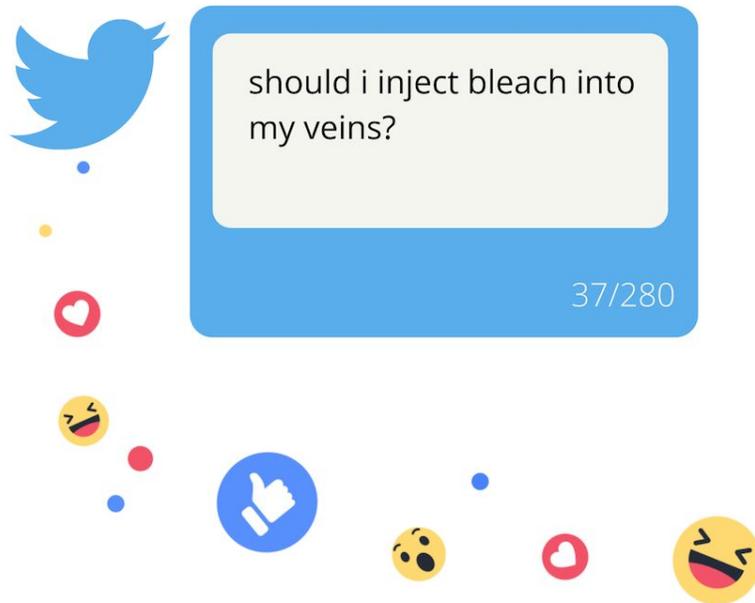


What's the Cure?

Misinformation and Platform Responses in the Era of COVID-19



Nick Abbott ('22)

Jacob Arkatov ('22)

Ryen Bani-Hashemi ('22)*

Casey Corcoran ('22)

David Dapaah-Afriyie ('22)*

Grant Fergusson ('22)

Genie Gorbonosov ('22)

Shira Kotz ('22)

Eric Lee ('22)

Adira Levine ('22)*

Jessica Li ('22)*

Katie Loftus ('22)

Amre Metwally ('22)

Carson Parks ('22)

Sasha Peters ('22)

Kate Pfeiffer ('22)

Katherine Shen ('22)

Sidarth Singh ('22)

Adam Toobin ('22)

Velo-Vincent van Houden ('22)

Table of Contents¹

Executive Summary	3
I. Understanding the Infodemic	4
Categorizing COVID-19 Misinformation	4
“Cures”	6
Fear	7
Domestic Politics	7
International Politics	8
Other Health-Related Misinformation	9
A Characterization of How COVID-19 Misinformation Has Tended to Spread	10
Social Media	11
Peer-to-Peer Messages	12
Websites	14
Traditional News Media	17
Discernible Impacts of COVID-19 Misinformation	21
Effect of COVID-19 Misinformation on Actual Practices	21
Effect of COVID-19 Misinformation on Attitudes and Beliefs	23
Effect of COVID-19 Misinformation on Media Coverage	24
Potential Interventions for Gauging the Impact of COVID-19 Misinformation	26
II. Platform Responses to the Infodemic	29
Comparing Platform Responses	29
Industry Collaboration and Communication	30
Product Interventions and Public Health Partnerships	31
Surfacing (and Signaling) Authoritative Sources to Users	31
Enabling Experts and Organizations to Communicate with Users Directly	31
Policy	32
Content Policy	32
Monetization Policy	33
Platform Responses through a “Public Health” Framework	33
Fire in a Crowded Theater: A Public Health Framework	34
The Whole World’s on Fire: A Global Response to a “Public Health” Crisis	35
Why Platforms Approach COVID-19 Differently	36
Further Proposed Interventions	39
Ensuring Users are Safe and Informed	42

¹ We are deeply grateful to Jonathan L. Zittrain, John Bowers, and the Harvard Law School Governing Digital Technology Spring 2020 class for their thoughtful feedback throughout this project.

Proposal 1: Share User Data with Health Organizations to Enable Targeted Counter-Narratives.	43
Proposal 2: Add Social Media Platforms to the Nation’s Emergency Alert System.	43
Proposal 3: Actively Try to Change User’s Habits.	44
Cross-Platform Collaboration to Stoke Innovation and Improve Existing Interventions	45
Proposal 1: Develop an Industry-wide COVID-19 Misinformation Database.	45
Proposal 2: Coordinate with Third-Party Fact-Checkers to Standardize COVID-19 Fact-Checking Policies.	46
Combating Misinformation Directly	48
Proposal: Increase Efforts to Actively Remove Misinformation, and Accounts that Regularly Disseminate It	48
Treating Misinformation as Public Health Data	49
Proposal: Share Health-related, Trend Information and Developments in Misinformation with Public Health Agencies.	49
Conclusion	50

Executive Summary

Social media companies have occupied particularly critical roles in the COVID-19 pandemic. Whether in disseminating information² or providing new entertainment options³ while billions of individuals are social distancing, the largest platforms have been stepping up in both predictable and novel ways. There are dozens of technology companies that could be analyzed in the scope of this report. However, because resource constraints inform the range of content moderation options available to smaller companies, this report instead focuses primarily on Facebook, Twitter, Instagram, and YouTube and their steps to combat coronavirus-related abuse on their platforms.

As the pandemic pushes more users to online spaces, a concurrent infodemic is pushing waves of false content to those same platforms.⁴ The dramatic rise in posts about miracle cures, virus conspiracy theories, and fake reports of draconian government action have led platforms to adopt new products and policies to combat misinformation.⁵ These interventions represent a substantial new assumption of responsibility for the accuracy of the content on their sites.

This report contains two sections. The first section aims to create a taxonomy of misinformation, detailing its forms, spread, and effects. The second section documents and compares how platforms have responded to misinformation during the pandemic, and proposes changes to improve platform responses to the infodemic.

Misinformation surrounding COVID-19 is unique both in terms of its content, motivation, and spread. Furthermore, the platforms are responding in novel ways. Though some argue that platforms are still not doing enough, this report suggests that the response represents a fundamentally new approach to public health challenges. Rather than treating health misinformation as an “exception” to their generally permissive content policies, platforms are

² Roose, Kevin, and Gabriel J. X. “The Coronavirus Revives Facebook as a News Powerhouse.” *The New York Times*, The New York Times, 23 Mar. 2020, www.nytimes.com/2020/03/23/technology/coronavirus-facebook-news.html.

³ Whateley, Dan. “A New 22-Page Report Breaks down How Livestream Video Has Surged in the Last Month on YouTube, Twitch, and Other Platforms. Here Are the 4 Key Takeaways.” *Business Insider*, Business Insider, 7 Apr. 2020, www.businessinsider.com/how-coronavirus-is-changing-livestream-viewer-habits-youtube-twitch-2020-4.

⁴ “UN Tackles 'Infodemic' of Misinformation and Cybercrime in COVID-19 Crisis.” United Nations, United Nations, www.un.org/en/un-coronavirus-communications-team/un-tackling-'infodemic'-misinformation-and-cybercrime-covid-19.

⁵ Statt, Nick. “Major Tech Platforms Say They're 'Jointly Combating Fraud and Misinformation' about COVID-19.” *The Verge*, The Verge, 17 Mar. 2020, www.theverge.com/2020/3/16/21182726/coronavirus-covid-19-facebook-google-twitter-youtube-joint-effort-misinformation-fraud.

assuming a new level of responsibility for the accuracy of the content on their sites generally. While the challenges of the moment may pose unique problems, this report also suggests that the motivations behind the new interventions will continue beyond the present moment. These precedents may encourage platforms to continue a more active approach to public health challenges in the future.

I. Understanding the Infodemic

COVID-19 has uniquely captured the world’s attention and directed it toward a singular threat. With this unity of focus has come a proliferation of false information and narratives related to the pandemic. This misinformation is highly varied in both form and content, encompassing everything from false cures and inaccurate representations of government policy to outlandish claims about dolphins returning to Venice. A number of publications⁶ have sought to develop running trackers of pandemic-related misinformation. Here we will go a step further, offering a taxonomy of the pandemic-related misinformation circulating today, followed by an assessment of its spread and impact.

A. Categorizing COVID-19 Misinformation

Behind any taxonomy should be a question in need of an answer – something that the clustering and arrangement of examples helps us to understand. Here, our question is this: what sorts of pandemic-related misinformation is most salient to the platforms, public health officials, and journalists working to promote public knowledge in service of harm mitigation and an effective public health response? To answer this question, we have sought to cluster categories of misinformation in terms of two key factors – harmfulness and believability.

⁶ Lytvynenko, Jane. “Here's A Running List Of The Latest Hoaxes Spreading About The Coronavirus.” *BuzzFeed News*, 24 Mar. 2020. <https://www.buzzfeednews.com/article/janelytvynenko/coronavirus-fake-news-disinformation-rumors-hoaxes>



Figure 1. Believability vs Harmfulness. This chart shows examples plotted according to how believable and harmful they are thought to be. The examples are sorted by color into the categories of cures, fear, domestic politics, international politics, and other health-related misinformation.

As shown in the provided chart in Figure 1, harmfulness and believability give rise to a spectrum from low salience (bottom left) to high salience (top right). The harmfulness axis estimates the harm caused by those acting on the piece of misinformation, and believability measures the likelihood of the public acting on or retransmitting a given piece of misinformation. For example, recommendations to eat more garlic to prevent the virus might not be useful to public health, but – even if their audiences accept them completely – are substantially less harmful than spoofed medical opinions recommending off-label usages of medicine with serious side effects. While claims that drinking bleach prevents the virus might, if followed, be even more dangerous than off-label drug use, they are also substantially less believable and easier to disprove. As such, narratives promoting off-label drug use may be more salient than the other narratives mentioned here, requiring more attention on the part of platforms, fact-checkers, and others.

The examples in the chart above are offered as one necessarily subjective illustration of how this framework could be useful in evaluating the level of attention a given category of misinformation poses. The types of misinformation we have elected to show are offered as an illustration of the framework, not as an empirically informed evaluation applicable across all contexts. People and organizations working to triage a response to pandemic-related misinformation should pair this framework with relevant empirical data to inform their

strategies. For further background information on the narratives used in Figure 1, please see Appendix A.

Below, we discuss representative examples of several substantive categories of misinformation widely observed over the past months. While misinformation within each of these categories can vary substantially in where it falls on the axes described above, the examples chosen offer case studies for how the harmfulness and believability of a given narrative might be assessed.

1. “Cures”

The mysteries surrounding COVID-19 have given rise to plenty of misinformation related to potential cures. For example, “cures” such as drinking bleach and eating garlic are equally unsubstantiated but unequally harmful.⁷ These cures range widely in terms of harmfulness from following through with the cure, and in the amount of people likely to believe this information, which is why the cures section is relatively dispersed across the graph.

Off-label medications have been misrepresented as proven “magic bullet” cures for COVID-19. Given that these narratives are both believable (relative to the others described above) and harmful, we have placed them at the top right of our salience grid. False narratives around off-label drug use are harmful because they can lead to hoarding, which may limit access to patients that need the drug for its intended use,⁸ and because the unprescribed use of medication can have serious short- and long-term side-effects, particularly where dosage is not carefully controlled. They might be more intuitively believable – though data would be needed to substantiate this claim – because the drugs they reference have been deemed fit for *some* therapeutic purpose, however unrelated to the pandemic at hand, and reasonably safe for human consumption under certain conditions.

⁷ Capatides, Christina. “Coronavirus cannot be cured by drinking bleach or snorting cocaine, despite social media rumors,” *CBS News*, March 9, 2020,

<https://www.cbsnews.com/news/coronavirus-drinking-bleach-cocaine-false-rumors-social-media>; World Health Organization. “Coronavirus disease (COVID-19) advice for the public: Myth busters,” 2020, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>

⁸ Raphelson, Samantha, and Robin Young. “Lupus Patient Fears Greater Shortages Of Hydroxychloroquine,” *WBUR*, 19 May 2020, www.wbur.org/hereandnow/2020/05/19/lupus-patient-hydroxychloroquine-shortages.

2. Fear

Other misinformation serves mainly to promote fear. The disruption caused by COVID-19 is unprecedented in many ways, and wild and unfounded rumors of national food shortages and extreme quarantine measures have spread quickly through many communication channels. Rumors citing insider information with official-sounding language have jumped from social media to text messages and even to word of mouth. For example, a fake message from the National Guard used official logos and appropriately-styled language to announce that a two week quarantine was imminent and that no one would be allowed to leave their home for any reason.⁹ The message concluded by urging the reader to buy a two week supply of all necessities and forward the message to their family and friends.¹⁰ Genuine concern for loved ones propels this kind of misinformation to an ever increasing audience, and the rumor gains credibility when family and friends inform each other compared to when individuals encounter the rumor online where information is subject to a higher level of suspicion.

Believable fear-mongering misinformation like the example above is highly salient, even if it does not result in direct bodily harm of the sort contemplated above. For example, rumors of supply shortages become self-fulfilling as concerned shoppers rush to hoard essentials like food, hand sanitizer, and paper products.¹¹ Empty shelves then validate the fear promoted by the misinformation. Eventually, the original fiction could even be viewed as credible, leaving the public cynical of official news sources.

3. Domestic Politics

COVID-19 misinformation narratives are also disrupting domestic politics. Relative to traditional political misinformation, some contemporary COVID-19 political misinformation narratives pose heightened risks in terms of harm and believability.

First, harms have escalated as COVID-19 has itself been politicized. Countless unfounded conspiracies range from COVID-19 being a method to push media ratings¹² to a plan

⁹ @KansasGuard. "There is a letter circulating on social media about the @DHSgov mobilizing the National Guard. This is a RUMOR!!! Currently, there are less than 60 Kansas National Guardsmen on state active duty in support of Kansans during the #Covid_19 Response. #KS RumorWatch." *Twitter*, 31 Mar. 2020, 12:22 p.m., <https://twitter.com/KansasGuard/status/1245038886184779777>

¹⁰ *Ibid.*

¹¹ Suthivarakom, Ganda. "Don't Overdo the Coronavirus Stockpiling." *The New York Times*, The New York Times, 31 Mar. 2020, www.nytimes.com/2020/03/31/smarter-living/wirecutter/dont-overdo-the-coronavirus-stockpiling.html.

¹² Gowen, Annie. "Coronavirus Deniers and Hoaxers Persist despite Dire Warnings, Claiming 'It's Mass Hysteria'." *The Washington Post*, WP Company, 19 Mar. 2020, www.washingtonpost.com/national/coronavirus-deniers-outbreak-hoax/2020/03/19/46bc5e46-6872-11ea-b313-df458622c2cc_story.html.

by the Democratic Party to win the presidential election.¹³ These narratives may be fueling in-person protests, as well as refusals to observe social distancing guidelines vital to stopping the virus' spread.¹⁴

Second, the believability of COVID-19 misinformation has been buoyed by statements from authority figures who – thanks to the media attention around the pandemic – have access to much more coverage and public attention than usual. For example, Texas' Lieutenant Governor, Dan Patrick, stoked controversy by arguing that a majority of elderly Americans would prefer to die than permit the virus to wipe out America's economic vitality.¹⁵ Similarly, at a crowded campaign rally in South Carolina, Donald Trump denied the existence of COVID-19 – calling it a “new hoax.”¹⁶ Shortly after his speech, the internet faced a frenzy of far-fetched posts and countless contrarian COVID-19 conspiracies.¹⁷

4. International Politics

Misinformation narratives are not confined to domestic politics. Stories that COVID-19 originated as a bioweapon, or from accidental leakage from a lab in Wuhan, have gained traction among some lawmakers.¹⁸ Conversely, Chinese officials have stoked a theory that the virus was brought to Wuhan by the US army in October 2019.¹⁹ Scientific evidence points convincingly to a natural origin.²⁰ The rapidly evolving pandemic leaves platforms with an uncomfortable choice: censor information – including from governments – with shaky credibility and risk being proven wrong later, or play a role in perpetuating divisive conspiracies. While they may not

¹³ *Ibid.*

¹⁴ Iati, Marisa, and Lateshia Beachum. “Georgia, South Carolina and Tennessee Governors Announce Plans to Ease Coronavirus Restrictions.” *The Washington Post*, WP Company, 21 Apr. 2020, www.washingtonpost.com/world/2020/04/20/coronavirus-latest-news/.

¹⁵ Beckett, Lois. “Older People Would Rather Die than Let Covid-19 Harm US Economy – Texas Official.” *The Guardian*, 24 Mar. 2020, www.theguardian.com/world/2020/mar/24/older-people-would-rather-die-than-let-covid-19-lockdown-harm-us-economy-texas-official-dan-patrick.

¹⁶ Palma, Bethania. “Did President Trump Refer to the Coronavirus as a 'Hoax'?” *Snopes.com*, Snopes Media Group Inc., 2 Mar. 2020, www.snopes.com/fact-check/trump-coronavirus-rally-remark/.

¹⁷ Chotiner, Isaac. “The Contrarian Coronavirus Theory That Informed the Trump Administration.” *The New Yorker*, Condé Nast, 30 Mar. 2020, www.newyorker.com/news/q-and-a/the-contrarian-coronavirus-theory-that-informed-the-trump-administration.

¹⁸ Taylor, Adam. “Analysis | What Caused the Coronavirus? A Skeptical Take on the Theories about the Outbreak's Chinese Origin.” *The Washington Post*, WP Company, 16 Apr. 2020, www.washingtonpost.com/world/2020/04/16/what-caused-coronavirus-skeptical-take-theories-about-outbreaks-chinese-origin/.

¹⁹ Westcott, Ben, and Steven Jiang. “Chinese Diplomat Promotes Coronavirus Conspiracy Theory.” *CNN*, Cable News Network, 14 Mar. 2020, www.cnn.com/2020/03/13/asia/china-coronavirus-us-lijian-zhao-intl-hnk/index.html.

²⁰ Andersen, Kristian G., et al. “The Proximal Origin of SARS-CoV-2.” *Nature Medicine*, vol. 26, no. 4, 2020, pp. 450–452., doi:10.1038/s41591-020-0820-9.

directly affect physical health, scientists have condemned the theories as creating fear and prejudice, which jeopardize global collaboration.²¹

5. Other Health-Related Misinformation

Early data showed that young people are not as affected by COVID-19.²² While it appears to be true that the elderly have a higher risk of serious symptoms, young people can fall seriously ill from the virus.²³ Even so, a sense of invulnerability among younger people²⁴ due to misinterpretations of early data meant that many neglected to observe social distancing measures, even as the pandemic spread.²⁵ In addition to increasing the number of symptomatic cases of COVID-19, such behavior also stands to promote spread among more vulnerable people by increasing the number of disease carriers.²⁶ Exaggerated narratives around resistance on the part of the young should likely be considered highly salient in terms of our framework. The fact that such narratives are not entirely without basis makes them substantially more believable – and harder to debunk. And their ability to undermine social distancing efforts, thus escalating the pandemic, make them potentially very harmful.

Other examples of health-related misinformation include narratives equating COVID-19 to the flu.²⁷ Data has shown that COVID-19 is more contagious and deadlier than the flu,²⁸ and has not yet been addressed with vaccines or approved treatments.²⁹

²¹ Calisher, Charles, et al. “Statement in Support of the Scientists, Public Health Professionals, and Medical Professionals of China Combatting COVID-19.” *The Lancet*, vol. 395, no. 10226, 19 Feb. 2020, doi:10.1016/s0140-6736(20)30418-9.

²² Azad, Arman, and Minali Nigam. “Yes, Young Adults Are Sick and Spreading Coronavirus -- but They Can Help Stop It.” *CNN*, Cable News Network, 20 Mar. 2020, www.cnn.com/2020/03/20/health/covid-19-young-adults-sick-spreading/index.html.

²³ Meltzer, Kerry Kennedy. “I’m Treating Too Many Young People for the Coronavirus.” *The Atlantic*, Atlantic Media Company, 26 Mar. 2020, www.theatlantic.com/ideas/archive/2020/03/young-people-are-not-immune-coronavirus/608794/.

²⁴ Gunia, Amy. “Millennials Aren’t Taking Coronavirus Seriously, WHO Official Warns.” *Time*, Time, 20 Mar. 2020, time.com/5807073/millennials-coronavirus-who/.

²⁵ Bella, Timothy. “‘If I Get Corona, I Get Corona’: Miami Spring Breakers Say Covid-19 Hasn’t Stopped Them from Partying.” *The Washington Post*, WP Company, 19 Mar. 2020, www.washingtonpost.com/nation/2020/03/19/coronavirus-spring-break-party/.

²⁶ Gunia, “Millennials Aren’t Taking Coronavirus Seriously, WHO Official Warns.”

²⁷ Montanaro, Domenico. “FACT CHECK: Trump Compares Coronavirus To The Flu, But It Could Be 10 Times Deadlier.” *NPR*, NPR, 24 Mar. 2020,

www.npr.org/sections/coronavirus-live-updates/2020/03/24/820797301/fact-check-trump-compares-coronavirus-to-the-flu-but-they-are-not-the-same.

²⁸ *Ibid.*

²⁹ *Ibid.*

B. A Characterization of How COVID-19 Misinformation Has Tended to Spread

The unique ambiguity of facts and consequent lack of control in the era of COVID-19, particularly with regard to its causes, symptoms, and cures, has prompted the spread of false “remedies” and explanations.³⁰ COVID misinformation is often composed of half-truths — falsehoods and facts packaged together within the same message or narrative.³¹ The absence of reliable information has exacerbated fear and anxiety surrounding the virus, driving the dissemination and consumption of misinformation. This report uses the term “misinformation” to describe the information spread with good intentions for instance, when individuals unwittingly spread false COVID-19 cures in an effort to keep their loved ones safe.³²

COVID-19 misinformation ranges from harmless — for instance, baseless garlic-based treatments — to harmful — such as racist conspiracies theories.³³ Although it is tempting to view individual pieces of misinformation as harmful or harmless, attempting to address case-by-case pieces of misinformation, as social media platforms often do, cannot account for the cumulative “drip” effects of misinformation. For instance, the conspiracy theory that Bill Gates created COVID-19 is harmless on its face but could lead to individuals refusing COVID vaccines down the line.³⁴ This section aims to elucidate the mechanisms by which COVID-19 misinformation has tended to spread through various types of media, which, by extension, suggests potential means of intervention.

³⁰ One fascinating phenomenon is the rising number of individuals who are seeking information directly from doctors, epidemiologists and researchers, usually through social media. For a sense of this phenomenon and the controversy associated with direct “doctor” to patient informational flows see: <https://www.chronicle.com/article/This-Harvard-Epidemiologist-Is/248557>

³¹ For example, a widely circulated false Stanford University study claimed that individuals who could hold their breath for more than ten seconds are free of coronavirus. Shortness of breath is a common symptom of COVID-19, so the misinformation in this falsified report is more believable.

³² Benkler, Yochai, et al. “The Propaganda Feedback Loop.” *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, 2018, pp. 6.

³³ Heilweil, Rebecca. “How the 5G Coronavirus Conspiracy Theory Went from Fringe to Mainstream.” *Vox*, 24 Apr. 2020, www.vox.com/recode/2020/4/24/21231085/coronavirus-5g-conspiracy-theory-covid-facebook-youtube.

³⁴ Wakabayashi, Daisuke, et al. “Bill Gates, at Odds With Trump on Virus, Becomes a Right-Wing Target.” *The New York Times*, The New York Times, 17 Apr. 2020, www.nytimes.com/2020/04/17/technology/bill-gates-virus-conspiracy-theories.html.

1. Social Media

Social media is generally a pervasive source of misinformation, and the COVID-19 pandemic presents no exception. Of a sample of 225 sources of misinformation gathered by The Reuters Institute, 88% of the sample appeared on social media platforms, contrasted with 9% appearing on TV and 8% appearing on news outlets.³⁵ And of the sources of misinformation making the rounds on social media, political figures' posts make up the bulk of engagement in a top-down spread of misinformation centered largely around unproven cures, nationalistic rhetoric blaming other countries for the virus, or rumors of a forthcoming imposition of martial law.³⁶ A significant contributing factor to the ecosystem of misinformation on social media is a pervasive carveout in platforms' Terms of Use that permit violations of the rules by political figures.³⁷ Often cloaked under a "newsworthiness" exemption, politicians have long been immune from restrictions on online speech.³⁸ But platforms have responded to the uniquely dangerous pandemic with a series of takedowns, in some cases under the justification of "physical harm."³⁹ Furthermore, social media platforms have sought to counter the spread of misinformation through free or subsidized ad-space for public health entities.⁴⁰

In addition to the far-reaching posts of politicians, a number of celebrities and some individuals with minor followings have been instrumental in the spread of misinformation as they seek two polar outcomes: to sow panic or in many cases to protect their loved ones. Celebrities like Woody Harrelson have posted about 5G as a cause of COVID-19,⁴¹ while Instagram's IGTV feature has also been used by average people to spread video conspiracy theories about 5G as a cause of COVID.⁴² Instead of removing false content,⁴³ Instagram and

³⁵ Brennen, J Scott, et al. *Types, Sources, and Claims of COVID-19 Misinformation. Types, Sources, and Claims of COVID-19 Misinformation.*

³⁶ *Ibid.*

³⁷ Clegg, Nick. "Facebook, Elections and Political Speech." *Facebook*, 24 Sept. 2019, about.fb.com/news/2019/09/elections-and-political-speech/; "About Public-Interest Exceptions on Twitter." *Twitter*, help.twitter.com/en/rules-and-policies/public-interest; Ohlheiser, Abby. "The One Word That Lets Politicians Get Away with Breaking the Rules on Social Media." *The Washington Post*, 25 Sept. 2019, www.washingtonpost.com/technology/2019/09/25/newsworthiness-one-word-that-lets-politicians-get-away-with-spreeding-misinformation-social-media/.

³⁸ Ohlheiser, Abby. "The One Word That Lets Politicians Get Away with Breaking the Rules on Social Media." *The Washington Post*, 25 Sept. 2019.

³⁹ Ingram, David. "Facebook, Twitter Bar Video of Brazilian President Endorsing Unproven Antiviral Drug." *NBC News*, 30 Mar. 2020; Hamilton, Isobel A. "Facebook and Twitter Blocked Videos from Brazilian President Jair Bolsonaro for Coronavirus Misinformation." *Business Insider*, 31 Mar. 2020.

⁴⁰ "Facebook Gives WHO Free Ads in Battle with Coronavirus Misinformation." *Reuters*, Thomson Reuters, 4 Mar. 2020.

⁴¹ Sorkin, Amy D. "The Dangerous Coronavirus Conspiracy Theories Targeting 5G Technology, Bill Gates, and a World of Fear." *The New Yorker*, 24 Apr. 2020.

⁴² See, e.g., https://www.instagram.com/tv/B9xbYrxh55_; See also https://www.instagram.com/tv/B-bgWn1Jkbl/?igshid=axummbpoxi59&fbclid=IwAR3k5WkT8OuO_KPedfp2PFC-6qo9Vmm_h-eNeCznjyFNPIJSZYsnyaDiNdg.

Facebook have vowed to reduce misinformation by downranking content rated false by third-party fact checkers.⁴⁴ Contributing to this complicated situation, many content moderators were sent home in the wake of COVID-19, leading Facebook to rely on machine learning systems to remove false information.⁴⁵ However, though this development has led to a number of false positives and innocuous information taken down,⁴⁶ Facebook has admitted it believes this outcome is acceptable in the wake of an aggressive response to COVID-19 misinformation.⁴⁷

Social media has been a critical source of misinformation with the power to impact the response of politicians, families, and individuals to an uncertain period of pandemic. In a time of uncertainty and fear, social media users are encountering a trove of unverified information with the seeming potential to save their lives and the lives of their loved ones. Without adequate takedown measures implemented by social media platforms, misinformation will continue to spread through networks. However, examining social media alone never presents a complete picture of the breadth of misinformation, as observations of social media communications only offers a partial view of the overall ecosystem. The other sources of misinformation are crucial to an adequate understanding of misinformation in the era of COVID-19.⁴⁸

2. Peer-to-Peer Messages

Hidden virality describes the phenomenon in which trusted parties in private networks are able to quickly spread false information to a wider network of users while remaining impenetrable to outside oversight.⁴⁹ Misinformation spread through peer-to-peer messages are able to achieve hidden virality since information shared through direct messaging apps are private and unmoderated, making it more difficult for false information to be disproved and more difficult to monitor how widespread the rumors are.⁵⁰ Peer-to-peer messages also take on an increased importance in the spread of misinformation when social media platforms like Facebook and Twitter have begun to crack down on COVID-19 misinformation.

⁴³ Kreps, Sarah, and Brendan Nyhan. "Coronavirus Fake News Isn't Like Other Fake News." *Foreign Affairs*, 7 Apr. 2020.

⁴⁴ "Combatting Misinformation on Instagram." *Instagram*, 16 Dec. 2019; "How Is Facebook Addressing False News through Third-Party Fact-Checkers?" *Facebook*.

⁴⁵ Kreps.

⁴⁶ Newton, Casey. "The Coronavirus Is Forcing Tech Giants to Make a Risky Bet on AI." *The Verge*, 18 Mar. 2020.

⁴⁷ Perry, Tekla S. "How Facebook Is Using AI to Fight COVID-19 Misinformation." *IEEE Spectrum*, 12 May 2020, spectrum.ieee.org/view-from-the-valley/artificial-intelligence/machine-learning/how-facebook-is-using-ai-to-fight-covid19-misinformation.

⁴⁸ Benkler, 71.

⁴⁹ Britt Pari and Joan Donovan, "Deepfakes and Cheap fakes: The Manipulation of Audio and Visual Evidence" in *Data and Society*, https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1.pdf

⁵⁰ WhatsApp is particularly difficult to monitor messages are encrypted and the very structure of the app makes it "relatively intimate" i.e. you need to know someone's phone number in order to add them as a contact and is thus largely used for direct contacts in Europe and the US.

Viral text messages prey on the fact that people often have an emotional, as opposed to a rational, relationship to information.⁵¹ Messages containing COVID misinformation stress the dangers of the illness before asking recipients to “please send and share with family and friends.”⁵² The language and structure of such texts play on the anxieties of the recipient, encouraging them to share the information out of concern for loved ones. Peer-to-peer messages are also particularly conducive to the spread of misinformation because they are able to exploit secrecy in both content and transmission. Many viral messages claim that the source of the information is an individual with some kind of special access to traditional sources of authority - for instance “my friend who works at the Department of Defense, who said there was a secret meeting.”⁵³ It is tremendously difficult to disprove (or prove) such information given how often official advice and government advice regarding COVID-19 is changing. Since traditional news sources are providing limited and less-than-certain information, it is difficult for individuals to find a reliable alternative source of information that one can use to cross-reference the claims made in viral messages.

Finally, when misinformation is coming directly from friends and family, it becomes more difficult to ignore or reject misinformation because it now comes from a trusted source as opposed to an unknown and possibly malicious actor. Individuals are more susceptible to misinformation contained in peer-to-peer messages because they are viewed in a context most individuals regard as safe, informal and trustworthy. Since most people use SMS/messaging apps to connect with friends, family and colleagues, misinformation is often mixed with important regular messages. This means that viral text messages do not need to utilize sensationalized language or headlines in order to be passed on. Furthermore, peer-to-peer messages are “delivered straight into your hand” and unless you change the default settings, automatically notifies the user, thereby creating a sense of immediacy.⁵⁴ Individuals who are traditionally not exposed or susceptible to misinformation are now at risk since false information is delivered straight to them.

⁵¹ Staff, NPR/TED. “Claire Wardle: Why Do We Fall For Misinformation?” *NPR*, NPR, 20 Mar. 2020, www.npr.org/2020/03/20/818299094/claire-wardle-why-do-we-fall-for-misinformation.

⁵² See Appendix B.

⁵³ Collins, Ben. “False Coronavirus Rumors Surge in 'Hidden Viral' Text Messages.” *NBCNews.com*, NBCUniversal News Group, 17 Mar. 2020, www.nbcnews.com/tech/tech-news/false-coronavirus-rumors-surge-hidden-viral-text-messages-n1160936

⁵⁴ Davies, Guy. “Coronavirus Misinformation on WhatsApp Is Going Viral, despite Steps to Combat Its Spread.” *ABC News*, ABC News Network, 24 Mar. 2020, abcnews.go.com/Health/coronavirus-misinformation-whatsapp-viral-steps-combat-spread/story?id=69688321.

3. Websites

Outside of traditional news/media/social media outlets, the public also relies on the vast amount of information available on the internet through message boards and personal websites. Message boards present a particular danger because of users' anonymity. Just as individuals are more susceptible to misinformation in peer-to-peer messages, readers are similarly susceptible to anonymous posts on online forums. For example, coordinated efforts on 4chan pushed a theory about vitamin C being able to protect against COVID-19. This narrative was then embraced by conspiracy theorists on Twitter and Facebook.⁵⁵ Reddit is similarly "vulnerable to disinformation flows,"⁵⁶ as evidenced by its misinformation-riddled subreddit r/Wuhan_Flu, which the platform has "quarantined,"⁵⁷ but not banned or eliminated. Reddit has been promoting its slightly more policed subreddit, r/CoronaVirus, but Reddit's absence of a platform-wide response has led to mental health subreddits implementing their own moderation rules.⁵⁸ While this allows for the spread of information generally, allowing unpaid, anonymous moderators to supervise and manage subreddits regarding COVID-19 always leaves open the risk that misinformation will make its way through the cracks.

Social media also amplifies the voices of conspiracy theorists and con artists looking to profit politically or monetarily from the pandemic. As of the beginning of March, personal blogs or political websites spreading misinformation (including conspiracy theories about the origin of COVID-19 or miracle cures for the virus) had garnered over ten times the social media engagement (shares, likes, comments, etc.) of government operated websites, including the CDC and WHO.⁵⁹ These websites include the hyperpartisan and conspiracy theory-filled sites of Zero Hedge, Mind Unleashed, Natural News, and WND.com.⁶⁰ They also include Jim Bakker of the

⁵⁵ Ariel Bogle, "5G And Anti-Vax Conspiracy Theorists Are Exploiting The Coronavirus Crisis," *Australian Broadcasting Corporation*, April 14, 2020, <https://www.abc.net.au/news/science/2020-04-15/coronavirus-5g-vitamin-c-anti-vaccine-conspiracy-theories-spread/12145096>

⁵⁶ Clea Skopeliti, "How Mental Health Subreddits Are Coping With The Coronavirus Infodemic," *First Draft News*, April 3, 2020, https://firstdraftnews.org/latest/how-mental-health-subreddits-are-coping-with-the-coronavirus-infodemic/?fbclid=IwAR0mVvHE10JZvp8z3Biak0xg6OtKuhJFBE_veX64ZUibiHyKyxt8Wk-i9vg

⁵⁷ Steven Asarch, "Reddit Trying to Fight Coronavirus Misinformation – Which Subreddit is the Best News Source?" *Newsweek*, March 3, 2020, <https://www.newsweek.com/reddit-fight-coronavirus-misinformation-subreddit-best-news-source-1490292>

⁵⁸ Clea Skopeliti, "How Mental Health Subreddits Are Coping With The Coronavirus Infodemic," *First Draft News*, April 3, 2020, https://firstdraftnews.org/latest/how-mental-health-subreddits-are-coping-with-the-coronavirus-infodemic/?fbclid=IwAR0mVvHE10JZvp8z3Biak0xg6OtKuhJFBE_veX64ZUibiHyKyxt8Wk-i9vg

⁵⁹ John Gregory, "The Coronavirus 'Infodemic' Is Real. We Rated The Websites Responsible For It," *Stat News*, February 28, 2020, <https://www.statnews.com/2020/02/28/websites-spreading-coronavirus-misinformation-infodemic/>

⁶⁰ *Ibid.*

Jim Bakker show and Alex Jones of the InfoWars website advertising on social media to buy their “miracle cures” for COVID-19 on their personal online stores.⁶¹

Dr. Kate Starbird, Associate Professor of Human Centered Design & Engineering at the University of Washington, emphasized that the spread of misinformation and disinformation is a “natural response to the uncertainty and anxiety that are inherent to crisis events,” especially like this pandemic.⁶² There is a “chain of influence,” according to Jonathan Morgan, the CEO of Yonder, an A.I. software company tracking the growth of online narratives, referring to the way that misinformation that starts in smaller communities gets picked up by communities desperate for answers.⁶³ The unique lack of information and fear surrounding the COVID-19 pandemic has amplified this effect further.

Dr. Claire Wardle of First Draft had created a trumpet model to describe the way that bad actors had effectively spread misinformation in the past (see Figure 2).⁶⁴ Conspiracy theories, disinformation, or even well-intentioned misinformation can appear on the anonymous web on forums like 4chan. Usually, this information gains traction by getting picked up by closed networks (i.e. group chats), conspiracy communities (i.e. Reddit), major social media platforms, and then the professional media outlets.

Sometimes, this acceleration is even faster, especially in the COVID-19 pandemic where information is scarce and the public is fearful. We also cannot ignore the impact that public figures have on the news cycle and, consequently, the general public’s consumption of misinformation.⁶⁵ For example, the 5G conspiracy theory was first promulgated by New Agers and QAnon followers.⁶⁶ Since then, celebrities have spread this information to their millions of

⁶¹ Kate Gibson, “Feds Order Alex Jones to Stop Selling Phony Coronavirus Cures,” *CBS News*, April 10, 2020
<https://www.cbsnews.com/news/alex-jones-infowars-accused-fake-coronavirus-cures-04-10-2020/>

⁶² Clea Skopeliti, “How Mental Health Subreddits Are Coping With The Coronavirus Infodemic,” *First Draft News*, April 3, 2020,
https://firstdraftnews.org/latest/how-mental-health-subreddits-are-coping-with-the-coronavirus-infodemic/?fbclid=IwAR0mVvHE10JZvp8z3Biak0xg6OtKuhJFBE_veX64ZUibiHyKyxt8Wk-i9vg

⁶³ Ariel Bogle, “5G And Anti-Vax Conspiracy Theorists Are Exploiting The Coronavirus Crisis,” *Australian Broadcasting Corporation*, April 14, 2020,
<https://www.abc.net.au/news/science/2020-04-15/coronavirus-5g-vitamin-c-anti-vaccine-conspiracy-theories-spread/12145096>

⁶⁴ Dr. Claire Wardle, “5 Lessons for Reporting in an Age of Disinformation,” *First Draft News*, December 27, 2018,
<https://firstdraftnews.org/latest/5-lessons-for-reporting-in-an-age-of-disinformation/>

⁶⁵ Fergal Gallagher, “Tracking Hydroxychloroquine Misinformation: How An Unproven COVID-19 Treatment Ended Up Being Endorsed By Trump,” *ABC News*, April 22, 2020
<https://abcnews.go.com/Health/tracking-hydroxychloroquine-misinformation-unproven-covid-19-treatment-ended/story?id=70074235>

⁶⁶ Harmeet Kaur, “The Conspiracy Linking 5G To Coronavirus Just Will Not Die,” *CNN*, April 9, 2020
<https://www.cnn.com/2020/04/08/tech/5g-coronavirus-conspiracy-theory-trnd/index.html>

followers on social media.⁶⁷ The story had been picked up by news outlets such as CNN⁶⁸ and BuzzFeed News,⁶⁹ and the more journalists covered the story, the more it fanned the flame. Even news stories debunking the conspiracy increased people’s awareness of the theory and magnified the original reach of the QAnon posters. Thus, the bad actors on the anonymous web got what they wanted.

4. Traditional News Media

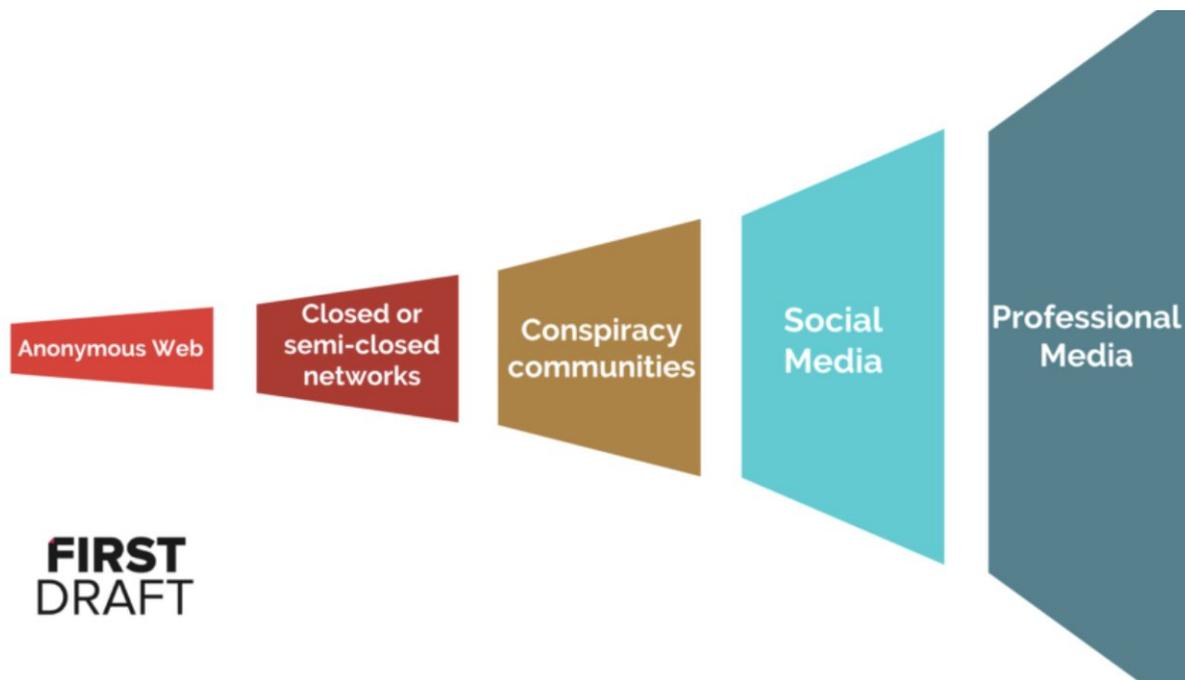


Figure 2. The ‘Trumpet of Amplification.’ This image describes the intentional spread of misinformation by bad actors. Disinformation created those with the malicious intent to deceive can be amplified through different news sources. When bad actors spread disinformation on the anonymous web or on conspiracy websites, they are often aiming to have the false information picked up by the media platforms and outlets with broader reach. Thus, while websites on their own may not have significant reach or impact, their impact increases when media outlets amplify and legitimize the information without first properly vetting the information. Note that this model of amplification looks slightly different to the model proposed by Benkler; see below.

⁶⁷ Solis, Jorge. “All the Celebrities Who Are Worried about the Side Effects of Using 5G and What They’ve Said about It.” *Newsweek*, 13 Apr. 2020, www.newsweek.com/celebrities-who-are-worried-about-side-effects-using-5g-what-theyve-said-about-it-1497546.

⁶⁸ *Ibid.*

⁶⁹ Ryan Broderick, “A Conspiracy Theory That 5G Is Causing The Coronavirus Is Spreading Alongside The Pandemic,” *Buzzfeed*, April 3, 2020 <https://www.buzzfeednews.com/article/ryanhatesthis/conspiracy-theory-5g-coronavirus-qanon>

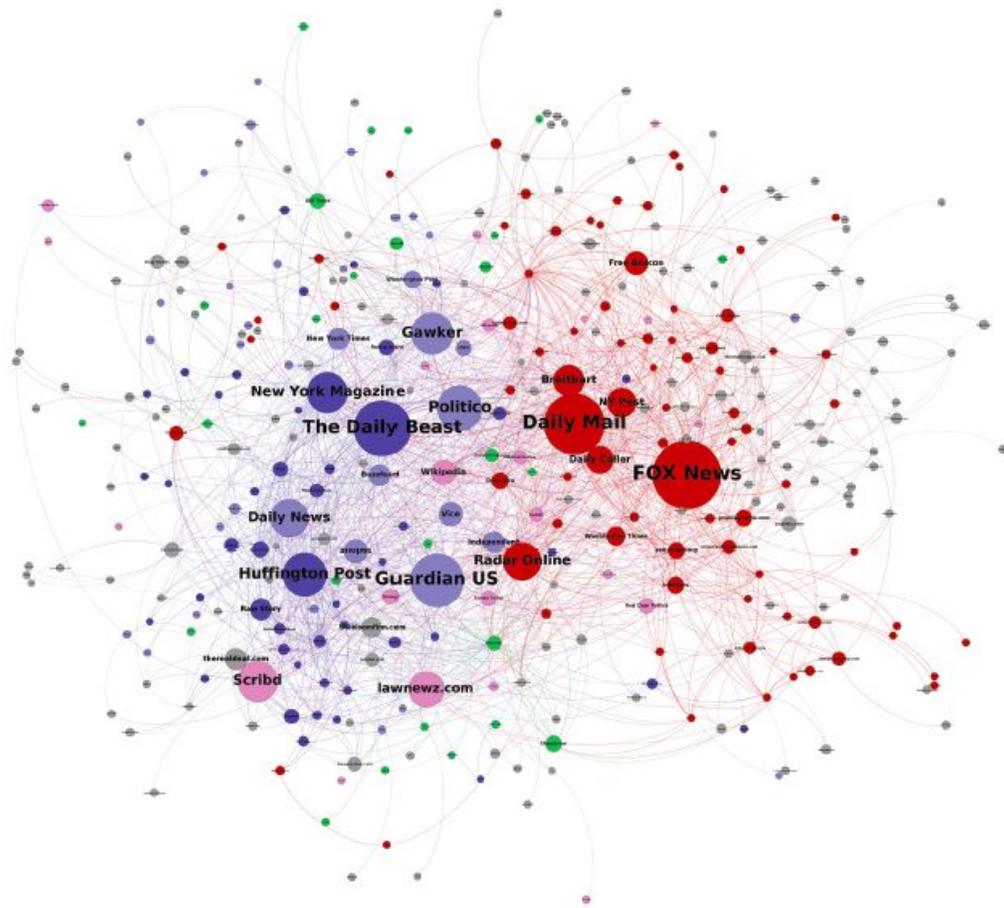


Figure 3. Model of Amplification. The model of amplification proposed by Benkler et al. rejects the bottom-up, trumpet approach.⁷⁰ Benkler et al. claims that right-wing narratives and disinformation are amplified by “credible” sites like Fox News and Daily Mail. In other words, Top media on the right operates differently from the rest of the media ecosystem because they participate in publishing propaganda and disinformation that is usually relegated to the peripheries. In this case, the network map reflects media sources reporting on “Trump rape” and “Clinton pedophilia” stories from 2015-2016.

The large demand for reliable and authoritative COVID-19 information has coincided with a significant increase in news consumption. Polling by the Reuters Institute found that news use has increased among respondents in all six countries that were studied.⁷¹ In a separate study, the Edelman Trust Barometer found that 70% of poll respondents follow COVID-19 news at least once a day (61% in the United States alone).⁷² The BBC News website, for example, has

⁷⁰ Benkler, Yochai, et al. “The Propaganda Feedback Loop.” *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Oxford University Press, 2018.

⁷¹ J. Scott Brennan, Felix Simon, Philip N. Howard, and Rasmus Kleis Nielsen, “Types, sources, and claims of COVID-19 misinformation,” (2020), <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.

⁷² Edelman Trust Barometer, “Special Report: Trust and the Coronavirus,” (2020), available at https://www.edelman.co.uk/sites/g/files/aatuss301/files/2020-03/2020%20Edelman%20Trust%20Barometer%20Coronavirus%20Special%20Report_0.pdf.

seen a huge uptick in traffic during the pandemic.⁷³ Thus, news organizations are positioned to play a large role in the COVID-19 infodemic.

Following journalistic standards is vital for mitigating the harms of the infodemic. However, news organizations face many obstacles in combating the spread of COVID-related misinformation. First, the unprecedented scale of the infodemic has pushed the fact-checking abilities of news organizations to the limit. Back in February 2020, the #CoronaVirusFacts alliance had already published over 500 fact-checks on COVID-19 claims.⁷⁴ The Reuters Institute found that between January and March, the number of English-language fact checks increased by more than 900%.⁷⁵ Professor Carley of Carnegie Mellon University's Center for Informed Democracy & Social-cybersecurity (IDEaS) has observed that, compared with previous misinformation events such as elections and natural disasters, she has seen “way, way more” misinformation related to COVID-19.⁷⁶

Second, news organizations are far outpaced by social media in the dissemination of information. As a 2018 study out of MIT suggests, false news is capable of spreading farther and faster than truth can.⁷⁷ Further, Scott Brennan of the Reuters Institute cautions that social media allows a small number of influential figures with large followings to have an outsize impact on the public discourse around COVID-19.⁷⁸ Thus, news organizations dealing with the COVID-19 infodemic are forced to play catch-up in countering misinformation spread.⁷⁹

Given the enormous amount of uncertainty about the COVID-19 and its status as an ever-developing story, even reputable news organizations are also prone to mistakes. For example, Reuters published a story stating that a company named Bodysphere had gotten FDA

⁷³ Amol Rajan, “Coronavirus and a fake news pandemic,” (2020), <https://www.bbc.com/news/entertainment-arts-51858555>.

⁷⁴ Cristina Tardáguila, “No race or religion can prevent coronavirus — don’t fall for these hoaxes,” (2020), <https://www.poynter.org/fact-checking/2020/no-race-or-religion-can-prevent-coronavirus-dont-fall-for-these-hoaxes/>

⁷⁵ J. Scott Brennan, Felix Simon, Philip N. Howard, and Rasmus Kleis Nielsen, “Types, sources, and claims of COVID-19 misinformation,” (2020), <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.

⁷⁶ Liz Reid, “There Is 'Way, Way More' Disinformation Related To Coronavirus Compared To Other Events,” (2020), <https://www.wesa.fm/post/there-way-way-more-disinformation-related-coronavirus-compared-other-events#stream/0>.

⁷⁷ Soroush Vosoughi, Deb Roy, and Sinan Aral, “The spread of true and false news online.” *Science*, vol. 359, no. 6380, 2018, pp. 1146-1151, doi:10.1126/science.aap9559.

⁷⁸ Jim Waterson, “Influencers among 'key distributors' of coronavirus misinformation,” (2020), <https://www.theguardian.com/media/2020/apr/08/influencers-being-key-distributors-of-coronavirus-fake-news#maincontent>.

⁷⁹ John Crowley, “Tackling misinformation during Covid-19: a journalistic and ethical imperative,” (2020), <https://ethicaljournalismnetwork.org/tackling-misinformation-during-covid-19-an-journalistic-and-ethical-imperative>

approval for a coronavirus test, which they had picked up from a Bodysphere press release issued on Business Wire.⁸⁰ After Bodysphere said that the press release was inaccurate, Reuters removed the story and issued a retraction.⁸¹ Although the retraction was issued quickly, the elevated consumption of news media during the pandemic is likely to amplify the impact of such errors.

Furthermore, well-intentioned attempts by journalists to address the info can extend the reach of misinformation. Claire Wardle of First Draft cautions that, even if peddlers of false or misleading information fail to have their claims endorsed by traditional news media, the act of debunking false or misleading claims itself increases the reach of those claims.⁸² Whitney Phillips refers to this phenomenon as the “oxygen of amplification.”⁸³ These issues have driven news organizations such as The Poynter Institute,⁸⁴ First Draft,⁸⁵ the Ethical Journalism Network⁸⁶ and Scientific American⁸⁷ to propose new ethical guidelines on responsible COVID-19 reporting. It remains to be seen whether these measures will be widely adopted and have a salutary impact on the infodemic.

⁸⁰ Sidney Smith, “Reuters retracts Coronavirus FDA test story. Was it April Fools’ hoax?,” (2020), <https://www.imediaethics.org/reuters-retracts-coronavirus-fda-test-story-was-it-april-fools-hoax/>.

⁸¹ Reuters, “ADVISORY-Story on FDA authorizing new test kit for coronavirus withdrawn,” (2020), <https://news.yahoo.com/u-fda-authorizes-two-minute-122110210.html>.

⁸² Claire Wardle, “5 Lessons for Reporting in an Age of Disinformation,” (2018), https://firstdraftnews.org/latest/5-lessons-for-reporting-in-an-age-of-disinformation/?fbclid=IwAR0-AQUsxxCkhFevR1DWNe917HGV5Mm-d086mtAv08PS18rvDE_zB6XMHzQ

⁸³ Whitney Phillips, “The Oxygen of Amplification,” (2018), <https://datasociety.net/library/oxygen-of-amplification/>.

⁸⁴ Al Tompkins, “How newsrooms can tone down their coronavirus coverage while still reporting responsibly,” (2020), <https://www.poynter.org/reporting-editing/2020/how-newsrooms-can-tone-down-their-coronavirus-coverage-while-still-reporting-responsibly/>.

⁸⁵ First Draft, “Coronavirus: Responsible reporting and ethics,” (2020), <https://firstdraftnews.org/long-form-article/coronavirus-responsible-reporting-and-ethics/>

⁸⁶ Hannah Storm, “Media ethics, safety and mental health: reporting in the time of Covid-19,” (2020), <https://ethicaljournalismnetwork.org/media-ethics-safety-and-mental-health-reporting-in-the-time-of-covid-19>.

⁸⁷ Bill Hanage, “How to Report on the COVID-19 Outbreak Responsibly,” (2020), <https://blogs.scientificamerican.com/observations/how-to-report-on-the-covid-19-outbreak-responsibly/>.

C. Discernible Impacts of COVID-19 Misinformation⁸⁸

While researchers have had success tracing the spread of COVID-19 misinformation online and offline, identifying and documenting the real-world impacts of misinformation has proven much more difficult. The virus situation is a fluid, developing matter in which the credibility of information fluctuates. With that in mind, maintaining a mindfulness of what is labeled as “misinformation” by fact-checkers, and what impacts are recorded is crucial.⁸⁹ In our summary of existing work and proposals for future study, we lay out three categories for the impact of misinformation—on actual practices, on attitudes and beliefs, and on media coverage—each of which presents distinct research challenges.

1. Effect of COVID-19 Misinformation on Actual Practices⁹⁰

While discovery of the impact that misinformation has had on actual practices is arguably most important for diagnosing the severity of the problem, it is also the area of impact that is most difficult to study. Linking misinformation exposure to actual behaviors such as social distancing presents practical and ethical challenges for researchers. Among the foremost work in this area is a working paper “Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic” published by a team of economists from Stanford and Harvard.⁹¹ This paper uses aggregated smartphone location data from SafeGraph to model the effect of partisanship on social distancing, concluding that, in March 2020, people in areas with more Republicans were less likely to engage in social distancing, controlling for other factors.⁹² While it is far from clear that partisan affiliation is predictive of one’s exposure to or belief in misinformation, this paper offers a useful template for future research. Although the use of SafeGraph data poses limitations—the data is available by census tract and not individually identifiable—it does provide a rich dataset that researchers could use if exposure to misinformation could be indexed by geographic area.

More specific studies of misinformation’s effect on actual behavior would rely on

⁸⁸ Many of the observations in this section are the product of discussions with Professor Yochai Benkler, whose insights were invaluable in understanding the impact of COVID-19 misinformation and potential areas for further study.

⁸⁹ Consider that, while viral conspiracies are addressed, low-level scams taking advantage of uninformed, guidance-seeking individuals may not garner significant attention, but nonetheless are impactful. *See, e.g.*, Witt, Paul. “COVID-19 Scam Reports, by the Numbers.” *Consumer Information*, Federal Trade Commission, 15 Apr. 2020, www.consumer.ftc.gov/blog/2020/04/covid-19-scam-reports-numbers.

⁹⁰ “Actual practices” meaning identifiable activities, e.g., increased facemask wearing.

⁹¹ Allcott, Hunt, et al. “Polarization and Public Health: Partisan Differences in Social Distancing during the Coronavirus Pandemic.” *NBER Working Paper* w26946 (2020).

⁹² “Stopping COVID-19 with New Social Distancing Dataset.” *SafeGraph*, SafeGraph Inc., 1 Apr. 2020, www.safegraph.com/blog/stopping-covid-19-with-new-social-distancing-dataset.

self-reporting and may bump up against privacy concerns. For example, YouGov has begun asking respondents about both consumption of COVID-related information and changed behaviors, including working from home and wearing face-masks.⁹³ In a similar vein, FiveThirtyEight has created a live COVID-related poll tracker, which contains a slew of polling measures that might be indexed to measures of misinformation.⁹⁴ Some researchers are also conducting their own surveys to ask respondents about their exposure to misinformation. For example, as documented in the working paper “Misinformation During a Pandemic,” University of Chicago researchers found that age 55+ viewers of *Tucker Carlson Tonight*—which is credited with treating the virus threat seriously from an early stage—on average adopted pro-mitigation behaviors like hand-washing and social distancing five days earlier than a comparable sample of viewers of *Hannity*, a show which has consistently downplayed the significance of the virus.⁹⁵

However, self-reported data—especially about health-related behaviors—raise serious reliability questions.⁹⁶ Yet, alternative, more reliable forms of data collection pose grave privacy concerns. For example, platforms like Facebook collect data on users who have clicked on, shared, or been exposed to misinformation about COVID-19. This data could be indexed to location data from mobile phone companies or third-party applications. However, that process would require linking datasets using individually identifiable information and open up a Pandora’s box of privacy problems that make such research unrealistic.

Nevertheless, there is some evidence of discrete impacts that COVID-19 misinformation has had at this point in time. Most notably, the virus has seen xenophobia and racism—charged by misinformation, disinformation, and consequently the lack of accurate information—translate into an increase in hate crimes against Asian Americans,⁹⁷ and overt anti-black sentiments in China.⁹⁸ To track the former, the Asian Pacific Policy and Planning Council (A3PCON), Chinese for Affirmative Action (CAA), and San Francisco State University’s Asian American Studies

⁹³ “The Economist/YouGov Poll March 15 - 17, 2020 - 1500 US Adult Citizens.” YouGov PLC, 17 Mar. 2020.

⁹⁴ “How Americans View The Coronavirus Crisis And Trump's Response.” *FiveThirtyEight*, 23 Apr. 2020, projects.fivethirtyeight.com/coronavirus-polls/.

⁹⁵ Bursztyn, Leonardo and Rao, Aakaash and Roth, Christopher and Yanagizawa-Drott, David, *Misinformation During a Pandemic* (April 19, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-44. Available at SSRN: <https://ssrn.com/abstract=3580487> or <http://dx.doi.org/10.2139/ssrn.3580487>

⁹⁶ Short, Meghan E et al. “How accurate are self-reports? Analysis of self-reported health care utilization and absence when compared with administrative data.” *Journal of occupational and environmental medicine* vol. 51,7 (2009): 786-96. doi:10.1097/JOM.0b013e3181a86671

⁹⁷ Yan, Holly, et al. “What’s Spreading Faster than Coronavirus in the US? Racist Assaults and Ignorant Attacks against Asians.” *CNN*, Cable News Network, 21 Feb. 2020, www.cnn.com/2020/02/20/us/coronavirus-racist-attacks-against-asian-americans/index.html.

⁹⁸ Lee, Kristine, and Simon Marks. “Coronavirus Ends China's Honeymoon in Africa.” *POLITICO*, 16 Apr. 2020, www.politico.com/news/2020/04/16/coronavirus-china-africa-191444.

Department have created an online reporting center called STOP AAPI Hate.⁹⁹ The express purpose of the site is to identify and provide aid to impacted communities; as of April 3rd, there have been over 1,100 reported incidents of COVID-related discrimination.¹⁰⁰ For the latter, Axios has detailed how misinformation-bred fear of Africans as carriers of the virus has led to discriminatory quarantine by force, evictions, and bans from certain public areas. COVID-related racial discrimination suggests that the dissemination of inaccurate information has had an anger-fueling and anger-orienting impact.¹⁰¹ In addition, as previously noted, there are baseless conspiracy theories purporting links between 5G signals and the virus' origination and spread,¹⁰² which have led to attacks on 5G masts in the United Kingdom.¹⁰³ Finally, promotion of hydroxychloroquine as a treatment—the efficacy of which has not been verified—has contributed to at least one death,¹⁰⁴ numerous overdoses, and reported shortages of the drug for lupus patients that rely on it.¹⁰⁵

2. Effect of COVID-19 Misinformation on Attitudes and Beliefs

Existing research on the effects of misinformation on attitudes and beliefs has illuminated what sorts of messages are getting through (see Figure 1) and through what media that information is being communicated (see Figure 2). One of the most interesting findings thus far comes a Pew Research Data poll on beliefs concerning the virus' origin.¹⁰⁶ Given the characterization of the spread of misinformation outlined above, it is unsurprising that those who mostly get their news from social media (27%) are more likely than those who rely on print

⁹⁹ “Asian American Pacific Islander (AAPI) Civil Rights Organizations Establishes STOP AAPI HATE Reporting Center.” *Stop AAPI Hate*, Asian Pacific Policy and Planning Council, www.asianpacificpolicyandplanningcouncil.org/asian-american-pacific-islander-aapi-civil-rights-organizations-establishes-stop-aapi-hate-reporting-center/.

¹⁰⁰ “STOP AAPI HATE Receives over 1,100 Incident Reports of Verbal Harassment, Shunning and Physical Assault in Two Weeks.” Asian Pacific Policy and Planning Council, 3 Apr. 2020.

¹⁰¹ Allen-Ebrahimian, Bethany, and Dave Lawler. “Complaints of Racism against Africans Mar China's Image.” *Axios*, 15 Apr. 2020,

www.axios.com/coronavirus-racism-discrimination-africa-china-326de397-a3f9-4993-9c8b-1dd303b593db.html.

¹⁰² Temperton, James. “How the 5G Coronavirus Conspiracy Theory Tore through the Internet.” *WIRED*, WIRED UK, 9 Apr. 2020, www.wired.co.uk/article/5g-coronavirus-conspiracy-theory.

¹⁰³ Hamilton, Isobel Asher. “Vandals Set 50 Cellphone Masts in the UK on Fire Because of a Conspiracy Theory Linking the Coronavirus with 5G.” *Business Insider*, Business Insider, 15 Apr. 2020, www.businessinsider.com/attacks-cellphone-towers-coronavirus-5g-conspiracy-2020-4.

¹⁰⁴ Beaumont, Peter, and Rebecca Ratcliffe. “Chloroquine: Trump's Misleading Claims Spark Hoarding and Overdoses.” *The Guardian*, Guardian News and Media, 25 Mar. 2020, www.theguardian.com/science/2020/mar/25/can-chloroquine-really-help-treat-coronavirus-patients.

¹⁰⁵ “MARCH 26 STATEMENT: Lupus Research Alliance Puts Plaquenil (Hydroxychloroquine) Shortage into Perspective for Lupus Community.” *Lupus Research*, Lupus Research Alliance, 26 Mar. 2020, www.lupusresearch.org/march-26-statement-lupus-research-alliance-puts-plaquenil-hydroxychloroquine-shortage-in-to-perspective-for-lupus-community/.

¹⁰⁶ “Election News Pathways 2020 Data Tool.” *Pew Research Center*, Pew Research Center, www.pewresearch.org/pathways-2020/covidcreate/platform_used_for_news/us_adults/.

(17%) or cable TV (22%) to believe the widely discredited theory that the virus was intentionally developed in a lab.¹⁰⁷ What is somewhat surprising is the age breakdown of responses to that same question—27% of those between 18-29 believed the theory, while only 15% of those 65 and over did. This finding flies in the face of previous research that finds the problem of misinformation sharing and consumption is much more acute among older populations.¹⁰⁸ A concerning implication of these findings is that an individual’s primary media source may render one more susceptible to believing misinformation and misconceiving the virus situation. While much more research is needed to buttress these findings, this data suggests that COVID-19 misinformation may be spreading differently—such as via Instagram or via text message—than previous strains of misinformation, such as fake news about the 2016 election spread on Facebook.

3. Effect of COVID-19 Misinformation on Media Coverage

In their 2018 book *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*, Yochai Benkler, Robert Faris, and Hal Roberts criticized a misplaced focus on actors like Russia’s Internet Research Agency and platforms like Facebook in evaluations of misinformation in the 2016 election.¹⁰⁹ Rather, based on exhaustive study of the publication and dissemination of four million news stories published between 2015 and 2017, they concluded that a right-wing media ecosystem—including Fox News, Breitbart, Daily Caller, and others—played a much larger role in shaping beliefs and contributing to hyper-polarized conceptions of reality during the 2016 election. These findings—along with data showing that consumers of Fox News (30%) are more likely to believe the virus was intentionally developed in a lab than consumers of sources like CNN (18%) or *The New York Times* (5%)—indicate that a more fruitful study of misinformation’s impacts requires understanding its interaction with media coverage and the right-wing media ecosystem in particular.¹¹⁰ This focus is particularly important given President Trump’s frequent amplification and adoption of theories promulgated in this ecosystem.

An instructive example of this phenomenon is the coverage of hydroxychloroquine as a potential treatment for COVID-19. Importantly, the discussions of the drug’s potential as a treatment did not start on the fringe, but emerged online following China’s inclusion of

¹⁰⁷ Barclay, Eliza. “The Conspiracy Theories about the Origins of the Coronavirus, Debunked.” *Vox*, Vox, 4 Mar. 2020, www.vox.com/2020/3/4/21156607/how-did-the-coronavirus-get-started-china-wuhan-lab.

¹⁰⁸ Guess, Andrew, Jonathan Nagler, and Joshua Tucker. “Less than you think: Prevalence and predictors of fake news dissemination on Facebook.” *Science advances* 5.1 (2019): eaau4586.

¹⁰⁹ Benkler, Yochai, Robert Faris, and Hal Roberts. *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press, 2018.

¹¹⁰ “Election News Pathways 2020 Data Tool.” *Pew Research Center*, Pew Research Center, www.pewresearch.org/pathways-2020/covidcreate/platform_used_for_news/us_adults/.

Chloroquine Phosphate in treatment guidelines in February¹¹¹ and the release of two¹¹² studies¹¹³ on hydroxychloroquine's effectiveness in mid-March. Subsequently, many Fox News personalities including Laura Ingraham and Sean Hannity began aggressively promoting the drug's effectiveness.¹¹⁴ A few days later, President Trump falsely claimed that the FDA had approved chloroquine as a treatment for COVID-19 and shortly thereafter met with Ingraham and two guests from her show to discuss the drug's effectiveness.¹¹⁵ In quick succession, despite widespread expert skepticism about the efficacy of chloroquine and its relative hydroxychloroquine as effective treatments,¹¹⁶ discussions about the drug moved from niche online and offline conversations within the medical community, morphed into definitive statements about their beneficial use on Fox News, and soon became proclamations from the White House concerning their efficacy. Following a report published by medical journal *The Lancet*, finding an increased mortality rate among patients treated with chloroquine and hydroxychloroquine,¹¹⁷ safety concerns prompted the WHO to temporarily suspend clinical trials of the two drugs on May 25; upon further review, trials resumed on June 3.¹¹⁸ The efficacy of the drugs for treating COVID-19 remains in question.

Another useful example of the interaction between online message boards, right-wing media sources, and conservative politicians is the spread of misinformation about the virus' origin. Rumors that the virus was developed as a bioweapon first gained traction in late January

¹¹¹ "Antimalarial Drug Confirmed Effective on COVID-19." *Xinhua*, Xinhua Net, 17 Feb. 2020, www.xinhuanet.com/english/2020-02/17/c_138792545.htm.

¹¹² Gautret, Philippe, et al. "Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial." *International journal of antimicrobial agents* (2020): 105949.

¹¹³ Liu, Jia, et al. "Hydroxychloroquine, a Less Toxic Derivative of Chloroquine, Is Effective in Inhibiting SARS-CoV-2 Infection in Vitro." *Nature News*, Nature Publishing Group, 18 Mar. 2020, www.nature.com/articles/s41421-020-0156-0.

¹¹⁴ Gogarty, Kayla, et al. "A Comprehensive Guide to Fox's Promotion of Hydroxychloroquine and Chloroquine." *Media Matters for America*, 16 Apr. 2020, www.mediamatters.org/coronavirus-covid-19/comprehensive-guide-foxs-promotion-hydroxychloroquine-and-chloroquine.

¹¹⁵ Dale, Daniel. "Fact Check: Trump Wrongly Claims FDA 'Approved' Drug Chloroquine to Treat the Coronavirus." *CNN*, Cable News Network, 20 Mar. 2020, edition.cnn.com/2020/03/19/politics/fact-check-chloroquine-trump-fda/index.html.

¹¹⁶ Irfan, Umair. "The Evidence for Using Hydroxychloroquine to Treat Covid-19 Is Flimsy." *Vox*, Vox, 7 Apr. 2020, www.vox.com/2020/4/7/21209539/coronavirus-hydroxychloroquine-covid-19-clinical-trial.

¹¹⁷ Mehra, Mandeep R., et al. "Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis." *The Lancet* (2020).

¹¹⁸ Lovelace, Berkeley, and William Feuer. "World Health Organization Resumes Coronavirus Trial on Malaria Drug Hydroxychloroquine after Examining Safety Concerns." *CNBC*, CNBC, 3 June 2020, www.cnbc.com/2020/06/03/world-health-organization-resumes-coronavirus-trial-on-malaria-drug-hydroxychloroquine-after-safety-concerns.html.

in *The Washington Times*.¹¹⁹ Around the same time, memes alleging the virus was produced in a Wuhan laboratory began circulating the Internet,¹²⁰ right-wing blog Zero Hedge accused a Chinese scientist of being responsible for the virus' spread,¹²¹ and many other sources, including Steve Bannon's podcast *War Room: Pandemic* started giving the theory air.¹²² These theories were then amplified by Senator Tom Cotton during a February 16 appearance on Fox News, who alleged that the virus may have originated in a laboratory in Wuhan.¹²³ This speculation became increasingly inflammatory as Rush Limbaugh, among others, proclaimed in late February that the virus was likely the product of a Chinese laboratory.¹²⁴ As this theory has continued to gain traction—even reaching mainstream publications like the *Washington Post*¹²⁵—it has left an imprint on many Americans, an estimated 23% of whom believe the virus was intentionally developed in a laboratory.¹²⁶

4. Potential Interventions for Gauging the Impact of COVID-19 Misinformation

More sophisticated, invasive technological interventions would illuminate and more effectively capture the impact of misinformation concerning COVID-19, but the storied and current state of play in the digital space may impede their adoption. As outlined above, misinformation is incredibly pervasive and, while there are brilliant researchers the world over diligently seeking to ascertain the effects of its spread, the gravity of its impact remains unclear. Additional insight about the information landscape would expedite the process of attaining a

¹¹⁹ Gertz, Bill. "Coronavirus May Have Originated in Lab Linked to China's Biowarfare Program." *The Washington Times*, The Washington Times, 26 Jan. 2020,

www.washingtontimes.com/news/2020/jan/26/coronavirus-link-to-china-biowarfare-program-poss/.

¹²⁰ Evon, Dan. "Is the 'Umbrella Corporation' Logo Oddly Similar to a Wuhan Biotech Lab's?" *Snopes.com*, Snopes Media Group, 29 Jan. 2020, www.snopes.com/fact-check/resident-evil-umbrella-coronavirus/.

¹²¹ Peters, Jay. "Markets Blogger Zero Hedge Suspended from Twitter after Doxxing a Chinese Scientist." *The Verge*, The Verge, 1 Feb. 2020, www.theverge.com/2020/1/31/21117663/twitter-zero-hedge-suspended-platform-manipulation-policy-doxxing-coronavirus.

¹²² Peltz, Madeline. "Steve Bannon Is Pushing Debunked Claims about the Coronavirus That Are Linked to His Billionaire Benefactor." *Media Matters for America*, 6 Feb. 2020, www.mediamatters.org/steve-bannon/steve-bannon-pushing-debunked-claims-about-coronavirus-are-linked-his-billionaire.

¹²³ Senator Tom Cotton. "February 16, 2020: Senator Cotton Joins Sunday Morning Futures with Maria Bartiromo." Online video clip. *Youtube*, 16 February 2020. Web.

https://www.youtube.com/watch?time_continue=1&v=ytGIkcCh7T8&feature=emb_title

¹²⁴ Staff, Media Matters. "Rush Limbaugh: 'The Coronavirus Is an Effort to Get Trump.'" *Media Matters for America*, 24 Feb. 2020, www.mediamatters.org/rush-limbaugh/rush-limbaugh-coronavirus-effort-get-trump.

¹²⁵ Ignatius, David. "Opinion | How Did Covid-19 Begin? Its Initial Origin Story Is Shaky." *The Washington Post*, WP Company, 2 Apr. 2020, www.washingtonpost.com/opinions/global-opinions/how-did-covid-19-begin-its-initial-origin-story-is-shaky/2020/04/02/1475d488-7521-11ea-87da-77a8136c1a6d_story.html.

¹²⁶ "Election News Pathways 2020 Data Tool." *Pew Research Center*, Pew Research Center, www.pewresearch.org/pathways-2020/covidcreate/main_source_of_election_news/us_adults/.

better understanding of the effects of COVID-related misinformation thus far, as well as inform forecasting about potential future, second-order effects. An increased cognizance of the misinformation spreading, the manner in which it spreads, and the consequences of its spread would enable online platforms and services to implement preventative measures to stem the misinformation flow, and promote verified, accurate information. With more accurate information comes clarity about the virus: what is known, what needs to be discovered, and how, on both macro- and micro-levels, the virus situation can be managed, and ultimately resolved. More data can be a source of greater clarity, but at what cost?

Imagine a no-holds-barred world in which privacy and freedom of speech concerns were tabled and discerning the impact of COVID-related misinformation was of paramount importance. What interventions could be used to most effectively pinpoint misinformation? One possibility is a comprehensive “digital contact tracing” system.¹²⁷ Paralleling contact tracing,¹²⁸ through interorganizational collaboration predicated on the free exchange of user data between cell carriers, social media platforms, search databases, and the CDC, online and offline activity data would be indexed to paint a more vivid picture of misinformation trails and their corollaries. For example, the system would be able to capture the timeline of a user’s consumption of content downplaying the severity of the virus situation, subsequent sharing of such content to their social networks, the responses of the user to online surveys expressing a lack of concern about the virus, and the user’s disregard of social distancing via location data monitoring. All of this information would be available to the collaborating entities, serving to guide policy interventions to emphasize the legitimacy and necessity of social distancing to that user, and like individuals. Although indicative of correlation more so than causation, data about sequential happenings would aid the effort to ascertain the impact of misinformation by identifying associated acts, self-reported attitudes, and media consumption.

Yet, despite the insightfulness of the digital contact tracing system detailed above, such a system for dealing with COVID-related misinformation could be problematic in the real world in light of considerations regarding privacy and freedom of speech. However, interventions that are

¹²⁷ The hypothetical “comprehensive digital contact tracing” system envisioned here is an integration application that monitors all mobile phone activity and accesses GPS location data. The app would be installed via an automatic cell software update, and have notification permission to regularly provide verified COVID-related information from the CDC. There would be no opt-out provision, so as to promote data collection, insight, and accurate information dissemination. Note that the “comprehensive digital contact tracing” system presented here differs from traditional, COVID-transmission-focused digital contact tracing that is increasingly being considered. *See, e.g.,* Zittrain, Jonathan. “Entering the Minefield of Digital Contact Tracing.” *Medium*, Berkman Klein Center Collection, 11 May 2020, medium.com/berkman-klein-center/entering-the-minefield-of-digital-contact-tracing-9c042941bb23. (evaluating the prospect of digital contact tracing).

¹²⁸ For a general overview of contact tracing, see: “Contact Tracing : Part of a Multipronged Approach to Fight the COVID-19 Pandemic.” *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 22 Apr. 2020, www.cdc.gov/coronavirus/2019-ncov/php/principles-contact-tracing.html.

similarly insightful yet more sensitive to the complexities of digital governance are possible and should be pursued.¹²⁹ Conceptions of digital governance have evolved to consider the rights of users, the public health impacts of the digital landscape's design, and the impact of systemic interventions upon both rights and health.¹³⁰ Through a framework mindful of user rights and public health, a balancing of a digital contact tracing system's treatment of data—which presents data privacy, data security, freedom of speech, and transparency concerns—with the perceived value of the information able to be attained is in order. That balance favoring the collective good over the rights of an individual's rights, especially in light of an emergency situation, is not a novel concept.¹³¹ Understanding the impact of COVID-related misinformation is important, but so is how that understanding is achieved.

¹²⁹ See Fairbank, N. A., Murray, C. S., Couture, A., Kline, J., & Lazarro, M. (2020, June 2). *There's An App for That: Digital Contact Tracing and Its Role in Mitigating a Second Wave* [PDF]. The Berkman Klein Center for Internet & Society at Harvard University.

¹³⁰ For a more comprehensive analysis of evolving perspectives on digital governance, see: Zittrain, Jonathan L. "Three Eras of Digital Governance." *Available at SSRN 3458435* (2019).

¹³¹ Phelan, Alexandra. "Explainer: National Emergency Declarations and COVID-19." *Just Security*, Just Security, 2 Apr. 2020, www.justsecurity.org/69190/explainer-national-emergency-declarations-and-covid-19/.

II. Platform Responses to the Infodemic

Given the landscape of COVID-19 misinformation and its spread over the internet and social media, platforms have deployed a number of responses, some carried over from past policies and others wholly new. This section aims to summarize and compare platform responses to the COVID-19 pandemic, with a particular focus on Twitter, Facebook, and YouTube. It concludes by putting forth proposals for how platforms might better respond to the infodemic.

A. Comparing Platform Responses

While there are meaningful differences in platform approaches, Twitter, Facebook, and YouTube have all adopted an active approach toward combating misinformation. This section shows that they have advanced their willingness to collaborate, adopt new products and partnership and implement broad public health policies to address the “infodemic.” In particular, they have engaged in industry collaborations, product interventions, and policy interventions.

		Platforms		
		Twitter	Facebook	YouTube
Responses	Industry Collaboration	✓	✓	✓
	Product Interventions	✓	✓	✓
	Content Policy: Misinformation and Real-World Harm	✓	✓	✓
	Ban on Exploitative Ads	✓	✓	✓
	Content Policy: Conspiracy Theories		✓	✓
	Restrictions on Advertisers	✓		✓
	Connecting Users and Experts Directly		✓	

Figure 4. Platform Responses, Compared. This matrix lists the different responses deployed by social media platforms in light of the pandemic as of April 2020. Checkmarks indicate that a platform had announced new or expanded policies and product features to deal specifically with the pandemic. Companies have been highly aligned across response types, with variations emerging either due to product differences or diverging attitudes on enforcement approaches.

1. Industry Collaboration and Communication

One of the most surprising platform responses throughout the crisis has been the speed with which companies have collaborated and communicated publicly. This is not, however, the first time technology companies have cooperated to address threats. In 2017, after growing concerns about the proliferation of terrorist material on social media, Twitter, Google, YouTube, Facebook, and Microsoft formed the Global Internet Forum to Combat Terrorism (GIFCT).¹³² Part of GIFCT’s vision was to more effectively tackle terrorist content online, envisioning two means of realizing this goal: (1) coordinating platform responses, and (2) supporting smaller companies that may traditionally lack larger technical, human, and financial resources to moderate their platforms through the creation of a hash-sharing database.¹³³

As industries, cities, and countries began quickly shutting down throughout February and March, technology platforms were much faster to react. Seven companies—Facebook, Reddit, Twitter, LinkedIn, Microsoft, Google, and YouTube—published a joint industry statement that expressly called out their efforts to “jointly” address fraud and misinformation and encouraged other companies to join.¹³⁴ This coordination is largely reminiscent of the 2017 terrorism scandals that plagued tech companies, though it is clear platforms have learned from past criticism and are addressing this newest security threat in record time.

Additionally, multiple platforms—chiefly Twitter,¹³⁵ YouTube,¹³⁶ and Facebook¹³⁷—have all attempted to signal¹³⁸ to their users and observers a potential rise in false positives, or erroneous content removals. This proactive communication arose in part because large platforms typically rely on an army of content moderators,¹³⁹ one that was recently sidelined as companies have sent more of their workforces home to slow the spread of the virus. With fewer human reviewers available, algorithms play a larger role in the review and enforcement process.¹⁴⁰

¹³² *Global Internet Forum to Counter Terrorism*, www.gifct.org.

¹³³ “Global Internet Forum to Counter Terrorism: Evolving an Institution.” *Global Internet Forum to Counter Terrorism*, www.gifct.org/about/.

¹³⁴ “Keeping People Safe and Informed About the Coronavirus.” *Facebook*, 21 Apr. 2020, about.fb.com/news/2020/04/coronavirus/#joint-statement.

¹³⁵ “An Update on Our Continuity Strategy during COVID-19.” *Twitter*, Twitter, blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html.

¹³⁶ “Protecting Our Extended Workforce and the Community.” *YouTube Creator Blog*, 16 Mar. 2020, youtube-creators.googleblog.com/2020/03/protecting-our-extended-workforce-and.html.

¹³⁷ Elizabeth Dwoskin, Nitasha Tiku. “Facebook Sent Home Thousands of Human Moderators Due to the Coronavirus. Now the Algorithms Are in Charge.” *The Washington Post*, WP Company, 24 Mar. 2020, www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/.

¹³⁸ Matsakis, Louise. “Coronavirus Disrupts Social Media’s First Line of Defense.” *Wired*, 18 Mar. 2020, www.wired.com/story/coronavirus-social-media-automated-content-moderation/.

¹³⁹ *Ibid.*

¹⁴⁰ Kreps.

2. Product Interventions and Public Health Partnerships

No analysis of platform responses would be complete without also examining the products themselves and the explosion of partnerships with public health organizations. The largest platforms have responded to this crisis not only through their content policies but also with the products themselves to connect users to public health sources, organizations, and experts.

a. Surfacing (and Signaling) Authoritative Sources to Users

The largest platforms all provide links to national organizations focused on health or public health when a search related to COVID-19 takes place. Links to the Centers for Disease Control (or national equivalent in other countries) appear in pop-ups or at the top of search results on Facebook, Instagram, Twitter, and YouTube.¹⁴¹ Google has created a knowledge panel in its search results pages that lists symptoms, statistics, treatment options, and preventative measures.¹⁴² Facebook also guides users to an information center that is displayed at the top of news feeds.¹⁴³

b. Enabling Experts and Organizations to Communicate with Users Directly

Given its unique position as a messaging service, WhatsApp has experimented with connecting experts directly to users. Through its coronavirus hub,¹⁴⁴ users can actually sign up to receive updates from health organizations or even interact with the World Health Organization's chatbot.¹⁴⁵ Reddit, for example, has launched in its popular Ask Me Anything subreddit a recurring series to bring users and public health experts in conversation on a weekly basis.¹⁴⁶ In the same vein, the World Health Organization, the CDC, and other public health organizations have received free advertising rights to communicate to users across platforms, as well.¹⁴⁷

¹⁴¹ Birnbaum, Emily, and Chris Mills Rodrigo. "Social Media Struggles to Counter Coronavirus Misinformation." *The Hill*, 2 Feb. 2020, thehill.com/policy/technology/480987-social-media-struggles-to-counter-coronavirus-misinformation.

¹⁴² Fingas, Jon. *Google Explains How It's Tackling the Coronavirus Outbreak*. Engadget, 6 Mar. 2020, www.engadget.com/2020-03-06-google-coronavirus-information.html.

¹⁴³ Hatmaker, Taylor. *Facebook Will Put a New Coronavirus Info Center on Top of the News Feed*. TechCrunch, 18 Mar. 2020, techcrunch.com/2020/03/18/facebook-coronavirus-information-center-zuckerberg/.

¹⁴⁴ *WhatsApp Coronavirus Information Hub*. www.whatsapp.com/coronavirus/.

¹⁴⁵ *Ibid.*

¹⁴⁶ *Expert Conversation on Coronavirus*. Reddit, 7 Apr. 2020, redditblog.com/2020/03/02/expert-conversation-on-coronavirus/.

¹⁴⁷ Elias, Jennifer. *Google Is Offering \$340 Million in Free Ads for Small Businesses as Part of Coronavirus Help Package*. CNBC, 27 Mar. 2020, www.cnbc.com/2020/03/27/google-offering-800m-coronavirus-help-package.html; *Facebook Gives WHO Free Ads in Battle with Coronavirus Misinformation*. Thomson Reuters, 4 Mar. 2020, www.reuters.com/article/us-health-coronavirus-facebook/facebook-gives-who-free-ads-in-battle-with-coronavirus-misinformation-idUSKBN20R14A.

3. Policy

Platforms have also adopted a number of policy approaches to govern the ways that content is disseminated and monetized. These policies, whether or not developed in response to the pandemic, bear heavily on the platforms' approaches to COVID-19 misinformation.

a. Content Policy

The letter of the platforms' content policies for this crisis are similar in their broad strokes. Many technology companies have augmented their misinformation policies by widening the policies to be more aggressive in addressing misleading content with the capacity either for real-world harm or likelihood of spreading the virus further. For example, videos promoting the consumption of bleach as a way to stay safe fall afoul of these policies.¹⁴⁸ Additionally, conspiracy theories related to the virus, its treatment options, or its origins also appear to be clear violations of platform community guidelines for the companies.¹⁴⁹

While platforms have been criticized in the past for having inconsistent¹⁵⁰—or in some cases, weak—enforcement approaches, it is worth noting the companies have been quite persistent, and sometimes aggressive, with COVID-19 misinformation enforcement. Twitter, for example, has not shied away from removing tweets from politicians,¹⁵¹ celebrities, and public figures¹⁵² that either deny the existence of the virus or guidance from experts or promote medical treatments that are unproven.

A study from the University of Oxford compared the efficacy of Facebook, Twitter, and YouTube's enforcement apparatuses, suggesting that while the letter of the policies might be substantially similar, the spirit (or enforcement) differs significantly.¹⁵³ Twitter was called out in the study as having far more COVID-19 misinformation on its platform compared to the other websites.¹⁵⁴ When asked about these results, a Twitter spokesperson said, "We're prioritizing the

¹⁴⁸ Frenkel, Sheera, et al. *Surge of Virus Misinformation Stumps Facebook and Twitter*. The New York Times, 8 Mar. 2020, www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html.

¹⁴⁹ Thomas, Elise. *As the Coronavirus Spreads, Conspiracy Theories Are Going Viral Too*. Foreign Policy, 14 Apr. 2020, foreignpolicy.com/2020/04/14/as-the-coronavirus-spreads-conspiracy-theories-are-going-viral-too/.

¹⁵⁰ Matsakis, Louise. *YouTube's Content Moderation Is an Inconsistent Mess*. *Wired*, Conde Nast, www.wired.com/story/youtube-content-moderation-inconsistent/.

¹⁵¹ Holroyd, Matthew. "Bolsonaro Busy Market Videos Removed amid Coronavirus Pandemic." *Euronews*, 31 Mar. 2020, www.euronews.com/2020/03/30/twitter-removes-videos-of-jair-bolsonaro-visiting-a-busy-market-during-coronavirus-pandemi.

¹⁵² "Twitter: Laura Ingraham Tweet Broke Rules against Coronavirus Misinformation." *POLITICO*, www.politico.com/news/2020/03/30/twitter-laura-ingraham-coronavirus-misinformation-155731.

¹⁵³ Brennen, Scott, et al. *Types, Sources, and Claims of COVID-19 Misinformation*.

¹⁵⁴ *Ibid.*

removal of content when it has a call to action that could potentially cause harm.”¹⁵⁵ Interestingly, statements from YouTube and Facebook representatives did not parse the types of content that would or would not be removed.¹⁵⁶ The lack of detail in these two companies’ statements suggests that their takedown philosophy is less forgiving than Twitter’s. As the swirl of information about the pandemic spreads online, it will be critical to keep up with content enforcement as the best indicator of how truthful companies are to the policies.

b. Monetization Policy

All three companies have cracked down on coronavirus-related advertising. The platforms have tried to address concerns that advertisements for masks, sanitizer, and other supplies are price gouging.¹⁵⁷ Twitter and YouTube, however, first announced they would forbid the monetization of any coronavirus-related content.¹⁵⁸ Both platforms ultimately softened their stance,¹⁵⁹ allowing advertisements on videos from certain creators that either receive direct support or complete certification programs.¹⁶⁰

Platforms have responded to the infodemic with a range of approaches. The following section compares the ways in which these interventions are different from those adopted in prior crises, analyzes how they may represent a fundamentally new approach to public health crises and considers why this new framework may be relevant for the future.

B. Platform Responses through a “Public Health” Framework

The COVID-19 pandemic is not the first time that platforms have come under scrutiny for the presence of content that threatens public health. Platforms have responded to these developments by adopting a variety of policies intended to provide a consistent framework for dealing with these issues. The sheer scope of the current pandemic and related misinformation, however, has challenged platforms’ ability to tailor their global responses to policies designed for what they perceived as narrow exceptions to their rules. As Evelyn Douek observed in *The Atlantic*, these narrow “public health” exceptions to platforms’ otherwise permissive content

¹⁵⁵ Timberg, Craig. *On Twitter, Almost 60 Percent of False Claims about Coronavirus Remain Online - without a Warning Label*. WP Company, 7 Apr. 2020, www.washingtonpost.com/technology/2020/04/07/twitter-almost-60-percent-false-claims-about-coronavirus-remain-online-without-warning-label/.

¹⁵⁶ *Ibid.*

¹⁵⁷ *Facebook Will Ban Certain Ads to Prevent Efforts to Exploit Coronavirus Fears*. The Guardian, 7 Mar. 2020, www.theguardian.com/world/2020/mar/07/facebook-mask-ad-ban-coronavirus.

¹⁵⁸ Birnbaum, Emily. *YouTube to Allow Ads on Coronavirus Videos*. The Hill, 12 Mar. 2020, thehill.com/policy/technology/487297-youtube-to-allow-ads-on-coronavirus-videos.

¹⁵⁹ *Ibid.*

¹⁶⁰ *Ibid.*

rules are expanding to apply in new and varied domains and may represent a fundamental transformation of how platforms view their roles as conduits of information.¹⁶¹

1. Fire in a Crowded Theater: A Public Health Framework

Mark Zuckerberg has likened the current crisis of misinformation to someone “shouting fire in a crowded theater.”¹⁶² He argues that this uniquely dangerous situation justifies his platforms in invoking established guidelines for dealing with public health crises and “our policy that does not allow content that’s going to cause imminent danger or physical risk.”¹⁶³ Similarly, Google has justified its removal of certain applications from the Play Store and its blocking of certain advertisements on COVID-19-related content as enforcing “long-standing content policies” blocking “medical or health-related content or functionalities that are misleading or potentially harmful.”¹⁶⁴

Platforms’ denials that they are changing anything about their misinformation policy—rather than implementing policies for public health crises that were already in place—bear some truth. Facebook, for example, has been criticized for not stepping up in public health crises made worse by false narratives: when the Ebola outbreak of 2018 began in the Democratic Republic of the Congo, the platform was blamed for allowing the spread of mis- and disinformation about the disease, including claims that the disease didn’t exist.¹⁶⁵ And historically, social media has had a bad rap when it comes to providing platforms for anti-vaxxers.¹⁶⁶ But social media platforms had already made shifts toward removal of medical and health-related misinformation months before COVID-19 started making headlines. In mid-2019, Twitter started directing users who searched for anti-vaccination content to the U.S. Department of Health and Human Services.¹⁶⁷ At the time that the COVID-19 pandemic started sweeping the globe, Facebook was a few months into rolling out their plan for addressing anti-vaccination claims.¹⁶⁸ The initiative promised to reduce

¹⁶¹ Douek, Evelyn. “The Internet’s Titans Make a Power Grab.” *The Atlantic*, Atlantic Media Company, 18 Apr. 2020, www.theatlantic.com/ideas/archive/2020/04/pandemic-facebook-and-twitter-grab-more-power/610213/.

¹⁶² Facebook. March 18 2020 Facebook Press Call. *Facebook*. 18 Mar. 2020, <https://about.fb.com/wp-content/uploads/2020/03/March-18-2020-Press-Call-Transcript.pdf>.

¹⁶³ *Ibid.*

¹⁶⁴ “Coronavirus: How We’re Helping.” *Google*, 6 Mar. 2020. <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>.

¹⁶⁵ “Ebola Responders in Congo Confront Fake News and Social Media Chatter.” *The New Humanitarian*, <https://www.thenewhumanitarian.org/news/2019/05/02/ebola-responders-congo-confront-fake-news-and-social-media-chatter>. Accessed 19 Apr. 2020.

¹⁶⁶ Hoffman, Beth L., et al. “It’s Not All about Autism: The Emerging Landscape of Anti-Vaccination Sentiment on Facebook.” *Vaccine*, vol. 37, no. 16, Apr. 2019, pp. 2216–23. *ScienceDirect*, doi:10.1016/j.vaccine.2019.03.003.

¹⁶⁷ Kelly, Makena. “Twitter Fights Vaccine Misinformation with New Search Tool.” *The Verge*, 14 May 2019. <https://www.theverge.com/2019/5/14/18623494/twitter-vaccine-misinformation-anti-vax-search-tool-instagram-face-book>.

¹⁶⁸ “Vaccine Misinformation: Statement by WHO Director-General on Facebook and Instagram.” *World Health Organization*,

rankings of pages and reject advertisements that promoted anti-vaccination messages, while directing users to information from WHO and the CDC.¹⁶⁹ By the time COVID-19 became a recognized global health emergency, social media platforms had already started formulating best practices for addressing medical misinformation as they faced smaller-scale medical emergencies. But they began to implement that response in a much larger and more aggressive way with the pandemic.

2. The Whole World's on Fire: A Global Response to a "Public Health" Crisis

The coronavirus pandemic is transforming the "public health" policies of the major platforms from the exception into the default. As described above, platforms have created "exceptions" from their general open standards for content that allow them to either take down prohibited content or promote posts from "authoritative" voices like the WHO in certain contexts. While these two features are at the core of current responses, platforms appear to have adopted a broader responsibility for the quality and accuracy of the content on their site. For what appears to be the first time, the major platforms are acknowledging a general obligation not only to avoid causing harm but to promote public health.

The application of interventions originally designed for exceptional circumstances to their broader sites suggests that platforms are no longer willing or able to finely tailor their responses to particular public health problems. Facebook has now installed a COVID-19 Information Center on every user's home feed, providing access to verified information before a user seeks it.¹⁷⁰ Previously, users in the United States had to search on Facebook for vaccine-related information to receive a prompt to visit the CDC for relevant information.¹⁷¹ YouTube now offers a similar "news shelf" featuring authoritative sources on its home page in many instances,¹⁷² while Twitter and Google users receive verified information when they conduct relevant searches.¹⁷³ Instead of simply serving their normal role providing users news

<https://www.who.int/news-room/detail/04-09-2019-vaccine-misinformation-statement-by-who-director-general-on-facebook-and-instagram>. Accessed 19Apr. 2020.

¹⁶⁹ "Combating Vaccine Misinformation." *About Facebook*, 7 Mar. 2019, <https://about.fb.com/news/2019/03/combating-vaccine-misinformation/>.

¹⁷⁰ Perez, Sarah. "Facebook Launches a Global Version of Its Community Help Feature in Response to the COVID-19 Pandemic." *TechCrunch*, TechCrunch, 31 Mar. 2020, techcrunch.com/2020/03/31/facebook-launches-a-global-version-of-its-community-help-feature-in-response-to-the-covid-19-pandemic/.

¹⁷¹ Howard, Jacqueline. "Facebook Debuts Vaccine Education Pop-up Windows." *CNN*, Cable News Network, 4 Sept. 2019, www.cnn.com/2019/09/04/health/facebook-vaccine-education-bn/index.html.

¹⁷² Alexander, Julia. "YouTube Will Start Displaying Trustworthy Coronavirus Videos on Its Homepage." *The Verge*, The Verge, 19 Mar. 2020, www.theverge.com/2020/3/19/21187413/youtube-coronavirus-covid-19-shelf-homepage-news-authority-misinformation.

¹⁷³ "Tech Companies Step Up Fight Against Coronavirus Misinformation Online." *Talking Points Memo*, 15 Apr. 2020, talkingpointsmemo.com/news/tech-companies-fight-misinformation-coronavirus-online.

from third parties, platforms have assumed the responsibility for ensuring that their users receive accurate information that aligns with public health objectives.

Moreover, since platforms have viewed the challenge of pandemic misinformation as a general threat, they have been able to adopt policies that affect all users, rather than only those directly impacted. WhatsApp eliminated the ability of users to forward messages to more than one other user.¹⁷⁴ Facebook is removing event pages in the United States promoting protests against social distancing policies¹⁷⁵ and even removed a video from Brazilian President Jair Bolsonaro despite their well-known reticence to censor political posts.¹⁷⁶ Platforms are leveraging their “data for good” programs to voluntarily make user data available as a public health resource.¹⁷⁷ Google and Apple have partnered to create an application for contact tracing with a common API.¹⁷⁸ Particularly in areas where they once tread with sensitivity, the willingness of platforms to prioritize confronting pandemic misinformation suggests they are emphasizing a new overarching goal: how best to promote “public health.”

The response to the spread of pandemic misinformation has raised the prospect that platforms will continue to apply this public health framework for evaluating content and policy decisions even once normal life has resumed.¹⁷⁹ Indeed, the fact that the major platforms all claim to be applying their standard “public health” responses to the current crisis suggests they will face significant pressure to apply the wide range of new interventions to future ones. As discussed below, the motivations that have pushed platforms to adopt an unprecedented response to pandemic misinformation may also militate for continuing their new approach and even extending its reach to new areas.

¹⁷⁴ Hern, Alex. “WhatsApp to Impose New Limit on Forwarding to Fight Fake News.” *The Guardian*, Guardian News and Media, 7 Apr. 2020,

www.theguardian.com/technology/2020/apr/07/whatsapp-to-impose-new-limit-on-forwarding-to-fight-fake-news.

¹⁷⁵ Wong, Julia Carrie. “Facebook Bans Some Anti-Lockdown Protest Pages.” *The Guardian*, Guardian News and Media, 21 Apr. 2020, www.theguardian.com/technology/2020/apr/20/facebook-anti-lockdown-protests-bans.

¹⁷⁶ Biller, Mauricio Savarese and David. “Facebook Deletes Bolsonaro's Posts Over Dismissal of Social Distancing.” *Time*, Time, 1 Apr. 2020, time.com/5813613/facebook-brazil-bolsonaro-coronavirus/.

¹⁷⁷ Newton, Casey. “Tech Giants Are Finding Creative Ways to Use Our Data to Fight the Coronavirus.” *The Verge*, The Verge, 7 Apr. 2020,

www.theverge.com/interface/2020/4/7/21209595/google-mobility-reports-facebook-disease-prevention-maps-tech-privacy-backlash.

¹⁷⁸ Etherington, Darrell. “First Version of Apple and Google's Contact Tracing API Should Be Available to Developers next Week.” *TechCrunch*, TechCrunch, 23 Apr. 2020, techcrunch.com/2020/04/23/first-version-of-apple-and-googles-contact-tracing-api-should-be-available-to-developers-next-week/.

¹⁷⁹ Douek, Evelyn. “The Internet's Titans Make a Power Grab.” *The Atlantic*, Atlantic Media Company, 18 Apr. 2020, www.theatlantic.com/ideas/archive/2020/04/pandemic-facebook-and-twitter-grab-more-power/610213/.

3. Why Platforms Approach COVID-19 Differently

As demonstrated above, platform policy regarding COVID-19 has differed from prior responses to misinformation crises. Several considerations about the pandemic help explain why this new approach has emerged: the relative ease with which platforms can detect COVID-19 misinformation compared to other forms of false or misleading content, the prospect of repairing reputational damage and regaining legitimacy, the possibility of cementing user dependency, and the general desire to limit the global impacts of COVID-19.

Platforms are able to detect health misinformation more easily and with less controversy relative to political mistruth.¹⁸⁰ Scientific or health topics generally tend toward this ease of detection—as they entail rigorous investigative procedures and a narrowly defined topic—in contrast with political truths that generally involve elements of majority rule and various constituent interests. It should be noted that this distinctiveness has been questioned in the context of COVID-19, with human rights advocates cautioning that protective measures may lead to framing critical journalism as misinformation and augmenting the public’s inability to weed out faulty scientific findings in the context of an “infodemic.”¹⁸¹ However, this distinction between scientific and political content has been wholeheartedly embraced by platforms with regard to COVID-19 misinformation and has informed their willingness to take more aggressive steps than in previous crises.¹⁸²

Indeed, the reputational damage incurred following recent controversies, such as Russian interference in the 2016 election,¹⁸³ also appears to be informing the current responses. The platforms appear to view the global crisis as a critical test — how well they respond to the pandemic could either restore the reputational damage incurred over previous crises or further alienate their users and public authorities.¹⁸⁴ Of course, misinformation has been a constant threat for years: a 2019 study found evidence that, in 70 countries, at least one political actor used

¹⁸⁰ Sarah Kreps and Brendan Nyhan, “Coronavirus Fake News Isn’t Like Other Fake News,” March 30, 2020, <https://www.foreignaffairs.com/articles/2020-03-30/coronavirus-fake-news-isnt-other-fake-news>.

¹⁸¹ Sarah Wild, “Citing Virus Misinformation, South Africa Tests Speech Limits,” April 3, 2020, <https://undark.org/2020/04/03/fake-news-south-africa-covid-19/>; World Health Organization, “Novel Coronavirus (2019-nCoV) Situation Report - 13,” February 2, 2020, https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6.

¹⁸² Ben Smith, “When Facebook Is More Trustworthy Than the President,” March 15, 2020, <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html>.

¹⁸³ Andrew Arnold, “Consumer Trust In Social Media Is Declining: Here’s How Brands Should Change Their Strategies,” July 29, 2018, <https://www.forbes.com/sites/andrewarnold/2018/07/29/consumer-trust-in-social-media-is-declining-heres-how-brands-should-change-their-strategies/#2f675e976b76>.

¹⁸⁴ Elizabeth Dwoskin, “Why the coronavirus pandemic is a test Facebook can’t afford to fail,” March 27, 2020, <https://www.washingtonpost.com/technology/2020/03/27/facebook-zuckerberg-coronavirus-test/>.

social media manipulation to influence domestic conversations and several countries used large social media platforms to target global audiences.¹⁸⁵ However, COVID-19 has also clearly triggered a dramatic increase.¹⁸⁶ A cyber-security company recently found a 667% increase in phishing emails due to the pandemic.¹⁸⁷ Almost one-fifth of the 100 million phishing emails Google blocks each day are related to COVID-19.¹⁸⁸ This enormous swell has helped sustain a general consensus regarding the need to limit exposure to harmful information and promote trusted, “objective” health authorities.¹⁸⁹ Thus, the challenge of COVID-19 misinformation presents a clear test for platforms to demonstrate that they have learned from previous mistakes—a test where at least certain courses of action appear broadly popular and uncomplicated by political dynamics. While control of speech in the political realm remains controversial, the consensus behind aggressively regulating COVID-19 information gives the platforms an opportunity to win back favor and demonstrate control, if only in a narrow realm.

In addition to salvaging their reputation by combating misinformation, the platforms can also use their enhanced response to ingrain themselves as legitimate news sources and trusted platforms. In this manner, they can expand and solidify their user base. The platforms have publicly partnered with accredited health organizations around the world, acting as a medium for trusted authorities.¹⁹⁰ WHO has consistently emphasized the danger of misinformation to their mission. In February, Director-General Tedros Adhanom Ghebreyesus stated, “We’re not just fighting an epidemic; we’re fighting an infodemic.”¹⁹¹ Accordingly, health organizations such as WHO, CDC, and AMA rely on the platforms to spread correct information and aggressively counter misinformation.¹⁹² This legitimacy would likely increase their potential user bases and the time users spend on their sites. In addition, in the event of controversy surrounding their misinformation policy, this emphasis on credible partners also creates a backdoor of abdication regarding platform actions by allowing the platforms to shift blame and decision-making authority to these organizations.

¹⁸⁵ Samantha Bradshaw and Philip Howard, “The Global Disinformation Order: 2019 Global Inventory of Organized Social Media Manipulation,” 2019,

<https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>.

¹⁸⁶ Joe Tidy, “Google blocking 18m coronavirus scam emails every day,” April 17, 2020,

<https://www.bbc.com/news/technology-52319093>.

¹⁸⁷ Tidy, Joe. “Google Blocking 18m Coronavirus Scam Emails Every Day.” *BBC News*, BBC, 17 Apr. 2020, www.bbc.com/news/technology-52319093.

¹⁸⁸ *Ibid.*

¹⁸⁹ Sarah Kreps and Brendan Nyhan, “Coronavirus Fake News Isn’t Like Other Fake News,” March 30, 2020, <https://www.foreignaffairs.com/articles/2020-03-30/coronavirus-fake-news-isnt-other-fake-news>.

¹⁹⁰ The Department of Global Communications, “UN tackles ‘infodemic’ and cybercrime in COVID-19 crisis,” <https://www.un.org/en/un-coronavirus-communications-team/un-tackling-%E2%80%98infodemic%E2%80%99-misinformation-and-cybercrime-covid-19>.

¹⁹¹ *Ibid.*

¹⁹² Patrice Harris, “COVID-19: The importance of science in an era of distrust and misinformation,” April 7, 2020, <https://www.ama-assn.org/press-center/press-videos/covid-19-importance-science-era-distrust-and-misinformation>.

Platforms also have an interest in limiting the impact of COVID-19 on society. Misinformation is directly tied to immediate health effects and the overall effectiveness of the global response. On the individual level, misinformation has killed dozens of people in Iran with false information about the protective benefits of methanol,¹⁹³ and on a community level it has harmed entire groups with false stories about African Americans being immune.¹⁹⁴ The platforms may desire to curb the devastating effects of COVID-19 for the good of society or in recognition that despite their size, they will be unable to escape pandemic unscathed.¹⁹⁵

While this is by no means a comprehensive list of motivations, all or any of these factors may drive the platforms' altered responses. They have contributed to an unprecedented response that will likely enhance the platforms' reputations and market share.

C. Further Proposed Interventions

Thus far, this report has analyzed efforts already taken by Facebook, Twitter, and Google to combat misinformation and disinformation surrounding COVID-19. Some efforts, like Google's efforts to remove misleading COVID-19 content,¹⁹⁶ have reinforced or extended existing removal policies. Others, like Facebook's provision of free advertisement space to national and international health organizations,¹⁹⁷ appear to be new interventions tailored to COVID-19. Despite the diverse array of interventions being implemented by platforms, however, crucial gaps remain within each platform's response to COVID-19 misinformation.

¹⁹³ Malekian, Somayeh. "Iran Confronts Deadly Alcohol Crisis in Midst of Dealing with Coronavirus." *ABC News*, 27 Mar. 2020,

abcnews.go.com/Health/iran-confronts-deadly-alcohol-crisis-midst-dealing-coronavirus/story?id=69842613.

¹⁹⁴ Valentina Pop, "EU, Tech Firms Renew Pact to Fight Coronavirus Disinformation," March 11, 2020,

https://www.wsj.com/articles/eu-tech-firms-renew-alert-system-to-fight-coronavirus-disinformation-11583946710?mod=article_inline; Jason Breslow, "Why Misinformation and Distrust Are Making COVID-19 More Dangerous For Black America, April 10, 2020,

<https://www.npr.org/sections/coronavirus-live-updates/2020/04/10/832039813/why-misinformation-and-distrust-is-making-covid-19-more-dangerous-for-black-amer>; Patrice Harris, "COVID-19: The importance of science in an era of distrust and misinformation," April 7, 2020,

<https://www.ama-assn.org/press-center/press-videos/covid-19-importance-science-era-distrust-and-misinformation>.

¹⁹⁵ Daisuke Wakabayashi, Tiffany Hsu, and Mike Isaac, "Even Google and Facebook May Face an Ad Slump," April 14, 2020,

<https://www.nytimes.com/2020/04/14/technology/coronavirus-google-facebook-advertising.html?action=click&module=moreIn&pgtype=Article®ion=Footer&action=click&module=MoreInSection&pgtype=Article®ion=Footer&contentCollection=Technology>.

¹⁹⁶ Pichai, Sundar. "Coronavirus: How we're helping." Google. 6 Mar. 2020,

<https://www.blog.google/inside-google/company-announcements/coronavirus-covid19-response/>.

¹⁹⁷ Zuckerberg, Mar. Facebook Coronavirus Update. *Facebook*. 3 Mar. 2020,

<https://www.facebook.com/4/posts/10111615249124441/?d=n>.

The purpose of this section is to highlight these gaps and consider additional interventions that may fill them. These gaps appear both in terms of the methods platforms use and the goals their current interventions are designed to achieve. As discussed above, platform responses involve interventions designed both to alter the platform’s product offerings (“product” interventions) as well as to alter how platforms manage their existing products and services (“policy” interventions). For example, a platform may develop a new contact tracing notification system (a “product” intervention) alongside a change in its privacy policy to allow users to opt in to the notification system (a “policy” intervention). Further, some of these interventions are designed to combat misinformation directly¹⁹⁸ (“sword” interventions) while others are designed to protect users from the harms resulting from misinformation that platforms cannot remove effectively (“shield” interventions). Typically, “sword” interventions target misinformation and its sources, while “shield” interventions target and protect a platform’s users. For example, an intervention that trains machine learning algorithms to detect and remove 5G conspiracy misinformation targets misinformation directly (a “sword” intervention), while an intervention that restricts the number of people with whom a user can share an article targets users and their safety directly (a “shield” intervention).

Regardless of the methods employed by platforms, a review of both existing platform responses and proposed interventions suggest that a successful platform response to COVID-19 misinformation must address four main goals:

1. Ensuring users are safe and informed;
2. Collaborating with others to stoke innovation and improve existing interventions;
3. Combating misinformation directly; and
4. Treating misinformation as public health data.

Platforms, both large and small, should carefully consider where gaps in their own responses may lie—and whether the projected benefits of a proposed intervention would outweigh its projected costs.

¹⁹⁸ François, Camille. “Actors, Behaviors, Content: A Disinformation ABC.” Transatlantic Working Group. 20 Sept. 2019, https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf.

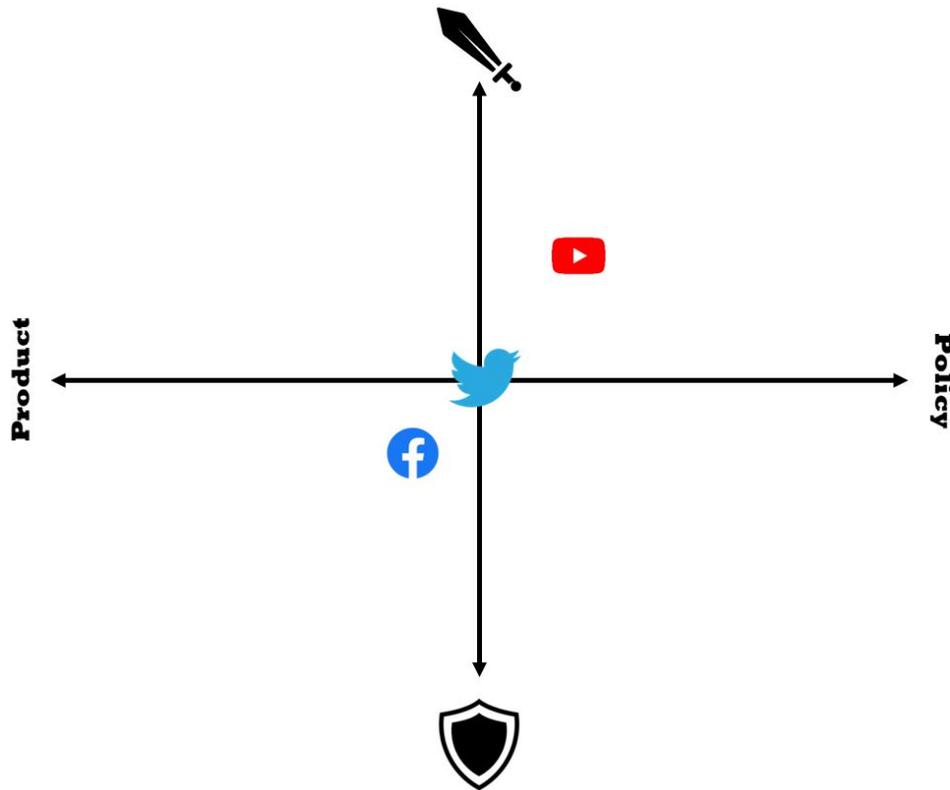


Figure 5. Platform Responses, Graphed. This graph plots the overall responses of Facebook, Twitter, and YouTube, discussed earlier, along two methodological dimensions: “product” versus “policy” interventions and “sword” versus “shield” interventions. The position of each platform within the graph is an estimate of that platform’s overall COVID-19 response as of April 2020, relative to other platform responses, and is based on independent research and interviews with platform employees. For example, because YouTube has banned all medically unsubstantiated and 5G-related COVID-19 misinformation¹⁹⁹ (a policy-based, “sword” intervention), it was placed farther toward the northeast quadrant compared to Facebook, which bans misinformation that risks imminent physical harm,²⁰⁰ but which has developed a COVID-19 information hub²⁰¹ (a product-based, “shield” intervention). Twitter, which has been more aggressive with content removal than Facebook²⁰² and has developed a variety of both policy and product interventions,²⁰³ is placed closer to the middle of the graph. Product interventions include changes to product interfaces or service options; policy interventions include content moderation,

¹⁹⁹ “Coronavirus: YouTube bans ‘medically unsubstantiated’ content.” *BBC News*, 22 Apr. 2020, <https://www.bbc.com/news/technology-52388586>.

²⁰⁰ Facebook. March 18 2020 Facebook Press Call. *Facebook*. 18 Mar. 2020, <https://about.fb.com/wp-content/uploads/2020/03/March-18-2020-Press-Call-Transcript.pdf>.

²⁰¹ Reichert, Corinne. “Facebook Messenger now has a COVID-19 information hub.” *CNET*, 26 Mar. 2020, <https://www.cnet.com/news/facebook-messenger-now-has-a-covid-19-information-hub/>.

²⁰² Peters, Jay. “Twitter will remove misleading COVID-19-related tweets that could incite people to engage in ‘harmful activity’.” *The Verge*, 22 Apr. 2020, <https://www.theverge.com/2020/4/22/21231956/twitter-remove-covid-19-tweets-call-to-action-harm-5g>.

²⁰³ “Coronavirus: Staying safe and informed on Twitter.” *Twitter Blog*, 3 Apr. 2020, https://blog.twitter.com/en_us/topics/company/2020/covid-19.html.

monetization, and fact-checking. Sword interventions aim to reduce misinformation; shield interventions aim to increase user protections against misinformation that has not been removed.

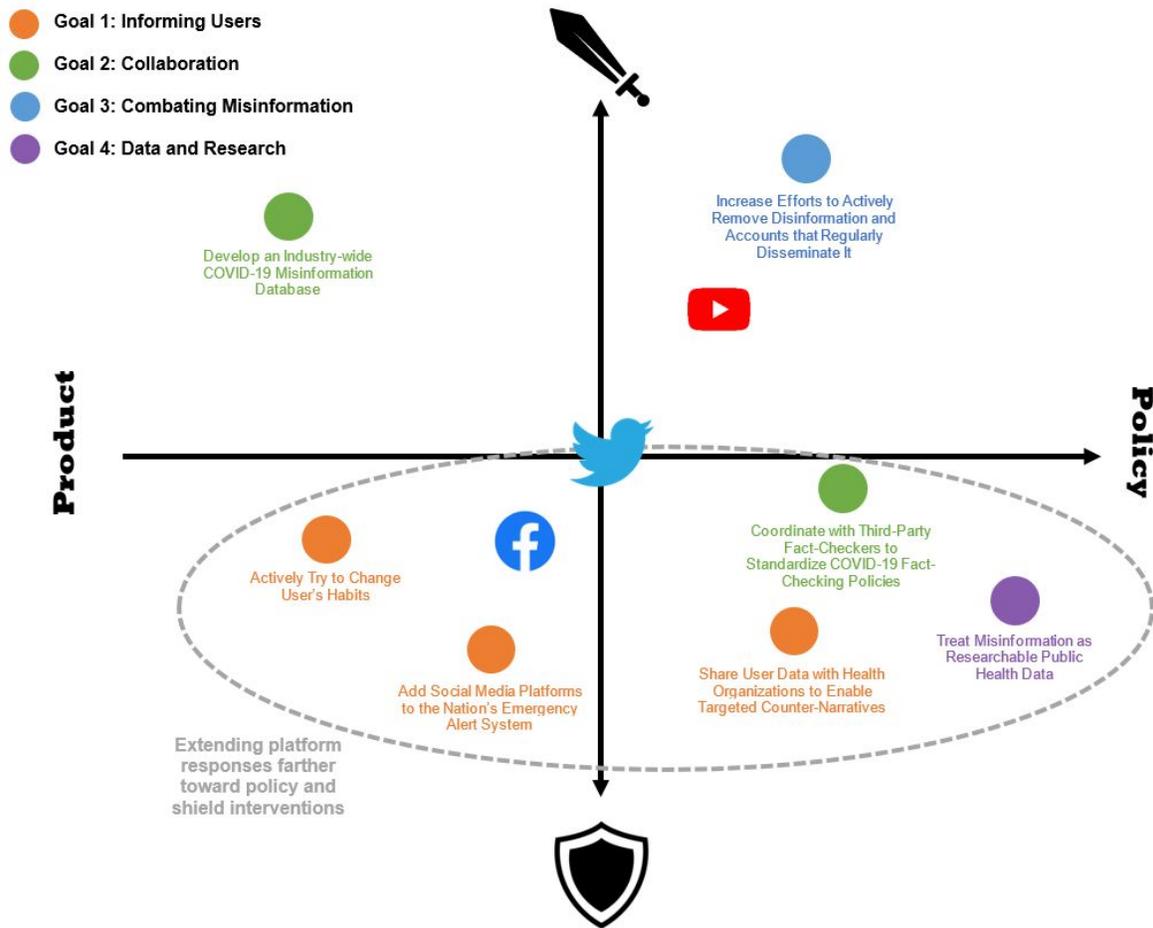


Figure 6. Proposed Interventions, Graphed. This graph overlays the previous platform response graph with one that plots each proposal discussed herein along the same two methodological dimensions. Each proposal's placement is estimated in relation to both the platforms' responses and other proposals, and should be viewed as a subjective indication of (1) the direction in which and (2) the degree to which adopting any one proposal would shift the average platform response. Note that proposals cover all four quadrants and are designed to (1) expand the range of interventions being considered to include more "policy" and "shield" responses and (2) extent existing "sword" responses used by platform responses further.

1. Ensuring Users are Safe and Informed

Content moderation debates centered on hate speech and election interference focus on stopping the spread of harmful content. However, in the current health crisis, platforms have taken on the additional role of affirmatively providing access to expert information. Should platforms choose to, the following steps would allow them to go even further, harnessing

existing tools to facilitate more effective communication between expert officials and at-risk users while promoting best health practices.

Proposal 1: Share User Data with Health Organizations to Enable Targeted Counter-Narratives.

Instead of directing users who have interacted with pandemic related misinformation to trusted sources that offer general advice, platforms could use the same data to group individuals who encountered misinformation into targetable groups based on the type of false narrative they engaged.²⁰⁴ Then, health organizations, some of which have already been given free advertisements, could craft specific counter-narratives based on the characteristics of the user and the type of misinformation being combatted.²⁰⁵ This strategy would likely be more effective than the current strategy of calling a piece of information false as research has shown the most successful debunking efforts involve providing repeated, fact based alternative narratives.²⁰⁶ It also outsources the fact heavy, case-specific inquiry of debunking claims at a time when platforms are being forced to rely on algorithms rather than human content moderators.²⁰⁷ However, the method would set a dangerous precedent of allowing platforms to label certain citizens as “misinformed” and passing that information on to official entities for correction. Additionally, even if this danger could be allayed if the program were tightly confined and tailored to the current crises, the advertising groupings would contain highly sensitive data revealing details about users that could be manipulated by bad actors. The platforms would need to ensure that such information would be shared only with highly trusted organizations and deleted within a reasonable time to avoid abuse.

Proposal 2: Add Social Media Platforms to the Nation’s Emergency Alert System.

The current U.S. government emergency alert system that allows federal, state, and local officials to send out public messages over popular communication mediums does not include social media platforms.²⁰⁸ Joan Donovan, the Director of the Technology and Social Change

²⁰⁴ Romm, Tony. “Facebook Will Alert People Who Have Interacted with Coronavirus 'Misinformation'.” *The Washington Post, WP Company*, 16 Apr. 2020, www.washingtonpost.com/technology/2020/04/16/facebook-will-alert-people-who-have-interacted-with-coronavirus-misinformation/.

²⁰⁵ “Combating COVID-19 Misinformation Across Our Apps.” *About Facebook*, 2 Apr. 2020, about.fb.com/news/2020/03/combating-covid-19-misinformation/.

²⁰⁶ Lewandowsky, Stephan, et al. “Misinformation and Its Correction.” *Psychological Science in the Public Interest*, vol. 13, no. 3, 2012, pp. 106–131., doi:10.1177/1529100612451018.

²⁰⁷ Elizabeth Dwoskin, Nitasha Tikku. “Facebook Sent Home Thousands of Human Moderators Due to the Coronavirus. Now the Algorithms Are in Charge.” *The Washington Post, WP Company*, 24 Mar. 2020, www.washingtonpost.com/technology/2020/03/23/facebook-moderators-coronavirus/.

²⁰⁸ Donovan, Joan. “Here's How Social Media Can Combat the Coronavirus 'Infodemic'.” *MIT Technology Review*, 10 Apr. 2020, www.technologyreview.com/2020/03/17/905279/facebook-twitter-social-media-infodemic-misinformation/.

Research Project at the Harvard Kennedy School's Shorenstein Center, says that social media platforms should be added to this alert distribution system.²⁰⁹ She explains that platforms could use the data they already have about their users to feed them localized and timely notifications.²¹⁰ Adding platforms to this system would provide another avenue for the timely dissemination of factual, public health information by a national public health agency in moments of crisis, but would also need to be used judiciously to avoid abuse or politicization.

Proposal 3: Actively Try to Change User's Habits.

Aza Raskin of the Center for Humane Technology argues that platforms need to move from merely passively providing expert information to actively persuading their users to adopt recommended health strategies.²¹¹ His ideas include showing users their friends who have opted to "stay home to save lives" and asking if they want to join them, as well as creating novelty photo filters that activate when the user is wearing a mask.²¹² Similar tactics have shown to be effective at changing user behavior to increase voter turnout in the past and could convince users to practice safe physical distancing.²¹³ Such a shift would require the platforms to take a normative stance on best practices which could put them at odds with its users and policy makers as public officials increasingly disagree on the right way forward and even health officials' guidance on best health practices can change.²¹⁴ Additionally, efforts by social media companies to use their platforms to control users have been met with understandable criticism in the past and led to warnings about the dangerous potential of these efforts if unchecked in the future.²¹⁵

²⁰⁹ *Ibid.*

²¹⁰ *Ibid.*

²¹¹ Technology, Center For Humane. "From Inform to Persuade: How Can Tech Step Up for Humanity?" *Medium, Center for Humane Technology*, 28 Mar. 2020, medium.com/center-for-humane-technology/from-inform-to-persuade-how-can-tech-step-up-for-humanity-6360c6c3d03e.

²¹² *Ibid.*

²¹³ Bond, Robert M., et al. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature*, vol. 489, no. 7415, 2012, pp. 295–298., doi:10.1038/nature11421.

²¹⁴ Shear, Michael and Sarah Mervosh. "Trump Encourages Protest Against Governors Who Have Imposed Virus Restrictions." *The New York Times*. 17 Apr. 2020, www.nytimes.com/2020/04/17/us/politics/trump-coronavirus-governors.html; Aubrey, Allison. "U.S. Government Discussing New Guidance On Wearing Face Masks In Public." *NPR*, 1 Apr. 2020, www.npr.org/2020/04/01/825499494/u-s-government-discussing-new-guidance-on-wearing-face-masks-in-public.

²¹⁵ Goel, Vindu. "Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry." *The New York Times*, 30 June 2014, www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html; Zittrain, Jonathan. "Facebook Could Decide an Election without Anyone Ever Finding Out." *New Statesman*, 3 June 2014, www.newstatesman.com/politics/2014/06/facebook-could-decide-election-without-anyone-ever-finding-out.

2. Cross-Platform Collaboration to Stoke Innovation and Improve Existing Interventions

While digital platforms have collaborated in the past,²¹⁶ the level of platform collaboration that has emerged in the wake of the pandemic is rare—and warrants a closer look. Increasing collaboration and communication both among digital platforms and between platforms and governmental entities may organically produce regulatory sandboxes: regulated ecosystems for rapid information-sharing and experimentation that may produce effective interventions more quickly (and cheaply) than purely competitive or state-regulated models.²¹⁷ However, platform collaboration also risks running afoul of (1) antitrust law,²¹⁸ (2) free speech protections,²¹⁹ and (3) data privacy.²²⁰ As platforms consider more collaborative interventions, they will need to consider not only how to balance projected costs and benefits, but also how to weigh different types of costs and benefits—from privacy to efficiency to health impact.

Proposal 1: Develop an Industry-wide COVID-19 Misinformation Database.

Digital platforms typically use internal content moderation policies to determine when to reduce the spread of content²²¹ or remove it outright.²²² However, platforms and governmental entities have collaborated in the past to develop industry-wide systems for removing two types of harmful content: child pornography²²³ and terrorist content.²²⁴ These systems use a combination of perceptual image hashing²²⁵ (also applicable to videos)²²⁶ and machine learning to

²¹⁶ Global Internet Forum to Counter Terrorism (GIFCT). *Joint Tech Innovation*. 8 Feb. 2020, <https://www.gifct.org/joint-tech-innovation/>.

²¹⁷ United Nations Secretary-General's Special Advocate for Inclusive Finance for Development. "Briefing on Regulatory Sandboxes." 14 Nov. 2018, <https://www.unsgsa.org/files/1915/3141/8033/Sandbox.pdf>.

²¹⁸ Overly, Steven, Leah Nylen, and Gabby Orr. "Why Silicon Valley's virus-era D.C. glow may not last." *Politico*. <https://www.politico.com/news/2020/03/25/silicon-valley-virus-era-dc-glow-147178>.

²¹⁹ Mchangama, Jacob, and Sarah McLaughlin. "Coronavirus Has Started a Censorship Pandemic." *Foreign Policy*. 1 Apr. 2020, <https://foreignpolicy.com/2020/04/01/coronavirus-censorship-pandemic-disinformation-fake-news-speech-freedom/>.

²²⁰ Hao, Karen. "Coronavirus is forcing a trade-off between privacy and public health." *MIT Technology Review*. 24 Mar. 2020, <https://www.technologyreview.com/2020/03/24/950361/coronavirus-is-forcing-a-trade-off-between-privacy-and-public-health/>.

²²¹ Zuckerberg, Mark. "A Blueprint for Content Governance and Enforcement." *Facebook*. 15 Nov. 2018, <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>.

²²² Hatmaker, Taylor. "Twitter broadly bans any COVID-19 tweets that could help the virus spread." *TechCrunch*. 18 Mar. 2020, <https://techcrunch.com/2020/03/18/twitter-coronavirus-covid-19-misinformation-policy/>.

²²³ Microsoft. "PhotoDNA." 3 May. 2020, <https://www.microsoft.com/en-us/photodna>.

²²⁴ Global Internet Forum to Counter Terrorism (GIFCT). *Joint Tech Innovation*. 8 Feb. 2020, <https://www.gifct.org/joint-tech-innovation/>.

²²⁵ See, e.g., Monga, Vishal, and Brian L. Evans. "Perceptual Image Hashing Via Feature Points: Performance Evaluation and Trade-Offs." *IEEE*. 16 Oct. 2006, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.637&rep=rep1&type=pdf>.

automatically remove harmful content: platforms and government actors (1) populate a database of images that they agree should be removed, (2) create hashes of each image, and (3) use machine learning algorithms to compare new content with the images in the database.

An industry-wide database of COVID-19 misinformation would expedite the training of AI algorithms designed to automatically detect misinformation for platforms to flag, restrict, or remove. To ensure that the benefits of such a system outweigh its costs, however, platforms will need to implement restrictions based on and in response to those used for terrorist content.²²⁷

First, the database should include only extreme examples of misinformation—content that is highly likely to cause actual harm and would likely be banned under platforms’ individual policies. Second, content that matches hashes in the database should *not* be immediately removed. Instead, algorithms trained on the database should flag COVID-19 misinformation for platforms to review under their own policies, balancing the benefits of removal with journalistic needs and free speech protections. Third, the COVID-19 database should be accessible to journalists and government regulators. Following concerns that the Global Internet Forum to Counter Terrorism (GIFCT), the entity in charge of the terrorist hash database, is not sufficiently transparent and accountable²²⁸ and to reduce the risk of censorship creep,²²⁹ any industry efforts to use hash databases should prioritize transparency and accountability. However, platforms may still need to restrict access to the general public to insulate the database from actors attempting to circumvent the database’s algorithms.

Proposal 2: Coordinate with Third-Party Fact-Checkers to Standardize COVID-19 Fact-Checking Policies.

While some platforms already partner with independent fact-checkers²³⁰ to flag misinformation, the quality and consistency of fact-check flagging²³¹ and removal policies²³²

²²⁶ Khelifi, Fouad, and Ahmed Bouridane. “Perceptual Video Hashing for Content Identification and Authentication.” *IEEE*. 21 Nov. 2017, http://nrl.northumbria.ac.uk/32873/1/paper_double.pdf.

²²⁷ Facebook. “Partnering to Help Curb the Spread of Online Terrorist Content.” *Facebook*. 5 Dec. 2016, <https://about.fb.com/news/2016/12/partnering-to-help-curb-spread-of-online-terrorist-content/>.

²²⁸ Díaz, Ángel. “Global Internet Forum to Counter Terrorism Transparency Report Raises More Questions Than Answers.” *Brennan Center for Justice*. 25 Sept. 2019, <https://www.brennancenter.org/our-work/analysis-opinion/global-internet-forum-counter-terrorism-transparency-report-raises-more>.

²²⁹ Citron, Danielle K. “Extremist Speech, Compelled Conformity, and Censorship Creep.” *Notre Dame L. Rev.* 2018, https://scholarship.law.bu.edu/cgi/viewcontent.cgi?article=1623&context=faculty_scholarship.

²³⁰ Facebook. “How is Facebook addressing false news through third-party fact-checkers?” *Facebook*. 2020, <https://www.facebook.com/help/1952307158131536>.

²³¹ Tardáguila, Cristina. “Google, Facebook and Twitter could do more to surface fact-checks about coronavirus.” *The Poynter Institute*. 6 Feb. 2020, <https://www.poynter.org/fact-checking/2020/google-facebook-and-twitter-could-do-more-to-surface-fact-checks-about-coronavirus/>.

vary widely among platforms. This inconsistency in platform responses creates opportunities for COVID-19 misinformation to spread more widely by (1) providing more opportunities for misinformation to prey on users' cognitive biases;²³³ (2) leading users to unduly trust unflagged information²³⁴ that may nonetheless be false or misleading; and (3) stoking distrust in fact-checkers such that misinformation is spread more widely.²³⁵

A global consortium of 48 fact-checking organizations, the International Fact-Checking Network (IFCN),²³⁶ is already working collaboratively to debunk COVID-19 misinformation, and many platforms have already committed to funding their efforts.²³⁷ However, funding will not be enough. To ensure that platforms' fact-checking efforts are effective at reducing the spread of COVID-19 misinformation, platforms should:

First, standardize COVID-19 fact-checking policies across the industry. Most platforms still use vague and inconsistent fact-checking policies, which can undermine the efforts of independent fact-checkers like those at the IFCN and create opportunities for COVID-19 misinformation to spread. Facebook, by contrast, uses IFCN fact-checking to flag COVID-19 misinformation and removes misinformation that is likely to cause harm.²³⁸ In line with their joint statement,²³⁹ major platforms should consider standardizing their fact-checking policies similarly: fact-checks should link to IFCN sources and platforms should remove content that national and international health organizations consider to be major health risks.

²³² Chakravorti, Bhaskar. "Social media companies are taking steps to tamp down coronavirus misinformation – but they can do more." 30 Mar. 2020, <https://theconversation.com/social-media-companies-are-taking-steps-to-tamp-down-coronavirus-misinformation-but-they-can-do-more-133335>.

²³³ Ciampaglia, Giovanni Luca, and Filippo Menczer. "Biases Make People Vulnerable to Misinformation Spread by Social Media." *Scientific American*. 21 Jun. 2018, <https://www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/>.

²³⁴ Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. "The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings." *Management Science*. 23 Apr. 2020, <https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2019.3478>.

²³⁵ Levin, Sam. "Facebook promised to tackle fake news. But the evidence shows it's not working." *The Guardian*. 16 May. 2017, <https://www.theguardian.com/technology/2017/may/16/facebook-fake-news-tools-not-working>.

²³⁶ Tardáguila, Cristina. "Coronavirus: Fact-checkers from 30 countries are fighting 3 waves of misinformation." *The Poynter Institute*. 28 Jan. 2020, <https://www.poynter.org/fact-checking/2020/coronavirus-fact-checkers-from-30-countries-are-fighting-3-waves-of-misinformation/>.

²³⁷ Goldshlager, Karen, and Orlando Watson. "Launching a \$1M Grant Program to Support Fact-Checkers Amid COVID-19." *Facebook*. 2 Apr. 2020, <https://www.facebook.com/journalismproject/coronavirus-grants-fact-checking>.

²³⁸ Jin, Kang-Xing. "Keeping People Safe and Informed About the Coronavirus." *Facebook*. 22 Apr. 2020, <https://about.fb.com/news/2020/04/coronavirus/#limiting-misinfo>.

²³⁹ Shu, Catherine, and Jonathan Shieber. "Facebook, Reddit, Google, LinkedIn, Microsoft, Twitter and YouTube issue joint statement on misinformation." *TechCrunch*. 16 Mar. 2020, <https://techcrunch.com/2020/03/16/facebook-reddit-google-linkedin-microsoft-twitter-and-youtube-issue-joint-statement-on-misinformation/>.

Second, implement fact-check flags on all COVID-19 content. To reduce the implied truth effect²⁴⁰ given existing constraints on fact-checking resources, platforms should expand their fact-checking responses to all COVID-19 content. Under this regime, platforms would use algorithms to flag all content related to COVID-19 with initial “unverified” tags until fact-checked by the IFCN. To expedite fact-checking efforts, trusted sources of COVID-19 information—e.g. national health organizations—could receive “verified” tags on their posts by default.²⁴¹

3. Combating Misinformation Directly

Some of the more forward-leaning approaches to combating misinformation involve taking steps to remove or reduce the spread of misinformation posted on platforms. Examples include removing misinformation from a platform altogether, disabling the account of a user posting misinformation, or limiting the distribution and virality of misinformation. Such action couples well with steps to promote truthful information from health authorities by attacking the problem of misinformation and the narrative it creates from two sides: making factual information more prevalent while reducing the incidence of false information.

Proposal: Increase Efforts to Actively Remove Misinformation, and Accounts that Regularly Disseminate It

Misinformation targeted for removal may range from information that platforms identify as false and harmful, such as COVID-19 “cures” encouraging users to ingest harmful substances, to scams that seek to profit off of the pandemic. Such content should be flagged and subjected to removal site-wide, including when similar or identical content is disseminated by multiple user accounts. Platforms should also suspend accounts regularly disseminating such harmful misinformation and scams. This may require an increase in the platform’s investment in human fact-checkers and a greater refinement of algorithms charged with identifying misinformation.²⁴²

While many platforms have begun undertaking such efforts, in some cases, portions of their sites have slipped through the cracks, such that misinformation posted in different areas of a platform may be subject to varying levels of fact-checking and a different likelihood of potential

²⁴⁰ Pennycook, Gordon, Adam Bear, Evan T. Collins, and David G. Rand. “The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings.” *Management Science*. 23 Apr. 2020, <https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2019.3478>.

²⁴¹ @TwitterSupport (Twitter Support). “PSA about what we’re doing to Verify Twitter accounts that are providing credible updates around #COVID19: we are working with global public health authorities to identify experts and have already Verified hundreds of accounts, but there’s more to do and we could use your help.” *Twitter*. 20 Mar. 2020, <https://twitter.com/TwitterSupport/status/1241155701822476288>.

²⁴² Frenkel, Sheera, Davey Alba, and Raymond Zhong. “Surge of Virus Misinformation Stumps Facebook and Twitter.” *New York Times*, <https://www.nytimes.com/2020/03/08/technology/coronavirus-misinformation-social-media.html>.

removal. Platform policies should be applied uniformly to user messages, group postings, and comment sections.

Content removal offers advantages over efforts to maintain the presence of misinformation while reducing its user traffic. Facebook itself has revealed that its algorithm tends to increase the visibility of posts that have received the most traffic, and sensationalist content tends to attract greater user engagement and virality than factual content.²⁴³ Yet even changing this algorithm to reduce the number of users who view false and harmful content about COVID-19 fails to address the problem of spreading untruthful public health information that may endanger lives. Platforms ought to consider more aggressively removing harmful misinformation and reserving their ability to reduce visibility of COVID-19 misinformation only for that content which is least harmful.

Removal of content and suspension of user accounts raise difficult issues for platforms, many of which commit to promoting user expression as part of their foundational principles. Misinformation removal policies may draw their support from other core platform tenets, such as promoting safety and public health. In balancing these values with curtailed expression, platforms may consider distinguishing COVID-19 misinformation from other types of online content. The exigencies of the worldwide pandemic and its severe health and safety implications may be explained to justify unique policies that platforms might decline to extend to other content areas.

4. Treating Misinformation as Public Health Data

Proposal: Share Health-related, Trend Information and Developments in Misinformation with Public Health Agencies.

Digital data has become a critical tool in the fight against COVID-19, and major platforms now find themselves in the midst of a paradigm shift in determining how to use their users' digital footprints to shape public health policy. Both the information and *misinformation* generated by activity on these platforms can play a critical role in building a modernized public health surveillance system to allow for more robust and effective containment of future outbreaks. "Information" here refers to data related to user movement and cellular connection, which, as demonstrated by endeavors like Facebook's Disease Prevention Maps,²⁴⁴ allows policy-makers and health professionals to understand the effectiveness of social distancing protocols and track how the disease is spreading. Aggregating this information across platforms

²⁴³ Zuckerberg, Mark. "A Blueprint for Content Governance and Enforcement." *Facebook*, <https://www.facebook.com/notes/mark-zuckerberg/a-blueprint-for-content-governance-and-enforcement/10156443129621634/>.

²⁴⁴ "Disease Prevention Maps." *Facebook Data for Good*, Facebook, dataforgood.fb.com/tools/disease-prevention-maps.

can be used in contributing to a larger repository of information required to create “a central data surveillance system to link laboratory data with population data and clinical measures, so that organizations across our health system will have the necessary information to track and predict outbreaks.”²⁴⁵

Yet, in addition to this *information*, major digital platforms can and should also share their internal understandings related to *misinformation* campaigns on their sites as well. Unfortunately, the problematic social behaviors caused by exposure to these fraudulent and misleading misinformation need to be tracked and accounted for in order to fully allow for effective public health surveillance. Specifically, understanding how misinformation spreads and influences behavior is integral to understanding the overall social dynamics driving population activity tracked by public health officials. As Laurent Hébert-Dufresne and Vicky Chuqiao Yang argue, platforms can effectively:

“contribute to disease modeling by sharing data related to the spread of myths and lies around public health emergencies. These data are now often part of the emergency itself and can greatly affect our forecasts and models of interventions. Until we stop treating epidemics as if they are happening in a vacuum, our models and forecasts will be incomplete. Epidemics occur within an ecosystem of diseases, information, and behaviors, all of which are integral to understanding the spread of COVID-19.”²⁴⁶

Thus, major digital platforms should include their internal findings on misinformation as part of the data they provide to government and health officials seeking to effectively monitor and combat the spread of COVID-19.

There are legitimate concerns regarding privacy and abuses of info that must be addressed if such extensive data sharing—with the purpose of being used as surveillance—is implemented. In order to prevent the data from being used for purposes other than pandemic-related responses, there are a number of checks available to curb potential abuses, including making sure that such “data is ferociously ring-fenced and siloed and not routinely augmented by credit card, CCTV, or mass immigration data.”²⁴⁷ Moreover, programs like

²⁴⁵ Kim, Jane J., and Michelle A. Williams. “Commentary: There Will Be Another Pandemic after the Coronavirus—and It’s Time to Start Preparing Now.” *Fortune*, 29 Mar. 2020, fortune.com/2020/03/29/coronavirus-pandemic-public-health-preparedness/.

²⁴⁶ Hébert-Dufresne, Laurent, and Vicky Chuqiao Yang. “Misinformation about an Outbreak Is Important Public Health Data.” *STAT*, 6 Apr. 2020, www.statnews.com/2020/04/07/misinformation-outbreak-is-important-public-health-data/.

²⁴⁷ Wright, Nicholas. “Coronavirus and the Future of Surveillance.” *Foreign Affairs*, 13 Apr. 2020, www.foreignaffairs.com/articles/2020-04-06/coronavirus-and-future-surveillance.

Facebook Disease Mapping rely on de-identified aggregate data²⁴⁸—demonstrating that protection of individual information while safeguarding the overall public good is a not mutually exclusive proposition. In sum, we cannot lose sight of the very real privacy concerns and potential sources of abuse that could follow from increased public health surveillance. Nonetheless, COVID-19 poses an unprecedented threat, and the information—and misinformation—collected during this current crisis will not only function to help contain this initial outbreak, but also serve as the foundation for establishing protocols and programs that will limit the damage unleashed by potentially ensuing waves of COVID-19 or other future pandemics.

Conclusion

COVID-19 misinformation has taken many forms and has adapted to the multitude of media by which it has spread. While the concrete impacts of COVID-19 misinformation are not entirely understood yet, it has contributed to the overall confusion surrounding the virus and interfered with effective responses to the virus outbreak. Understanding the spread and impact of misinformation, as well as platform steps thus far, is critical to the design and implementation of successful platform responses that strengthen public health and safety.

²⁴⁸ Morrison, Sara. “The Government Might Want Your Phone Location Data to Fight Coronavirus. Here's Why That Could Be Okay.” *Vox*, 18 Mar. 2020, www.vox.com/recode/2020/3/18/21184160/government-location-data-coronavirus.

Appendix A (Citations for Graph)

Cures:

- Capatides, Christina. “Coronavirus cannot be cured by drinking bleach or snorting cocaine, despite social media rumors,” *CBS News*, March 9, 2020, <https://www.cbsnews.com/news/coronavirus-drinking-bleach-cocaine-false-rumors-social-media/>.
- Dupuy, Beatrice. “Drinking Water Won’t Prevent the Coronavirus,” *AP News*, February 28, 2020, <https://apnews.com/afs:Content:8551030197>
- Gorvett, Zaria. “Can You Kill Coronavirus with UV Light?” *BBC*, 27 Mar. 2020, <https://www.bbc.com/future/article/20200327-can-you-kill-coronavirus-with-uv-light>.
- Lavery, Jim. “What Do We Have to Lose with Hydroxychloroquine? Evidence That Might Save Lives.” *The Hill*, 16 Apr. 2020, <https://thehill.com/opinion/healthcare/493035-what-do-we-have-to-lose-with-hydroxychloroquine-evidence-that-might-save>
- Przybyla, Heidi. “Family of New York Woman Blames Hydroxychloroquine Combo for Fatal Heart Attack.” *NBCNews.com*, 17 Apr. 2020, <https://www.nbcnews.com/health/health-news/family-new-york-woman-blames-hydroxychloroquine-combo-fatal-heart-attack-n1185451>.
- Whoriskey, Peter. “Herbs. Silver. Salt. Toothpaste. As Coronavirus Fears Rise, Marketers Pitch Remedies.” *The Washington Post*, WP Company, 1 Apr. 2020, <https://www.washingtonpost.com/business/2020/04/01/scams-coronavirus-remedies/>.
- World Health Organization. “Coronavirus disease (COVID-19) advice for the public: Myth busters,” 2020, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>

Fear:

- Schraer, Rachel and Lawrie, Eleanor “Coronavirus: Scientists brand 5G claims 'complete rubbish.'” *BBC News*, 15 Apr. 2020, <https://www.bbc.com/news/52168096>.
- “What Misinformation Has to Do With Toilet Paper.” *The Atlantic*, uploaded by Biran Gersten, 19 Mar. 2020, <https://www.theatlantic.com/video/index/608209/toilet-paper-shortage>
- “False claim: Putin released lions on the streets of Russia to enforce coronavirus lockdown.” *Reuters*, 25 Mar. 2020, <https://www.reuters.com/article/uk-fact-check-putin-lions-coronavirus-loc/false-claim-putin-released-lions-on-the-streets-of-russia-to-enforce-coronavirus-lockdown-idUSKBN21C371>
- Elassar, Alaa. “One dangerous coronavirus ‘self-check test’ is circulating on social media. Here’s why you should avoid it.” *CNN*, 17 Mar. 2020, <https://www.cnn.com/2020/03/16/us/coronavirus-myths-debunking-holding-breath-10-seconds-trnd/index.html>

Domestic Politics:

- Chang, Julie. “San Francisco Chinatown Affected By Coronavirus Fears, Despite No Confirmed Cases.” *NPR*, NPR, 26 Feb. 2020, www.npr.org/2020/02/26/809741251/san-francisco-chinatown-affected-by-coronavirus-fears-despite-no-confirmed-cases.
- Chiu, Allyson. “Trump Has No Qualms about Calling Coronavirus the 'Chinese Virus.' That's a Dangerous Attitude, Experts Say.” *The Washington Post*, WP Company, 20 Mar. 2020, www.washingtonpost.com/nation/2020/03/20/coronavirus-trump-chinese-virus/.
- Chotiner, Isaac. “The Contrarian Coronavirus Theory That Informed the Trump Administration.” *The New Yorker*, Condé Nast, 30 Mar. 2020, www.newyorker.com/news/q-and-a/the-contrarian-coronavirus-theory-that-informed-the-trump-administration.
- Gowen, Annie. “Coronavirus Deniers and Hoaxers Persist despite Dire Warnings, Claiming 'It's Mass Hysteria'.” *The Washington Post*, WP Company, 19 Mar. 2020, www.washingtonpost.com/national/coronavirus-deniers-outbreak-hoax/2020/03/19/46bc5e46-6872-11ea-b313-df458622c2cc_story.html.
- Iati, Marisa, and Lateshia Beachum. “Georgia, South Carolina and Tennessee Governors Announce Plans to Ease Coronavirus Restrictions.” *The Washington Post*, WP Company, 21 Apr. 2020, www.washingtonpost.com/world/2020/04/20/coronavirus-latest-news.
- Palma, Bethania. “Did President Trump Refer to the Coronavirus as a 'Hoax'?” *Snopes.com*, Snopes Media Group Inc., 2 Mar. 2020, www.snopes.com/fact-check/trump-coronavirus-rally-remark/.

International Politics:

- Andersen, Kristian G., et al. “The Proximal Origin of SARS-CoV-2.” *Nature Medicine*, vol. 26, no. 4, 2020, pp. 450–452., doi:10.1038/s41591-020-0820-9.
- Calisher, Charles, et al. “Statement in Support of the Scientists, Public Health Professionals, and Medical Professionals of China Combatting COVID-19.” *The Lancet*, vol. 395, no. 10226, 19 Feb. 2020, doi:10.1016/s0140-6736(20)30418-9.
- Taylor, Adam. “Analysis | What Caused the Coronavirus? A Skeptical Take on the Theories about the Outbreak's Chinese Origin.” *The Washington Post*, WP Company, 16 Apr. 2020, www.washingtonpost.com/world/2020/04/16/what-caused-coronavirus-skeptical-take-theories-about-outbreaks-chinese-origin/.
- Westcott, Ben, and Steven Jiang. “Chinese Diplomat Promotes Coronavirus Conspiracy Theory.” *CNN*, Cable News Network, 14 Mar. 2020, www.cnn.com/2020/03/13/asia/china-coronavirus-us-lijian-zhao-intl-hnk/index.html.

Other Health-Related:

- Azad, Arman, and Minali Nigam. “Yes, Young Adults Are Sick and Spreading Coronavirus -- but They Can Help Stop It.” CNN, Cable News Network, 20 Mar. 2020, www.cnn.com/2020/03/20/health/covid-19-young-adults-sick-spreading/index.html.
- Bella, Timothy. “‘If I Get Corona, I Get Corona’: Miami Spring Breakers Say Covid-19 Hasn't Stopped Them from Partying.” The Washington Post, WP Company, 19 Mar. 2020, www.washingtonpost.com/nation/2020/03/19/coronavirus-spring-break-party/.
- Elassar, Alaa. “One Dangerous Coronavirus 'Self-Check Test' Is Circulating on Social Media. Here's Why You Should Avoid It.” CNN, Cable News Network, 17 Mar. 2020, www.cnn.com/2020/03/16/us/coronavirus-myths-debunking-holding-breath-10-seconds-trnd/index.html.
- Gunia, Amy. “Millennials Aren't Taking Coronavirus Seriously, WHO Official Warns.” Time, Time, 20 Mar. 2020, www.time.com/5807073/millennials-coronavirus-who/.
- Meltzer, Kerry Kennedy. “I'm Treating Too Many Young People for the Coronavirus.” The Atlantic, Atlantic Media Company, 26 Mar. 2020, www.theatlantic.com/ideas/archive/2020/03/young-people-are-not-immune-coronavirus/608794/.
- Montanaro, Domenico. “FACT CHECK: Trump Compares Coronavirus To The Flu, But It Could Be 10 Times Deadlier.” NPR, NPR, 24 Mar. 2020, www.npr.org/sections/coronavirus-live-updates/2020/03/24/820797301/fact-check-trump-compares-coronavirus-to-the-flu-but-they-are-not-the-same.

Appendix B

