-states-social-media-statistics [https://perma.cc/JW7J-MHJ6] [https://web.archive.org/web/20230303203917/https://www.theglobalstatistics.com/united-states-social-media-statistics]. ("Facebook is the most used social media platform in the United States, with 74.20% of the internet users enrolled. In numbers, it is 227.94 active users. It saw a statistically significant growth since 2019. This tells that the reach is growing rapidly. US is the biggest market for Facebook, both in terms of views and subscribers. The second most liked Social Media Platform by the Americans is Instagram with 186.47 million users[.]").

245. Nextdoor has about 38 million U.S. users in 2022. *See id.*

246. *See id.* ("FB Messenger is the most popular Messenger App in the US with 187.70 million active users. Facebook-owned FB Messenger has 61.10% of the country's total internet users. The second most popular on the list is iMessage, an instant messaging service developed by Apple Inc., with 40.20% penetration. It has 123.49 million active users. The third is Snapchat (118.89 million), which is really popular among teenagers, has 38.70% users. Forth in the list of 2022 social media chat apps is WhatsApp with 28.60% penetration.").

to participate in the review process. The review panel would then host cross-examining sessions, giving the officials the opportunity to clarify the three following aspects of their companies' algorithms.

i. Transparency

The review panel may inquire into whether and how the social media company in question has publicized its disinformation-related algorithms. In particular, the panel may ask the company to provide a textual description of each algorithm employed, covering its objective or objectives; its main operating principles, the data used, and weighting given to each criterion; its effects; whether it is customizable by users; and the number of changes made to it during the year.[247]

Based on this written description and company officials' verbal clarification, the panel would then determine whether the company's disclosure meets the transparency requirement. For example, after reviewing social media companies' disclosure of information on their algorithms in 2020, the French authority criticized them for making statements that were too broad to be understood without complete contextual knowledge of platform operations.[248] More specifically, it pointed out that Microsoft and Facebook had focused their reports merely on "algorithms used to combat disinformation in general or that related to the COVID-19 health crisis" rather than supply comprehensive information that would allow meaningful comparison of platform efforts.[249]

ii. Intelligibility

The panel may also review whether the social media company's disclosure of algorithmic information is suitably intelligible. The intelligibility requirement, according to the French authority, refers to information concerned with the various principles underlying algorithm operations.[250] China's Recommendation Algorithm Provisions explicitly require platforms to publish "the basic principles, purposes, and main operating mechanisms of algorithmic recommendation services in an appropriate manner."[251]

247. COMBATTING THE DISSEMINATION OF FALSE INFORMATION ON ONLINE PLATFORMS: AN EVALUATION OF THE APPLICATION AND EFFECTIVENESS OF MEASURES IMPLEMENTED BY OPERATORS IN 2019, CONSEIL SUPÉRIEUR L'AUDIOVISUAL 80 (2020), https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-les-infox-le-CSA-publie-son-premier-bilan [https://perma.cc/PP5G-2WDD] [https://web.archive.org/web/20230309165004/https://www.csa.fr/Informer/Toutes-les-actualites/Actualites/Lutte-contre-les-infox-le-CSA-publie-son-premier-bilan].

248. *Id.* at 37.

249. *Id.*

250. *See* 2019 CSA Report, *supra* note 247, at n.46 (including "elements relating to the purpose of the processing for which the algorithms have been programmed . . . rules defining this processing, the main characteristics of their implementation, the data processed and their sources, the processing parameters and their weighting").

251. *Recommendation Algorithm Provisions*, *supra* note 180, at art. 16.

Such principles of operation include information on the data used by recommendation algorithms and where this information is made available. In particular, the principles should offer users modification or personalization settings, such as an option to opt out of recommendation algorithms' tracking of their preferences and data. Platforms should also provide users with a tool for real-time interaction with platform operators to obtain more personalized and precise information.[252] The French authority has shed light on how a review panel might assess whether a social media company's articulation of these principles has resulted in the intelligibility of its algorithms. After reviewing the information submitted to it, the French authority in its 2019 report identified LinkedIn, the Wikimedia Foundation, and Verizon Media as operators that provided no declaration or indication that algorithm personalization settings were available.[253] It also made a statement of intent to look more closely at the "usage and user feedback data that will be provided by these operators" to further assess the impact of user organization choices "on improving knowledge and understanding by users of the algorithmic recommendation of content."[254]

iii. Accountability

After inquiring into the transparency and intelligibility requirements, the review panel may further consider whether the social media company in question has taken measures to make its algorithms capable of detecting and removing disinformation. To prove such capability, the company may submit data relating to the performance of its algorithms and the virality of the content detected, as well as provide information on how the algorithms work in suppressing such content.[255] For example, in its 2020 report, the French authority identified Google, Snapchat, and Twitter as providers of sparse data in relation to their moderation algorithms, and Microsoft, LinkedIn, and Verizon Media as providers of no information, although it offered no further detail.[256]

Moreover, social media companies should also be required to submit information showing how they have worked together with fact-checking institutions to filter out disinformation.[257] Companies that offer applications with generative algorithms should be further required to demonstrate that their applications are capable of detecting and deleting disinformation.

252. *Id.*

253. *Id.* at 36.

254. *Id.*

255. *Id.* at 34.

256. *Id.*

257. *See* YALE INFORMATION SOCIETY PROJECT, FIGHTING FAKE NEWS 10, https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf [https://perma.cc/YFF2-VU6F] [https://web.archive.org/web/20230309174202/https://law.yale.edu/sites/default/files/area/center/isp/documents/fighting_fake_news_-_workshop_report.pdf] ("To the extent statements are labeled as false, it is preferable for content distributors to present fact checks as the product of the organization (like an unattributed editorial), rather than relying on individual journalists, speaking in their personal capacities, to do so.").

c. Oversight by the FCC

In addition to setting up review panels, what other roles should the FCC play in the ADRS? The FCC is a governmental agency that regulates communications by radio, television, wire, satellite, and cable across the United States.[258] In particular, as the primary authority in U.S. communications law, the FCC has the power to "[r]evis[e] media regulations so that new technologies flourish alongside diversity and localism" and "[d]evelop[] and implement[] regulatory programs."[259] It also maintains jurisdiction over broadband access, fair competition, radio frequency use, *media responsibility*, public safety, and homeland security. [260]

The FCC thus has the *legal* power to make rules and regulations dealing with algorithmic disinformation, as well as the *policy-making* power to ensure that social media companies take sufficient responsibility for effectively tackling such disinformation. Relying upon such broad-based powers, the FCC can take the following actions to implement the ADRS.

First, the FCC can make new rules to govern the operation of the ADRS. The FCC adopts new communications rules through the so-called "notice and comment" process. Under that process, it gives the public notice that it is considering adopting or modifying rules on a particular subject and seeks comments from the public. It then examines the comments received in developing final rules.[261] The FCC can avail itself of this process to adopt a new set of rules for the ADRS. The rules may elucidate such matters as the form of the review panels, the subjects of the review process, the issuance of review reports, social media companies' obligations to comply with the reports, and the penalties to be imposed on them if they fail to comply.

Second, the FCC should take charge of the logistics for implementing the ADRS. Well before each review exercise, the FCC should organize meetings to provide panel members with up-to-date information on algorithms and disinformation and offer them ample opportunities to interact and share their thoughts and expertise. The FCC should also arrange the schedule and location for each review, and subsequently liaise with the panel experts on the issuance of their report.

Third, the FCC should take measures to improve the ADRS and promote other potential regulatory schemes. It may consider organizing a major conference every five years to examine the system's operation, inviting stakeholders such as policymakers, technology professionals, legal experts, journalists, and media activists to participate and put forward proposals on how to make the ADRS a better mechanism for fighting

258. *The FCC's Mission*, FCC, https://www.fcc.gov/about/overview [https://perma.cc/947U-ZHA4] [https://web.archive.org/web/20230218181041/https://www.fcc.gov/about/overview].

259. *What We Do*, FCC, https://www.fcc.gov/about-fcc/what-we-do [https://perma.cc/9B84-BQDB] [https://web.archive.org/web/20230218181132/https://www.fcc.gov/about-fcc/what-we-do].

260. FEDERAL COMMUNICATIONS COMMISSION, 2008 PERFORMANCE AND ACCOUNTABILITY REPORT (SEPTEMBER 2008).

261. *Rulemaking Process*, FCC, https://www.fcc.gov/about-fcc/rulemaking-process [https://perma.cc/Y6NR-ZLTF] [https://web.archive.org/web/20230218181309/https://www.fcc.gov/about-fcc/rulemaking-process].

algorithmic disinformation. As the FCC will gain disinformation-regulation experience and insights from handling panel reports, conferences, and other activities, it should be proactive in making recommendations to Congress on how to legislate anti-disinformation laws.

Fourth, the FCC should oversee social media companies' compliance with the expert reports and impose penalties for any compliance failures. After the FCC receives the expert reports, it should first review their fairness. If the recommendations presented in the reports are deemed fair, the FCC should then require the social media companies in question to take concrete actions to follow the recommendations and update it in a timely manner on the results of those actions. The FCC would then review whether the company had met its compliance requirements and, if not, order it to make expeditious corrections and impose penalties according to its anti-disinformation rules.

Finally, the FCC should collaborate with relevant governmental agencies to augment the effectiveness of the ADRS. Because disinformation can harm trade and electoral operations, the FCC should make every effort to fight algorithmic disinformation in close partnership with other regulatory agencies such as the Federal Trade Commission (FTC) and Federal Election Commission (FEC). For example, the FCC may invite officials from the FTC and FEC to review the fairness of the expert reports and the results of social media companies' compliance actions with respect to trade- and election-related disinformation. Such a dynamic partnership would create a more effective institutional framework to combat disinformation through concerted regulatory interventions.

With respect to financial support, the FCC's expenditures on implementing the ADRS should be modest, primarily because panel members would serve on a voluntary, unpaid basis, an engagement model that would also ensure impartiality and genuine interest on panel members' part. Expenditures on hosting panel reviews and reform conferences and enforcing expert reports would also likely be modest. Therefore, the proposed ADRS is a cost-effective system.

Figure 1 illustrates the holistic operation of the ADRS.

Figure 1. Algorithmic Disinformation Review System

2. Advantages of the ADRS

To improve its content policy and enforcement, Meta approved the establishment of an Oversight Board in November 2018, consisting of 20 founding members who are experts in free speech, Internet governance, and human rights.[262] The Oversight Board is by nature a dispute resolution system that allows Meta users to appeal content moderation decisions made by Meta officials. After hearing an appeal, the Board determines whether the decision should be upheld or overturned.[263] However, the Board does not review the validity of Meta's content moderation rules, and nor does it make recommendations on how such rules could be improved.[264]

Although the ADRS would actively engage experts, it would be fundamentally different from the Meta Oversight Board. First and foremost, it would be a regulatory regime created by a governmental agency, not an internal dispute resolution system established by a private company. The ADRS would require social media companies to make their disinformation-related algorithms transparent and intelligible to the public, thereby paving the way toward more effective curbs on algorithmic disinformation. Its aim would be to alleviate the disinformation problems caused by the development and application of algorithms as black boxes through the acquisition and release of information on how such algorithms create and spread disinformation.

Second, the ADRS would ensure that social media companies have broader, more forward-looking accountability for their involvement in creating and/or spreading disinformation. On a case-by-case basis, the Meta Oversight Board holds Facebook and Instagram responsible only for removing or recovering content subject to their moderation systems. The ADRS would not deal with the removal or recovery of a particular piece of news. Rather, drawing on review panels' engagement and expertise, it would delve into the systemic problems caused by a social media company's algorithms and order their rectification if necessary. Failure to comply would result in an FCC-imposed penalty. Such penalties are intended to motivate social media companies to take forward-looking measures to curb disinformation without requiring them to look backwards to remove disinformation that has already been disseminated.

If designed and implemented in the manner proposed, the ADRS can promote the platform economy in the public interest. The rise of the platform economy has hugely benefited social media companies with skyrocketing revenues accrued from online advertising services and their public listings' financial contributions to corporate development.[265] As a result, companies such as Google and Meta are among the world's

262. Elizabeth Culliford, *Facebook Names First Members of Oversight Board That Can Overrule Zuckerberg*, REUTERS (May 6, 2020, 1:12 PM), https://www.reuters.com/article/us-facebook-oversight-idUSKBN22I2LQ [https://perma.cc/NSY5-6PPK] [https://web.archive.org/web/20230218181401/https://www.reuters.com/article/us-facebook-oversight-idUSKBN22I2LQ].

263. Kate Klonick, *The New Governors: The New Governors: The People, Rules, and Processes Governing Online Speech*, *supra* note 24, at 1603.

264. *Id.*

265. *See* JAMIE BARTLETT, THE PEOPLE VS TECH: HOW THE INTERNET IS KILLING DEMOCRACY (AND HOW WE SAVE IT) 1 (2018) ("In the coming few years either tech will destroy democracy and the social order

most financially powerful.[266] Complying with the ADRS would require of these companies only modest expenditure, with a correspondingly modest decrease in profits. Hence, the ADRS would not have significantly negative effects on the platform economy that is so crucial to the United States and many other countries. Rather, it would encourage the platform economy to develop in the public interest through the generation and dissemination of far less disinformation.[267]

The ADRS can also raise public awareness of the problems caused by the rapid spread of disinformation and pervasive use of algorithms on social media networks and empower social media users to take part in addressing those problems.[268] As suggested above, if the ADRS is implemented as envisioned the FCC will call for applications from social media users and select suitable applicants to serve on review panels. This arrangement is likely to arouse users' interest in learning more about the impact of disinformation and encourage them to seek out information on algorithmic transparency and accountability. The application and selection processes are also very likely to attract the media spotlight, thereby inducing public discourse on disinformation and algorithms. The more such discourse there is, the more the public will be engaged in brainstorming ways to curb algorithmic disinformation.[269]

However, the ADRS may give rise to concerns over its potential to adversely affect freedom of expression.[270] It should be noted that it would target only disinformation that was willfully created and disseminated to cause harm to a person, institution, or

as we know it, or politics will stamp its authority over the digital world. It is becoming increasingly clear that technology is currently winning this battle, crushing a diminished and enfeebled opponent.").

266. Stephen Johnston, *Largest Companies 2008 vs. 2018, A Lot Has Changed*, MILFORD (Jan. 31, 2018), https://milfordasset.com/insights/largest-companies-2008-vs-2018-lot-changed [https://perma.cc/D7JM-PAP8]
[https://web.archive.org/web/20230218181604/https://milfordasset.com/insights/largest-companies-2008-vs-2018-lot-changed] ("Technology companies not only dominate our daily lives (how many times have you checked your iPhone today?) but also the ranking of world's biggest companies.").

267. *See* Niva Elkin-Koren et al., *Social Media as Contractual Networks: A Bottom Up Check on Content Moderation*, 107 IOWA L. REV. 987, 991 (2022) ("Users of social media platforms are important stakeholders in the platform economy.").

268. *See id.* at 994 (arguing that it is very important to "enable users to restrain platforms' discretion and safeguard their private interests"); NICHOLAS NICOLI & PETROS IOSIFIDIS, DIGITAL DEMOCRACY, SOCIAL MEDIA AND DISINFORMATION 73–76 (2021) (discussing the importance of empowering social media users).

269. *See* Nabiha Syed, *Real Talk About Fake News: Towards a Better Theory for Platform Governance*, 127 Yale L.J. F. 337 (2017) ("For example, once we understand amplification as a relevant concept, we should account for the psychology of how people actually come to believe what is true—not only through rational deliberation, but also by using familiarity and in-group dynamics as a proxy for truth."); Gilad Abiri & Johannes Buchheim, *Beyond True and False: Fake News and the Digital Epistemic Divide*, 29 MICH. TELECOMMS. & TECH L. REV. 59, 62–68 (2022).

270. *See, e.g.*, Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1357 (2018) (cautioning that technology "corporations' power over public discourse today is so concentrated and far-reaching that it resembles and arguably surpasses state power within its sphere"); Alan Z. Rozenshtein, *Surveillance Intermediaries*, 70 STAN. L. REV. 99, 105 (2018) (concluding that "we've created a new generation of *surveillance intermediaries*: large, powerful companies that stand between the government and our data and, in the process, help constrain government surveillance").

society at large.[271] Its regulation of disinformation would therefore be in line with the Supreme Court's First Amendment jurisprudence.[272] For example, the Court has proclaimed that "[f]alse statements of fact are particularly valueless [because] they interfere with the truth-seeking function of the marketplace of ideas."[273] Therefore, false statements "are not protected by the First Amendment in the same manner as truthful statements"[274] given that they involve a "legally cognizable harm associated with a false statement."[275]

IV. CONCLUSION

In the social media age, technology companies' advertising policies and sophisticated algorithms have made the dissemination of disinformation quick, easy, and profitable. As this article shows, by distorting reality, algorithmic disinformation causes and inflames social, economic, and political problems.

Unfortunately, the United States and many other countries have yet to adopt laws that directly regulate algorithmic disinformation, although France and China recently adopted laws containing legal rules requiring social media companies to publicize their disinformation-related algorithms and make them intelligible to users. However, as this Article has also revealed, these laws are still in their infancy. Administrative agencies are required to better enforce them, and users need to be robustly engaged in the policing of disinformation.

Absent effective legal regulation, the dangers of algorithmic disinformation are likely to continue to escalate, with untold harm to society, widespread mental distress, and even the loss of lives.[276] My multi-stakeholder approach constitutes a response to the urgent need to design a regulatory mechanism to combat algorithmic disinformation.[277] It advocates for the adoption of transparency, intelligence, and accountability as the three major principles of the new regulatory mechanism. To further implement these principles, it also calls for the establishment of an algorithmic

271. Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J.L. & TECH. 388, 396 (2020) ("We need, in short, to find ways to discourage the spread of statements that are at once false and damaging.").

272. *See* Michael P. Goodyear, *Priam's Folly: United States v. Alvarez and the Fake News Trojan Horse*, 73 STAN. L. REV. ONLINE 194 (2021) ("As shown by the utter failure of the *Alvarez* decision to plan for or restrain the fake news disasters of the 2016 and 2020 elections and COVID-19, robust protections for the vast majority of false statements are not in the best interests of the United States.").

273. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 52 (1988).

274. Brown v. Hartlage, 456 U.S. 45, 60 (1982).

275. United States v. Alvarez, 567 U.S. 709, 718 (2012).

276. Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 HARV. J. L. & TECH. 388, 395–96 (2020) ("Some falsehoods can hurt or even ruin individual lives. For all these reasons, it is sensible to hope that social norms and even laws will chill them.").

277. Melissa De Witte et al., *'Regulation Has To Be Part of The Answer' To Combating Online Disinformation, Barack Obama Said at Stanford Event* (Apr. 21, 2022), https://news.stanford.edu/2022/04/21/disinformation-weakening-democracy-barack-obama-said [https://perma.cc/36CV-485U] [https://web.archive.org/web/https://news.stanford.edu/2022/04/21/disinformation-weakening-democracy-barack-obama-said].

disinformation review system to effectively empower administrative oversight and dynamically engage social media users.