



**PUBLIC  
KNOWLEDGE**

# HOW ARE PLATFORMS RESPONDING TO THIS PANDEMIC?

What Platforms Are Doing to Tackle  
Rampant Misinformation



**Public Knowledge Presents:  
What Are Digital Platforms Doing to Counter Pandemic Misinformation?**

**FINAL REPORT**

Given the enormous role they now play in distributing news and information, it is important that the major technology platforms help consumers navigate between fact and fiction, between helpful and harmful information distributed on the internet. The quality of information available at the onset of the coronavirus epidemic, it's not an exaggeration to say, became a matter of life and death. We wanted to know: are the digital platforms doing their part to help us work through the difficulties engendered by the Coronavirus crisis? Are they setting a new standard of corporate behavior to prevent the dangerous flow of disinformation, or falling down on the job? And perhaps most importantly, what should our expectations be after the pandemic is over?

To answer these questions, beginning in early March and for three months in 2020, Public Knowledge tracked public information about the platforms' approaches to countering misinformation about the pandemic, as well as analysis and commentary about them from a wide variety of academic, journalistic, policy, and civil society organizations. Our goal was to help inform policy perspectives about the digital platforms and their approaches to countering misinformation.<sup>1</sup>

Here's what you'll find on the site:

- **Overview:** provides a general view of the scale of misinformation about the infodemic.
- **Dangers of the Infodemic:** describes the harms and types of misinformation we address.
- **Insights & Commentary:** provides analysis and commentary about the platforms' efforts, organized by week of our reporting period (weekly reports were issued on Wednesdays).
- **Case Studies:** provides a comprehensive summary of what each platform company communicated specifically in response to the coronavirus, organized by the date of their communication.

---

<sup>1</sup> Many frameworks define disinformation as knowingly false content meant to deceive. Given the difficulty of assessing the intent of information sharing for CV-19, we followed the Reuters Institute practice and used the term misinformation throughout this report to refer broadly to any type of false information – including disinformation – except when other terms are used in direct quotes or footnotes.

## Overview

An army of at least 48 fact-checking organizations from 30 countries began working in early March to debunk false information about the 2019 novel coronavirus. The collaborative project, coordinated by the International Fact-Checking Network, will be active for as long as the disease spreads worldwide and can be followed on social media channels through two hashtags, #CoronaVirusFacts and #DatosCoronaVirus.<sup>2</sup>

All of the digital platforms established dedicated websites, hashtags, and handles to communicate their response and actions in regard to the COVID-19 crisis (you'll find many marked in footnotes), and all of the platforms formed partnerships with the WHO, CDC and/or other health organizations as sources of current, accurate content. As you'll see, these partnerships became perhaps the most important enabler of the platforms' efforts.

One of the most notable elements in the platforms' response is greater use of Artificial Intelligence and machine learning systems for content moderation, due to the need to allow social distancing for their employees and contractors. Twitter<sup>3</sup>, Google<sup>4</sup> and Facebook<sup>5</sup> all issued communications noting that greater reliance on AI and machine learning systems would result in mistakes, especially "false positives" (legitimate content being removed or blocked from being shared as posts or comments).<sup>6</sup>

The Coronavirus Misinformation Tracking Center hosted by NewsGuard found 187 (as of early June, up from 132, then 141) sites — 82% of which NewsGuard had identified as unreliable prior to the coronavirus outbreak — that are publishing false claims about the coronavirus.<sup>7</sup> NewsGuard provides credibility ratings and detailed "Nutrition Labels" for thousands of news and information websites. You can read detailed reviews of each website tracked by NewsGuard for more information.

---

2

<https://www.poynter.org/fact-checking/2020/coronavirus-fact-checkers-from-30-countries-are-fighting-3-ways-of-misinformation/>

3

[https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html)

<sup>4</sup><https://youtube-creators.googleblog.com/2020/03/protecting-our-extended-workforce-and.html>

5

<https://www.reuters.com/article/us-health-coronavirus-google/vanishing-youtube-videos-google-expects-ai-errors-as-coronavirus-empties-offices-idUSKBN2133BM>

6

<https://www.theverge.com/2020/3/17/21184445/facebook-marking-coronavirus-posts-spam-misinformation-covid-19>

<sup>7</sup> <https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/>

## Dangers of the Infodemic

### Harms of Misinformation

Early in the pandemic the World Health Organization's director-general ~~has~~ said that "disinformation is as dangerous as COVID-19"<sup>8</sup>. Based on the analysis we have seen, the primary harms of misinformation related to COVID-19 are likely:

- Fear mongering and increasing panic and angst
- Threat to the physical safety of individuals
- Limiting the effectiveness of official efforts to curb the pandemic
- Fostering racism

### Categories of Misinformation

Efforts to counter misinformation need to address both their source and the type of misinformation they represent. The primary types of misinformation we saw related to the COVID-19 pandemic are:

- potential cures
- factors that contribute to disease transmission
- conspiracies about the root cause of the disease<sup>9</sup>

The primary sources of information are:

- both U.S. and foreign clickbait sites, some of which have simply been repurposed for reporting on the pandemic
- celebrities in the entertainment, news and political realm, both online and offline
- bad political actors, both in the U.S. and internationally

These sources may create or amplify information, or both.

Successful strategies for countering misinformation will involve both moderation and removal of misinformation, and empowering the creation and distribution of accurate and reliable information.

---

<sup>8</sup> <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf>

<sup>9</sup> This includes rumors related to the impact of 5G technology on the disease and whether Bill Gates, George Soros, the military, or other "elites" have had a role in creating it.

## Insights & Commentary

Updates on analysis and commentary about platforms' efforts to counter misinformation about the COVID-19 pandemic are reported in reverse chronological order.

### JUNE 3

Forbes reported on studies showing that even early in the pandemic, 1 in 4 videos on YouTube contained false and misleading health information. Since then, the proliferation of videos with false, misleading or harmful information about the COVID-19 pandemic has only intensified, despite social media companies' attempts to rein them in. And unfortunately, very few of the videos containing accurate, verified information came from recognized health authorities like WHO and CDC. The main recommendation coming from the studies was that health authorities, academic medical hospitals and centers should recognize YouTube as an important platform to convey accurate medical information during health crises, using the same urgency and emotion in use by conspiracy theorists and bad actors. However, they also recommended that individual social media users think of YouTube, TV news, and partisan sites as the "sugar and fat" of a news diet: "Here are the protein and vegetables of news: city and national newspapers, public radio and established online publications."<sup>10</sup>

An article in Slate described Wikipedia as potentially the best future source of the historical record of evolving knowledge about the pandemic, since one of its signature design elements is citation and retention of every editorial change. This feature means it inherently addresses the letter sent by 75 signatory organizations in early April asking social media companies and content-sharing platforms to preserve all data that they have blocked or removed during the COVID-19 pandemic and make it available for future research. That means future researchers would be using Wikipedia as a primary source (it is actually a tertiary source and rarely cited for that reason).<sup>11</sup>

A previous report noted that Adam Schiff (D-Pasadena), the Chairman of the House Intelligence Committee, had written letters to several media outlets asking them to describe the actions they were taking to address coronavirus misinformation on their respective platforms. He had asked them to proactively inform users who engage with harmful coronavirus-related misinformation before it can be removed and to direct them to authoritative, medically accurate resources. This week he released their written responses. He said, "I appreciate the steps each platform is taking to reduce Coronavirus misinformation and connect users with authoritative health resources. While it is more effective to limit engagement with harmful content and provide context in real time and before users interact with it, that is not always possible given the scale of these platforms. When unwitting users do engage with false content that could harm them or

---

<sup>10</sup>

<https://www.forbes.com/sites/tarahaelle/2020/05/31/1-in-4-youtube-coronavirus-videos-had-bad-information-long-before-pandemic/#68c6893e768d>

<sup>11</sup> <https://slate.com/technology/2020/05/wikipedia-coronavirus-information-future-historians.html>

their families, they should be informed. As we look ahead to this year's election and beyond, the platforms' investment and responsiveness to misinformation about Coronavirus will be gravely tested, and the health of our society and democracy along with it."<sup>12</sup>

## MAY 27

MIT's Technology Review and NPR - and then many others - reported on a study "associated with" Carnegie Mellon that concluded nearly half of Twitter accounts pushing to reopen America may be bots. The original "statement" about the research described a huge upswell of Twitter bot activity since the start of the coronavirus pandemic, amplifying medical disinformation and the push to reopen America. Across US and foreign elections, natural disasters, and other politicized events, the researchers noted, the level of bot involvement is normally between 10 and 20%. But in this study of 200M tweets, the researchers found that bots may account for between 45 and 60% of Twitter accounts discussing COVID-19. The researchers have begun to analyze Facebook, Reddit, and YouTube to understand how disinformation spreads between platforms. The work was described as being "in the early stages", but already revealing some unexpected patterns.<sup>13</sup>

Yes, but...several researchers and analysts specializing in misinformation tracking and analysis quickly contended that the "statement" was actually a press release, with neither a peer-reviewed study nor a pre-print available for review (more on that below). The release offered no description of data sets or methodology, and was offered by two researchers with a previously loose method for assigning an account to a bot. Twitter also contested the conclusions in the report.<sup>14</sup>

In a report released on May 20, the strategic communications division of the European diplomatic corps, the European External Action Service (EEAS), noted a drop in volume of COVID-19 misinformation on digital platforms but said "it is clear that much more needs to be done." "The work of independent media and fact-checkers is crucial to deliver reliable and authoritative information about the pandemic," said the report, which covers observations and assessments from the period of April 23 to May 18. It also noted that threats and harassment against fact-checkers and fact-checking organizations are being observed. The publication was a follow-up to three previous reports in March and April in which EU monitors identified a "trilateral convergence of disinformation narratives" being promoted by China, Iran, and Russia on the pandemic to undermine public trust. (For anyone with a particular interest in

---

12

<https://www.pasadenanow.com/main/schiff-receives-responses-from-social-media-platforms-regarding-coronavirus-misinformation/>

13

<https://www.technologyreview.com/2020/05/21/1002105/covid-bot-twitter-accounts-push-to-reopen-america/>

<sup>14</sup> <https://twitter.com/evelyndouek/status/1264956411785641986>

misinformation themes from foreign state and non-state actors, this is a good source.)<sup>15</sup>

Executives from Twitter, Facebook and Google are to be recalled by British Parliament in the first week of June to demand more information about the steps they are taking to stamp out coronavirus misinformation. The chair of the Digital, Culture, Media and Sport Committee for the U.K. accused the technology companies of failing to show “clarity and openness” in an earlier evidence session last month. The committee “remains concerned” about the role social media influencers including celebrities and politicians have played in promoting conspiracy theories about the pandemic online.<sup>16</sup>

Steak-umm, the decades-old maker of thin-sliced frozen beef used in cheesesteak sandwiches, has garnered considerable attention on social media during the coronavirus crisis with a Twitter thread warning people to question their news sources amid a torrent of misinformation about the virus. The company implored people to “be careful in our media consumption” and reminded the public that it is crucial to “follow a range of credentialed sources for both breaking news and data collection.” The meat purveyor’s Twitter thread has generated about 13,000 retweets, over 48,000 likes, and hundreds of comments - and a supposedly “dead brand” (by its own admission) has gotten a lot of positive attention.<sup>17</sup>

A research study conducted at Institute for Health System Innovation & Policy at Boston University leveraged volunteer fact checking to identify misinformation about COVID-19 in social media. Identifying emerging health misinformation for CV-19 is a challenge because its manner and type are often unknown. However, many social media users correct misinformation when they encounter it. These researchers implemented a strategy that detected emerging health misinformation by tracking *replies* that seem to provide accurate information. The strategy was more efficient than keyword-based search in identifying COVID-19 misinformation about antibiotics and a cure. It is one of several studies showing the potential value of crowd-sourcing and community engagement to counter misinformation.<sup>18</sup>

Several outlets have accompanied articles about the platforms’ role in countering misinformation with advice for digital users. One of the most recent and simplest offered these five measures to

---

15

<https://euvsdisinfo.eu/eeas-special-report-update-short-assessment-of-narratives-and-disinformation-around-the-covid19-pandemic-updated-23-april-18-may/>

16

<https://www.telegraph.co.uk/technology/2020/05/20/facebook-twitter-google-face-mps-covid-19-misinformation/>

17

<https://www.wsj.com/articles/steak-umm-emerges-as-unlikely-coronavirus-misinformation-watchdog-11586300943>

18

<https://misinforeview.hks.harvard.edu/article/leveraging-volunteer-fact-checking-to-identify-misinformation-about-covid-19-in-social-media/>

prevent the spread of misinformation (or to prompt the platforms to do so):<sup>19</sup>

1. Be critical when you look at social media.
2. Don't leave false information in your online networks. You can politely ask the person who shared it to remove it.
3. Report the false information to the platform administrators.
4. When in doubt, take the time to verify the shared information.
5. Make more noise than people who share false information.

## MAY 20

Washington Post reported on the myriad new workarounds being used to thwart companies taking a tougher line on misinformation during the pandemic. "More and different actors" are participating in tactical innovation, including using work productivity tools like Google Drive and exploiting the Internet Archive, a critical resource for researchers as reported in last month's report. Actors include links to Drive or the Archive in social media posts, or post parts of removed videos with the most troubling parts edited out, or use CDC links to try to get content upranked by algorithms.<sup>20</sup>

Mozilla published a thoughtful piece highlighting key challenges and opportunities to improve platform regulation in regard to content moderation; it included the limitations of automation and filtering, the gaps in transparency and consistency of rules, and some of the people and organizations surfacing social, technical and legal alternatives. Mozilla used a storytelling approach to help illustrate the need for balance among competing interests including public pressure, desire for regulation, and the impact of content moderation on the physical and mental health of human moderators.<sup>21</sup>

An article in Brookings Institution's TechStream raised the question of whether it is worth considering whether the near-blanket liability protections granted to social media companies for content posted on their platforms should apply to questions of public health. Their conclusion, based on past regulation of deceptive drug ads: "Chipping away at liability protections has emerged as the favorite tool of Washington to hold big platforms to account, but it is a blunt instrument that legislators should be wary of deploying....it could prove hard to enforce and a disaster to implement." Their rationale: 1) it is unclear that FOSTA-SESTA, the last attempt to temper 230, worked as intended; there is a good argument to be made that the FOSTA-SESTA bill has actually endangered sex workers by driving their business even further underground, and it has not stopped the ads. 2) the boundaries of public health information are blurry. Who will decide where wellness begins and health ends, particularly when top administration officials

---

<sup>19</sup>

<https://theconversation.com/5-ways-to-help-stop-the-infodemic-the-increasing-misinformation-about-coronavirus-137561>

<sup>20</sup>

<https://www.washingtonpost.com/technology/2020/05/20/misinformation-coronavirus-plandemic-workaround/>

<sup>21</sup> <https://foundation.mozilla.org/en/blog/when-content-moderation-hurts/>

are themselves spreading health misinformation. 3) the worst offenders are sometimes older forms of media, including partisan media. Any new regulation would have to encompass them too in order to really protect public health -- something that the current administration and Republicans in Congress would be loath to do.<sup>22</sup>

Preprints - non-peer-reviewed reports on scientific studies - are being weaponized during the pandemic. Preprints are meant to help scientists find and discuss new findings in real time, which is especially important during a pandemic. They generally carry a warning label: "This research has yet to be peer reviewed." To a scientist, this means it's provisional knowledge — maybe true, maybe not, and not yet passed through the primary means of academic quality control. But for partisan news media, anything carrying the mark of a respected institution counts as knowledge, particularly when it reinforces the day's talking points.<sup>23</sup>

Mainstream media coverage has added to the problem of digital content moderation and compounded the challenges of conveying accurate health information, according to the Harvard Global Health Institute. At many major news outlets, reporters and editors with no medical or public health training were quickly reassigned to cover the unfolding pandemic and are scrambling to get up to speed with complex scientific terminology, methodologies, and research, and then identify, as well as vet, a roster of credible sources. Because many are not yet knowledgeable enough to report critically and authoritatively on the science, they can sometimes lean too heavily on traditional journalism values like balance, novelty, and conflict. In doing so, they lift up outlier and inaccurate counterarguments and hypotheses, unnecessarily muddying the water. Then there is the problem of political bias. This has been especially true at right-leaning media outlets, which have largely repeated news angles and viewpoints promoted by the White House and the president on the progress of the pandemic and the efficacy of the administration's response, boosting unproven COVID-19 treatments and exaggerating the availability of testing and safety equipment and prospects for speedy vaccine development.<sup>24</sup>

One in four of the most popular English-language YouTube videos about the coronavirus contains misinformation, according to a study in the journal *BMJ Global Health*. For the study, researchers from the University of Ottawa analyzed 69 of the most widely-viewed English language videos from a single day in March and found 19 contained non-factual information, garnering more than 62 million views. Misinformation, according to the researchers, included any video that contained false information on the transmission, symptoms, prevention strategies, treatments and epidemiology of the coronavirus. Internet news sources were most likely to misinform, though entertainment, network and internet news outlets were all sources of misinformation, according to the study. *None of the most popular professional and government videos contained misinformation.*<sup>25</sup>

---

<sup>22</sup> <https://www.brookings.edu/techstream/should-there-be-a-public-health-exemption-for-section-230/>

<sup>23</sup> <https://www.nytimes.com/2020/05/14/opinion/coronavirus-research-misinformation.html>

<sup>24</sup>

<https://news.harvard.edu/gazette/story/2020/05/social-media-used-to-spread-create-covid-19-falsehoods>

<sup>25</sup> <https://www.nbcnews.com/health/health-news/live-blog/2020-05-13-coronavirus-news-n1205916>

Researchers at the George Washington University developed a first-of-its-kind map to track the vaccine conversation among 100 million Facebook users during the height of the 2019 measles outbreak. The new study and its "battleground" map reveal how distrust in establishment health guidance could jeopardize public health efforts to protect populations from COVID-19 and future pandemics through vaccinations. They discovered that, while there are fewer individuals with anti-vaccination sentiments on Facebook than with pro-vaccination sentiments, there are nearly three times the number of anti-vaccination communities on Facebook than pro-vaccination communities. This allows anti-vaccination communities to become highly entangled with undecided communities, while pro-vaccination communities remain mostly peripheral. In addition, pro-vaccination communities which focused on countering larger anti-vaccination communities may be missing medium-sized ones growing under the radar. The researchers also found anti-vaccination communities offer more diverse narratives around vaccines and other established health treatments -- promoting safety concerns, conspiracy theories or individual choice, thus increasing the chances of influencing individuals in undecided communities. Pro-vaccination communities, on the other hand, mostly offered monothematic messaging typically focused on the established public health benefits of vaccinations. In their study, the GW researchers proposed several different strategies to fight against online disinformation, including influencing the heterogeneity of individual communities to delay onset and decrease their growth and manipulating the links between communities in order to prevent the spread of negative views. "Instead of playing whack-a-mole with a global network of communities that consume and produce (mis)information, public health agencies, social media platforms and governments can use a map like ours and an entirely new set of strategies to identify where the largest theaters of online activity are and engage and neutralize those communities peddling in misinformation so harmful to the public," Dr. Johnson said.<sup>26</sup>

A survey conducted by Flixed, a site that helps viewers manage their media consumption, reported that Facebook was the most-used platform for news related to the coronavirus. In the survey, 35.8% of people use Facebook as their primary social media platform for news about the pandemic. The social media network was followed in popularity by Twitter (17.0%), YouTube (16.3%), Reddit (12.4%), and other platforms (12.7%) for news about the virus. 5.8% reported that they do not use social media. However, a majority of people (57.6%) who tended to use Reddit for their primary pandemic news source were more likely to report worsening mental health since the start of 2020. Among Facebook users, 41.6% reported a decline in mental health, followed by Twitter (43.0%), YouTube (32.2%), and other platforms (32.5%). The lowest likelihood of worse mental health during the pandemic was reported by those who did not use social media (26.2%).<sup>27</sup>

## MAY 13

An article in Axios highlighted the challenges the digital platforms face in distinguishing coordinated information warfare (disinformation) from false claims spread by people who sincerely believe them (misinformation). Although Facebook, YouTube and Twitter have gotten

<sup>26</sup> <https://www.sciencedaily.com/releases/2020/05/200513111440.htm>

<sup>27</sup> <https://flixed.io/insights/covid19mediadiet/>

better at spotting and stopping disinformation, two factors make it challenging to “stem the coronavirus misinformation tide”: it’s a new disease and there’s a lot we don’t actually know for sure, making it hard for content moderators to draw clear distinctions between what’s true and what’s not, and 2) enough business and political leaders have lined up in opposition to the scientific consensus that fringe positions have moved into the mainstream.<sup>28</sup>

An article in Brookings Institution’s TechStream maintains that the greatest information problem introduced by the pandemic is less one of *moderation* (of identifying and removing content that is demonstrably false and/or harmful), and more one of *mediation* (identifying what information is credible, when, and how to communicate these changes). By identifying good sources of information and highlighting them, platforms can reduce the need to address bad information that is quickly gaining visibility and engagement over algorithmically determined spaces. However, we must ask whether we trust tech companies to play this role of reconciling the user-generated internet with hierarchies of knowledge production.<sup>29</sup>

Researchers in the Technology and Social Change Research Project at Harvard Kennedy’s Shorenstein Center found that pandemic conspiracy theorists are using the Internet Archive’s Wayback Machine to promote “zombie content” that evades moderators and fact-checkers on the digital platforms. Even after content is initially removed from platforms, versions of it saved on the Internet Archive’s Wayback Machine can flourish on Facebook with high engagement with its links. Some people use the Internet Archive to evade blocking of banned domains in their home country, but it is not simply about censorship. Others are seeking to get around fact-checking and algorithmic demotion of content. The research shows one way harmful conspiracies permeate private pages and groups on Facebook, and that health misinformation circulates in spaces where journalists, independent researchers, and public health advocates cannot assess it or counterbalance these false claims with facts.<sup>30</sup>

An army of bot accounts linked to an alleged Chinese government-backed propaganda campaign is spreading disinformation on social media about coronavirus and other topics, according to a London-based researcher. The accounts have been used to promote content attacking critics of the Chinese government and to spread conspiracy theories blaming the U.S. for the origins of the virus, according to Benjamin Strick, who specializes in analyzing information operations on social media websites. Based on the number and velocity of the campaign on Facebook and Twitter he believes it is a state-backed Chinese campaign.<sup>31</sup>

---

28

<https://www.axios.com/coronavirus-misinformation-foe-tech-a5b347e9-99d6-4d4c-9232-02e405253427.html>

29 <https://www.brookings.edu/techstream/covid-19-misinformation-is-a-crisis-of-content-mediation/>

30

<https://www.technologyreview.com/2020/04/30/1000881/covid-hoaxes-zombie-content-wayback-machine-disinformation/>

31

<https://www.bloomberg.com/news/articles/2020-05-12/china-s-disinformation-campaign-targets-virus-and-businessman>

International nonprofit research center First Draft noted that a unique challenge facing journalists and fact-checkers around COVID-19 is that “what was true last month was true then, but now we know something different. Each week there is something new, and this is really challenging to communicate to the public.” Another challenge is that content once focused in online communities has now gone into closed messaging apps like WhatsApp, Facebook Messenger, Signal and Telegram. They report “better behavior and citizenship from all of the [major] platforms over the past six months” including efforts from the platforms to raise up quality information and root out problematic posts. First Draft asks for more collaboration across platforms and transparency about their rules and their enforcement of the rules.<sup>32</sup>

The Washington Post reported on new research that indicates “observational correction” - correction that occurs on social media where people can observe other people being corrected - can be a highly effective strategy for countering misinformation on digital platforms. The research also found that people are engaging in the practice, and that attitudes about this kind of correction were highly positive. The latter two points were true for people across the political spectrum. The research suggests the most effective response when witnessing misinformation is to provide credible information that the misinformation is incorrect — and offering facts in response. A fact check from an independent journalistic organization or information from a credible organization like the American Medical Association is particularly effective. Making these types of corrections even if someone else has already done so reinforces the true message to those who see it. It’s also important to note that people generally agree that correction is appropriate, and that it’s a shared responsibility. This may make people on social media more comfortable with correcting others, and more likely to engage in it more often. The approach likely works because 1) the correction occurs in proximity (temporally and spatially) to the original misinformation, increasing the likelihood that people hadn’t had a chance to absorb the misinformation at all and 2) witnessing someone else being corrected may be less threatening than being corrected directly, but with all the same benefits. When highly trusted groups like the CDC directly respond to users sharing misinformation on social media, people are likely to believe the correction. Correction can also come from social media platforms themselves. For example, Facebook uses its “related articles” function to display debunking information from third-party fact-checkers.<sup>33</sup>

A peer-reviewed study about how search engines disseminate information about COVID-19 found that different search engines (Baidu, Bing, DuckDuckGo, Google, Yandex, and Yahoo) prioritize specific categories of information sources, such as government-related websites or alternative media - even in non-personalized search results. It also found that source ranking within the same search engine is subjected to randomization, which can result in unequal access to information among users. The degree of randomization varied, and could mean that different users are exposed to different information. This can be detrimental when society urgently needs to access consistent and accurate information. If we assume that a major driver

---

<sup>32</sup> <https://cpj.org/blog/2020/05/first-draft-aimee-rinehart-fact-checking-coronavirus-misinformation.php>

<sup>33</sup>

<https://www.washingtonpost.com/politics/2020/05/07/americans-are-fighting-coronavirus-misinformation-social-media/>

of randomization is the maximization of user engagement by testing different ways of ranking search results, it would mean companies' private interests directly interfere with the people's rights to access accurate and verifiable information. The study raises important questions like, what is "good" information, who should decide on its quality and can these decisions be applied univocally?<sup>34</sup>

Beginning May 4, all the social media networks struggled to completely remove or stop the spread of a particularly virulent viral video featuring a well-known vaccine conspiracist, which contains false, misleading or unproven claims about Covid-19. The roughly 26-minute "Plandemic Movie" video claims to be an excerpt of a larger documentary to be released this summer and contains claims about the origins of the virus and how it spreads. In a matter of hours, the video became one of the most widespread pieces of coronavirus misinformation, drawing millions of views across major technology platforms.<sup>35</sup> While it received an enormous amount of media attention, several sources noted that the "plandemic" conspiracy theory is small in its online spread in comparison to the Bill Gates conspiracy theory and the disinfectant conspiracy theory.<sup>36</sup>

A Washington Post research team working with a student at Stanford University mapped the spread of a particular story that originated with an article on Medium; the article said the health risks of COVID-19 were overstated and that social distancing would hurt the economy. The spread of the story demonstrated some of the challenges associated with countering misinformation:

- A small group of key social media influencers can amplify the spread of misleading information and boost the long-term profile of previously obscure authors.
- Social media platforms like Twitter interact quickly with other media like cable news; Fox News personalities played a key role in spreading the story.
- Most important, science is being politicized. The article was originally published by a Medium channel associated with the Lincoln Network, a conservative nonprofit organization, and spread by media personalities with the same political orientation. While the downstream spreaders and amplifiers probably weren't intentionally sharing misleading information, their choices presumably reflected their political priorities.<sup>37</sup>

## MAY 6

---

<sup>34</sup>

<https://misinforeview.hks.harvard.edu/article/how-search-engines-disseminate-information-about-covid-19-and-why-they-should-do-better/>

<sup>35</sup>

<https://www.cnbc.com/2020/05/07/facebook-youtube-struggling-to-remove-plandemic-conspiracy-video.html> (and many others)

<sup>36</sup>

<https://www.axios.com/coronavirus-conspiracy-theories-220bab78-9024-4209-ab2d-6daeb97427be.html>

<sup>37</sup>

<https://www.washingtonpost.com/politics/2020/05/08/this-covid-19-misinformation-went-viral-heres-what-we-learned/>

On April 30, Rep. Adam Schiff (D-CA) sent a letter to Google, YouTube, and Twitter urging the platforms to explicitly notify users when they've engaged with misinformation about the coronavirus. Schiff wrote to Google CEO Sundar Pichai, YouTube CEO Susan Wojcicki, and Twitter CEO Jack Dorsey, saying it's not enough to remove or downgrade harmful or misleading content about the pandemic, but that it's critical to ensure that users who saw the content have access to correct information as well. Facebook recently announced plans to display messages to any users who have engaged with harmful coronavirus-related misinformation.<sup>38</sup>

On April 30, the New York Times reported that Facebook, Twitter and YouTube have declined to remove President Trump's statements in a White House briefing that disinfectants and ultraviolet light were possible treatments for the virus, saying he did not specifically direct people to pursue the unproven treatments. His remarks immediately found their way onto Facebook, Instagram and other social media sites, and people rushed to defend the president's statements as well as mock them. That led to a mushrooming of other posts, videos and comments about false virus cures with UV lights and disinfectants that the companies have largely left up.<sup>39</sup>

Wikipedia can be a guide to the big commercial platforms on how to moderate misinformation, according to an article in Wired. While all the major platforms are trying to cleanse their sites of dangerous disinformation, they are doing so by relying in part on familiar, passive tools like acting when others flag dangerous content. Wikipedia shows that extreme circumstances, especially when related to public health, require different, more stringent rules, not better application of existing rules. For example, "you have to cite everything you write", meaning legitimate sourcing filters out mistakes and lies.<sup>40</sup>

An investigation co-led by BBC Click and the UK counter-extremism think-tank Institute of Strategic Dialogue indicated how both extremist political and fringe medical communities have tried to exploit the pandemic online. A review of 150K public Facebook posts sent by 38 far-right groups and pages since January identified five distinct communities, united by the topic of discussion: immigration, Islam, Judaism, LGBT, and Elites. For the first four of these, the scale of activity hadn't increased in volume overall since the lockdown. However, the fifth and largest community - the one concerning the "elites" like Jeff Bezos and Bill Gates - had shown a significant spike in activity during the lockdown. Discussions included the relationship of these "elites" to the "deep state", and their alleged role in causing the pandemic or using a lockdown as a tool of social control.<sup>41</sup>

## APRIL 29

A US national survey conducted during the early days of the COVID-19 spread showed that,

---

38

<https://www.theverge.com/2020/4/30/21243026/facebook-twitter-youtube-coronavirus-covid-19-misinformation-adam-schiff>

<sup>39</sup> <https://www.nytimes.com/2020/04/30/technology/trump-coronavirus-social-media.html>

<sup>40</sup> <https://www.wired.com/story/how-wikipedia-prevents-spread-coronavirus-misinformation/>

<sup>41</sup> <https://www.bbc.com/news/technology-52490430>

above and beyond respondents' political party, mainstream broadcast and print media use (e.g., NBC News, the New York Times) correlated with *more accurate* information about the disease's lethality and/or *more accurate* beliefs about protection from infection. Conservative media use (e.g., Fox News) correlated with conspiracy theories including believing that some in the CDC were exaggerating the seriousness of the virus to undermine the presidency of Donald Trump. Exposure to online outlets such as Google News and Yahoo News correlated with lower belief in the efficacy of regular hand washing and avoiding contact with symptomatic individuals. Exposure to sources such as Facebook, Twitter or YouTube was positively correlated with belief in the efficacy of vitamin C, the belief that the CDC was exaggerating the threat to harm President Trump, and the belief that the virus was created by the US government. The report put forward five recommendations<sup>42</sup>:

1. Proactively put forward communication about disease prevention before a crisis.
2. Focus on debunking beliefs that are considered salient in the population (10% or more) in order to avoid inadvertently increasing awareness of the problematic claim.
3. Establish a baseline for monitoring social media interventions.
4. Place public service announcements, encourage hyperlinks to the CDC information pages, and seek interviews on social (and conservative) outlets whose audiences are less knowledgeable, more misinformed, or more accepting of conspiracy theories.
5. Encourage or fund newspapers to take down paywalls on coverage of medical crises.

According to new research from Carnegie Mellon University, nearly half the “people” talking about the coronavirus pandemic on Twitter are not actually people, but bots. They are feeding Twitter with harmful, false story lines about the pandemic, including some inspiring real-world activity, such as the theory that 5G towers cause COVID-19, or state-sponsored propaganda from Russia and China that falsely claims the U.S. developed the coronavirus as a bioweapon or that American politicians are issuing “mandatory” lockdowns. In many ways, the bots are acting in ways that are consistent with the story lines that are coming out of Russia or China, according to researchers. The Carnegie Mellon team identified more than 100 false narratives relating to coronavirus worldwide, which they divided into six different categories: cures or preventative measures, weaponization of the virus, emergency responses, the nature of the virus (like children being immune to it), self-diagnosis methods, and feel-good stories, like dolphins returning to Venice's canals.<sup>43</sup>

China, Iran, and Russia are each using the COVID-19 pandemic as an opportunity to spread disinformation related to the United States, according to a State Department report viewed by Business Insider. The messaging from each government aligns with the others, the report says. They include the baseless narratives that the coronavirus is an American bioweapon and is

---

42

<https://misinforeview.hks.harvard.edu/article/the-relation-between-media-consumption-and-misinformation-at-the-outset-of-the-sars-cov-2-pandemic-in-the-us/>

43

[https://www.vice.com/en\\_asia/article/dygnwz/if-youre-talking-about-coronavirus-on-twitter-youre-probably-a-bot](https://www.vice.com/en_asia/article/dygnwz/if-youre-talking-about-coronavirus-on-twitter-youre-probably-a-bot)

being spread by US troops, that the US is scoring political points from the crisis, and that all three governments — unlike the US — are managing the crisis well, according to the document. The report, which is produced by the department's Global Engagement Center, is not public. The report makes the case that propaganda from the three governments have converged as coronavirus has spread. Some of the information is produced by state-run media, and some has been put out by the governments, the report says.<sup>44</sup>

On the April 28 edition of his show on Fox News, Tucker Carlson complained about the "ludicrous" measures taken by leading technology companies to combat misinformation about the coronavirus pandemic. Referring primarily to YouTube's removal of a highly circulated video featuring Dr. Daniel Erickson, in which Erickson alleged that doctors were encouraged to link deaths to COVID-19 to amplify concerns about the pandemic, Carlson accused the service of "cracking down on free expression" and "banning dissent" from medical orthodoxy. Carlson maintained that "the big technology companies are using this tragedy to increase their power over the population".<sup>45</sup>

## APRIL 22

One of the most widely circulated stories of the week, from the Associated Press, took a positive view of platforms' efforts to counter COVID-19 misinformation. Under the headline "Tech companies step up fight against bad coronavirus info", the article noted, "Facebook, Google and others have begun using algorithms, new rules and factual warnings to knock down harmful coronavirus conspiracy theories, questionable ads and unproven remedies that regularly crop up on their services — and which could jeopardize lives. Health officials, critics and others who have long implored the tech companies to step up their response to viral falsehoods have welcomed the new effort, saying the platforms are now working faster than ever to scrub their sites of coronavirus misinformation".<sup>46</sup>

As noted in previous editions, media and opinion writers have begun to speculate about whether the platforms' efforts to counter misinformation about COVID-19 could - or should - be expanded beyond the epidemic. For example, this week in Politico, Paul Barrett, deputy director of the NYU Stern Center for Business and Human Rights, noted, "While this system might be imperfect, it shows that fact-checking is possible, and it works...what I'm urging is an expansion of fact-checking, not to serve a partisan agenda, but to limit the amount of mis- and disinformation polluting American public life. That's just a public service."<sup>47</sup>

In The Atlantic, Evelyn Douek, an Affiliate at Harvard's Berkman Klein Center for Internet and Society, notes that the platforms have rewritten "a silent constitution that bound users and the

---

44

<https://www.businessinsider.com/russia-china-iran-propaganda-us-disinformation-narratives-echo-each-other-2020-4>

<sup>45</sup> <https://www.foxnews.com/media/tucker-carlson-youtube-coronavirus-censorship-science>

<sup>46</sup> <https://apnews.com/88ccd8d2714998cb06fb88639c271af6>

<sup>47</sup> <https://www.politico.com/news/agenda/2020/04/14/social-media-coronavirus-184438>

platforms themselves”, exposing just how much power they can exercise when they decide to do so. First, many platforms have adopted new rules specifically addressing coronavirus-related content. Second, enforcement during the state of emergency is swift and blunt, largely the work of automated tools. Third, even with these sweeping new rules and blunter enforcement, platforms have been suspending their usual due-process protections. She notes that while the tech companies’ actions in the current crisis may deserve praise, they also raise important questions about checks or constraints, whether these emergency powers are temporary, and the role of oversight.<sup>48</sup>

U.N. Secretary-General Antonio Guterres said the world is fighting a “dangerous epidemic of misinformation” about the coronavirus and has announced a UN initiative to disseminate the facts and science about the pandemic. The United Nations Communications Response initiative will “flood the internet with facts and science while countering the growing scourge of misinformation — a poison that is putting even more lives at risk.”<sup>49</sup>

Ad tech executives - representing the marketers whose advertising fuels the platforms’ business model - diverge on the responsibility and capability of Google and Twitter to monitor the veracity of the content they host during the crisis. Some execs said brands should seek safer environments than the social giants, others expressed a measured amount of trust in Google and Twitter’s capacity to root out fake news, and still others said preventing false content is impossible and up to readers to sort through. One VP of insight said, “As an advertiser, the safer bet is to partner with trusted and reliable sites that have teams of journalists vetting every piece of content”. A media VP claimed Google and Twitter are aware of the harm misinformation can cause on this issue, meaning advertisers can trust that coronavirus-related content will be “closely monitored,” though another noted that the content vetting process will be imperfect. But an agency COO referred to the “fundamental insurmountability” of eliminating all false content on platforms teeming with information when the truth itself changes day to day.<sup>50</sup>

One of the most-reported updates of the week was Facebook’s blog post (noted in the chart below) that they would start showing messages in News Feed to people who have liked, reacted or commented on harmful misinformation about COVID-19 that was subsequently removed. It contradicted past reports in which the company has said it is “challenging” to reliably identify and notify everyone who had been incidentally exposed to (in that case) foreign propaganda.<sup>51</sup>

---

48

<https://www.theatlantic.com/ideas/archive/2020/04/pandemic-facebook-and-twitter-grab-more-power/610213/>

49

<https://thehill.com/changing-america/well-being/longevity/492918-un-kicks-off-initiative-to-combat-coronavirus>

50

<https://streetfightmag.com/2020/04/17/ad-tech-exec-says-weigh-in-on-coronavirus-ads-google-and-twitter-and-misinformation>

51

<https://www.theguardian.com/technology/2017/nov/23/facebook-to-tell-users-if-they-interacted-with-russia-troll-army>

The design of the message came under a lot of criticism for being too soft.<sup>52</sup> It says: “Help friends and family avoid false information about COVID-19.” It then invites them to share a link to the WHO’s myth-busting site, as well as a button that will take the user to the site directly. Users are never directly told they have engaged with misinformation, or why they got the message, and the WHO site currently debunks 19 different hoaxes. However, Facebook maintains the “nudging” approach is based on past research that people were more likely to share posts that had been labeled “disputed”, and that the design and language may evolve.<sup>53</sup> The debate shows the challenges of designing methods of countering misinformation without research or evidence of their effectiveness.

Several outlets reported on a study by activist organization Avaaz specifically focused on Facebook’s response to the infodemic.<sup>54</sup> Avaaz examined over 100 pieces of misinformation content in six different languages about the virus that were rated false and misleading by reputable, independent fact-checkers and could cause public harm. After noting “the commendable efforts Facebook’s anti-misinformation team has applied to fight this infodemic”, and that “the company’s efforts to combat the problem had steadily improved”, Avaaz claimed the platform’s current policies were insufficient and did not protect its users. Specifically, Avaaz found that Facebook’s approaches are 1) incomplete; 2) delayed; 3) English language-focused; and 4) unable to address how individual stories (“mother stories”) mutate and spread (“babies”). They shared their findings directly with Facebook along with recommendations for more overt notification to users who have interacted with misinformation. Avaaz’ report was considered instrumental in Facebook’s decision to add messages in News Feed (see above), but Avaaz noted that “the step doesn’t reflect the full gamut of what we would like to see them do.”<sup>55</sup>

Avaaz also published an academic study they say proves that providing social media users who have seen false or misleading information corrections from fact-checkers can significantly decrease belief in disinformation. The research, commissioned by Avaaz, was conducted by Dr. Ethan Porter of George Washington University and Dr. Tom Wood of Ohio State University, authors of “False Alarm: The Truth About Political Mistruths in the Trump Era”, which also showed corrections can reduce the share of inaccurate beliefs. In the test, which used a hyper-real visual model of Facebook, “correcting the record” reduced belief in disinformation by half, on average, and worked across party affiliation and political ideology.

Avaaz maintains that showing fact-checked corrections to every single user exposed to viral disinformation is currently one of the strongest defenses we have against coordinated disinformation campaigns.<sup>56</sup>

---

52

<https://www.vox.com/recode/2020/4/16/21223972/facebook-coronavirus-hoaxes-warning-misinformation-avaaz>

53

<https://www.theverge.com/interface/2020/4/17/21223742/coronavirus-misinformation-facebook-who-news-feed-message-avaaz-report>

54 [https://secure.avaaz.org/campaign/en/facebook\\_coronavirus\\_misinformation](https://secure.avaaz.org/campaign/en/facebook_coronavirus_misinformation)

55 <https://www.politico.com/news/2020/04/16/facebook-fake-news-coronavirus-190054>

56 [https://secure.avaaz.org/campaign/en/correct\\_the\\_record\\_study/](https://secure.avaaz.org/campaign/en/correct_the_record_study/)

A group of scholars and nonprofit organizations have asked web platforms to keep track of the content they're removing during the coronavirus pandemic so they can make it available to researchers studying how online information affects public health. The signatories — including Access Now, the Committee to Protect Journalists, and EU DisinfoLab — sent an open letter to social media and content sharing services, urging them to preserve data even as they remove misinformation. The letter urges companies to preserve content that is removed from the service, including accounts, posts, and videos. It also encourages them to keep records of the removal process itself, like whether a takedown was automated or received human oversight, whether users tried to appeal the takedown, and whether content was reported but left online. Some of that information could be included in public transparency reports, and other pieces could be released specifically to researchers.<sup>57</sup>

Research results in a working paper from the Becker Friedman Institute of Economics at the University of Chicago “indicate that provision of misinformation in the early stages of a pandemic can have important consequences for how a disease ultimately affects the population”. The study, focused on news coverage of the novel coronavirus by Hannity and Tucker Carlson Tonight, both on Fox News, showed that greater exposure to Hannity, who originally dismissed the risks associated with the virus relative to Tucker Carlson Tonight, is associated with a greater number of county-level cases and deaths.<sup>58</sup>

In an effort to combat misinformation, doctors and health professionals have taken to social media. Using popular platforms such as tik tok, Snapchat, Instagram, and Facebook, health professionals are increasingly creating interesting content that also helps spread accurate health information.<sup>59</sup>

The U.S. State Department has assessed that Russia, China and Iran are mounting increasingly intense and coordinated disinformation campaigns against the U.S. relating to the outbreak of the new coronavirus. All three countries are using state-controlled media, social media and government agencies and officials to disseminate information to domestic audiences and global audiences alike that denigrates the U.S. and spreads false accounts, the State Department report says.<sup>60</sup>

## APRIL 15

All updates this week were listed under individual platforms.

## APRIL 8

---

<sup>57</sup>

<https://www.theverge.com/2020/4/22/21231061/cpj-access-now-researchers-letter-social-media-platform-coronavirus-preserve-misinformation-data>

<sup>58</sup> <https://bfi.uchicago.edu/working-paper/2020-44/>

<sup>59</sup> <https://www.phillymag.com/be-well-philly/2020/04/17/social-media/>

<sup>60</sup>

<https://www.wsj.com/articles/u-s-adversaries-are-accelerating-coordinating-coronavirus-disinformation-report-says-11587514724>

Reuters reported on April 7 that India has told Facebook and TikTok to remove users that spread misinformation about coronavirus. India is concerned about videos intended to mislead Muslims. This request came after Voyager Infosec, a Delhi-based digital analytics firm, identified a pattern of misinformation videos using religious beliefs to justify defying health advisories.<sup>61</sup>

On April 7, Washington Post reported on a study of 225 pieces of English-language misinformation (88% of which appeared on social media platforms) rated false or misleading by fact-checkers between January and the end of March 2020, conducted by Oxford Internet Institute, Reuters Institute for the Study of Journalism and the Oxford Martin School. The researchers found that misinformation about COVID-19 comes in many different forms, from many different sources, and makes many different claims. It most frequently reconfigures existing or true content rather than fabricating it wholesale, and where it is manipulated, it is edited with simple tools (not deep fakes or other AI-based tools). High-level politicians and celebrities produced or spread only 20% of the misinformation, but that content attracted a large majority of all social media engagements. The most common claims of misinformation concern the actions or policies that public authorities are taking to address COVID-19, and the spread of the virus through individual communities (e.g., date of first case or claims of where it came from). There was very significant variation from company to company: while 59% of false posts remained active on Twitter with no direct warning label, the number was 27% for YouTube and 24% for Facebook.<sup>62</sup> That was the primary focus of the Post's coverage.<sup>63</sup>

In the April 6 episode of GZero Media, Ben Smith, media columnist at The New York Times and former head of BuzzFeed News, described the platforms' efforts to counter mis- and misinformation as "aggressive" and "confident", in part due to the "clarity" of direction provided by world health organizations. But he also said he believes they will return to a position of "neutrality" once the obvious risk of harm has passed. He agreed with host Ian Bremmer that the platforms would emerge from the crisis as "trusted institutions", particularly in contrast to the "dysfunction" of government and "polarization" of traditional media. In the same episode, Danny Rogers of The Global Disinformation Index described the two major themes of current misinformation as "bunk cures" and "weaponized conspiracy theories", including some specifically designed to inflame racial tensions. His reference to coronavirus as "the Super Bowl of disinformation" has also been widely quoted.<sup>64</sup>

An April 3 article in TechCrunch also made the case that the tech companies "may face a respite from focused criticism, particularly with the industry leveraging its extraordinary resources to pitch in with COVID-19 relief efforts". The article noted that platforms have been

---

61

<https://www.reuters.com/article/us-health-coronavirus-india-socialmedia/india-asks-tiktok-facebook-to-remove-users-spreading-coronavirus-misinformation-idUSKBN21P2QI>

62 <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation#menu>

63

<https://www.washingtonpost.com/technology/2020/04/07/twitter-almost-60-percent-false-claims-about-coronavirus-remain-online-without-warning-label/>

64 <https://www.gzeromedia.com/covering-a-pandemic-covid-19-and-the-news-media>

“uncharacteristically transparent” about the shifts the pandemic is creating within their own workflows, including greater use of artificial intelligence, and speculated that social media companies will “have a fresh appreciation for the value of human efforts”. Unfortunately, we may ultimately need to “take companies’ word” for the effectiveness of their efforts.<sup>65</sup>

On April 2, Free Press circulated a petition asking recipients to “immediately tell Facebook, Twitter and YouTube to save lives by doing everything they can to shut down disinformation now” (this was separate from a letter asking the FCC to provide guidelines for broadcast coverage of the pandemic). Free Press asked the platforms to “invest more in people — not algorithms — to identify and remove harmful content and to equip moderators to do this work safely from home”.<sup>66</sup>

Regulators, privacy advocates and others in the U.S. (and Europe) are wrestling with the tension between consumer privacy and the degree of surveillance required to address the pandemic, including for critical contact tracing and assessment of the effectiveness of policies designed to fight it. Some advocates appear willing to make trade-offs (assuming the data is essential and effective, as well as anonymized and aggregated<sup>67</sup>) but worry about data retention, and use beyond the pandemic, as well as issues of equity and discrimination.<sup>68, 69</sup>

In general there is strong support for the platforms’ reliance on health organizations such as the WHO and CDC to guide their content moderation practices. However, an opinion from the Cato Institute highlighted the potential pitfalls of “outsourcing truth” to external authorities for the legitimacy of their moderation practices. They used two examples from Twitter: when some users advocated wearing masks contrary to those authorities, Twitter did not remove their posts. Twitter also tolerated a falsehood from a Chinese government official because it was unlikely to cause immediate harm. Both decisions to tolerate posts that contradicted expert beliefs came from Twitter rather than external authorities. Cato called for more transparency and justification for how Twitter’s own values inform its use of expert knowledge.<sup>70</sup>

---

<sup>65</sup> <https://techcrunch.com/2020/04/03/the-pandemic-is-already-reshaping-techs-misinformation-crisis/>

<sup>66</sup> [https://act.freepress.net/sign/justice\\_covid\\_disinfo\\_platforms](https://act.freepress.net/sign/justice_covid_disinfo_platforms)

<sup>67</sup>

[https://www.washingtonpost.com/opinions/privacy-amid-a-pandemic-may-look-a-little-different/2020/04/03/e1c10c90-74eb-11ea-87da-77a8136c1a6d\\_story.html](https://www.washingtonpost.com/opinions/privacy-amid-a-pandemic-may-look-a-little-different/2020/04/03/e1c10c90-74eb-11ea-87da-77a8136c1a6d_story.html)

<sup>68</sup> <https://www.politico.com/news/2020/04/05/coronavirus-personal-data-privacy-165066>

<sup>69</sup> <https://spectator.us/privacy-death-final-triumph-big-tech/>

<sup>70</sup> <https://www.cato.org/blog/outsourcing-fact-checking-reputation-it>

## APRIL 1

New research from Pew Research Center shows that Americans who get news mostly from social media are least likely to follow COVID-19 coverage; most likely to say they have seen at least some misinformation about the pandemic; fare among the poorest at answering a question about when a vaccine might be available; were more likely than most to say that news sources have exaggerated the threat posed by the virus, and were also slightly more likely than those who turn to other pathways for their news to say that the virus was created in a lab, either intentionally or unintentionally.<sup>71</sup>

While some of the platforms are gaining praise for their efforts to manage misinformation, there are also calls for them to do more. Bhaskar Chakravorti, Dean of Global Business at The Fletcher School at Tufts University and an economist who tracks digital technology's use worldwide, identified three ways to evaluate the companies' responses to the pandemic<sup>72</sup>:

- Are they applying the techniques they have meticulously designed to anticipate the user's experience, hold their attention and influence their actions and behaviors related to the pandemic?
- Are they enforcing responsible advertising policies, including closing loopholes, setting clear industry-wide principles and enforcing firm policies to avoid fraud and misdirection?
- Are they providing data to public health authorities (including geographic information, data about people's movements, high-resolution population density maps, search and location data, trends analyses, depending on the platform) and independent researchers - without compromising privacy?

Efforts by the platforms to remove misinformation are making the news when they involve notable political and media figures. On Friday, March 27, Google confirmed the removal of the Infowars Android app from its Play Store, after the app posted a video in which Alex Jones disputed the need for social distancing, shelter in place, and quarantine efforts meant to slow the spread of the novel coronavirus.<sup>73</sup> Over the past few days Twitter has removed Tweets from Rudy Giuliani and Charlie Kirk<sup>74</sup>, actress and activist Alyssa Milano<sup>75</sup>, conservative magazine The Federalist (which also had its account restricted), and Laura Ingraham<sup>76</sup> for violating its terms of service when they posted false information about cures for the coronavirus. Facebook

71

<https://www.journalism.org/2020/03/25/americans-who-primarily-get-news-through-social-media-are-least-likely-to-follow-covid-19-coverage-most-likely-to-report-seeing-made-up-news/>

72

<https://fletcher.tufts.edu/news-events/news/social-media-companies-are-taking-steps-tamp-down-coronavirus-misinformation-they>

<sup>73</sup> <https://www.wired.com/story/google-bans-infowars-android-app-alex-jones-coronavirus/>

74

<https://www.mediaite.com/news/twitter-deletes-rudy-giuliani-tweet-featuring-coronavirus-misinformation-and-false-attack-on-gov-whitmer/>

<sup>75</sup> <https://www.vanityfair.com/news/2020/03/conservative-pumping-out-coronavirus-misinformation>

<sup>76</sup> <https://variety.com/2020/digital/news/twitter-fox-news-laura-ingraham-delete-coronavirus-1203549252/>

removed a post from Brazilian President Jair Bolsonaro in which he claimed that hydroxychloroquine is working to cure the virus<sup>77</sup>. Conversely, Twitter came under fire for NOT removing a post from Elon Musk, founder of Tesla and SpaceX, with a false assertion about the coronavirus, and for removing one from John McAfee, founder of the eponymous security solutions company, only after it had been widely shared<sup>78</sup>. And as of March 28 Facebook maintained that The Federalist's post did not violate its policies.<sup>79</sup>

There are mixed views in the media about whether the platforms' aggressive practices for managing misinformation about the coronavirus should extend beyond the pandemic, including into political content and the electoral process. An article in Foreign Affairs noted, "The platforms' approach to pandemic information has been aggressive, effective, and necessary—but it cannot and should not be applied to politics." The article notes that false stories about health information are easier to detect; enable more effective moderation by artificial intelligence; are easier to establish evidentiary standards; more conducive to a consensus; and are far less subjective or likely to provoke controversy. Further, it notes, "False speech about politics is a necessary byproduct of living in a free society (unless it runs afoul of carefully circumscribed laws against libel and slander). Identifying false claims about politics is a laborious affair that requires difficult judgments about the nature of truth. As a result, the social consensus in favor of reducing political misinformation on social media is more limited."<sup>80</sup>

Some reports, including a New York Times article, have expressed concern that the platforms' all-consuming efforts to combat misinformation related to the virus will deter their ability to address new practices being used by both domestic and foreign players to influence the 2020 election. Strains on their technical infrastructure and the challenges of coordinating "a vast election effort spanning multiple teams and government agencies" from employees' homes may increase the difficulty.<sup>81</sup>

This week there was a flurry of articles encouraging digital literacy; that is, encouraging consumers of media to improve their own ability to detect and avoid sharing of misinformation. Sources ranged from The Daily Tar Heel, the student newspaper at the University of North Carolina at Chapel Hill, which asked its community to "do our part to flatten the misinformation curve and fight fake news whenever we see it"<sup>82</sup>, to Atlantic magazine, which noted: "Here's

---

<sup>77</sup> <https://techcrunch.com/2020/03/30/facebook-removes-bolsonaro-video/>

<sup>78</sup>

<https://theconversation.com/social-media-companies-are-taking-steps-to-tamp-down-coronavirus-misinformation-but-they-can-do-more-133335>

<sup>79</sup>

[https://www.huffpost.com/entry/facebook-coronavirus-infection-parties-misinformation-policy\\_n\\_5e7faf6cc5b6cb9dc1a16f36](https://www.huffpost.com/entry/facebook-coronavirus-infection-parties-misinformation-policy_n_5e7faf6cc5b6cb9dc1a16f36)

<sup>80</sup> <https://www.foreignaffairs.com/articles/2020-03-30/coronavirus-fake-news-isnt-other-fake-news>

<sup>81</sup> <https://www.nytimes.com/2020/03/29/technology/facebook-google-twitter-november-election.html>

<sup>82</sup> <https://www.dailytarheel.com/article/2020/03/fake-news-covid-0330>

How to Fight Coronavirus Misinformation: Send this to the person in your life who needs to read it".<sup>83</sup> The Advisory Board, a research organization for leaders in the healthcare industry, offered a guide to why misinformation spreads so readily, and how to spot it and deter its spread.<sup>84</sup>

The Global Disinformation Index (GDI), a U.K. research organization that provides the advertising community with non-partisan and independent ratings of a site's disinformation risks, released a new report this week specifically focused on coronavirus disinformation sites. It showed that ad tech players continue to serve up ads and provide ad revenue streams to known disinformation sites peddling coronavirus conspiracies. That means they are placing unknowing brands' advertisements (in the survey these included Amazon Prime Video, Hyundai, Jeep, Samsung, Wayfair, Spotify and others) on websites with false claims about the pandemic. In a sample survey of nearly 50 sites carrying coronavirus conspiracies in the U.S., U.K. and Germany, GDI found Google provided ad services to 86% of these sites. (Google responded, "Similar to past reports, this report is flawed. GDI doesn't detail how it defines disinformation, nor does it provide the full list of domains examined.")<sup>85</sup> For more information about GDI's methodology, see the footnote below.<sup>86</sup>

On March 25, a group of attorneys general led by Pennsylvania Attorney General Josh Shapiro sent letters to Amazon, Walmart, Craigslist, Facebook and eBay asking them to create and enforce policies aimed at preventing price-gouging of products related to COVID-19. "We believe you have an ethical obligation and duty to help your fellow citizens in this time of need by doing everything in your power to stop price gouging in real-time". The letters cite several examples of price-gouging uncovered by reporters, including a 2-liter bottle of hand sanitizer being sold on Craigslist for \$250, and an 8-ounce bottle being sold on Facebook for \$40. The attorneys general are asking for the platforms' voluntary cooperation.<sup>87</sup>

On March 26, companies from across the tech industry joined with health organizations in the #BuildForCOVID19 global hackathon. WHO, scientists from the Chan Zuckerberg Biohub and experts from other industries will be joined by teams from Facebook, Slack, Pinterest, TikTok, Twitter, WeChat, Giphy, Slow Ventures and more to build tools to help tackle some of the health, economic and community challenges coming from the outbreak.

The White House will join forces with major tech companies to pool supercomputing resources in the battle against the novel coronavirus. The initiative is meant to help researchers gain access to computing power to help "discover new treatments and vaccines." IBM helped launch

---

<sup>83</sup> <https://www.theatlantic.com/ideas/archive/2020/03/heres-how-fight-coronavirus-misinformation/608914/>

<sup>84</sup> <https://www.advisory.com/daily-briefing/2020/03/31/misinformation>

<sup>85</sup> [https://disinformationindex.org/wp-content/uploads/2020/03/AdsAdTech-CoronaVirusSites\\_GDI.pdf](https://disinformationindex.org/wp-content/uploads/2020/03/AdsAdTech-CoronaVirusSites_GDI.pdf)

<sup>86</sup>

[https://disinformationindex.org/wp-content/uploads/2019/12/GDI\\_Index-Methodology\\_Report\\_Dec2019.pdf](https://disinformationindex.org/wp-content/uploads/2019/12/GDI_Index-Methodology_Report_Dec2019.pdf)

<sup>87</sup>

<https://www.mediapost.com/publications/article/349008/tech-platforms-urged-to-crack-down-on-covid-19-rel.html>

the “COVID-19 High Performance Computing Consortium,” and Amazon, Microsoft and Google also confirmed they would participate.<sup>88</sup> A related article noted, “The coronavirus crisis is testing whether Big Tech can restore its reputation in Washington after years of backlash and scrutiny. Companies are presenting themselves as advocates of the U.S. government in its efforts to stop the spread of the virus, keep unemployment down, push out critical public health information and support workers financially. But if the industry doesn’t tread carefully – especially to preserve users’ privacy – companies risk coming out of the crisis looking like villains rather than heroes.”<sup>89</sup>

In an article in Lawfare, Evelyn Douek, an S.J.D candidate at Harvard Law School studying online regulation of speech, called out how different the platforms’ approaches have been for the pandemic, including greater transparency, greater use of interest balancing and greater use of artificial intelligence. She noted, “There are three lessons to be learned from this. First, platforms should be commended for being upfront about the likely effects of this change in approach. But they should also be collecting data about this natural experiment and preparing themselves to be equally as transparent about the actual effects of the change. Second, platforms and lawmakers should remember these announcements in the future. The candid announcements from platforms in recent weeks about the costs of relying on AI tools should be nailed to the door of every legislative body considering such an approach. Finally, regulators and academics need to recognize that these announcements are really just an extreme version of the choices that platforms are making every day. Content moderation at scale is impossible to perform perfectly—platforms have to make millions of decisions a day and cannot get it right in every instance. Because error is inevitable, content moderation system design requires choosing which kinds of errors the system will err on the side of making.” She further noted that “the pandemic is shaping up to be a formative moment for tech companies...we’re asking them to step up, but we also need to keep thinking about how to rein them in”.<sup>90</sup>

Some advocates are asking the platforms to go further: use their massive capabilities in segmentation, microtargeting and personalization to deliver “21st century PSA’s”: personalized information that motivates the most effective actions by consumer group to flatten the curve. In an op-ed in Wired, Tristan Harris, the president and cofounder of the Center for Humane Technology, writes, “This emergency, this moment, calls for a fundamentally new approach to technology—to abandon the myth of neutral metrics and engagement, and restructure technology to prioritize this corrective lens that can help save millions of lives.” He recommends the platforms: change internal performance measures to emphasize the highest-priority actions; use their ad targeting capabilities (including the ranking of posts within feeds, notification delivery, and group suggestions) to personalize behavioral cues; use their persuasive powers of

---

88

<https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2020/03/23/the-technology-202-president-trump-unveils-new-supercomputing-initiative-to-fight-coronavirus/5e779fcc88e0fa101a74f123/>

89

<https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2020/03/18/the-technology-202-coronavirus-crisis-tests-silicon-valley-s-relationship-with-washington/5e710a7d88e0fa101a74bb47/>

<sup>90</sup> <https://www.lawfareblog.com/covid-19-and-social-media-content-moderation>

“social proof”, social norms and signaling (as they have tested with voting); and use their capabilities in localization to assist in relief coordination and coalition-building.<sup>91</sup>

In a related editorial on Medium, Tristan and CHT co-founder Aza Raskin ask the platforms to move from “informing” to “persuading”. They defined five persuasion principles - many of which are already the basis of persuasive ads on the platforms - that could “accelerate life-saving choices”<sup>92</sup>:

-  Social Proof (we do what others do)
-  Make the future feelable (through compelling graphics)
-  ♀ Make it personal (show our actions’ impact on our Friends)
-  ✓ Make it concrete (quizzes, surveys, checklists)
-  👁️ Social comparison (we compare ourselves to others)

## MARCH 27

Joan Donovan, the newly-appointed Research Director of Harvard Kennedy School’s Shorenstein Center and one of the world’s leading analysts of how internet misinformation is seeded and spread, notes that the platforms may in fact be getting better at maintaining a “rolling index” of “tripwire” terms that redirect users to reputable sources of information. However, they are perpetually behind the culture. For example, younger consumers colloquially refer to the virus as ‘rona or “the ‘rona” in social media posts and these are not detected; searches and shares related to the “Chinese virus” exploded after the term was used by President Trump; and text messages and posts about “martial law” and the possibility of a national quarantine were widely shared after several states called upon the National Guard to distribute food or medical supplies.<sup>93</sup>

Some reports also note that misinformation about the virus is significantly more challenging to detect and remove because it is being created and shared so widely by regular social media users: “it is sporadic, not networked”. Most time and investment in countering misinformation has gone into systems to detect, monitor and combat sophisticated digital misinformation campaigns coming from coordinated, state-backed campaigns, often in attempts to influence elections or sow discord.<sup>94</sup> However, there are also reports - most disputed by the relevant governments - that Russian and Chinese media have deployed disinformation campaigns to

---

<sup>91</sup> <https://www.wired.com/story/opinion-this-is-silicon-valleys-chance-to-step-up-for-humanity/>

<sup>92</sup>

[https://medium.com/@HumaneTech\\_/from-inform-to-persuade-how-can-tech-step-up-for-humanity-6360c6c3d03e](https://medium.com/@HumaneTech_/from-inform-to-persuade-how-can-tech-step-up-for-humanity-6360c6c3d03e)

<sup>93</sup> Lecture in “Politics of the Press” class at Harvard Kennedy School, March 22, 2020.

<sup>94</sup>

<https://www.politico.com/news/2020/03/12/social-media-giants-are-fighting-coronavirus-fake-news-its-still-spreading-like-wildfire-127038>

worsen the impact of the coronavirus, generate panic and sow distrust.<sup>95</sup>

Independent researchers have struggled to track misinformation traveling from person to person, or through closed groups of people, through email or texts that are not seen by the general public. Text messages are particularly difficult for independent researchers to trace, especially when messages — like some recent texts about a national quarantine in the U.S. - are delivered as graphical images as opposed to words that computers can more easily analyze. Those pushing misinformation may be changing tactics away from social media to thwart the major platforms' efforts to catch and block falsehoods. The sophistication of the campaign about the U.S. quarantine resulted in an interagency effort — involving the NSC, FBI, intelligence agencies, the Department of Homeland Security and the State Department — to determine who is behind the apparent disinformation campaign<sup>96</sup>. Encrypted messaging services such as WhatsApp and iMessage and private groups on Facebook are probably also among the greatest sources of misinformation, but impossible or challenging for researchers to monitor in real time.

A preliminary analysis of the online conversation surrounding the coronavirus pandemic prepared by social network analysis firm Graphika suggests that conservative and right-wing voices played an outsized role in spreading mis- and disinformation online about the coronavirus pandemic worldwide. Graphika produced a set of three global network maps that capture the mainstream global conversation around coronavirus at monthly intervals as part of an effort to map and analyze what the World Health Organization called an “infodemic”. Their key findings were:<sup>97</sup>

1. The online conversation about the pandemic has become more complex over time; a large “mega cluster” of US right-wing accounts became diminished by the mainstreaming of the coronavirus conversation online over time.
2. Particularly in the US, Italy, and France, more right-wing accounts are involved in the conversation and these accounts are more active in their engagement than their leftwing counterparts.
3. A number of groups are leveraging the conversation around coronavirus to propagate racism and anti-immigration sentiment, or to draw attention to immigration policy in their respective countries.
4. At first, conspiracy theories appeared to revolve around the causes of the outbreak, but as the pandemic continued to spread, conspiratorial content has become more closely focused on governmental responses to the outbreak.
5. Habitual sharers of health misinformation increased their share of voice in the coronavirus in February; the data highlights a serious distrust of established and official sources of health information.

---

<sup>95</sup>

<https://www.reuters.com/article/us-health-coronavirus-disinformation/russia-deploying-coronavirus-disinformation-to-sow-panic-in-west-eu-document-says-idUSKBN21518F>

<sup>96</sup> <https://www.washingtonpost.com/technology/2020/03/16/disinfo-texts-trump-quarantine/>

<sup>97</sup> [https://public-assets.graphika.com/reports/Graphika\\_Report\\_Covid19\\_Infodemic.pdf](https://public-assets.graphika.com/reports/Graphika_Report_Covid19_Infodemic.pdf)



6. Narratives designed to stoke geopolitical tensions, including some seeming to originate with the Kremlin, seem focused on undermining trust in global institutions and drawing attention to the failures of other governments, predominantly the Chinese response.

The core narratives of the misinformation efforts were: racism and xenophobia, conspiracy communities, health misinformation, and geopolitical tensions. Conversely, the spread of #FlattenTheCurve is a small case study in the efficacy of science communication that attests to the positive, and likely life-saving, impact of credible information online.

## Case Studies

Digital Platform	Approaches to Disinformation Mitigation
Priority Level 1	
Facebook/Facebook Messenger	<ul style="list-style-type: none"> <li>On May 21, said it would allow many employees to work from home permanently. Mark Zuckerberg told workers during a staff meeting that was live streamed on his Facebook page that within a decade as many as half of the company's more than 48,000 employees would work from home. Employee compensation will be adjusted based on the cost of living in the locations where workers choose to live. However, it was pointed out that none of the articles on the topic referred to contracted workers for content moderation, whose working conditions before and during the pandemic have come under criticism.<sup>98</sup></li> <li>On May 15, a group of Democratic senators sent a letter to Facebook urging the company to take steps to curb coronavirus misinformation that is not in English. The letter from Sens. Bob Menendez (D-N.J.), Mazie Hirono (D-Hawaii) and Elizabeth Warren (D-Mass.) notes that more than 60 million Americans speak another language and cites a study from Avaaz (noted in previous reports) that Facebook fails to issue warning labels on content in those languages at a much higher rate.<sup>99</sup></li> <li>On May 12, released its biannual Community Standards Enforcement Report which provided the latest metrics on the number of controversial posts removed from the platform. The company took action against more than 2 billion pieces of content violating its community standards between January and March 2020 and an upsurge in content promoting hate and violence was particularly noteworthy. Facebook attributed the sharp increase in hate speech removals to technology improvements for automatically identifying images and text.<sup>100</sup> However, AI hasn't played as big a role in handling coronavirus misinformation. Facebook has instead relied primarily on human reviewers at over 60 partner fact-checking</li> </ul>

<sup>98</sup> <https://www.nytimes.com/2020/05/21/technology/facebook-remote-work-coronavirus.html>

<sup>99</sup>

<https://thehill.com/policy/technology/498017-democratic-senators-urge-facebook-to-curb-coronavirus-misinformation-in> and also

<https://www.cnet.com/news/senators-demand-facebook-detail-how-its-combating-non-english-covid-19-misinformation/>

<sup>100</sup>

<https://www.forbes.com/sites/niallmccarthy/2020/05/13/facebook-removes-record-number-of-hate-speech-posts-infographic>

organizations. Only once a person has flagged something, such as an image with a misleading headline, do AI systems take over to search for identical or similar items and automatically add warning labels or take them down. The team hasn't yet been able to train a machine-learning model to find new instances of disinformation itself.<sup>101</sup>

- On May 6, announced the names of the first members of its Oversight Board for content moderation appeals and policy (covered by many outlets).
- On April 22, removed “pseudoscience” from the list of categories advertisers can use to target people, following an investigation from The Markup. The social media platform’s ad portal showed more than 78M Facebook users were interested in pseudoscience, The Markup reported. We’ve removed this targeting option to prevent potential abuse in ads,” a Facebook spokesperson confirmed in an email.<sup>102</sup>
- On April 22, announced extra steps to make Pages and accounts with large audiences more transparent: Facebook will provide the location of high-reach Facebook Pages [and Instagram accounts] on every post they share, so people have more information to help them gauge the reliability and authenticity of the content they see in their feeds. This step, primarily focused on protecting election integrity, will be piloted in the US, starting specifically with Facebook Pages [and Instagram accounts] that are based outside the US but reach large audiences based primarily in the US. It builds on a past effort to include the primary country location of the people who manage a Page.
- On April 21, announced that Facebook was taking steps to bring content reviewers back on a voluntary basis to assist with moderation.<sup>103</sup>
- On April 20, on Good Morning America, Mark Zuckerberg stated that protests of stay-at-home orders in violation of state social distancing rules organized through Facebook qualify as “harmful misinformation.”<sup>104</sup> This does not impact all lockdown protests - only those that are prohibited by government's guidance on social distancing.

<sup>101</sup>

<https://www.technologyreview.com/2020/05/12/1001633/ai-is-still-largely-baffled-by-covid-misinformation/>

<sup>102</sup>

<https://themarkup.org/coronavirus/2020/04/23/want-to-find-a-misinformed-public-facebooks-already-done-it>

<sup>103</sup> <https://about.fb.com/news/2020/04/coronavirus/>

<sup>104</sup>

<https://thehill.com/homenews/media/493649-zuckerberg-says-stay-at-home-protests-organized-through-facebook-qualify-as>

- On April 20, announced it is shutting down event pages for anti-quarantine protests in states with stay-at-home orders amid the coronavirus pandemic. A company spokesperson told Politico that it has removed protesters' event pages and messages encouraging rallies in California, New Jersey, and Nebraska after state governments advised those events are currently prohibited by law. "...events that defy government's guidance on social distancing aren't allowed on Facebook," the spokesperson said. A Facebook representative also told CNN that it's in discussions with four other states — New York, Wisconsin, Ohio, and Pennsylvania — to determine if planned protests would violate those states' stay-at-home orders.<sup>105</sup>
- On April 16, announced in a blog post that "in the coming weeks" it will begin notifying users who have interacted with misinformation about the novel coronavirus that is subsequently removed from the platform. Users who liked, reacted to or commented on potentially harmful debunked content will see a message in their news feeds directing them to the World Health Organization's Myth busters page. There, the WHO dispels some of the most common falsehoods about the pandemic.<sup>106</sup> Facebook said the design in its blog post is an "early version" and that it's also testing "more explicit" variations.<sup>107</sup>
- Facebook is also adding a new section to its coronavirus information center dedicated to vetted articles that focus on debunking misinformation, and said that during March it applied warning labels on about 40 million posts related to the pandemic, based on roughly 4,000 articles reviewed by its third-party fact checkers.
- On April 14, announced that an interactive Health Alert interactive service in partnership with the WHO already on WhatsApp is being expanded to Facebook Messenger. Messenger's 1.3B monthly active users will be able to ask questions and get quick answers about the virus.<sup>108</sup>
- On April 7, Consumer Reports ran an article about a "test" designed to see how well Facebook is policing coronavirus-related advertising. Seven ads,

105

<https://www.syracuse.com/coronavirus/2020/04/coronavirus-facebook-shuts-down-anti-quarantine-protest-events-at-request-of-states.html>

<sup>106</sup> <https://about.fb.com/news/2020/04/covid-19-misinfo-update/>

107

<https://www.vox.com/recode/2020/4/16/21223972/facebook-coronavirus-hoaxes-warning-misinformation-avaaz>

108

<https://messengernews.fb.com/2020/04/14/world-health-organization-launches-messenger-experience-to-help-deliver-accurate-information-on-covid-19/>

including two that described dangerous claims Nick Clegg (below) had specifically said would trigger rejection, were approved and scheduled to run on Facebook. Only one ad, with an image of a stock shot of a respirator-style face mask, was rejected, suggesting Facebook is using image recognition—one of its strong suits—to flag posts. Consumer Reports pulled all the ads from the schedule before they had the chance to actually appear. When contacted, Facebook confirmed all the ads violated their policies and closed the fake account created to run them.<sup>109</sup>

- On April 1, COO Sheryl Sandberg participated in a podcast, “Skimm’d from The Couch”, in which she maintained that the lessons learned from its numerous mistakes over the past couple of years has prepared it to deal with the crises caused by the coronavirus pandemic. She said, “Some of the troubles we had and some of the mistakes we made over the past years, we worked so hard to correct those and set ourselves up, they’re serving us well”.<sup>110</sup>
- On March 30, in response to an investigation by Politico<sup>111</sup>, Facebook said it removed private groups that spread falsehoods from recommendations shown to users about communities that they may want to join. Within these invite-only online communities, the company also has added an information box that links directly to official government advice.
- On March 30, pledged \$100M to support journalists; \$25M is in the form of an emergency grant for local news through the Facebook Journalism Project (the first round went to 50 local newsrooms in the US and Canada) and \$75M will go towards “additional marketing spend to move money over to news organizations around the world.”<sup>112</sup>.
- In India on March 30, launched a Corona Helpdesk Chatbot on its Messenger platform to raise awareness and provide authentic and accurate information in order to debunk fake news about coronavirus. Users can go to the Facebook page on MyGov Corona Hub and start a chat in either English or Hindi by typing ‘Get Started’, which will prompt

---

109

<https://www.consumerreports.org/social-media/facebook-approved-ads-with-coronavirus-misinformation/>

110

<https://www.cnbc.com/2020/04/01/facebooks-sandberg-learning-from-our-mistakes-prepared-us-to-deal-with-the-coronavirus.html>

<sup>111</sup> <https://www.politico.eu/article/facebook-misinformation-fake-news-coronavirus-covid19/>

<sup>112</sup> <https://about.fb.com/news/2020/03/coronavirus/#joint-statement>

them to either type in a query or choose from a list of frequently asked questions.<sup>113</sup>

- On March 26, Facebook Messenger introduced Coronavirus Community Hub, offering tips, authoritative information and other resources, weeks after launching a similar information hub on WhatsApp, its other messaging service.<sup>114</sup>
- On March 25, Nick Clegg, FB's vice president of global affairs and communications, in an interview with NPR, again distinguished their approach to the virus as distinct from how they may handle political misinformation in the future: "...in a democracy with an independent press and with the claims and counterclaims that politicians make about each other, we think it's very important that private companies should allow voters for themselves to make their own judgments about what politicians are saying, about the future of their country, whereas when it comes to a medical pandemic, as I say, underpinned by science and by authoritative institutions such as the CDC and the WHO and others, it's, of course, much easier for us to act under the strict expertise and guidance from those institutions themselves."<sup>115</sup>
- On March 25, in an effort to encourage social distancing, Facebook communicated that they released a series of Physical Distancing Profile Frames that allow people to place frames around their profile pictures with various phrases and hashtags like: Stay Home Save Lives, #Quaranteam, #StayHome<sup>116</sup>
- On March 23, Facebook began enlisting outside developers to create ways its Messenger service can help health organizations; they invited software experts to take part in an online "hackathon" aimed at creating ways to use Messenger to ease social-distancing and deliver accurate information about the pandemic.<sup>117</sup>
- On March 18, Facebook announced that it is launching a Coronavirus

113

<https://www.thedrum.com/news/2020/03/30/facebook-launches-chatbot-and-news-hub-india-fight-against-misinformation>

114

<https://techcrunch.com/2020/03/26/facebook-community-hub-messenger-fight-coronavirus-misinformation/>

115

<https://www.npr.org/2020/03/25/821591134/how-facebook-wants-to-handle-misinformation-around-the-coronavirus-epidemic>

<sup>116</sup> Email dated March 25 from Shaarik Zafar, Public Policy at Facebook, to partners and coalition members

117

<https://www.freemalaysiatoday.com/category/business/2020/03/23/facebook-enlists-its-messenger-to-fight-coronavirus-misinformation/>

Information Center, featured at the top of NewsFeed, to provide a central place for people to get the latest news and information.<sup>118</sup>

- Facebook is limiting the number of people a user can forward a message to in Messenger in order to curb the spread of misinformation.<sup>119</sup>
- Although Facebook doesn't usually remove content, it is removing coronavirus misinformation because it could cause physical harm.<sup>120</sup>
- Introduced a pop-up at the top of search results that directs users to credible health information<sup>121</sup>
- Sending alerts to users who have shared or attempted to share misleading content.<sup>122</sup>
- From the outset, Facebook has tried to distinguish its approach on the pandemic to how it may act in future cases of misinformation. In an interview with NY Times, Mark Zuckerberg noted, "When you're dealing with a pandemic, a lot of the stuff we're seeing just crossed the threshold...So it's easier to set policies that are a little more black and white and take a much harder line."<sup>123</sup>
- Partnered with International Fact-Checking Network (IFCN) to support the fact-checking community by broadening the #CoronaVirusFacts Alliance, the COVID-19 related misinformation effort, with a budget of \$1 million in grants<sup>124</sup>. USA Today has joined Facebook's third-party fact checking network.
- Prohibiting ads from making health or medical claims related to the coronavirus in product listings on commerce surfaces, including those listings that guarantee a product will prevent someone from contracting it<sup>125</sup>
- Banning ads and commercial listings for medical masks, hand sanitizer, surface disinfecting wipes, and COVID-19 testing kits, and prohibiting exploitative tactics in ads<sup>126</sup>
- Providing free ads to global health organizations; in their updates they describe giving WHO "as many free ads as they need" and "millions in ad credits to other health authorities"<sup>127</sup>

<sup>118</sup> <https://about.fb.com/news/2020/03/coronavirus/#coronavirus-info-center>

<sup>119</sup> [https://www.phonearena.com/news/Facebook-Messenger-limit-forwarding\\_id123205](https://www.phonearena.com/news/Facebook-Messenger-limit-forwarding_id123205)

<sup>120</sup>

<https://www.campaignasia.com/article/is-facebook-coping-with-the-steady-wave-of-covid-19-misinformation/458277>

<sup>121</sup> *Id.*

<sup>122</sup>

<https://www.wwd.com/business-news/media/facebook-instagram-twitter-coronavirus-social-media-1203541823/>

<sup>123</sup> <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html>

<sup>124</sup>

<https://www.poynter.org/fact-checking/2020/flash-grants-of-up-to-50k-are-now-available-for-fact-checkers-fighting-coronavirus-misinformation/>

<sup>125</sup> <https://about.fb.com/news/2020/03/coronavirus/#banning-ads>

<sup>126</sup> <https://about.fb.com/news/2020/03/coronavirus/>

<sup>127</sup> *Id.*

	<ul style="list-style-type: none"> <li>• Adopted a 3 tiered approach to mitigating misinformation: (1) "disrupting economic incentives because most false news is financially motivated;" (2) "building new products to curb the spread of false news;" and (3) "helping people make more informed decisions when they encounter false news."<sup>128</sup></li> </ul>
Google	<ul style="list-style-type: none"> <li>• On June 1, the Tech Transparency Project blog published the results of a test which showed that 40% of the news and information sites that NewsGuard, an independent watchdog group, identified as publishing "materially false information about the virus" (97 out of 224) used Google's DoubleClick or AdSense tools to display third-party ads. In some cases, the placements contradicted the policies of other Alphabet divisions; for example YouTube deleted the channel of David Icke, the conspiracy theorist, citing his "continued violation" of policies that prohibit the spread of misinformation about the coronavirus pandemic. Yet Google has not banned Icke from using its advertising system to monetize his website, where he promotes the same conspiracy theories featured in his YouTube videos.<sup>129</sup></li> <li>• On April 22, changed its advertising policies to require all advertisers, not just those with political affiliations, to complete an identity verification program to show ads on its platforms, which include Google Search, Google News, and YouTube. Advertisers will need to submit personal identification, business incorporation documents or other information that proves who or what they are and the country they operate in. Google will start rolling out the policy with U.S.-based advertisers in a phased approach and will release it internationally over the next few years.<sup>130</sup></li> <li>• On April 16, various tech sites reported that Google detects and blocks an average 18M daily malware and phishing emails related to COVID-19, representing almost 20% of the total 100M daily phishing messages. Meanwhile, there are more than 240M coronavirus-related spam messages a day. For the most part, existing malware campaigns have just been repurposed to take advantage of the current pandemic. That commonality helps Google's Safe Browsing system flag and warn against nefarious links in Gmail, Chrome, and other services that use the API.<sup>131</sup></li> <li>• On April 6, announced it is launching a COVID-19 hub on Google News; the hub organizes news from authoritative global and local sources to help readers access the latest guidance regarding the virus. You can select which regions you want to be updated on in particular. It can also highlight stories by topics such as the economy, health, and travel, and emphasizes</li> </ul>

<sup>128</sup> <https://www.facebook.com/facebookmedia/blog/working-to-stop-misinformation-and-false-news>

<sup>129</sup> <https://www.techtransparencyproject.org/articles/google-profiting-coronavirus-conspiracy-sites>

<sup>130</sup> <https://www.axios.com/google-advertisers-identity-97d6a071-2e19-45fe-9f82-c9a151e1861e.html>

<sup>131</sup> <https://9to5google.com/2020/04/16/gmail-coronavirus-spam/>

results from local news most relevant to you, including tweets from local authorities.<sup>132</sup>

- On April 2, in a letter to partners and clients, communicated it would adjust the enforcement of its sensitive events policy, which blocks ads that try to capitalize on short-term events like natural disasters. Google first allowed health information PSA ads from government entities, supported by \$250 million in donated ad grants (see below), then explored ways to support ads from hospitals, medical providers, government entities, and NGOs, and will extend it to additional sectors, including politics, “as soon as we’re able to do so safely”.<sup>133</sup>
- On April 2, announced it would spend \$6.5 million toward fighting the spread of misinformation around the coronavirus pandemic; funding will go toward fact-checkers, news organizations and nonprofits around the world. Part of the money will go toward fellowships at Stanford University to help reporters covering the coronavirus pandemic, and to groups supporting the ability of journalists to access research around the virus. Google Trends will be made more widely available for reporters, health care workers and law enforcement, with funding to train journalists on how to spot health misinformation<sup>134</sup>.
- On March 31, the impact of Google’s ban on ads capitalizing on the virus on political messaging was called out in an article in Protocol<sup>135</sup>. As a result of the policy many political and advocacy campaigns are focusing paid ads on Facebook.
- On March 27, announced a new \$800M commitment to support small- and medium-sized businesses (SMBs), health organizations and governments, and health workers.<sup>136</sup> Of this, an estimated \$610M takes the form of “ad credits” and “cloud computing services”; these in-kind donations have been met with some skepticism.<sup>137</sup>
- On March 21, CNBC reported that Google (and Apple) have added features to their voice assistants to provide users with a step-by-step

132

<https://thenextweb.com/google/2020/04/06/google-launches-coronavirus-news-hub-to-keep-reliable-coverage-in-one-place/>

<sup>133</sup> Email from Chanelle Hardy, Policy Partnerships and Strategic Outreach, Google, to Chris Lewis, Public Knowledge, April 2, 2020

134

<https://thehill.com/policy/technology/490865-google-to-invest-65-million-to-fight-coronavirus-misinformation>

<sup>135</sup> <https://www.protocol.com/google-coronavirus-ad-ban-democrats>

136

<https://blog.google/inside-google/company-announcements/commitment-support-small-businesses-and-crisis-response-covid-19>

<sup>137</sup> <https://www.vox.com/recode/2020/3/31/21197652/coronavirus-philanthropy-donations-google-cisco>

questionnaire if they ask variations of, “Hey Siri, do I have the coronavirus?”<sup>138</sup>

- On March 20, launched [google.com/covid19](https://www.google.com/covid19), offering education, prevention and local resources related to the novel coronavirus. At the top of the site, there’s an information box describing virus symptoms, treatment and prevention tips from the World Health Organization. Google also included links to state departments of health across the U.S., search trends related to COVID-19, and other resources for individuals.<sup>139</sup>
- Working to quickly remove any content that claims to prevent the coronavirus in place of seeking medical treatment.<sup>140</sup>
- Searches related to the virus trigger an “SOS Alert,” with news from mainstream publications including National Public Radio, followed by information from the U.S. Centers for Disease Control and Prevention and the World Health Organization displayed prominently.
- Promoting WHO’s “Do the Five” campaign on the homepage to raise awareness of simple measures people can take to slow the spread of the disease.
- Expanding Knowledge Panels to include CV-19 symptoms, prevention and treatments<sup>141</sup>
- May eventually need to address its deal with Twitter to show tweets in search results, especially for queries about live and recent events.<sup>142</sup>
- Google Ads: claims they are blocking all ads capitalizing on the coronavirus; have blocked “tens of thousands” over the last six weeks. Also helping WHO and government organizations run PSA ads.<sup>143</sup>
- Google Play: prohibits developers from capitalizing on sensitive events; existing policy prohibits apps that feature medical or health-related content or functionalities that are misleading or potentially harmful.<sup>144</sup>

138

<https://www.cnbc.com/2020/03/21/apple-updated-siri-to-help-people-who-ask-if-they-have-coronavirus.html>

139

<https://www.bloomberg.com/news/articles/2020-03-21/google-coronavirus-website-launches-a-week-after-trump-tout>

<sup>140</sup> <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>

<sup>141</sup> <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>

142

<https://www.bloomberg.com/news/articles/2020-03-10/dr-google-scrubs-coronavirus-misinformation-on-search-youtube>

<sup>143</sup> <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>

<sup>144</sup> <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>

## Twitter

- On June 3, in response to press inquiries about its fact-checking of President Trump’s tweet, clarified that it will not tackle all misinformation on its platform but will instead focus on posts with the “highest potential for harm”. Twitter will “provide context, not fact-checking”; “prioritize based on the highest potential for harm”; and “focus on manipulated media, civic integrity, and COVID-19”. Flagged tweets will link to a page that will show “factual statements,” “counterpoint opinions” or “ongoing public conversation around the issue.” Likelihood, severity and type of potential harm — along with reach and scale — factor into their decisions.<sup>145</sup>
- On May 26, for the first time applied a fact-checking notice to tweets from President Trump: two tweets about the potential for fraud involving mail-in ballots. The move was based on a policy announced May 11 (below) to apply fact-checking labels about the coronavirus and other disputed issues subject to misinformation, including the election. The warning label was used as an extension of Twitter’s policy against misleading COVID-19 information, because the mail-in ballots Trump was referring to were being used because of stay-at-home restrictions related to the virus. This marked the first time Twitter has applied the fact-checking label to a message about non-Covid news - or to one from President Trump.<sup>146</sup>
- On May 20, announced it is testing a new feature that allows a user to select, tweet by tweet, who could respond to their post. A new permissions button that appears in the bottom left corner of a tweet will allow users to choose whether they want any one (the default), "people you follow", or "only people you mention" to weigh in (people can still quote-tweet but it won't be in the original user's feed). While created in part to deal with bullying and harassment, it may be relevant to the spread of misinformation: it may prevent people from correcting misinformation in a tweet, or turn reply threads into partisan zones, reinforcing the original misinformation.<sup>147</sup>
- On May 11, announced in a company blog that the company would start affixing labels and warnings to tweets containing information about COVID-19 that goes against the advice and knowledge of public health experts. That will include tweets where the misinformation is contained in media, like videos and images. The company said, “during active conversations about disputed issues, it can be helpful to see additional context from trusted sources” .Twitter says it will use "internal systems to proactively monitor content related to COVID-19," as well as "rely on

<sup>145</sup>

<https://thehill.com/policy/technology/500946-twitter-to-prioritize-posts-with-highest-potential-for-harm-for-censorship> (note: language of footnote is original to the source)

<sup>146</sup>

<https://www.wsj.com/articles/widower-asks-twitter-to-pull-trumps-false-tweets-on-death-of-joe-scarborough-aide-11590516200>

<sup>147</sup> <https://mashable.com/article/twitter-limits-who-can-respond-to-tweets/>

trusted partners" for assessing the risk level of tweets. These labels will link to a Twitter-curated page or external trusted source containing additional information on the claims made within the Tweet. The announcement broadens a policy to label or remove misinformation that could cause physical harm or widespread panic to content where harm is not as obviously imminent.<sup>148</sup> Most coverage of the announcement focused on a response to questions from reporters in which executive Yoel Roth clarified that the company's new policy would apply to the tweets and content of world leaders — including President Trump.

- On May 8, pushed back on an assertion from the US State Department Friday that it was "highly probable" that the Chinese government coordinated networks of Twitter accounts to disseminate disinformation related to the coronavirus outbreak, saying their initial review of the accounts in question does not support the government's claims. An initial review from Twitter of more than 5,000 accounts turned over to them by the State Department found that numerous accounts belong to government entities, nongovernmental organizations, and journalists.<sup>149</sup>
- On April 22, broadened its guidance on "unverified claims that incite people to engage in harmful activity, could lead to the destruction or damage of critical 5G infrastructure, or could lead to widespread panic, social unrest, or large-scale disorder". Material of this nature may be considered in violation of Twitter's policies.<sup>150</sup>
- On April 8, briefly locked the account of online personalities Diamond and Silk over a tweet that violated the company's rules against coronavirus misinformation.<sup>151</sup>
- On April 6, Ad Age reported (via Ad Exchange) that Twitter had lifted its ban on coronavirus-related advertising. Twitter will allow marketers to feature their pandemic responses in paid tweets. Twitter initially banned COVID terms to forestall misinformation, but now feels that "the messaging that brands and businesses can provide to the world ... are going to be positively received".<sup>152</sup>
- On March 23, Daily Beast and The Hill reported that Twitter was leaving up misinformation about U.S. origins of the virus attributed to the Chinese government.<sup>153</sup> Twitter confirmed only that "official government accounts engaging in conversation about the origins of the virus and global public

148

[https://blog.twitter.com/en\\_us/topics/product/2020/updating-our-approach-to-misleading-information.html](https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information.html)

149 <https://www.cnn.com/2020/05/08/politics/china-twitter-disinformation-bots-state-department/index.html>

150 [https://blog.twitter.com/en\\_us/topics/company/2020/covid-19.html#unverifiedclaims](https://blog.twitter.com/en_us/topics/company/2020/covid-19.html#unverifiedclaims)

151 <https://www.politico.com/news/2020/04/08/diamond-silk-twitter-coronavirus-175914>

152 <https://www.adexchanger.com/ad-exchange-news/monday-06042020/>

153

<https://thehill.com/policy/technology/489061-twitter-says-coronavirus-disinformation-spread-by-chinese-officials-does>

conversation about potential emergent treatments will be permitted, unless the content contains clear incitement to take a harmful physical action". They also acknowledged "we will not be able to take enforcement action on every Tweet that contains incomplete or disputed information about COVID-19".<sup>154</sup>

- On March 20, announced it will accelerate its notoriously opaque process for verification (the blue checkmark) for accounts that are providing credible updates around COVID-19. They are prioritizing verification for Twitter accounts that have an email address associated with an authoritative organization or institution. They explained steps academics and others who work for public health organizations or academic institutions can take in order to ensure their accounts can be verified.<sup>155</sup>
- On March 19, banned paid ads and commerce listings for masks, hand sanitizer, and surface disinfecting wipes to help protect against inflated prices and predatory behavior. Will be ramping up automated enforcement for ads and commerce the week of March 23. If Twitter sees abuse around these products in organic posts, they will also be removed.<sup>156</sup>
- On March 19, announced it will implement a new policy on "synthetic and manipulated media," primarily focused on the 2020 election. When users scroll through posts, they may see a new labeling system: a blue exclamation point and the words "manipulated media" underneath content the platform believes to have been tampered with or deceptively shared (deepfakes, "cheepfakes", low-tech editing).<sup>157</sup>
- On March 18, expanded its safety rules; it will no longer allow content that could place people at a higher risk of transmitting COVID-19, including "Denial of expert guidance," "Encouragement to use fake or ineffective treatments, preventions, and diagnostic techniques," and "Misleading content purporting to be from experts or authorities."<sup>158</sup> It will also take action against posts alleging that any particular group or nationality is more or less susceptible to coronavirus.<sup>159</sup>
- Expanding its search feature to provide information about COVID-19 from credible sources. In January, it launched the COVID-19 Search Prompt in 64 countries and 20 languages in partnership with the CDC.<sup>160</sup>
- Increasing its use of machine learning and automation to act on potentially

<sup>154</sup> <https://twitter.com/TwitterSafety/status/1242221132956061697>

<sup>155</sup> <https://twitter.com/TwitterSupport/status/1241155701822476288>

<sup>156</sup> <https://twitter.com/robleathern/status/1240741322438266880>

<sup>157</sup>

[https://blog.twitter.com/en\\_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html](https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html)

<sup>158</sup> <https://twitter.com/TwitterSafety/status/1240418440982040579>

<sup>159</sup> <https://thehill.com/policy/technology/488442-twitter-expands-coronavirus-misinformation-policy>

<sup>160</sup>

<https://wwd.com/business-news/media/facebook-instagram-twitter-coronavirus-social-media-1203541823/>

	<p>abusive and manipulative content, (i.e., using tech to make more enforcement calls). Due to the risk of errors, will not permanently suspend any accounts based solely on automated enforcement systems; will layer in human checks.</p> <ul style="list-style-type: none"> <li>• Instituting a global content severity triage system to prioritize rule violations with biggest risk of harm and reduce the burden on people to report them.</li> <li>• Executing daily quality assurance checks on content enforcement processes to ensure agility in responding to this rapidly evolving, global issue.</li> <li>• Engaging partners around the world to ensure escalation paths remain open and urgent cases can be addressed.</li> <li>• Reviewing the Twitter Rules in the context of COVID-19 and considering ways in which they may need to evolve to account for new account behavior.</li> <li>• Dedicated COVID-19 page at top of timeline with accurate, current information.</li> <li>• Using @TwitterSafety for updates on policies<sup>161</sup>.</li> </ul>
YouTube	<ul style="list-style-type: none"> <li>• On May 21, was accused of ‘censorship’ after removing a video of a former World Health Organisation chief saying coronavirus could ‘burn out’ before a vaccine is found. YouTube has since said the deletion was an error and that the video had been reinstated. Following the incident, YouTube said that it was sifting through massive numbers of videos on its site looking for harmful fake news and occasionally its algorithms made mistakes.<sup>162</sup></li> <li>• On May 20, updated its Community Guidelines to include a page specifically on COVID-19 misinformation. Its COVID-19 Medical Misinformation Policy doesn't allow content about COVID-19 that poses a serious risk of egregious harm, or that spreads medical misinformation that contradicts the World Health Organization (WHO) or local health authorities’ medical information about COVID-19. This is limited to content that contradicts WHO or local health authorities’ guidance on:             <ul style="list-style-type: none"> <li>○ Treatment</li> <li>○ Prevention</li> <li>○ Diagnostic</li> <li>○ Transmission</li> </ul> </li> </ul> <p>For a detailed description of its policy see the footnote.<sup>163</sup></p>

<sup>161</sup>

[https://blog.twitter.com/en\\_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html](https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html)

<sup>162</sup>

<https://www.telegraph.co.uk/news/2020/05/21/youtube-accused-censorship-removing-video-claiming-covid-19/>

<sup>163</sup> <https://support.google.com/youtube/answer/9891785>

- In the first week of May, multiple media sources referred to YouTube “finally” deleting conspiracy theorist David Icke's official channel, having repeatedly warned him not to violate its policies. Icke has spread misinformation for years via social and traditional media, and his YouTube channel had close to 1M subscribers at the time of its deletion. The decision to “terminate” Icke’s channel seems to have been Icke linking COVID-19 symptoms to 5G mobile networks, in violation of YouTube's policies.<sup>164</sup>
- On April 28, communicated to partners that YouTube is expanding the use of “fact check information panels” - a feature launched in Brazil and India last year — to the United States. The panels are designed to highlight relevant, third-party fact-checked articles above search results for relevant queries. A fact check information panel will appear when there is a relevant fact checked article available from an eligible publisher, and it will only show when people search for a specific claim.<sup>165</sup>
- In her first interview on the topic on April 20, chief executive Susan Wojcicki said YouTube would remove anything it deems “medically unsubstantiated”. The move follows YouTube banning conspiracy theories falsely linking Covid-19 to 5G networks. “Anything that would go against World Health Organization recommendations would be a violation of our policy”, she noted, and added that YouTube had seen a 75% increase in demand for news from “authoritative” sources.<sup>166</sup>
- On April 20, the Wall Street Journal reported that Alphabet wants to substantially limit the information a key auditor of YouTube can share about the risks of advertising on the video service. The auditor, OpenSlate, is refusing to sign a contract that would prevent it from reporting to ad clients when ads have run in videos with sensitive subject matter, including information about COVID-19.<sup>167</sup>
- On April 7, in an interview with Axios, Chief Product Officer Neal Mohan said YouTube has been focused on a twofold approach: making authoritative information more prominent and aggressively removing policy-violating content. An information panel on its home page linking to national health agencies' websites represents the first time YouTube has linked to a text site rather than a video. It has expanded its existing medical misinformation policies that prohibit promoting false cures or encouraging people not to see a doctor to include barring promoting actions that go against recommendations from national health authorities.

<sup>164</sup> <https://www.businessinsider.com/youtube-david-icke-channel-2020-5>

<sup>165</sup> <https://youtube.googleblog.com/2020/04/expanding-fact-checks-on-youtube-to-united-states.html>

<sup>166</sup> <https://www.bbc.com/news/technology-52388586>

<sup>167</sup>

<https://www.wsj.com/articles/youtube-spars-with-auditor-over-transparency-of-advertising-risks-11587340250?mod=searchresults&page=1&pos=2>

Unlike Facebook and Twitter, YouTube's policies are entirely focused on the content of a video and not who is doing the speaking.<sup>168</sup>

- On April 6, confirmed it will remove or reduce recommendations for videos that falsely link 5G to the virus after reports of people setting phone masts on fire and attacking phone company workers. It will actively remove videos that breach its policies for CV, but content that is simply conspiratorial about 5G mobile communications networks, without mentioning coronavirus, is still allowed on the site.<sup>169</sup>
- On April 2, expanded monetization of content mentioning or featuring COVID-19 to all creators and news organizations, assuming they follow both Advertiser-Friendly and Community Guidelines.<sup>170</sup> Subsequent to this change, the Tech Transparency Project, a not-for-profit watchdog organization, reported the company was running advertisements with videos pushing herbs, meditative music, and potentially unsafe over-the-counter supplements as cures for Covid-19. When notified of these, Google removed 4 of the videos noting that the other 3 were not misinformation but “wellness”-related.<sup>171</sup>
- On March 19, communicated that they will create a new “row” of verified videos on its homepage for displaying trustworthy videos about coronavirus. It will pull from a list of authoritative news outlets and local health authorities that upload to YouTube and are more reliable than just general videos on the subject uploaded by random users. Videos are generated algorithmically, with hundreds of different signals being used to help pick videos. This technique of boxing out videos, known as a shelf, has been used by the platform in the past to help promote legitimate information as world news events unfold.<sup>172</sup>
- On March 16, confirmed that they “will start relying more on technology to help do some of the work normally done by reviewers, which may result in some accidental removal of content that does not violate our policies”.<sup>173</sup>
- On March 11, communicated a reversal of its Sensitive Events Policy:

<sup>168</sup>

<https://www.axios.com/youtube-coronavirus-misinformation-videos-google-d9ce89cb-0de0-4f50-8a25-5923e078a858.html>

<sup>169</sup> <https://www.businessinsider.com/youtube-delete-5g-coronavirus-conspiracy-2020-4>

<sup>170</sup>

[https://support.google.com/youtube/answer/9777243?p=covid19\\_updates&visit\\_id=637194868739116880-4122912225&rd=1](https://support.google.com/youtube/answer/9777243?p=covid19_updates&visit_id=637194868739116880-4122912225&rd=1)

<sup>171</sup>

<https://www.theguardian.com/technology/2020/apr/03/youtube-coronavirus-treatments-profit-misinformation>

<sup>172</sup>

<https://www.dailymail.co.uk/sciencetech/article-8135515/YouTube-adding-section-verified-videos-coronavirus-stop-spread-misinformation.html>

<sup>173</sup> <https://support.google.com/youtube/thread/33987650>

	<p>YouTube was previously de-monetizing videos that included “more than a passing mention” of COVID-19, but now will allow monetization for videos that discuss the coronavirus “on a limited number of channels”<sup>174</sup></p> <ul style="list-style-type: none"> <li>• Added the CDC’s YouTube channel as a featured channel on its homepage; giving priority to authoritative sources in search.<sup>175</sup></li> <li>• Using the homepage to direct users to the WHO or other locally relevant authoritative organizations.<sup>176</sup></li> <li>• Pulled ads from videos that discuss COVID-19<sup>177</sup>, while donating ad inventory to governments and NGOs in impacted regions to use for education and information.<sup>178</sup></li> <li>• According to Bloomberg, recent YouTube searches for specific coronavirus conspiracies showed videos debunking those untruths.<sup>179</sup></li> </ul>
Amazon	<ul style="list-style-type: none"> <li>• On May 26, four joint FDA/FTC letters for the first time specifically call out that companies making claims about a product alleged to be a COVID cure, treatment, or preventative product participated in the Amazon Associates program. These companies, “as an Amazon associate, earn[s] commissions by promoting the sale of [their respective products] with claims on [the company’s] website representing or implying that the products can mitigate, prevent, treat, diagnose, or cure COVID-19.”<sup>180</sup></li> <li>• On May 13, posted a blog post noting, “we need a strong federal anti-price gouging law.” The framework they recommend includes it should 1) kick in immediately when the federal government declares a public health crisis or national emergency; 2) establish clear pricing standards, define who and what are covered by the law, and ensure strong enforcement authority; 3) define pricing prohibitions as “unconscionable or grossly excessive or unconscionably excessive” compared to a reasonable and relevant reference; 4) apply to all levels of the supply chain so that retailers and resellers are not forced to bear price gouging increases by manufacturers and suppliers; 5) apply to the party who actually sets the price of a product; 6) clearly define the scope of products that would apply to the statute and 7) ensure strong enforcement authority.<sup>181</sup></li> </ul>

<sup>174</sup> <https://youtube-creators.googleblog.com/2020/03/coronavirus-update-on-creator-support.html>

<sup>175</sup>

<https://www.wired.com/business-news/media/facebook-instagram-twitter-coronavirus-social-media-1203541823/>

<sup>176</sup> <https://blog.google/inside-google/company-announcements/coronavirus-covid19-response/>

<sup>177</sup>

<https://www.bloomberg.com/news/articles/2020-03-10/dr-google-scrubs-coronavirus-misinformation-on-search-youtube>

<sup>178</sup> Ibid

<sup>179</sup>

<https://www.bloomberg.com/news/articles/2020-03-10/dr-google-scrubs-coronavirus-misinformation-on-search-youtube>

<sup>180</sup> <https://www.natlawreview.com/article/fda-and-ftc-continue-to-trace-and-fight-fraud-related-to-covid-19>

<sup>181</sup> <https://blog.aboutamazon.com/policy/its-time-for-congress-to-establish-a-federal-price-gouging-law>

- On May 5, agreed to partner with Pfizer, 3M, Citi, Alibaba, Merck and other companies along with the Homeland Security Investigations (HSI) branch of Immigration and Customs Enforcement (ICE) to identify and stop the sale of counterfeit coronavirus-related products. Operation Stolen Promise, which was launched to eliminate illegal activity related to the pandemic, will focus on banned pharmaceuticals and medical supplies, websites defrauding consumers, and other criminal activities involving trade or financial systems, according to the statement.<sup>182</sup>
- On April 26, confirmed to Reuters that it is piloting the use of video conference calls to verify the identity of merchants who wish to sell goods on its websites, in a new plan to counter fraud.<sup>183</sup>
- Temporarily prioritizing household staples, medical supplies, and other high-demand products coming into fulfillment centers; discretionary products must be fulfilled by vendors.<sup>184</sup>
- Restricted who can sell certain in-demand items (e.g., face masks, hand sanitizer) and said it pulled more than 500,000 high-priced listings and banned more than 2,000 sellers due to price-gouging.<sup>185</sup>
- In a response to a letter from Sen. Ed Markey (D-MA), also said they are 1) proactively monitoring the marketplace for unfair prices and “aggressively” enforcing its fair pricing policy; 2) working with state attorneys general to prosecute bad actors; 3) using their price monitoring methods to continuously compare prices submitted by sellers with current and historic prices within and outside of Amazon; 4) using machine learning models, in part, to detect price gouging, but has recently stepped up human monitoring as price gouging has spread in light of the coronavirus. This includes “an additional dedicated team” that’s working “24 hours, seven days a week” to spot price gouging on items like protective masks and hand sanitizers. Amazon has also responded to inquiries from several attorneys general.
- Instituted additional manual audits of products in stores due to the

182

<https://thehill.com/policy/national-security/department-of-homeland-security/496220-amazon-pfizer-3m-partner-with-ice-to>

183

<https://www.deccanherald.com/business/amazon-tests-screening-new-merchants-for-fraud-via-video-calls-in-pandemic-830360.html>

184

<https://www.vox.com/recode/2020/3/17/21183310/amazon-coronavirus-fba-inventory-seller-vendor-restrictions>

185

<https://www.cnbc.com/2020/03/06/amazon-removes-530000-products-amid-coronavirus-price-gouging.html>

	<p>increased risk of price gouging from unscrupulous sellers seeking to evade automated systems and take advantage of customers.<sup>186</sup></p> <ul style="list-style-type: none"> <li>Relies on a mix of automated tools and human moderation to enforce its policies, including price gouging, false listings, and review hijacking.<sup>187</sup></li> </ul>
<p><b>Priority Level 2</b></p>	
<p>Instagram</p>	<ul style="list-style-type: none"> <li>On May 28, Fox Business reported that Instagram has been testing labels on posts related to coronavirus that link to the Centers for Disease Control and Prevention's website. Instagram confirmed it is running a small test connecting people with resources from the CDC on posts mentioning COVID-19. Instagram also prompts users with a notification that reads, "Looking for coronavirus info? See the latest information from the [CDC] so you can help prevent the spread of novel coronavirus (COVID-19)" when they look up hashtags related to the virus. The notification then prompts users with two options: "Go to cdc.gov" or "See Posts."<sup>188</sup></li> <li>On May 18, introduced Guides, a way to more easily discover recommendations, tips and other content from favorite creators, public figures, organizations and publishers on Instagram. Due to the COVID-19 pandemic, Instagram's first Guide is about wellness content. It enables users to connect with expert organizations to share resources on how to look after your well-being, maintain connection with others or manage anxiety or grief. To view a Guide, visit the profile of participating creators and tap the middle icon to view their Guides. In the coming days, you'll also be able to access Guides within the Explore tab.<sup>189</sup></li> <li>On May 13, launched a set of new features primarily intended to fight online bullying, but also to more generally create a more positive experience on the platform. Users will be able to 1) delete comments in bulk, as well as block or restrict multiple accounts that post negative comments; 2) set the tone for their account and engage with their community by pinning a select number of comments to the top of their</li> </ul>

<sup>186</sup>

<https://www.cnbc.com/2020/03/06/amazon-removes-530000-products-amid-coronavirus-price-gouging.html>

<sup>187</sup> <https://www.wired.com/story/amazon-sellers-hijack-listings-face-masks-coronavirus/>

<sup>188</sup>

<https://www.foxbusiness.com/technology/instagram-adds-coronavirus-cdc-links-to-posts-mentioning-the-virus>

<sup>189</sup> <https://about.instagram.com/blog/announcements/supporting-well-being-with-instagram-guides/>

comments thread; and 3) manage who can tag or mention you on Instagram.<sup>190</sup>

- On April 22, announced they will provide the location high-reach Instagram accounts on every post they share, so people have more information to help them gauge the reliability and authenticity of the content they see in their feeds. This step, primarily focused on protecting election integrity, will be piloted in the US, starting specifically with Instagram accounts that are based outside the US but reach large audiences based primarily in the US.
- On March 25, Facebook communicated that in an effort to encourage social distancing, IG had launched the “Stay Home” sticker, modeled on one designed to encourage voting. It shows how others are social distancing by compiling accounts that you follow, that are using the sticker, in your shared story<sup>191</sup>.
- Also noted that IG has added a search module to connect people with accurate information; began downranking fact-checked misinformation in Feed and Stories; and is removing COVID-19 accounts and content from both search results and the Explore tab of recommendations, unless posted by a credible health organization. Karina Newton, Instagram's head of policy, notes that this requires overhauling the recommendation algorithm to look at the entire spectrum of possible queries related to coronavirus, including misspellings, “type-ahead” search results, and 180 languages, and weeding out accounts with names and bio information that would ordinarily correspond to those queries<sup>192</sup>.
- On March 17, unveiled a paid partnership with WHO and @DudeWithSign, an influencer account with 6.4M followers, to “create informative memes” that raise awareness of and direct followers to accurate information and updates on COVID-19.
- In countries impacted by the virus, added a call-out at the top of homescreen feed showing links to information from the World Health Organization and local health ministries.
- Preventing users from searching for COVID-19-related augmented reality effects unless they were made in partnership with legitimate health organizations.
- Sending potentially false information to fact checkers.
- Removing known harmful misinformation related to COVID-19, and when someone taps on a hashtag related to COVID-19, showing resources from WHO, CDC, and local health authorities<sup>193</sup>

<sup>190</sup> <https://about.instagram.com/blog/announcements/continuing-our-work-to-fight-online-bullying>

<sup>191</sup> Email from Shaarik Zafar, noted above.

<sup>192</sup> <https://www.protocol.com/coronavirus-instagram-tiktok-whatsapp-response>

<sup>193</sup> <https://twitter.com/InstagramComms/status/1235984308994703360>

	<ul style="list-style-type: none"> <li>• Blocking and restricting hashtags used to spread misinformation; banning ads exploiting the situation<sup>194</sup></li> <li>• “Thinking through a longer term solution” to help connect people searching COVID-19 related terms with credible information; in the meantime, showing the accounts of leading health organizations in these searches to better connect people to credible resources.<sup>195</sup></li> </ul>
Pinterest	<ul style="list-style-type: none"> <li>• On March 24, Pinterest announced it is “moving up [the] launch of the Today tab, a source of daily inspiration with curated topics and trending Pins that makes it easy to explore popular and timely ideas”. The previous weekend, Pinterest experienced an all-time high around the world with more saves and searches on the platform than any other weekend in its history. The new tab, which will live above the home feed, will include expert information from the World Health Organization and Centers for Disease Control on topics like hand washing during the coronavirus epidemic.</li> <li>• On March 24, Pinterest also noted that searches for “calming quotes” doubled and “stress relief” and “stress quotes” have tripled, so they have expanded “compassionate search” from the apps to the web. When users search for terms like “stress relief” they will see emotional wellbeing activities from emotional health experts.<sup>196</sup> Pinterest claims these interactions are kept private and aren’t used to inform recommendations or ads.<sup>197</sup></li> <li>• On March 24, VentureBeat reported that Pinterest uses AI to fight misinformation about COVID-19 on its platform, leveraging a machine learning system to identify pins it believes violate its health misinformation policy. Its model finds keywords or text associated with misinformation and blocks pins with that language, while at the same time identifying visual representations associated with medical misinformation.</li> <li>• Pinterest’s normal strict approach to combating health misinformation likely results in less content about the coronavirus than on any other major social platform. Anyone searching for terms related to the virus is sent to a sparse page with content from the World Health Organization. Users will see a message noting, “Pins about this topic often violate our Community Guidelines, which prohibit harmful medical misinformation. Because of</li> </ul>

<sup>194</sup> <https://twitter.com/InstagramComms/status/1235984311163158529>

<sup>195</sup> <https://twitter.com/InstagramComms/status/1235984311163158529>

<sup>196</sup> <https://newsroom.pinterest.com/en/post/introducing-the-today-tab-for-daily-inspiration>

<sup>197</sup>

<https://venturebeat.com/2020/03/24/pinterest-launches-today-tab-with-curated-topics-fights-coronavirus-misinformation-with-ai/>

	<p>this, we've limited search results to Pins from internationally-recognized health organizations."<sup>198</sup></p> <ul style="list-style-type: none"> <li>• Normal misinformation policy prohibits promotion of false cures, or the saving of content that includes advice where there may be immediate and detrimental effects on a Pinner's health or on public safety. Pinterest relies on information from nationally and internationally recognized institutions, including the CDC and WHO, to help determine if content violates these guidelines.<sup>199</sup></li> <li>• Added a banner to coronavirus search results in January to warn about misinformation and rolled out a custom results page in February.</li> </ul>
Snapchat	<ul style="list-style-type: none"> <li>• On June 3, confirmed the platform will no longer promote President Trump's account on its "Discover" page of curated content, after Trump tweeted comments that some suggested glorified violence amid racial justice protests. Notably, Snapchat is taking action on the president's account for comments he made elsewhere; Snapchat has not yet indicated that the president has violated any of its policies on its own platform. Snap CEO Evan Spiegel said, "Our Discover content platform is a curated platform, where we decide what we promote. We have spoken time and again about working hard to make a positive impact, and we will walk the talk with the content we promote on Snapchat."<sup>200</sup></li> <li>• On Thursday, March 25, rolled out a Covid-19 Myth Busting game.<sup>201</sup></li> <li>• Launched the Stay At Home Challenge through its acquisition Zenly that allows users to compete against each other using geo-location data. Whoever can stay home for the most amount of time wins.<sup>202</sup></li> <li>• Released a snapchat filter that allows users to share snaps with vetted info on how to prevent the spread of COVID-19.<sup>203</sup></li> <li>• Content platform, Discover, is curated and confined to the most trusted news organizations; do not offer an open news feed where unvetted publishers or individuals have an opportunity to broadcast misinformation. Now highlighting health information shared by its Discover partners, including NBC's Stay Tuned, Sky News, The Wall Street Journal, The</li> </ul>

<sup>198</sup>

<https://www.buzzfeednews.com/article/craigsilverman/pinterest-is-blocking-coronavirus-searches-and-people-are>

<sup>199</sup> <https://help.pinterest.com/en/article/health-misinformation>

<sup>200</sup>

<https://www.axios.com/snapchat-discover-trump-account-bbdfa6b8-f0cb-4c01-9a9d-c3dcb9b3de4e.html>

<sup>201</sup> <https://mashable.com/article/snapchat-coronavirus-myth-busting-game/>

<sup>202</sup> <https://techcrunch.com/2020/03/24/zenly-stay-home/>

<sup>203</sup> <https://www.snap.com/en-US/news/post/safety-first/>

	<p>Washington Post, CNN and NowThis.<sup>204</sup></p> <ul style="list-style-type: none"> <li>• Normal guidelines prohibit sharing content that deceives or deliberately spreads false information.</li> <li>• Now offering accurate Coronavirus-related coverage through vetted partners across seven countries and in six different languages.<sup>205</sup></li> <li>• Working closely with WHO and CDC to ensure Snapchatters have all the latest information from the experts themselves; working with WHO to develop custom content to answer questions from the Snapchat community.</li> <li>• Last week worked with WHO to launch a Filter promoting their safety tips and guidelines, available to all Snapchatters in the US, UK, Canada, Australia and more.</li> </ul>
TikTok	<ul style="list-style-type: none"> <li>• Hosting livestreams with the WHO to disseminate accurate information about COVID-19<sup>206</sup></li> <li>• Donating prominent in-feed ad space to trusted organizations and local health authorities<sup>207</sup></li> <li>• On March 18, introduced the group of technology and safety experts that will make up its Content Advisory Council (this initiative predates CV-19 but is included here for completeness).<sup>208</sup></li> <li>• Offers a library of information from WHO on the in-app landing page</li> <li>• Partnered with WHO on an informational page with trustworthy information (accessed from the Discover tab, appears when users search for coronavirus-related topics, and linked from videos that may relate to coronavirus).</li> <li>• Users who use hashtags related to coronavirus get an in-app notice that provides easy links to WHO's website and reminds users to report content that violates Community Guidelines.</li> <li>• Normal guidelines clearly prohibit misinformation intended to deceive or mislead the public; will remove misinformation that could cause harm to an individual's health or wider public safety.</li> <li>• Donating prominent in-feed ad space to trusted organizations and local health authorities; list organizations who now have TikTok accounts, including WHO, Red Cross, and UNICEF.</li> <li>• Ads policy does not allow ads that reference coronavirus, including when</li> </ul>

<sup>204</sup>

<https://forbusiness.snapchat.com/blog/coronavirus-how-snapchatters-are-reacting-to-the-news-and-staying-informed>

<sup>205</sup>

<https://forbusiness.snapchat.com/blog/coronavirus-how-snapchatters-are-reacting-to-the-news-and-staying-informed>

<sup>206</sup> <https://www.tiktok.com/safety/resources/covid-19?lang=en&appLaunch=web>

<sup>207</sup> <https://www.tiktok.com/safety/resources/covid-19?lang=en&appLaunch=web>

<sup>208</sup> <https://newsroom.tiktok.com/en-us/introducing-the-tiktok-content-advisory-council>

	<p>promoting products or services, to create a sense of fear, or to cause widespread offense.<sup>209</sup></p>
<p>Reddit</p>	<ul style="list-style-type: none"> <li>• On May 3, just one day after launching a new chat room feature, the company reversed course and pulled it entirely following an outcry from the site’s moderators. “Start Chatting” was designed to serve as “a new avenue to connect with each other during these difficult times”. However, moderators flooded the comments of the announcement post, worried that the feature, which moderators would have no control or visibility into, would draw the kind of abuse they work hard to keep out of their online communities. Other moderators overseeing subreddits that reach vulnerable groups, such as survivors of abuse, were concerned their groups would be targeted by trolls or scammers.<sup>210</sup></li> <li>• On April 28, updated the report flow for moderators to include “misinformation.” This allows subreddit moderators to track misinformation within their subreddit community and report users that post misinformation to Reddit’s investigations team.<sup>211</sup></li> <li>• Launched an AMA Series on Coronavirus. Reddit communities will host AMA sessions with medical professionals, health organizations, and authoritative voices on coronavirus and its impact.<sup>212</sup></li> <li>• Along with Twitter, removed links to a post from The Federalist suggesting “controlled voluntary infection” to the virus.<sup>213</sup></li> <li>• Relative to other platforms, it has faced considerable criticism for its lax approach to fighting coronavirus misinformation, appearing to leave most moderation activities in the hands of unpaid moderators. The platform does not have any explicit policy against health misinformation.<sup>214</sup></li> <li>• Featured a banner on its front page for a week directing users to an r/AskScience Megathread where they could find appropriate resources and authoritative content.</li> <li>• Allows posts on the science behind the virus to be tagged as peer-reviewed or not on the message board related to the virus.</li> <li>• Featured a homepage banner promoting the r/Coronavirus community as a source of timely discussion and the Centers for Disease Control as a</li> </ul>

<sup>209</sup> <https://www.tiktok.com/safety/resources/covid-19>

<sup>210</sup> <https://www.yahoo.com/lifestyle/reddit-pulls-start-chatting-after-angering-moderators-231655861.html>

<sup>211</sup>

[https://www.reddit.com/r/ModSupport/comments/g21ub7/misinformation\\_and\\_covid19\\_what\\_reddit\\_is\\_doing/](https://www.reddit.com/r/ModSupport/comments/g21ub7/misinformation_and_covid19_what_reddit_is_doing/)

<sup>212</sup> <https://redditblog.com/2020/03/02/expert-conversation-on-coronavirus/>

<sup>213</sup>

[https://www.huffpost.com/entry/facebook-coronavirus-infection-parties-misinformation-policy\\_n\\_5e7faf6cc5b6cb9dc1a16f36](https://www.huffpost.com/entry/facebook-coronavirus-infection-parties-misinformation-policy_n_5e7faf6cc5b6cb9dc1a16f36)

<sup>214</sup>

<https://firstdraftnews.org/latest/how-social-media-platforms-are-responding-to-the-coronavirus-infodemic/>

	<p>source for authoritative resources.</p> <ul style="list-style-type: none"> <li>• Users see a pinned message on search results for terms related to Coronavirus that directs them to CDC.gov for the latest information</li> <li>• May apply a quarantine to communities that contain hoax or misinformation content; quarantine removes the community from search results, warns the user that it may contain misinformation, and requires an explicit opt-in<sup>215</sup></li> </ul>
WhatsApp	<ul style="list-style-type: none"> <li>• On May 4, Poynter Institute’s International Fact-Checkers Network (IFCN) said it is launching a bot on WhatsApp to curb fake information related to coronavirus. The bot’s aim is to let people quickly check myths about the disease in 70 countries. IFCN said 80 fact-checking organizations from 74 countries have identified more than 4,000 coronavirus-related hoaxes since January. To use the bot, the user saves +1 (727) 2912606 as a contact or clicks on <a href="http://poy.nu/ifcnbot">http://poy.nu/ifcnbot</a> on an iPhone, sends a “hi”, and gets a menu which includes options to see the latest fact-checks, search by keyword, see tips on fighting misinformation, or link to other validated fact-checking organizations. The bot was financed with a \$1M grant from WhatsApp.<sup>216</sup></li> <li>• On April 7, imposed a new limit on message forwarding: if a user receives a frequently forwarded message – one which has been forwarded more than five times – they will only be able to send it on to a single chat at a time. The change does not completely prevent widespread forwarding, since a user can repeatedly hit the forward button.<sup>217</sup> (In 2019, WhatsApp reduced the number times a message could be forwarded: users could only forward one message to <u>five</u> chats. Group sizes were limited to 256 members. In 2018, WhatsApp reduced the number of times a message could be forwarded from 250 chats to <u>20</u>). Forwarded messages and chain messages are labeled with a double arrow icon.</li> <li>• On April 6, a group of Democratic senators sent a letter to Mark Zuckerberg saying Facebook hasn’t done enough to stop coronavirus misinformation on WhatsApp. Sens. Mazie Hirono (Hawaii), Kamala D. Harris (Calif.), Richard J. Durbin (Illinois) and Robert Menendez (N.J.) acknowledged the encryption challenge but called on Zuckerberg to take more drastic steps, including a prompt asking a user whether they have</li> </ul>

<sup>215</sup> <https://redditblog.com/2020/03/02/expert-conversation-on-coronavirus/>

<sup>216</sup> <https://9to5mac.com/2020/05/04/poynter-institute/>

<sup>217</sup>

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

	<p>verified a message is accurate before they forward it to another contact.<sup>218</sup></p> <ul style="list-style-type: none"> <li>• On March 20, launched WHO Health Alert, a chatbot developed with the World Health Organization that gives people (2B, in 4 languages) access to accurate information about COVID-19<sup>219</sup>.</li> <li>• WhatsApp is partnering with governments to create official WhatsApp accounts users can reach out to for accurate information about the virus, including India's "MyGov Corona Helpdesk" and the WHO's WhatsApp business account.<sup>220</sup></li> <li>• WhatsApp announced that it is donating \$1 million to the CoronaVirusFacts and DatosCoronaVirus alliance, created by IFCN in January this year to further COVID-19 fact checking efforts.<sup>221</sup></li> <li>• Users can forward messages to special accounts that can verify information, and world leaders, including in Ireland, are asking their citizens to stop spreading misinformation on the app.</li> </ul>
Apple	<ul style="list-style-type: none"> <li>• Released a Covid-19 website and app, developed in partnership with the Centers for Disease Control and Prevention, the White House's coronavirus task force, and the Federal Emergency Management Agency, which ask questions to screen for Covid-19 symptoms and point the public to resources for information on testing and physical distancing.</li> <li>• "Closely monitoring" new apps related to Covid-19 and updated its developer guidelines to note that only "recognized entities" — such as government organizations, health-focused nongovernmental organizations, medical or educational institutions, or "companies deeply credentialed in health issues" — should submit such apps.</li> </ul>

218

<https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2020/04/07/the-technology-202-democratic-senators-blast-whatsapp-s-handling-of-coronavirus-falsehoods/5e8b4bb8602ff10d49ade091/>

219

<https://www.who.int/news-room/feature-stories/detail/who-health-alert-brings-covid-19-facts-to-billions-via-whatsapp>

220

<https://www.firstpost.com/tech/news-analysis/covid-19-impact-how-whatsapp-google-microsoft-youtube-instagram-are-spreading-awareness-among-users-8178051.html>

221

<https://www.firstpost.com/health/whatsapp-launches-coronavirus-info-hub-will-donate-1-million-for-checking-fake-news-on-covid-19-pandemic-8168651.html>

## Recommendation

After three months of tracking platform efforts to counter misinformation about the novel coronavirus, we want to put forward a specific policy solution -- a “superfund for the internet” -- for countering misinformation flowing over digital platforms.

### The Politicization of the Pandemic

As background: even early in the pandemic, some hypothesized that the platforms would be more aggressive about addressing misinformation, in part to try to counter a growing “techlash narrative” and the growing risk of regulation, but also because the pandemic presented a fundamentally different kind of information challenge. For example, Claire Wardle, the co-founder and leader of First Draft, the world’s foremost nonprofit focused on research and practice to address mis- and disinformation, believed “There are no two sides with coronavirus, so they don’t have people on the other side saying: ‘We want this,’ the way you do with anti-vaxxers or political misinformation. [The platforms] are freer to act.”<sup>222</sup>

Facebook CEO Mark Zuckerberg, who has been notoriously resistant to moderation of political content on the platform, also noted, “The difference between good and bad information is clearer in a medical crisis than in the world of, say, politics. When you’re dealing with a pandemic, a lot of the stuff we’re seeing just crossed the threshold. So it’s easier to set policies that are a little more black and white and take a much harder line.”<sup>223</sup>

But that’s not how it played out. Very quickly, information about the novel coronavirus pandemic became every bit as politicized as what we normally consider highly partisan topics. This information -- about the severity of the virus, its origins, treatments, and whether or how to “open America,” for a few examples -- has been subject to the same patterns of creation and distribution as political content designed to sow division and undermine democratic institutions. That includes content created or shared by trolls and bots, foreign interference, and amplification by partisan players. The only good news about this nightmare scenario is that it makes the pandemic a very appropriate model for how platforms should manage other types of misinformation, including overtly political misinformation. In fact, they may be one and the same, since there is now strong evidence of the danger to democracy posed by pandemic-related disinformation from foreign parties, which is being used to weaken democratic checks on power or interfere with elections<sup>224</sup>.

---

<sup>222</sup>

<https://www.theguardian.com/world/2020/apr/10/tech-giants-struggle-stem-infodemic-false-coronavirus-claims>

<sup>223</sup> <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html>

<sup>224</sup> <https://foreignpolicy.com/2020/04/30/2020-election-interference-russian-coronavirus-disinformation/>

## How Digital Platforms Have Countered Pandemic Misinformation

Although we tracked all the major platforms through three months of the pandemic, considerations about information quality and the media are generally seen as specific to companies that engage largely in open (non-encrypted) information distribution such as Google, YouTube, Facebook, and Twitter. For this reason, the focus of this perspective is primarily on these platforms.

Simply put, based on our tracking, the efforts of these platforms to counter misinformation about the pandemic far exceed anything they've done -- or said they could do -- in the past. In the process, they have undermined their own past arguments for an arguably lax approach, ranging from "It's too hard!" to "Free Speech!" to "It's not our job." But besides their willingness to change their posture and user experience design in favor of content moderation, there's one thing that has enabled their approaches: more extensive partnering with other organizations, whose authoritative content and information analysis has enabled them to check sources, up- and down-rank content, direct people who've experienced misinformation to debunking sites, and understand what kinds of misinformation may create the greatest harm. Those relationships have allowed the platforms to deploy some of the most proven strategies for countering misinformation<sup>225</sup>, regardless of content or context. These include countering with accurate information, evaluating the source, avoiding binary solutions, setting priorities for remediation, and increasing the salience<sup>226</sup> of accuracy. And to some extent, it's been the platforms' partnerships with trusted sources of authoritative information -- including the WHO, CDC, and fact checking organizations -- that has allowed them to act so aggressively without appearing politically biased.

### A Policy Solution for Misinformation: "A Superfund for the Internet"

In general, we give these platforms credit for their efforts during the coronavirus crisis thus far. But we got a preview of what may happen to all that good work after the crisis when Mark Zuckerberg said, in that same interview<sup>227</sup>, that it was "hard to predict" how things would play out after the pandemic, and reiterated that the kind of threats posed by misinformation about the virus were "in a different class" (though many weeks later, he may not still feel that way).

---

<sup>225</sup>

<https://www.axios.com/what-experts-say-works-for-combating-coronavirus-misinformation-146084b8-1312-403d-b948-3e190f16b7f3.html>

<sup>226</sup> <https://psyarxiv.com/3n9u8/>

<sup>227</sup> <https://www.nytimes.com/2020/03/15/business/media/coronavirus-facebook-twitter-social-media.html>

Given how much of the misinformation problem is generated through the pervasive reach, speed and power of digital platforms, we believe it is critical that the effective strategies described above become fully embedded with the major information distribution platforms. We would like to see the platforms themselves, accountable to independent expert bodies established through legal mandate, master the process of identifying, minimizing, and helping the public navigate disinformation -- without interfering with Constitutionally-protected speech rights. This is particularly necessary in contexts where the quality of information is of high stakes, where spread of mis- and disinformation is virulent and destructive, and where salience or engagement is high. Given the risk for harm, we believe the platforms' efforts shouldn't be reliant on their continued good will and philanthropy, or conducted in the absence of oversight.

Given the enormous and now proven value of information analysis to support public health and institutions, we can imagine, and are now developing, a solution in which platforms are compelled to invest much more in the tools and approaches that work. We're thinking of this as a trust fund, or "superfund," modeled on the 1980 Superfund for clearing toxic waste sites. Unlike other, similar concepts, (like here<sup>228</sup> and here<sup>229</sup> and here<sup>230</sup> and here<sup>231</sup> and here<sup>232</sup> and here<sup>233</sup> and here<sup>234</sup> and here<sup>235</sup>) though, we don't believe a punitive "tax" on advertising revenue -- which isn't really the direct source of the problem -- is the preferable approach. We favor an approach of value creation, since the pandemic has given us such a powerful model for its benefits. It has essentially created a market in which the platforms have more demand for -- and journalistic organizations have more supply of -- information cleansing services. The platforms should pay for these services to help to clear the toxic junk from their platforms, at a fair price. It's an exchange of real value that would preclude any assumption, expectation, or threat of editorial influence. In doing so, we can provide an essential new revenue stream to local journalistic organizations and information analysts who also help protect our public and democratic institutions.

---

<sup>228</sup> <https://www.freepress.net/policy-library/beyond-fixing-facebook>

<sup>229</sup> <https://www.nytimes.com/2019/05/06/opinion/tax-facebook-google.html>

<sup>230</sup>

<https://www.theverge.com/2019/8/27/20835018/bernie-sanders-targeted-advertising-tax-google-facebook-journalism>

<sup>231</sup> <https://knightcolumbia.org/content/the-case-for-digital-public-infrastructure>

<sup>232</sup> <https://www.berggruen.org/activity/renewing-democracy-in-the-digital-age/>

<sup>233</sup> <https://hbr.org/2020/03/journalisms-market-failure-is-a-crisis-for-democracy>

<sup>234</sup>

[https://www.gmfus.org/sites/default/files/Kornbluh%20Goodman%20\\_5%20steps%20to%20combat%20the%20infodemic.pdf](https://www.gmfus.org/sites/default/files/Kornbluh%20Goodman%20_5%20steps%20to%20combat%20the%20infodemic.pdf)

<sup>235</sup>



## Not “Back to Normal” for Misinformation

Just like so many other aspects of life during the pandemic, we shouldn't expect -- or allow -- the platforms to go “back to normal” when the crisis is over. As the whole world has gone online for working, learning, telehealth, and entertainment, the platforms' power has only grown, and with it, their responsibility and accountability to the public. We need both a policy framework<sup>236</sup> and specialized regulatory authority<sup>237</sup> to limit their anti-competitive behaviors, protect Americans' privacy, and to stop or slow the spread of disinformation online. A superfund for the internet, which also fosters reputable journalism, is the next step.

---

<sup>236</sup> <https://www.publicknowledge.org/blog/the-right-way-to-regulate-digital-platforms/>

<sup>237</sup> <https://actnow.io/Qx4TqVG>