

## Notes

# A Parallel Infodemic: Multifaceted Approaches to Online Public Health Mis- and Disinformation During the COVID-19 Pandemic

ERIN HUTCHINS<sup>†</sup>

*During the COVID-19 pandemic, communities congregated in online spaces more than ever before. While some people found solidarity online, many others found snippets of false information regarding COVID-19's origin, transmission, and preventative measures. Inaccurate public health information originated long before the COVID-19 pandemic, but it thrived as the uncertainty around daily living dragged on. The pandemic prompted a conversation about who, if anyone, is responsible for deciphering and regulating the spread of false and misleading information. This Note presents two methods in which inaccurate information can be redressed. First, Section 230 of the Communications Decency Act—which provides sweeping immunity for social media platforms—could be carved out to impose liability if inaccurate public health information is available on public forums. In addition to, or in lieu of, this approach, online platforms could be held liable if they do not develop a sufficient content-moderation system. Second, social media platforms could use Corporate Social Responsibility as a tool to amend their existing policies to remove false information, flag misleading information as potentially untruthful or confusing, and flag truthful information as accurate and verified by third-party fact-checkers. Unflagged articles and posts would alert the user that the information has yet to be rigorously reviewed. No solution should work alone; rather, all can work together to coerce social media platforms to be receptive and answerable to their users and broader society.*

---

<sup>†</sup> J.D. Candidate 2022, University of California, Hastings College of the Law; Senior Production Editor, *Hastings Law Journal*. Thank you to my husband, family, and friends for their endless support, as well as Professor Chimène Keitner for her guidance and expertise throughout this process. Finally, thank you to all the editors of *Hastings Law Journal* for their hard work and invaluable feedback on this Note.

## TABLE OF CONTENTS

INTRODUCTION .....	1541
I. WHERE DOES PUBLIC HEALTH MIS- AND DISINFORMATION ORIGINATE?.....	1542
II. FROM AUTISM SPECTRUM DISORDER TO COVID-19: THE PROLIFERATION OF INACCURATE INFORMATION.....	1544
III. THE NEW MIS- AND DISINFORMATION: THE COVID-19 “INFODEMIC” .....	1546
IV. CARVEOUTS AND WORKAROUNDS: SECTION 230.....	1548
V. CORPORATE SOCIAL RESPONSIBILITY: CLEARING UP THE VAGUENESS ON SOCIAL MEDIA PLATFORMS’ POLICIES .....	1553
VI. SO WHERE DO WE GO FROM HERE? .....	1556
VII. RECOMMENDATIONS.....	1559
CONCLUSION.....	1561

## INTRODUCTION

In late April 2021, *The New York Times* published an article titled “No, Other People’s Covid Vaccines Can’t Disrupt Your Menstrual Cycle,” debunking a claim shared widely on social media that ignores fundamental biology.<sup>1</sup> This claim touts that vaccine particles can leave the body of a vaccinated person and be transferred to a nearby unvaccinated person—much like a nonsmoker could inhale secondhand smoke.<sup>2</sup> If a menstruating or pregnant person were nearby, these vaccine particles would threaten to disrupt a menstrual cycle or cause a miscarriage, as the theory goes.<sup>3</sup> A private school in Miami, Florida barred vaccinated teachers from interacting with students for fear that these teachers may inadvertently tamper with students’ reproductive systems.<sup>4</sup> As more vaccines are being offered globally, false or misleading information concerning the vaccine’s effects on the human body threaten to prolong the COVID-19 pandemic.<sup>5</sup> Now, more than ever, the proliferation of inaccurate information on social media deserves legislators’ and the public’s attention.

Inaccurate information about public health is not a product of the COVID-19 pandemic—rather, public health falsehoods have existed for decades. While mis- and disinformation are distinct concepts, they both center around spreading and consuming inaccurate information on online platforms.<sup>6</sup> The ease of liking and sharing news articles and other snippets of commentary allows for the spread of inaccurate information in fractions of a second. While some falsehoods may have minimally harmful impact on internet users, others can threaten public health from their ability to affect enough consumers’ opinions to materially lower vaccination rates. Inaccurate information surged during the COVID-19 pandemic concerning its spread, treatment, and threat of devastation. The pandemic prompted a conversation about whose job it is, if anyone’s, to ensure online users can distinguish between false and true information and further curtail dissemination of false information.

This Note addresses two methods to partially solve this problem. First, Section 230 of the Communications Decency Act,<sup>7</sup> which provides sweeping immunity for online providers from content posted and shared by users, can be

---

1. Maggie Astor, *No, Other People’s Covid Vaccines Can’t Disrupt Your Menstrual Cycle*, N.Y. TIMES (Apr. 29, 2021), <https://www.nytimes.com/2021/04/29/technology/covid-vaccine-period-conspiracy.html>.

2. *Id.*

3. *Id.*

4. Patricia Mazzei, *A Private School in Miami, Citing False Claims, Bars Vaccinated Teachers from Contact with Students*, N.Y. TIMES, <https://www.nytimes.com/2021/04/26/us/florida-centner-academy-vaccine.html> (May 2, 2021).

5. Other examples of inaccurate information include that mRNA vaccines alter human DNA, the vaccine causes infertility, and the vaccine uses microchip surveillance technology. See John Gregory, *The Top COVID-19 Vaccine Myths Spreading Online*, NEWSGUARD, <https://www.newsguardtech.com/special-report-top-covid-19-vaccine-myths/#infertility> (Sept. 28, 2021, 2:40 PM).

6. See *infra* Part II.

7. “Section 230” is used as a shorthand for section 230 of the Communications Decency Act of 1996 (a common term for Title V of the Telecommunications Act of 1996).

carved out in part to have a public health exemption.<sup>8</sup> Under this framework, social media platforms *could* be liable if they were to knowingly propagate inaccurate information about public health. Second, and the primary focus of this Note, social media providers can amend their guidelines and prohibit the posting of false news articles or inaccurate commentary on COVID-19's spread, statistics, and treatment. This idea, as an aspect of Corporate Social Responsibility, would shift the responsibility of moderating content away from government regulators or users onto the platforms themselves. It does not have a basis in "hard law"; rather, platforms would realize a sense of responsibility given their tremendous role in the dissemination of news and information. Neither of these approaches needs to work alone. Instead, they can operate simultaneously and reinforce one another to create online spaces with minimal public health falsehoods.

#### I. WHERE DOES PUBLIC HEALTH MIS- AND DISINFORMATION ORIGINATE?

The terms misinformation and disinformation offer different perspectives on the spread of falsified information. Misinformation, in the context of public health, can be understood as information that does not align with the consensus in the scientific community regarding a phenomenon.<sup>9</sup> Information that falls under this label is persistently evolving as medicine and scientific information advance.<sup>10</sup> Misinformation permeates into many people's day-to-day lives because they tend to trust the advice of their friends, family, and broader communities, which are generally lay, secondary sources of information.<sup>11</sup>

Disinformation, on the other hand, presents a more sinister approach to the circulation of false information. Disinformation refers to the deliberate or coordinated effort to spread inaccurate information to advance one's reputation, power, or wealth.<sup>12</sup> Disinformation is *intentional*, meaning the person circulating the erroneous information knows its falseness but nonetheless publicizes it to others for personal gain.<sup>13</sup> Well-known examples include food industry lobbyists who fund research to push misleading health information to shape public opinion on the health of the food groups they represent.<sup>14</sup> While it is important to note the differences between these two terms, the differences are not meaningful to understand the principal arguments of this Note because it

---

8. Communications Decency Act, 47 U.S.C. § 230 (1996).

9. Briony Swire-Thompson & David Lazer, *Public Health and Online Misinformation: Challenges and Recommendations*, 41 ANN. REV. PUB. HEALTH 433, 434 (2020).

10. *Id.*

11. Kimberly Rodgers & Nnandi Massac, *Misinformation: A Threat to the Public's Health and the Public Health System*, 26 J. PUB. HEALTH MGMT. PRAC. 294, 294 (2020).

12. Swire-Thompson & Lazer, *supra* note 9, at 435.

13. *Id.*

14. Marion Nestle, *Food Lobbies, the Food Pyramid, and U.S. Nutrition Policy*, 23 INT'L. J. HEALTH SERVS. 483, 483 (1993).

simply groups mis- and disinformation together as false or misleading information.

The history of false information in public health is long and troubling. Prior to social media, journalists, news anchors, and other writers served as gatekeepers between information and the general public.<sup>15</sup> Now, however, anyone with internet access can post, consume, and share without a filtering intermediary.<sup>16</sup> Most mis- and disinformation is posted anonymously and then shared by others, allowing the original poster to evade accountability from public backlash.<sup>17</sup> A Twitter user, for example, can post commentary on a false news story and avoid repercussions by enjoying the anonymity granted by a username.

This problem is further exacerbated by “influence bots,” which are anonymous, nonhuman agents that can spread mis- and disinformation.<sup>18</sup> These bots tend to go unnoticed and undisturbed by other users by displaying a profile picture and listing a false name and location.<sup>19</sup> These bots imitate human behavior by liking, replying, or retweeting other users’ posts, causing many users to dismiss bots as humans.<sup>20</sup> Studies suggest that these bots have been employed by companies that want to promote falsified health information to promote their products (for example, vitamin supplements and cigarettes).<sup>21</sup> In addition to spreading false content for the financial gain of its promoters, bots also pose a risk in skewing efforts to understand public sentiment by presenting themselves as human.<sup>22</sup> Information is seen as more credible if the user understands the original poster to be human.<sup>23</sup>

The massive amount of information available on the internet posted and spread by both bots and human users presents users with the problem of information overload.<sup>24</sup> Information overload can cause objective facts to become less influential to consumers, whereas inaccurate information that toys with its consumers’ emotions and established beliefs is pervasive in the shaping of public opinion.<sup>25</sup> Confirmation bias—that is, information that supports consumers’ long-standing beliefs—makes it difficult to debunk mis- or

---

15. Katherine J. Igoe, *Establishing the Truth: Vaccines, Social Media, and the Spread of Misinformation*, HARV. SCH. OF PUB. HEALTH (July 10, 2019), <https://www.hsph.harvard.edu/ecpe/vaccines-social-media-spread-misinformation/>.

16. See Swire-Thompson & Lazer, *supra* note 9, at 435.

17. Igoe, *supra* note 15.

18. *Id.*

19. Jon-Patrick Allem & Emilio Ferrara, *Could Social Bots Pose a Threat to Public Health?*, 108 AM. J. PUB. HEALTH 1005, 1005 (2018).

20. *Id.*

21. *Id.* at 1006.

22. Amelia M. Jamison, David A. Broniatowski & Sandra Crouse Quinn, *Malicious Actors on Twitter: A Guide for Public Health Researchers*, 109 AM. J. PUB. HEALTH 688, 688 (2019).

23. See *id.*

24. See *id.*; Lauren Vogel, *Viral Misinformation Threatens Public Health*, 189 CANADIAN MED. ASS’N. J. 1567, 1567 (2017).

25. *Id.*

disinformation that aligns with consumers' prior-held narratives regardless of the veracity of those narratives.<sup>26</sup> Users are unlikely to think critically about a false news story if it mirrors their beliefs about a topic, and many users are unwilling to spend their time and mental energy consuming information that suggests their beliefs are untrue. For example, in late 2016, a satirical article that falsely claimed President Obama banned the Pledge of Allegiance generated over two million interactions.<sup>27</sup> Without investigating its veracity, opponents of President Obama shared the article because it reaffirmed their distaste for his politics and their perception of him as being anti-American.<sup>28</sup>

An astounding 80% of internet users seek public health information online.<sup>29</sup> Among these users, 70% report that information they discovered on the internet influences their healthcare choices.<sup>30</sup> While some public good may be achieved by the massive amount of health information online, this public health information is engendering the threat of "digital pandemics," that is, "the rapid[,] far-reaching spread of unrestricted and scientifically inaccurate health information across the Web through social networks."<sup>31</sup> This massive dump of inaccurate information in the last few decades has caused a number of public health crises, mostly notably the false idea that scientific evidence does not support vaccine safety and efficacy.<sup>32</sup>

## II. FROM AUTISM SPECTRUM DISORDER TO COVID-19: THE PROLIFERATION OF INACCURATE INFORMATION

Vaccines can be understood as "victims of their own success."<sup>33</sup> As the dangers posed by vaccine-preventable diseases in developed countries have dramatically diminished, consternation about vaccine side effects has become more commonplace than the diseases against which the vaccine protects.<sup>34</sup> Vaccine misinformation often involves discussion about a "rhetoric of doubt" regarding possible vaccine risks, coupled with people spreading their personal experiences with vaccine side effects to their social groups.<sup>35</sup> These social

---

26. Wen-Ying Sylvia Chou, Anna Gaysynsky & Joseph N. Cappella, *Where We Go from Here: Health Misinformation on Social Media*, 110 AM. J. PUB. HEALTH S273, S273 (2020).

27. *Fake News: Separating Truth from Fiction: Four Fake News Examples*, VALENCIA COLL. LIBR. GUIDES, <https://libguides.valenciacollege.edu/c.php?g=612299&p=4251645> (last updated May 4, 2022, 11:12 AM); Caroline Wallace, *Obama Did Not Ban the Pledge*, FACTCHECK.ORG (Sept. 2, 2016), <https://www.factcheck.org/2016/09/obama-did-not-ban-the-pledge/>.

28. See Wallace, *supra* note 27.

29. Vogel, *supra* note 24, at 1567; Anna Kata, *A Postmodern Pandora's Box: Anti-Vaccination Misinformation on the Internet*, 28 VACCINE 1709, 1709 (2010).

30. Kata, *supra* note 29, at 1709.

31. Brittany Seymour, Rebekah Getman, Avinash Saraf, Lily H. Zhang & Elsbeth Kalendarian, *When Advocacy Obscures Accuracy Online: Digital Pandemics of Public Health Misinformation Through an Antifluoride Case Study*, 105 AM. J. PUB. HEALTH 517, 517 (2015).

32. *Id.* at 517.

33. Kata, *supra* note 29, at 1709.

34. *Id.*

35. *Id.*

groups and communities significantly pressure other group members' decisions on whether to vaccinate themselves or their children.<sup>36</sup> Further, the ease of posting on the internet allows inaccurate information on vaccines to proliferate *between* communities.<sup>37</sup> Pervasive misinformation about vaccines is alarming because high levels of vaccination are necessary to create meaningful immunity to vaccine-preventable diseases across populations.<sup>38</sup> These vaccine skeptic and anti-vaccine coteries can develop into echo chambers where community consensus about vaccines prevails over independent decisions and research.

The most pervasive inaccurate narrative about vaccines is that inoculation in children can cause Autism Spectrum Disorder (ASD).<sup>39</sup> ASD is a developmental disability that may result in notable social, communicative, and behavioral challenges.<sup>40</sup> The hypothesized link between ASD and vaccines began in 1998 with a study by British gastroenterologists describing eight children who developed their first symptoms of ASD within one month of receiving an MMR vaccine.<sup>41</sup> Despite consequential issues with the study that rendered the study scientifically meaningless, its results engendered deep fear within parental communities.<sup>42</sup> Several studies from experts around the world have followed in the decades since 1998 that confirm vaccines do not cause ASD.<sup>43</sup> These studies, in tandem with the biological impossibility that vaccines “overwhelm” children’s immune systems, conclusively confirmed that there is no link between vaccination and ASD.<sup>44</sup>

Nonetheless, the false causal link between vaccines and ASD has thrived in online communities. Some of the inaccurate information spread about the link between ASD and vaccines emanates from online communities where parents encourage others to adhere to anti-vaccination narratives.<sup>45</sup> Many parents rely on information from other parents and use anti-vaccination rhetoric from these social media groups as predicates when making their own decisions.<sup>46</sup> If one parent were to be skeptical about vaccinating their children, the support of other parents with similarly held beliefs (even if inaccurate) can serve to confirm

---

36. *Id.*

37. *Id.*

38. *Social Media Perpetuates Vaccine Misinformation*, 39 HEALTHCARE LEADERSHIP R. 12, 12 (2020).

39. *See Autism and Vaccines*, CTRS. FOR DISEASE CONTROL & PREVENTION, <https://www.cdc.gov/vaccinesafety/concerns/autism.html> (last updated Dec. 1, 2021).

40. *Id.*

41. Jeffrey S. Gerber & Paul A. Offit, *Vaccines and Autism: A Tale of Shifting Hypotheses*, 48 CLINICAL INFECTIOUS DISEASES 456, 456 (2009).

42. *Id.*

43. *Id.* at 457.

44. *Id.* at 460.

45. Shankar Vedantam, *When It Comes To Vaccines And Autism, Why Is It Hard To Refute Misinformation?*, NPR (July 22, 2019, 5:01 AM), <https://www.npr.org/2019/07/22/744023623/when-it-comes-to-vaccines-and-autism-why-is-it-hard-to-refute-misinformation>.

46. *Id.*; Sarah Emerson, *Anti-Vax Groups Thrive on Facebook as Nationwide Coronavirus Vaccinations Begin*, ONEZERO (Dec. 17, 2020), <https://onezero.medium.com/anti-vax-groups-thrive-on-facebook-as-nationwide-coronavirus-vaccinations-begin-76470a5d392c>.

“prior convictions and to disregard data that does not conform to what [they] already believe.”<sup>47</sup>

While anti-vaccination misinformation is prolific among online parenting communities, public health misinformation is generally more widespread and ever-present. As COVID-19 swept through communities throughout the world in 2020 and through to the time of publication of this Note, inaccurate information regarding how the virus works, how it spreads, and how its vaccines would affect people expanded across many communities on social media platforms. The term “infodemic”—referring to “an overabundance of [both accurate and inaccurate] information that makes it hard for people to find trustworthy sources and reliable guidance when they need it”—helps elucidate the rise of mass information surrounding COVID-19.<sup>48</sup> An infodemic includes mis- and disinformation seeking to deliberately or negligently spread untruthful information that can, in turn, be damaging to people’s physical and mental wellbeing, increase stigmatization of particular communities, and cause poorer public health outcomes.<sup>49</sup>

### III. THE NEW MIS- AND DISINFORMATION: THE COVID-19 “INFODEMIC”

The COVID-19 pandemic is remarkable in that it is the first public health pandemic where social media platforms and other information technologies have been effectively and ineffectively used to keep people informed, educated, and connected to one another.<sup>50</sup> While COVID-19 was still a novel crisis in early 2020, social media platforms notified communities to shelter in place, if possible, and avoid contact with members outside the household.<sup>51</sup> The spread of information through online channels quickly became an avenue to disseminate false information about COVID-19, as people became fearful, confused, skeptical, or restless about ongoing COVID-19 protocols.<sup>52</sup> As the virus spread and lockdowns began, online news consumption skyrocketed.<sup>53</sup> For example, daily use of BBC News doubled in March 2020, and use of social media for news consumption jumped by 9% among people under 35 from January to April 2020.<sup>54</sup>

---

47. Vendantam, *supra* note 45.

48. *Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm from Misinformation and Disinformation*, WORLD HEALTH ORG. (Sept. 23, 2020), <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation>.

49. *Id.*

50. *Id.*

51. Md Saiful Islam, Tonmoy Sarkar, Sazzad Hossain Khan, Abu-Hena Mostofa Kamal, S. M. Murshid Hasan, Alamgir Kabir Dalia Yeasmin, Mohammad Ariful Islam, Kamal Ibne Amin Chowdhury, Kazi Selim, Anwar, Abrar Ahmad Chughtai & Holly Seale, *COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis*, 103 AM. J. TROPICAL MED. & HYGIENE 1621, 1621 (2020).

52. *See id.* at 1623–28.

53. *Id.* at 1621.

54. Nic Newman, *Executive Summary and Key Findings of the 2020 Report*, DIGIT. NEWS REP., <https://www.digitalnewsreport.org/survey/2020/overview-key-findings-2020/> (last visited July 1, 2022).

Among the most common type (89% of reports related to the infodemic) of false information disseminated throughout the COVID-19 pandemic was rumors, which includes “any unverified and instrumentally relevant claims, statements, and discussion.”<sup>55</sup> Such rumors included those related to COVID-19 symptoms, transmission, mortality, and prevention, such as eating garlic and avoiding spicy foods to reduce risk of contracting COVID-19.<sup>56</sup> While some false information may not cause any concrete harm to oneself or others, other false information can be particularly damaging, such as the promotion of injurious stigmas about China and Chinese people and dangerous remedies to prevent or treat the virus.<sup>57</sup> Some forms of false information resulted in observable violence, such as the arson attacks of 5G infrastructure following inaccurate hypotheses that hold COVID-19 is linked to 5G signals.<sup>58</sup> This false idea—which stipulated that 5G signals caused changes in people’s bodies that made them more vulnerable to contracting the virus—demonstrates how online hypotheses can translate to real-life, concrete harm.<sup>59</sup>

While some effects of mis- and disinformation may be minimal, the aggregation of their false guidance, in tandem with more egregious forms of spreading inaccurate information, can result in long-standing effects in communities globally. Mis- and disinformation has literally cost lives during the COVID-19 pandemic, with misinformed individuals impugning immunization campaigns and foregoing diagnostic tests and life-saving medical treatment.<sup>60</sup>

These egregious consequences resulting from mis- and disinformation cannot be ignored by policymakers or social media giants. While addressing the problem can come in myriad ways, two disparate ways will be addressed below. The first approach involves the responsibility of legislators. Since 1996, Section 230 has generally provided immunity for website platforms from content posted by their users.<sup>61</sup> This singular piece of internet legislation has shielded social media companies from liability for decades, and its reform could contribute to a stronger incentive for social media platforms to correct or delete misleading and false information.

The second approach focuses on Corporate Social Responsibility (CSR) and puts the pressure on the social media companies themselves. Under a CSR model, these platforms would be obligated, given their role in the spreading of information among social groups, to ensure inaccurate information is corrected. This form of self-regulation has the power to curb the spread of inaccurate

---

55. Islam et al., *supra* note 51, at 1622, 1624.

56. *Id.* at 1622.

57. *Id.* at 1626–27.

58. Paul Butcher, *COVID-19 as a Turning Point in the Fight Against Disinformation*, 4 NATURE ELECS. 7, 7 (2021); Adam Satariano & Davey Alba, *Burning Cell Towers, Out of Baseless Fear They Spread the Virus*, N.Y. TIMES (Apr. 11, 2020), <https://www.nytimes.com/2020/04/10/technology/coronavirus-5g-uk.html>.

59. Satariano & Alba, *supra* note 58.

60. *Managing the COVID-19 Infodemic*, *supra* note 48.

61. Communications Decency Act, 47 U.S.C. § 230 (1996).

information because social media platforms would take it upon themselves to intervene with users' posts if the content of those posts contains false material. CSR makes internet platforms accountable to themselves, shareholders, and the public, and, if the platform is large and visible enough, strong internal policies can set an example for similar industries.<sup>62</sup>

While the methods of curtailing the consumption of inaccurate information need not stand alone, they each focus the conversation on different responsible parties. One must ask who—either legislators or large social media platforms—should supervise the internet to prevent the spread of inaccurate, and often harmful, information.

#### IV. CARVEOUTS AND WORKAROUNDS: SECTION 230

Section 230 is a part of the Communications Decency Act of 1996, and it separates the provider from the user by establishing that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.”<sup>63</sup> Jeff Kosseff, a cybersecurity professor at the U.S. Naval Academy's Cyber Science Department, refers to the quoted words in the preceding sentence as the “twenty-six words that created the [i]nternet.”<sup>64</sup> Section 230 provides online internet providers general immunity from the content posted by their users, as users themselves would be liable for their posts.<sup>65</sup> Users, however, may be anonymous or bots, which makes it cumbersome to attribute tortious or criminal liability.<sup>66</sup>

Prior to its passage, internet service providers were responsible for all content posted on the sites they hosted.<sup>67</sup> This task proved to be intractable for the internet in 1996, and thus many providers opted to take forums down rather than attempt to comb through all user-posted content.<sup>68</sup> Section 230 provides opportunities for platforms to host forums without fearing liability. While this problem provides fuel for pro-Section 230 advocates, its twenty-six words have been a hot topic in recent decades with fierce arguments on both sides.

Section 230's controversy is in part attributable to it being a product of the early-internet era. The internet in the mid-1990s was barebones: few people could send and receive email and log into chatrooms with a smattering of

---

62. Jason Fernando, *Corporate Social Responsibility*, INVESTOPEDIA, <https://www.investopedia.com/terms/c/corp-social-responsibility.asp> (last updated Feb. 20, 2022).

63. 47 U.S.C. § 230.

64. Heidi Tworek, *Should There Be a Public Health Exemption for Section 230?*, BROOKINGS (May 19, 2020), <https://www.brookings.edu/techstream/should-there-be-a-public-health-exemption-for-section-230/>.

65. *See* 47 U.S.C. § 230.

66. *See infra* Part II.

67. Dan Patterson, *What is "Section 230," and Why Do Many Lawmakers Want to Repeal It?*, CBS NEWS (Dec. 16, 2020, 10:59 AM), <https://www.cbsnews.com/news/what-is-section-230-and-why-do-so-many-lawmakers-want-to-repeal-it/>.

68. *Id.*

friends.<sup>69</sup> Google was not yet introduced until late 1998.<sup>70</sup> A few newspapers began to put some news online, but there was not yet a place to share these articles or comment on them.<sup>71</sup> In 1996, a mere twenty million Americans had access to the internet, and many users paid to access it by the hour.<sup>72</sup> The stark differences between today's internet and that of the mid-1990s help illustrate that a statute from the internet's nonage does not properly address modern-internet problems.

Section 230 jurisprudence arises from platforms where users could post their thoughts to either known friends or to anyone with internet access. Its effects have been very controversial. For example, victims of a terrorist attack could not recover from Facebook (now owned by Meta) for the platform allowing its users to publish material supporting the terrorist organization that instigated the attack.<sup>73</sup> Similarly, plaintiffs could not recover from AOL after it had a delayed response in taking down false and defamatory information, failed to post retractions of the statements, and refused to prevent posts of a similar nature thereafter.<sup>74</sup> These plaintiffs would have to refocus their recoveries on the users who posted the information, and, considering social media platforms have deeper pockets than most users, these plaintiffs are likely to recover a much smaller amount of damages, if any.

Internet users today have forums like YouTube, Facebook, Twitter, Yelp, TikTok, and Instagram to post their opinions, photos, videos, and restaurant reviews. Users can share news articles and bits of false information to wide-reaching audiences in ways that were beyond the capabilities of the mid-1990s internet. Section 230, however, is not wholly without limits. In 2018, Congress made a significant carveout to its grant of general immunity through the near-unanimous, bipartisan passage of the FOSTA-SESTA bill.<sup>75</sup> This bill—a merger of a House bill titled Fight Online Sex Trafficking Act (FOSTA) and a Senate bill titled Stop Enabling Sex Traffickers Act (SESTA)—allows plaintiffs to recover from platforms if their users post content soliciting prostitution, including consensual sex work.<sup>76</sup> Similarly, in February 2021, congressional Democrats proposed a carveout called the Safeguarding Against Fraud,

---

69. Farhad Manjoo, *Jurassic Web: The Internet of 1996 is Almost Unrecognizable Compared with What We Have Today*, SLATE (Feb. 24, 2009, 5:33 PM), <https://slate.com/technology/2009/02/the-unrecognizable-internet-of-1996.html>.

70. *From the Garage to the Googleplex*, GOOGLE, <https://about.google/our-story/> (last visited July 1, 2022).

71. Manjoo, *supra* note 69; Nicholas Carlson, *Presenting: This Is What the Internet Looked Like in 1996*, BUS. INSIDER (Apr. 15, 2014, 9:51 AM), <https://www.businessinsider.com/the-coolest-web-sites-from-1996-2014-4>.

72. Manjoo, *supra* note 69.

73. *See* *Force v. Facebook, Inc.*, 934 F.3d 53 (2d Cir. 2019).

74. *See* *Zeran v. Am. Online, Inc.*, 129 F.3d 327 (4th Cir. 1997).

75. Tworek, *supra* note 64.

76. Aja Romano, *A New Law Intended to Curb Sex Trafficking Threatens the Future of the Internet as We Know It*, VOX (July 2, 2018, 1:08 PM), <https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom>.

Exploitation, Threats, Extremism and Consumer Harms (SAFE TECH), which suggests imposing liability on platforms for criminal behavior organized or facilitated on their platforms.<sup>77</sup> While these acts have many challenges, they demonstrate that carveouts to Section 230 are both possible and actively considered by legislators.<sup>78</sup>

Mis- and disinformation carveouts to Section 230 may be possible as well. In March 2021, Mark Zuckerberg, CEO of Meta (the parent company for Facebook), suggested in his testimony before the House Energy and Commerce Committee that in lieu of granting platforms sweeping immunity, platforms should be required to demonstrate that “they have systems in place for identifying unlawful content and removing it.”<sup>79</sup> Zuckerberg proposed penalizing social media platforms that cannot provide evidence of their unlawful-content detection system, while keeping Section 230’s sweeping immunity in place.<sup>80</sup> His offered “adequate system” would be proportionate to the size of the platform.<sup>81</sup>

Zuckerberg’s proposal to modify the *processes* around Section 230 could provide a sort of carveout for public health mis- and disinformation. If social media platforms were required to implement a content moderation system (proportionate to their respective infrastructure) to evade liability, these platforms would be incentivized to develop robust infrastructure to catch the mis- and disinformation posted by their users, as determined by third-party fact-checkers. Zuckerberg’s approach would keep the integrity of Section 230 in place while concurrently imposing some form of liability on platforms that fail to regulate the false or misleading information posted on their digital spaces. It would transfer the immunity contained in Section 230 from something that is given to all platforms without qualification to a privilege that is *earned* by platforms that implement and enforce moderation systems. This approach seems to be a promising way to crack down on Section 230’s grant of immunity if platforms fail to properly moderate COVID-19 mis- and disinformation while additionally maintaining Section 230’s spirit of a free internet.

Some commentators fear, however, that this approach may do very little to address the spread of mis- and disinformation on large platforms. Facebook, with its sizable content moderation system, may not have to implement many changes but would nonetheless appease Congress’s apprehension of Section

---

77. Meaghan E. Mixon, *The Misinformation of Capitol Hill: Section 230 and the Weaponization of Social Media*, PILLSBURY INTERNET + SOC. MEDIA (Feb. 18, 2021), <https://www.internetandtechnologylaw.com/section-230-misinformation-social-media-capitol-hill/>.

78. Tworek, *supra* note 64; Karol Markowicz, Opinion, *Congress’ Awful Anti-Sex-Trafficking Law Has Only Put Sex Workers in Danger and Wasted Taxpayer Money*, BUS. INSIDER (July 14, 2019, 5:38 AM), <https://www.businessinsider.com/fosta-sesta-anti-sex-trafficking-law-has-been-failure-opinion-2019-7>.

79. Dylan Byers, *Zuckerberg Calls for Changes to Tech’s Section 230 Protections*, NBC NEWS (Mar. 24, 2021, 6:44 AM), <https://www.nbcnews.com/news/rcna486>.

80. *Id.*

81. *Id.*

230.<sup>82</sup> Congresswoman Anna Eshoo, a Representative from California whose district includes Silicon Valley, expressed concerns that Zuckerberg’s suggestion is a “masterful distraction” from substantive Section 230 reform.<sup>83</sup> However, if Zuckerberg’s suggestion were to be augmented to hold large platforms, such as Facebook, to a significantly higher bar than the status quo, it may have potential to effectively force large platforms to rid their sites of mis- and disinformation. The result of carving out Section 230 could be accomplished if implemented in a manner that ensures *all* platforms meaningfully change their current content moderation systems.<sup>84</sup> Importantly, Google CEO Sundar Pichai rejected Zuckerberg’s approach, and then Twitter CEO Jack Dorsey kept notably silent about Section 230 in his testimony.<sup>85</sup>

President Joe Biden has openly supported “immediately” revoking Section 230 “[f]or Zuckerberg and other platforms.”<sup>86</sup> He backed his strong opposition to Section 230 by arguing that “[Facebook] is not merely an internet company. It is propagating falsehoods [it] know[s] to be false.”<sup>87</sup> In July 2021, he told reporters Facebook is “killing people,” but he later walked back the statement and urged the platform to do something about the “outrageous misinformation about the vaccine.”<sup>88</sup> While many politicians on all sides of the political spectrum have suggested reforming Section 230, President Biden has been a long-time, fierce critic of its liability shield.<sup>89</sup>

82. Kris Holt, *Mark Zuckerberg Proposes a ‘Thoughtful Reform’ of Section 230*, ENGADGET (Mar. 24, 2021), <https://www.engadget.com/mark-zuckerberg-section-230-proposal-facebook-google-twitter-hearing-163105917.html>.

83. Lauren Feiner, *Facebook’s Suggestion to Reform Internet Law is a ‘Masterful Distraction,’ Says Silicon Valley Congresswoman*, CNBC (Mar. 24, 2021, 6:21 PM), <https://www.cnbc.com/2021/03/24/facebook-section-230-suggestion-masterful-distraction-rep-eshoo.html>.

84. See Byers, *supra* note 79.

85. Nicolás Rivero, *Jack Dorsey is Leaving Mark Zuckerberg to Fight Section 230 Alone*, QUARTZ (Apr. 19, 2021), <https://qz.com/1989105/jack-dorsey-is-leaving-mark-zuckerberg-to-fight-section-230-alone/>.

86. Editorial Board, *Joe Biden: Former Vice President of the United States*, N.Y. TIMES (Jan. 17, 2020), <https://www.nytimes.com/interactive/2020/01/17/opinion/joe-biden-nytimes-interview.html?smid=nytcore-ios-share>.

87. Makena Kelly, *Joe Biden Wants to Revoke Section 230*, THE VERGE (Jan. 17, 2020, 10:29 AM), <https://www.theverge.com/2020/1/17/21070403/joe-biden-president-election-section-230-communications-decency-act-revoke>.

88. Nandita Bose & Elizabeth Culliford, *Biden Says Facebook, Others ‘Killing People’ by Carrying COVID Misinformation*, REUTERS (July 16, 2021, 5:09 PM), <https://www.reuters.com/business/healthcare-pharmaceuticals/white-house-says-facebooks-steps-stop-vaccine-misinformation-are-inadequate-2021-07-16/>; Betsy Klein, Maegan Vazquez & Kaitlan Collins, *Biden Backs Away from His Claim that Facebook Is ‘Killing People’ by Allowing Covid Misinformation*, CNN (July 19, 2021, 8:31 PM), <https://www.cnn.com/2021/07/19/politics/joe-biden-facebook/index.html>.

89. *Id.* Despite the President Biden’s criticisms of Section 230, the Department of Justice intends to defend the constitutionality of Section 230 in a series of big tech lawsuits filed by former president Donald Trump. Josh Gerstein, *Justice Department to Defend Tech Protections Biden Denounced*, POLITICO (Nov. 18, 2021, 8:21 PM), <https://www.politico.com/news/2021/11/18/justice-department-tech-protections-biden-522990>. Trump filed suit against big tech companies because he sees Section 230 as a tool “to coerce platforms into censoring their opponents.” Donald J. Trump, Opinion, *Donald J. Trump: Why I’m Suing Big Tech*, WALL ST. J. (July 8, 2021, 12:31 PM), <https://www.wsj.com/articles/donald-j-trump-why-im-suing-big-tech-11625761897>. Unlike President Biden, Trump is not concerned about spreading wrong information; rather, he is concerned that “[B]ig

If Section 230 were not revoked but instead carved out, the carveout itself may face immense challenges. Who, precisely, is to decide what is accurate information and what is a falsehood? Even if a board of medical professionals were to evaluate these sorts of questions, it may not be feasible for them to provide a sound determination on each piece of public health news or commentary. Further, would the government assess whether a platform has violated the carveout by assessing a platform's *own* independent evaluators of public health mis- and disinformation? Daphne Keller, a fellow at Stanford Law's Cyber Policy Center, expressed concern that even with Section 230 in its current form, many platforms are overresponsive to potentially illegal content by granting a significant number of users' take-down requests, because it threatens the free flow of speech and expression in online spaces.<sup>90</sup> If Section 230 were to be carved out for a public health exemption, this "if in doubt, take it down" approach many companies use may exacerbate the threat to free speech.<sup>91</sup>

Surely, the claim about menstrual cycle disruptions caused by standing nearby a vaccinated person would fall well within the realm of blatantly inaccurate information.<sup>92</sup> However, claims that better fit into the misleading category—such as the suggestion that experts do not yet know the COVID-19 vaccine's long-term effects<sup>93</sup>—straddle the line of what should be considered a prohibited falsehood.

As these questions and concerns linger, perhaps a less exacting approach to address the regulation and moderation of public health mis- and disinformation could depend on the platform themselves. Corporate Social Responsibility (CSR) shifts the burden from legislators onto the platforms to ensure internet users are not consuming or spreading false information surrounding COVID-19 and other public health issues. CSR avoids the question of the government determining what type of speech is admissible (and thus steer clear of thorny First Amendment issues and other constitutional challenges

---

Tech giants] are manipulating and controlling the political debate itself," causing "flagrant attack[s] on free speech." *Id.*

90. Daphne Keller, *Empirical Evidence of Over-Removal by Internet Companies Under Intermediary Liability Laws: An Updated List*, STAN. L. SCH.: CTR. FOR INTERNET & SOC'Y (Feb. 8, 2021, 5:11 AM), <https://cyberlaw.stanford.edu/blog/2021/02/empirical-evidence-over-removal-internet-companies-under-intermediary-liability-laws>.

91. *See id.*

92. *See infra* Part I.

93. A common misperception about new vaccines is that their long-term effects are not yet known; however, adverse vaccine side effects almost always present in the first few weeks or months. Holly Yan, *Covid-19 Vaccine Myths: These Reasons for Not Getting a Shot Don't Hold Up. In Fact, They'll Set the US Back*, CNN (July 19, 2021, 2:38 PM), <https://www.cnn.com/2021/04/28/health/covid-vaccine-myths-debunked/index.html>. The longest time an adverse side effect has presented itself is six weeks. *Id.* While not blatantly false to claim long-term side effects are unknown, it is a misleading claim because it suggests adverse side effects may arise many months or even years after the vaccine is administered. *Id.*

inevitably tied to Section 230 reform).<sup>94</sup> Since CSR is not “hard law” for platforms to follow, it presents a solution to deter mis- and disinformation while evading the tricky legal and political challenges presented by Section 230 reform.

#### V. CORPORATE SOCIAL RESPONSIBILITY: CLEARING UP THE VAGUENESS ON SOCIAL MEDIA PLATFORMS’ POLICIES

CSR is a visionary concept that refers to corporate social or environmental goals meant for companies to better the societies in which they take up space and on which they depend. Its primary focus is harmonizing a business’s social and environmental activities with its purposes, culture, and values.<sup>95</sup> Companies can achieve a “triple-bottom-line approach” by balancing economic, social, and environmental initiatives while simultaneously addressing shareholders’ financial expectations.<sup>96</sup> While CSR aims to develop societal or environmental good, it differs from pure philanthropy in that it balances competitive advantages, such as positive brand reputation, enhanced employee productivity, and strengthened customer loyalty, rather than making decisions with the cardinal intent of achieving altruistic outcomes, sometimes at the cost of finances, optics, or efficiency.<sup>97</sup> CSR does not compromise economic performance; rather, it intends to capitalize on customers feeling like they are doing their part for social or environmental good.<sup>98</sup>

Environmental CSR can be evidenced by Johnson & Johnson using wind power to provide safe drinking water to communities around the globe and aiming to source 35% of all the company’s energy needs from renewable energy methods.<sup>99</sup> Social change through CSR can be observed in Pfizer creating initiatives to raise awareness for non-infectious diseases<sup>100</sup> and provide healthcare for women and children around the world.<sup>101</sup> Google is among the companies that have tackled both social and environmental CSR through its

94. Daphne Keller discusses six constitutional challenges tied to Section 230 reform, most of which arise from the First Amendment. Daphne Keller, *Six Constitutional Hurdles for Platform Speech Regulation*, STAN. L. SCH.: CTR. FOR INTERNET & SOC’Y (Jan. 22, 2021, 6:50 AM), <https://cyberlaw.stanford.edu/blog/2021/01/six-constitutional-hurdles-platform-speech-regulation-0>.

95. V. Kasturi Rangan, Lisa Chase & Sohail Karim, *The Truth About CSR*, HARV. BUS. REV. (Jan.–Feb. 2015), <https://hbr.org/2015/01/the-truth-about-csr>.

96. *What is CSR?*, U.N. INDUS. DEV. ORG., <https://www.unido.org/our-focus/advancing-economic-competitiveness/competitive-trade-capacities-and-corporate-responsibility/corporate-social-responsibility-market-integration/what-csr> (last visited July 1, 2022).

97. *Id.*

98. *Sixteen Brands Doing Corporate Social Responsibility Successfully*, DIGIT. MKTG. INST. (Sept. 1, 2021), <https://digitalmarketinginstitute.com/blog/corporate-16-brands-doing-corporate-social-responsibility-successfully>.

99. *Id.*

100. *World Antibiotics Awareness Week: Don’t Spread the Infection, Spread the Word*, PFIZER LTD. (Dec. 11, 2018), <https://www.pfizer.co.uk/world-antibiotic-awareness-week> [<https://web.archive.org/web/20201129160507/https://www.pfizer.co.uk/world-antibiotic-awareness-week>].

101. *Sixteen Brands Doing Corporate Social Responsibility Successfully*, *supra* note 98.

CEO taking a stand against President Donald Trump's anti-Muslim comments and its data center using 50% less energy when compared to other large-scale data centers.<sup>102</sup>

CSR is a form of "soft law"; that is, CSR is not required by any statutes, regulations, or caselaw.<sup>103</sup> Rather, it is an obligation resulting from consumers' expectations and industry norms, shown by an astounding 85% of companies S&P 500 Index publishing CSR reports in recent years.<sup>104</sup> While Google, for example, has numerous stakeholders, its most important stakeholder is its users, as their behaviors define its popularity and, consequently, its business value.<sup>105</sup> Google responds to the demands of users by having expansive CSR efforts.<sup>106</sup> Large platforms, via accountability of their users, can make moderation of false information surrounding the COVID-19 pandemic and other public health issues a piece of a larger picture of CSR, or they can make it a key aspect of their business model.

Some social media companies have used CSR-based approaches to tackle the rise of mis- and disinformation during the COVID-19 pandemic by regulating what COVID-19 information can be shared. Since February 2021, Facebook has removed some false claims about COVID-19 following consultations with the World Health Organization (WHO), such as COVID-19 being man-made, the ineffectiveness of the vaccine, it being safer to get the disease than the vaccine, and claims that vaccines are dangerous or cause ASD.<sup>107</sup> Among Facebook's prohibited content are ads that suggest vaccines are unsafe or inconsistent with guidance from local health authorities; although, the platform does allow conversation around vaccine legislation.<sup>108</sup> If a seemingly false claim is posted once, it will be reviewed by a third-party fact-checker and labeled false if adjudged false.<sup>109</sup> Pages or groups are removed from the platform for repeatedly sharing debunked information.<sup>110</sup>

Facebook has further promoted correct information by favoring reputable sources when a user searches for COVID-19 news and updates, sending users a notification if they once interacted with a source found to be false, providing

---

102. *Id.*; Jonas Sickler, *Powerful Corporate Social Responsibility (CSR) Examples*, REPUTATIONMANAGEMENT.COM (July 2, 2020), <https://www.reputationmanagement.com/blog/corporate-social-responsibility-examples/>.

103. Elizabeth George, *Can Corporate Social Responsibility Be Legally Enforced?*, FORBES (Oct. 11, 2019, 9:58 AM), <https://www.forbes.com/sites/uhenergy/2019/10/11/can-corporate-social-responsibility-be-legally-enforced/?sh=c806c903d449>.

104. *Id.*

105. Pauline Meyer, *Google Stakeholders & Corporate Social Responsibility (CSR)*, PANMORE INST. (Jan. 28, 2017), <http://panmore.com/google-stakeholders-corporate-social-responsibility-csr-analysis>.

106. *Id.*

107. Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, META (May 26, 2021, 3:30 PM), <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>.

108. *Id.*; *Advertising Policies: Vaccine Discouragement*, FACEBOOK, [https://www.facebook.com/policies/ads/prohibited\\_content/vaccine\\_discouragement](https://www.facebook.com/policies/ads/prohibited_content/vaccine_discouragement) (last visited July 1, 2022).

109. Rosen, *supra* note 107.

110. *Id.*

truthful and accessible information in its Coronavirus Information Center, and urging people to click “Learn More” to be redirected to WHO via pop-ups throughout the site.<sup>111</sup> The platform is willing to take down mis- and disinformation spread by regular users, public figures, and political figures, which deviates from their usual practice of not fact-checking political figures’ speech.<sup>112</sup> Facebook pledged \$100 million to support journalists through the Facebook Journalism Project, with the money split between local news stations, news publishers’ marketing, and fact-checking organizations.<sup>113</sup> Facebook, joined by Google and Twitter, is working with a coalition of governments to assist British fact-checking charity Full Fact in rolling out a plan to minimize the spread of inaccurate information on all social media platforms.<sup>114</sup>

Twitter’s approach has been similar. Twitter holds firm that it will flag or remove false or misleading information about the origin and transmission of the virus; the efficacy or safety of vaccines or other preventative measures; official health authority regulations, restrictions, or exemptions; and the virus’s rates of transmission or death.<sup>115</sup> The policies are unclear regarding what variety of posts would be removed and which would be merely flagged.<sup>116</sup> Like Facebook, Twitter uses a strike system to determine when an account should be temporarily locked or permanently suspended.<sup>117</sup> Twitter *does* allow, however, fierce debate, opinion, and satire about the pandemic (so long as it is not misleading); counterspeech to misleading or false information; first-person accounts or other personal anecdotes; and public debate about the effectiveness of COVID-19 research and prevention measures.<sup>118</sup> Twitter’s enforcement of the policy has resulted in the removal of thousands of Tweets and the investigation of millions of accounts.<sup>119</sup>

YouTube works with similar measures by taking down videos that present false or misleading information about COVID-19, which YouTube ascertains as

---

111. *Id.*

112. Facebook took down Brazilian President Jair Bolsonaro’s claim the COVID-19 is a hoax in March 2020. Jen Patja Howell, *The Lawfare Podcast: Kate Klonick and Alina Polyakova on Pandemics, Platform Governance and Geopolitics*, LAWFARE (Apr. 9, 2020, 2:50 PM), <https://www.lawfareblog.com/lawfare-podcast-kate-klonick-and-alina-polyakova-pandemics-platform-governance-and-geopolitics>.

113. Kang-Xing Jin, *Keeping People Safe and Informed About the Coronavirus*, META (Dec. 18, 2020), <https://about.fb.com/news/2020/12/coronavirus/#news-industry-investment>.

114. Alex Hern, *Tech Giants Join with Governments to Fight Covid Misinformation*, THE GUARDIAN (Nov. 20, 2020, 10:16 AM), <https://www.theguardian.com/technology/2020/nov/20/tech-giants-join-governments-fight-covid-misinformation-facebook-twitter-google>.

115. *COVID-19 Misleading Information Policy*, TWITTER, <https://help.twitter.com/en/rules-and-policies/medical-misinformation-policy> (last visited July 1, 2022).

116. *Id.*

117. *Id.*

118. *Id.*

119. Paul Ziobro, *Twitter Expands Use of Enforcement System to Covid-19 Falsehoods*, WALL ST. J. (Mar. 1, 2021, 3:57 PM), <https://www.wsj.com/articles/twitter-expands-use-of-enforcement-system-to-covid-19-falsehoods-11614631575>.

information contrary to expert consensus.<sup>120</sup> TikTok is combatting COVID-19 vaccine disinformation by directing users to their COVID-19 information hub when users search for information, and the platform rolled out a vaccine tag that presents a “Learn more about COVID-19 vaccines” clickable banner when users post videos about the COVID-19 vaccine.<sup>121</sup> TikTok partnered with scientists and public health experts to promote scientifically accurate information in an accessible format, and it uses third-party fact-checkers to flag further sources of disinformation posted throughout the platform.<sup>122</sup>

## VI. SO WHERE DO WE GO FROM HERE?

Despite these seemingly comprehensive CSR-based efforts to regulate mis- and disinformation and conspiracy theories surrounding COVID-19’s transmission, prevalence, and preventative measures, users are nonetheless presented with false claims on all platforms. These platforms evade liability for false information posted to their forums via Section 230; thus, they hold free rein over whether they choose to allow or disallow their users to post false or misleading information. Although these policies contribute to steering users to accurate information, this CSR-based approach is not entirely sufficient to address the mass spread of mis- and disinformation on their platforms.

First, it is difficult to track down how rigidly these policies are enforced. When platforms introduce new policies, they are “essentially grading their own work.”<sup>123</sup> Many outside contractors employed to catch false or misleading information were unable to work adequately from home throughout the pandemic, and thus this work shifted to Artificial Intelligence (AI) tools that are prone to high error rates in content moderation.<sup>124</sup> Transparency reports often fail to adequately capture whether existing human moderators and current AI technology are sufficient to properly enforce content moderation policies—especially the amped-up policies established in response to the COVID-19 pandemic.<sup>125</sup>

For example, Instagram, a platform owned by Meta, rolled out a zero-tolerance policy for false COVID-19 information, but many users were fed easily searchable false claims about COVID-19’s source and the false link

---

120. *COVID-19 Medical Misinformation Policy*, YOUTUBE HELP, [https://support.google.com/youtube/answer/9891785?hl=en&ref\\_topic=9282436](https://support.google.com/youtube/answer/9891785?hl=en&ref_topic=9282436) (last visited July 1, 2022).

121. Kevin Morgan, *Taking Action Against COVID-19 Vaccine Misinformation*, TIKTOK (Dec. 15, 2020), <https://newsroom.tiktok.com/en-gb/taking-action-against-covid-19-vaccine-misinformation>.

122. *Id.*

123. Evelyn Douek, *The Year that Changed the Internet*, THE ATLANTIC (Dec. 28, 2020), <https://www.theatlantic.com/ideas/archive/2020/12/how-2020-forced-facebook-and-twitter-step/617493/>.

124. Evelyn Douek, *COVID-19 and Social Media Content Moderation*, LAWFARE (Mar. 25, 2020, 1:10 PM), <https://www.lawfareblog.com/covid-19-and-social-media-content-moderation/>; Howell, *supra* note 112.

125. Douek, *supra* note 123.

between vaccines and ASD.<sup>126</sup> Between September and November 2020, Instagram recommended over one hundred posts with misinformation to accounts in the United Kingdom in its “Suggested Posts” and “Explore” tabs that encourage users to find content relevant to their interests.<sup>127</sup> One hundred COVID-19 conspiracy accounts reaching an audience of over six million people were identified on the platform.<sup>128</sup>

Facebook, also owned by Meta, has been accused of failing to be transparent with collected data on mis- and disinformation. Whistleblower Frances Haugen disclosed internal research reports, slide decks, and employee discussion threads to the U.S. Securities and Exchange Commission and Congress, including multiple studies about the types of users who are most likely to share false information about COVID-19.<sup>129</sup> Facebook’s internal data show that anti-vaccine posts stemmed from a relatively small number of accounts, and internal documents showed that Facebook employees could determine how many views a post with misinformation received.<sup>130</sup> The platform formerly refused to provide this information and other data in response to pleas from congressmembers in July 2021.<sup>131</sup> The documents demonstrate how deeply Facebook understands the spread of false information, and it may not be forthcoming with all of its data to lawmakers or the public.

Further, many platforms are vague regarding under what circumstances a post is flagged (in other words, kept on the platform but tagged with a banner labeling it as false or misleading information) and under what circumstances a post is removed entirely. Most policies seem to reserve discretion to the platform on whether their content moderation infrastructure will opt to label the post with a banner or whether it determines the information to be so egregiously false that it is removed altogether. The policies fail to differentiate between false and misleading information; rather, it is seemingly a case-by-case determination. This approach to content moderation could cause a *misleading* post to be removed, while a *blatantly false* post could simply be labeled with a banner.

---

126. Shannon Bond, *Instagram Suggested Posts to Users. It Served Up COVID-19 Falsehoods, Study Finds*, NPR (Mar. 9, 2021, 12:01 AM), <https://www.npr.org/2021/03/09/975032249/instagram-suggested-posts-to-users-it-served-up-covid-19-falsehoods-study-finds>.

127. *Id.*

128. Jasper Jackson, Alexandra Heal & Tom Wall, *Facebook ‘Still Too Slow to Act on Groups Profiting from Covid Conspiracy Theories’*, THE GUARDIAN (Apr. 11, 2021), <https://www.theguardian.com/technology/2021/apr/11/facebook-still-too-slow-to-act-on-groups-profiting-from-covid-conspiracy-theories>.

129. Gerrit De Vynck, Cat Zakrzewski & Cristiano Lima, *Facebook Told the White House to Focus on the ‘Facts’ About Vaccine Misinformation. Internal Documents Show It Wasn’t Sharing Key Data*, WASH. POST (Oct. 28, 2021, 7:00 AM), <https://www.washingtonpost.com/technology/2021/10/28/facebook-covid-misinformation/>.

130. *Id.*

131. *Id.* The platform stated it has “nothing to share . . . outside of what Mark has said publicly” in response to Representatives Jan Schakowsky and Anna Eshoo’s letter to Mark Zuckerberg. Cristiano Lima, *What Facebook Knew About Covid-19 Misinformation – and Didn’t Tell Congress*, WASH. POST (Oct. 28, 2021, 9:00 AM), <https://www.washingtonpost.com/politics/2021/10/28/what-facebook-knew-about-covid-19-misinformation-didnt-tell-congress/>.

If some posts that contain false or misleading information are left up and labeled with a banner, then how effective are these banners in discouraging users to read, believe, and share misleading and false COVID-19 information? Yale researchers observed that tagging false news stories as “disputed by third-party fact-checkers” has a minimal impact on whether users perceive the headlines as true.<sup>132</sup> The percentage of participants who rated false headlines as accurate only marginally decreased with headlines that contain a warning banner compared to headlines that did not contain a banner.<sup>133</sup> Some warnings on mis- and disinformation can even serve to *increase* the lie’s spread. This “backfire effect” has been observed when ideologically conservative groups are presented with false information labeled with a banner noting its untruthfulness.<sup>134</sup> These groups engaged in motivated reasoning resulting from a distrust of both traditional media and independent, third-party fact-checkers.<sup>135</sup>

Another problem posed by flagging false or misleading information is that users believe all other information not labeled with a banner is truthful, even if it is entirely false.<sup>136</sup> Dubbed as an “implied truth effect,” many posts and articles pass through content moderation infrastructures and are considered truthful and reshared, even if they indeed contain false or misleading information.<sup>137</sup> This finding is significant because many posts containing COVID-19 mis- or disinformation go untagged entirely due to the vast quantity of mis- and disinformation posted on social media platforms.<sup>138</sup> A leaked email from Facebook revealed that it takes an average of three days for a story to be labeled as false or misleading, and thus mis- and disinformation would not be flagged in the infancy of its posting when it is most likely to be read and shared.<sup>139</sup> Thus, *any* warnings posted to false or misleading information may suggest all information remaining is truthful.<sup>140</sup> The presence of any banners that are visible to users suggests that articles without banners *must* be truthful if users work with the presumption that all or virtually all content goes through a moderation system. If no banners were present, the user would be left to their own devices to evaluate each post to determine its veracity.

---

132. Gordon Pennycook, Adam Bear, Evan T. Collins & David Gertler Rand, *The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings*, 66 MGMT. SCI. 4944, 4954 (2020).

133. *Id.* at 4948.

134. Brendan Nyhan & Jason Reifler, *When Corrections Fail: The Persistence of Political Misperceptions*, 32 POL. BEHAV. 303, 323 (2010).

135. *Id.*

136. Pennycook et al., *supra* note 132, at 4955.

137. *Id.* at 4944.

138. *Id.* at 4945.

139. Craig Silverman, *Facebook Says Its Fact Checking Program Helps Reduce the Spread of a Fake Story by 80%*, BUZZFEED: NEWS (Oct. 11, 2017, 4:15 PM), [https://www.buzzfeed.com/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its?utm\\_term=.nujMkKrND#.jq11pDN7y](https://www.buzzfeed.com/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its?utm_term=.nujMkKrND#.jq11pDN7y); Pennycook et al., *supra* note 132, at 4954.

140. Pennycook et al., *supra* note 132, at 4955.

The existing efforts in place are not doing enough to be a true CSR-based approach to combat the spread of COVID-19 mis- and disinformation. CSR, however, is a simpler method to address public health mis- and disinformation because it is a form of “soft law” that evades tricky legal questions and inevitable litigation that would tie up Section 230 reform. Notwithstanding the existence of responsive CSR-based policies, more stringent regulations need to be put in place by all large platforms through consumer pressure to ensure users are not exposed to COVID-19 mis- and disinformation.

## VII. RECOMMENDATIONS

If a problem is enormous, its solutions should match its size. Kate Klonick, fellow at the Information Society Project at Yale Law School, summed up the COVID-19 pandemic well: as a “time when the rest of the world went out, but the WiFi stayed on.”<sup>141</sup> The pandemic ended the days in which online speech is no different than ordinary speech.<sup>142</sup> Online platforms became a primary forum of communication during the COVID-19 pandemic. Speech once reserved for a public town square now exists within the four corners of a screen, and platforms are in a unique position of power to ensure our discourse is free, but that it hits a backstop when public health is put at serious risk. COVID-19 is markedly different from other disinformation—as Zuckerberg said regarding strict moderation for COVID-19 posts, “You can’t yell fire in a crowded theater.”<sup>143</sup> Even if the pandemic ends, vaccine and disease transmission fears will linger and taint future public health crisis responses, and the results may be deadly.<sup>144</sup>

To address these problems posed by online speech, first, social media platforms’ policies should explicitly state under what circumstances posts and user accounts are *removed* and under what circumstances posts are merely *flagged*.<sup>145</sup> Blatantly false information should be removed entirely at fast speeds to intercept early exposure to vulnerable users. Misleading information, on the other hand, could be affixed with a banner warning users that its content may be deceptive, confusing, or untrue. Information identified as verified and truthful could be labeled with a banner that confirms its truth and accuracy through third-party fact-checkers. Users would be aware that any news article they encounter without a banner has not yet been reviewed by independent fact-checkers, and thus they should carefully approach the headline and accompanying text. This approach would not prove to be any more burdensome than existing structures because, as it stands now, unlabeled posts on most platforms suggest the article

---

141. Howell, *supra* note 112.

142. *Id.*

143. Douek, *supra* note 124.

144. *Id.*

145. There is concern, however, that this approach is overresponsive and will remove or flag posts that do not contain false or misleading information. Keller, *supra* note 90. However, content moderation is “impossible to perform perfectly,” and, given the egregious consequences of false information during the COVID-19 pandemic, there is no choice but to err on the side of over-removal and flagging. Douek, *supra* note 123.

is either found truthful (and thus reviewed) or unreviewed. A banner on reviewed truthful posts would serve to distinguish between truthful and untruthful posts for the user.

Second, when users log onto social media platforms, the platforms' mis- and disinformation policies should be in-view and alert users about the meanings of banners. If users attempt to share news articles without opening the full article (in other words, sharing information based on headline alone), users could be prompted with a pop-up that suggests they read the article before sharing.<sup>146</sup> This pop-up would ideally encourage users to reflect upon their understanding of the news article and decide whether they believe the information to be truthful and reliable.

Third, social media platforms could create regular impact assessments to provide transparent data about how their regulation of COVID-19 mis- and disinformation affects public health outcomes, such as understanding of the disease, beliefs about transmission, and public discourse about vaccination.<sup>147</sup> Impact assessments provide transparency between data and corresponding decision-making, and they can facilitate dialogue between stakeholders, decision-makers, and users.<sup>148</sup> Allowing academics to study impact assessments would be useful to guide future policymaking around content moderation.<sup>149</sup> Impact assessments could detail existing mis- and disinformation policies and their effectiveness at stopping the spread of inaccurate information. Further, they could anticipate the effects of proposed changes to policies, such as the suggestions discussed above.

Section 230 does not incentivize the adoption of *any* of these responsible policies on social media platforms.<sup>150</sup> Rather, CSR is the driving vehicle for social media platforms to instill responsibility unto themselves to care for their users and broader society. Since CSR comes from users demanding platform accountability and not "hard law," it is devoid of the complicated legal challenges presented by Section 230 reform.<sup>151</sup> The COVID-19 pandemic has shown that nothing is "innate, inevitable, or immutable about platforms as they

---

146. Twitter ran a trial in June 2020 for Android devices that prompted users to read an article before retweeting. James Vincent, *Twitter Is Bringing its 'Read Before You Retweet' Prompt to All Users*, VERGE (Sept. 25, 2020, 7:08 AM), <https://www.theverge.com/2020/9/25/21455635/twitter-read-before-you-tweet-article-prompt-rolling-out-globally-soon>. Users opened articles 40% more often after being prompted, and users opening articles before retweeting increased by 33%. *Id.*

147. Existing transparency reports do not sufficiently capture whether human and AI content moderation can enforce content moderation policies. Douek, *supra* note 123.

148. Sara Bjørn Aaena, Ivar Lyhnea & Helle Nielsen, *The Use of Social Media in Impact Assessment: Experiences Among National Infrastructure Developers in Denmark*, 36 IMPACT ASSESSMENT & PROJECT APPRAISAL 456, 456 (2018). Most current transparency reports about content moderation.

149. Evelyn Douek, S.J.D. candidate at Harvard Law School, argues platforms should collect data on their COVID-19-era policies and be transparent about their effects. Douek, *supra* note 124. Further, platforms should take efforts to fix the problems AI content moderation presents for future sources of mis- and disinformation. *Id.*

150. *See infra* Part IV.

151. Section 230 reform poses numerous constitutional and other legal challenges. Keller, *supra* note 94.

currently exist”); rather, platforms can and should respond to present needs.<sup>152</sup> All communities benefit from the spread of truthful, accurate, and reliable information. Mis- and disinformation stemming from the COVID-19 pandemic threatens vaccination rates, adherence to public health regulations and recommendations, and the duration and severity of the pandemic.<sup>153</sup> Platforms should shoulder responsibility through CSR to deliver accurate information to their users and contribute to positive public health outcomes throughout all communities.

#### CONCLUSION

A vaccinated person cannot and will not disrupt the menstrual cycle of an unvaccinated person.<sup>154</sup> So long as false information can be shared online, some users will inevitably believe and accordingly make dangerous decisions. Content moderation is essential to lifting the world out of the deadly COVID-19 pandemic, but all approaches to regulating online speech present unique challenges.

Section 230 reforms are numerous. First, a Zuckerberg-esque approach could be taken where the integrity of Section 230 remains with its sweeping immunity, but platforms could be held liable if they do not have sufficient content moderation systems (proportionate to each platform’s size) in place to filter through mis- and disinformation. Second, akin to FOSTA-SESTA, Section 230 could be carved out to hold social media platforms liable when they do not adequately filter through public health mis- and disinformation posted by their users. Both approaches would serve to curb the large amounts of inaccurate COVID-19 information posted by users by incentivizing platforms—through the threat of fines and litigation—to be held accountable if their forums are used as a vehicle to spread inaccurate information. However, Section 230 presents thorny constitutional challenges that can be avoided by shifting social pressure onto platforms through CSR.

CSR presents more opportunity to hinder the consumption and spread of COVID-19 mis- and disinformation. Most social media platforms follow a similar trend where they present vague policies surrounding COVID-19 mis- and disinformation, and they consistently fail to answer when content is removed and when it is merely flagged. Further, flagging some information can mislead users into assuming all unflagged posts are truthful. Platforms could amend their policies for social good by entirely removing all *false* information, flagging all *misleading* information as potentially untruthful or confusing, and flagging all

---

152. Douek, *supra* note 123.

153. Sahil Loomba, Alexandre de Figueiredo, Simon J. Piatek, Kristen de Graef & Heidi J. Larson, *Measuring the Impact of COVID-19 Vaccine Misinformation on Vaccination Intent in the UK and USA*, 5 NATURE HUM. BEHAV. 337, 344 (2021); Zapan Barua, Sajib Barua, Salma Aktar, Najma Nabir & Mingze Li, *Effects of Misinformation on COVID-19 Individual Responses and Recommendations for Resilience of Disastrous Consequences of Misinformation*, 8 PROGRESS IN DISASTER SCI. 1, 5–6 (2020).

154. *See infra* Part I.

truthful information as verified and accurate. Any posts left unflagged would alert the user that the post has yet to be reviewed by third-party fact-checkers. Further, social media platforms should regularly post impact assessments that reflect the effectiveness of past mis- and disinformation policies, as well as an anticipated trajectory for proposed changes to their policies. Policies should also be transparent and easily accessible to all users.

Finally, none of these solutions needs to stand alone. Rather, all solutions should operate in concert to limit the spread of COVID-19 mis- and disinformation most effectively. The pandemic has forced people to congregate virtually more than ever before, and these public forums need to be responsive and accountable to their users and the public.