2020

Misinformation Mayhem: Social Media Platforms' Efforts to Combat Medical and Political Misinformation

Dawn C. Nunziato
George Washington University Law School, dnunziato@law.gwu.edu

Recommended Citation
19 First Amendment L. Rev. ___ (2020).

Misinformation Mayhem:
Social Media Platforms' Efforts to Combat Medical and Political Misinformation

– Dawn Carla Nunziato[1]

INTRODUCTION

Social media platforms today are playing an ever-expanding role in shaping the contours of today's information ecosystem.[2] The events of recent months have driven home this development, as the platforms have shouldered the burden and attempted to rise to the challenge of ensuring that the public is informed – and not misinformed – about matters affecting our democratic institutions in the context of our elections, as well as about matters affecting our very health and lives in the context of the pandemic. This Article examines the extensive role recently assumed by social media platforms in the marketplace of ideas in the online sphere, with an emphasis on their efforts to combat medical misinformation in the context of the COVID-19 pandemic as well as their efforts to combat false political speech in the 2020 election cycle. In the context of medical misinformation surrounding the COVID-19 pandemic, this Article analyzes the extensive measures undertaken by the major social media platforms to combat such misinformation. In the context of misinformation in the political sphere, this Article examines the distinctive problems brought about by the microtargeting of political speech and by false political ads on social media in recent years, and the measures undertaken by major social media companies to address such problems. In both contexts, this Article examines the extent to which such measures are compatible with First Amendment substantive and procedural values.

Social media platforms are essentially attempting to address today's serious problems alone, in the absence of federal or state regulation or guidance in the United States. Despite the major problems caused by Russian interference in our 2016 elections, the U.S. has failed to enact regulations prohibiting false or misleading political advertising on social media – whether originating from foreign sources or domestic ones – because of First Amendment, legislative, and political impediments to such regulation. And the federal government has failed miserably in its efforts to combat COVID-19 or the medical

---

[1] William Wallace Kirkpatrick Research Professor and Professor of Law, The George Washington University Law School; Co-Director, Global Internet Freedom Project. I am extremely grateful to Nathaniel Christiansen, Chris Frascella, Conor Kelly, and Ken Rodriguez for providing excellent research and library assistance in connection with this article, to Kierre Hannon for excellent administrative assistance, and to Associate Dean for Research and Faculty Development Thomas Colby and Interim Dean Chris Bracey for academic and financial support of my research.

[2] See, e.g., Kate Klonick, The New Governors: The People, Rules, and Processes Governing Online Speech, 131 Harv. L. Rev. 1598 (2018).

misinformation that has contributed to the spread of the virus in the U.S. All of this essentially leaves us (in the United States, at least) solely in the hands, and at the mercy, of the platforms themselves, to regulate our information ecosystem (or not), as they see fit.

The dire problems brought about by medical and political misinformation online in recent months and years have ushered in a sea change in the platforms' attitudes and approaches toward regulating content online. In recent months, for example, Twitter has evolved from being the non-interventionist "free speech wing of the free speech party"[3] to designing and operating an immense operation for regulating speech on its platform – epitomized by its recent removal[4] and labeling[5] of President Donald Trump's (and Donald Trump, Jr.'s) misleading tweets. Facebook for its part has evolved from being a notorious haven for fake news in the 2016 election cycle[6] to standing up an extensive global network of independent fact-checkers to remove and label millions of posts on its platform – including by removing a post from President Trump's campaign account, as well as by labeling 90 million such posts in March and April 2020, involving false or misleading medical information in the context of the pandemic. Google for its part has abandoned its hands-off approach to its search algorithm results and has committed to removing false political content in the context of the 2020 election[7] and to serving up prominent information by trusted health authorities in response to COVID-19 related searches on its platforms.[8]

These approaches undertaken by the major social media platforms are generally consistent with First Amendment values, both the substantive values in terms of what constitutes protected and unprotected speech, and the procedural values, in terms of process accorded to users whose speech is restricted or otherwise subject to action by the platforms. As I discuss below, the platforms have removed speech that is likely to lead to imminent harm and have generally been more aggressive in responding to medical misinformation

---

[3] *See* e.g., Marvin Ammori, The 'New' New York Times: Free Speech Lawyering in the Age of Google and Twitter, 127 Harv. L. Rev 2259 (2014); Josh Halliday, *Twitter's Tony Wang: 'We Are the Free Speech Wing of the Free Speech Party'*, THE GUARDIAN (Mar. 22, 2012), https://www.theguardian.com/media/2012/mar/22/twitter-tony-wang-free-speech.

[4] *See* Arjun Kharpal, *Twitter Removes an Image Tweeted by Trump for Violating Its Copyright Policy*, CNBC (Jul. 2, 2020), https://www.cnbc.com/2020/07/02/twitter-removes-trump-image-in-tweet-for-violating-copyright-policy.html.

[5] *See* Elizabeth Dwoskin, *Trump Lashes Out at Social Media Companies After Twitter Labels Tweets with Fact Checks*, WASH. POST (May 27, 2020), https://www.washingtonpost.com/technology/2020/05/27/trump-twitter-label/.

[6] See text accompanying notes x - y.

[7] See text accompanying notes x - y.

[8] See text accompanying notes x - y.

than political misinformation. This approach tracks First Amendment substantive values, which accord lesser protection for false and misleading claims regarding medical information than for false and misleading political claims. The platforms' approaches generally adhere to First Amendment procedural values as well, including by specifying precise and narrow categories of what speech is prohibited, providing clear notice to speakers who violate their rules regarding speech, applying their rules consistently, and according an opportunity for affected speakers to appeal adverse decisions regarding their content.

While the major social media platforms' intervention in the online marketplace of ideas is not without its problems and not without its critics, this Article contends that this trend is by and large a salutary development – and one that is welcomed by the vast majority of Americans[9] and that has brought about measurable improvements in the online information ecosystem.[10] Recent surveys and studies show that such efforts are welcomed by Americans[11] and are moderately effective in reducing the spread of misinformation and in improving the accuracy of beliefs of members of the public.[12] In the absence of effective regulatory measures in the United States to combat medical and political misinformation online,[13] social media companies should be encouraged to continue to experiment with

[9] See text accompanying notes x - y.
[10] See text accompanying notes x - y.
[11] See text accompanying notes x - y.
[12] See text accompanying notes x - y.
[13] The United States government's hands-off approach to combating misinformation on social media stands in stark contrast to the aggressive approach undertaken by the European Union. The European Commission in 2018 adopted an aggressive Action Plan against Disinformation as part of the Commission's continued attempt to crack down on fake news. *See Roadmaps to Implement the Code of Practice on Disinformation*, EUROPEAN COMMISSION (Jul. 7, 2020) , https://ec.europa.eu/digital-single-market/en/news/roadmaps-implement-code-practice-disinformation. The Action Plan focuses on improving detection of disinformation; implements a coordinated alert system among E.U. institutions and member states to enable them to better respond to disinformation in real time; and calls for the adoption of a self-regulatory Code of Practice on Disinformation. Several major social media platforms have now signed on to the EU Code of Practice on Disinformation, including Facebook, Google, Twitter, Microsoft, and, most recently, TikTok. The E.U. Code of Practice on Disinformation seeks to ensure transparency in political advertising, the removal of fake accounts, identifying non-human interactions, cooperating with fact-checkers to detect disinformation, and promoting fact-checked content. Each of the platforms that have signed on to the Code presented detailed plans setting forth the tools that it commits to deploy against disinformation on its platform. *See Code of Practice on Disinformation*

developing and deploying even more effective measures to combat such misinformation, consistent with our First Amendment substantive and procedural values.

I. Platforms' Efforts to Address Medical Misinformation in the Context of the Pandemic

In recent months, arguably the most important challenge for social media platforms has been responding to the rampant spread of medical misinformation in the context of the COVID-19 pandemic. With a significant portion of the global community under some kind of lockdown order (one third of the global population, by one estimate[14]), Internet connectivity – and Internet content – are playing a more significant societal role than ever. In contrast to their previous hands-off position, the major platforms have risen to the challenge and have taken decisive action in response to medical misinformation in the context of the pandemic. The predominant focus across platforms has been on the curbing of false information, especially that which tends to encourage the spread of imminently harmful information about the virus. The platforms' actions taken in response to COVID-19-related medical misinformation have generally been more aggressive than their response to misinformation in the political arena, which is consistent with First Amendment substantive values that accord lesser protection for false and misleading statements of fact than for false and misleading political claims (as I discuss below). And

---

One Year On: Online Platforms Submit Self-Assessment Reports, EUROPEAN COMMISSION (Oct. 29, 2019), https://ec.europa.eu/commission/presscorner/detail/en/STATEMENT_19_6166.
See Facebook's commitment and roadmap for implementation:
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54789;
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54788;
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62681;
Twitter's commitment and roadmap for implementation:
https://blog.twitter.com/en_us/search.html?q=disinformation;
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54784;
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54780;
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62682;
Google/YouTube's commitment and roadmap for implementation: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54785
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=54781
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=62680; and
TikTok's commitment and roadmap for implementation:
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68231
https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=68231

[14] Juliana Kaplan, Lauren Frias and Morgan Mcfall-Johnson, *Our Ongoing List of How Countries Are Reopening And Which Ones Remain Under Lockdown*, BUS. INSIDER, available at https://www.businessinsider.com/countries-on-lockdown-coronavirus-italy-2020-3?r=DE&IR=T (last accessed July 21, 2020).

the platforms' actions in the context of medical misinformation generally track First Amendment substantive values by prohibiting false and imminently harmful information. In general, the platforms have undertaken extensive measures to remove imminently harmful false medical information (e.g., posts that advocate drinking bleach to cure COVID-19), while taking less severe measures – including labeling/countering with counter-speech, and/or reducing the reach of, less harmful or misleading medical information (e.g., posts that tout conspiracy theories claiming that Dr. Anthony Fauci created the virus). Although the platforms' efforts thus far are commendable, what is needed at this point is for them to act much more quickly to remove harmful false and misleading medical misinformation before it goes viral, as I discuss below.

Facebook's Response to Medical Misinformation

Facebook has responded to the rampant spread of misinformation on its platform in the context of the pandemic by: removing harmful false/misleading posts related to COVID-19 – including, recently, posts by President Trump; labeling other, less harmful posts; and issuing strong warnings to those who have shared or reacted to such posts. In addition, Facebook has made prominently available its Coronavirus Information Center, a repository of curated, expert information about the virus.

Facebook's current COVID-19 misinformation policy as follows:

> We remove COVID-19 related misinformation that could contribute to imminent physical harm. We've removed harmful misinformation since 2018, including false information about the measles in Samoa where it could have furthered an outbreak and rumors about the polio vaccine in Pakistan where it risked harm to health aid workers. Since January [2020], we've applied this policy to misinformation about COVID-19 to remove posts that make false claims about cures, treatments, the availability of essential services or the location and severity of the outbreak. We regularly update the claims that we remove based on guidance from the WHO and other health authorities. For example, we recently started removing claims that physical distancing doesn't help prevent the spread of the coronavirus. We've also banned ads and commerce listings that imply a product guarantees a cure or prevents people from contracting COVID-19. … For claims that don't directly result in physical harm, like conspiracy theories about the origin of the virus, we continue to work with our network of over 55 fact-checking partners covering over 45 languages to debunk these claims…. Once a post is rated false by a fact-checker, we reduce its distribution so fewer people see it, and we show strong warning labels and notifications to people who still come across it, try to share it or already have.[15]

[15] https://about.fb.com/news/2020/03/combating-covid-19-misinformation/

In a surprising move in August 2020, Facebook implemented its COVID-19 misinformation policy to delete a post from President Trump's campaign account, in which Trump can be heard to say on video, in the context of re-opening schools, that children are "almost immune" to the coronavirus.[16] While Facebook does not frequently remove medical misinformation, its Community Standards allow for removal of misinformation that contributes to the risk of physical harm or imminent violence. In response to guidance from external experts, including the World Health Organization and local health authorities, Facebook now requires the removal of false claims about: the existence or severity of COVID-19, how to prevent COVID-19, how COVID-19 is transmitted (including false claims that certain racial groups are immune to the virus), cures for COVID-19, and access to or the availability of essential services.[17]

Facebook has broadened its work with certified independent fact-checking organizations[18] as part of its effort to curb the spread of medical misinformation, adding eight new dedicated fact-checking partners and "expanding [its] coverage to more than a dozen new countries."[19] Facebook's approach to medical misinformation generally focuses less on removing false content and more on reducing the distribution of medical misinformation once one of its independent fact-checking partners has rated it as false.[20] To the everyday user, Facebook's approach takes the form of warning displays on posts

---

[16] Heather Kelly, *Facebook, Twitter Penalize Trump for Posts Containing Coronavirus Misinformation*, WASH. POST (Aug. 5, 2020), https://www.washingtonpost.com/technology/2020/08/05/trump-post-removed-facebook/

[17] *See Facebook's Civil Rights Audit – Final Report*, July 8, 2020.

[18] "To reduce the spread of misinformation and provide more reliable information to users, we partner with independent third-party fact-checkers globally who are certified through the non-partisan International Fact-Checking Network (IFCN)." *Partnering with Third-Party Fact-Checkers*, Facebook: Journalism Project (Mar. 23, 2020), https://www.facebook.com/journalismproject/programs/third-party-fact-checking/selecting-partners (last accessed July 21, 2020). IFCN verified signatories: *Verified Signatories of the IFCN Code of Principles*, https://ifcncodeofprinciples.poynter.org/signatories (last accessed July 21, 2020).

[19] Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, FACEBOOK NEWSROOM (Apr. 16, 2020), https://about.fb.com/news/2020/04/COVID-19-misinfo-update/. Facebook has also expanded the program to Instagram and now boasts "more than 60 fact-checking partners covering more than 50 languages around the world." Guy Rosen, *Investments to Fight Polarization*, FACEBOOK NEWSROOM (May 27, 2020), https://about.fb.com/news/2020/05/investments-to-fight-polarization/.

[20] Guy Rosen, *An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19*, FACEBOOK NEWSROOM (Apr. 16, 2020), https://about.fb.com/news/2020/04/COVID-19-misinfo-update/.

that have been deemed false, with Facebook issuing 40 million such warnings in March 2020 and 50 million in April 2020.[21] Facebook claims that when people see such warning labels, "95% of the time they did not go on to view the original content."[22] In addition, in response to the pandemic, Facebook has revised its content reviewer guidance to make clear that claims such as that people of certain races or religions have the virus, created the virus, or are spreading the virus violate Facebook's hate speech policies.[23]

As part of its general counterspeech approach – of presenting users with accurate information in response to false and misleading information, as opposed to by censoring false information – Facebook has also taken the step of reaching out to users who have interacted with (i.e., reacted to or commented on) medical misinformation related to COVID-19 and connecting those users with responses to common "myths" about COVID-19 that have been identified and addressed by the World Health Organization and inviting these users to share the link with others.[24] See notice below.

The most common of the myths shared on Facebook tend to suggest ineffective or potentially harmful remedies for COVID-19, such as drinking bleach or disinfectant, or taking unproven and potentially harmful drugs such as hydroxychloroquine.[25] Other myths

[21] *Id.*
[22] *Id.*
[23] *Id.*
[24] *Id.*
[25] *Coronavirus Disease Advice for the Public: Mythbusters*, World Health Organization, https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters (last accessed July 21, 2020).

commonly seen on the platform are those claiming that measures scientifically proven to contain the spread of the virus – such as social distancing – are ineffective.[26]

In attempting to combat medical misinformation in the context of the pandemic, Facebook has also stood up its Coronavirus Information Center.[27] This feature, which Facebook initially placed prominently at the top of the News Feed (so that it was immediately visible upon opening the platform) serves as a rolling collection of relevant updates about the pandemic as this information became available from both national and global health authorities.[28]

Facebook's approach to combating medical misinformation on its platform[29] is heading in the right direction but is plagued by unacceptable delays. A comprehensive

---

[26] Some recent research has highlighted the parallels in the sharing of disinformation in the current pandemic with the dissemination of information on supposed "cures" via newspapers during the 1918 flu pandemic. As Elizabeth Zetland, a researcher at MyHeritage, puts it, "You were meant to cook 12 onions, get the juice and drink it the day afterwards, and that would protect you from the flu." *See* Suyin Haynes, *'You Must Wash Properly.' Newspaper Ads From the 1918 Flu Pandemic Show Some Things Never Change*, TIME (Mar. 27, 2020), available at https://time.com/5810695/spanish-flu-pandemic-coronavirus-ads/. Newspapers were quick to urge individuals to wear or make their own masks; the Red Cross, in an ad placed in the Daily Gazette of Berkeley, California, called anyone not wearing a mask "a dangerous slacker." *See id.*

[27] https://about.fb.com/news/2020/07/coronavirus

[28] Kang-Xing Jin, "Launching the Coronavirus Information Center on Facebook", *Keeping People Safe and Informed About the Coronavirus*, FACEBOOK NEWSROOM (updated July 16, 2020) https://about.fb.com/news/2020/07/coronavirus/#coronavirus-info-center.

[29] Facebook's approach to combating medical misinformation also encompasses its response to protests involving stay-at-home measures that authorities have deemed necessary to curb the spread of the pandemic. Thus far, the company's response to such protests has been inconsistent. Facebook has removed posts organizing anti-stay-at-home protests in California, New Jersey, and Nebraska after determining - in consultation with state officials - that the protests violated the states' social distancing rules. Donie O'Sullivan and Brian Fung, *Facebook Will Take Down Some, But Not All, Posts Promoting Anti-Stay-at-Home Protests*, CNN POLITICS (Apr. 20, 2020), https://www.cnn.com/2020/04/20/politics/facebook-COVID-shutdown-protests/index.html. In Pennsylvania, however, an anti-lockdown group with more than 66,000 members promoted a lockdown protest scheduled to take place in Harrisburg, without any action from Facebook. *Id.* Facebook's efforts in this area are sporadic and appear to lack a coherent strategy. In New Jersey, for example, state officials had not specifically requested that Facebook take down content promoting anti-lockdown events, but Facebook staff had been "communicating about the issue" with the governor's staff. *Id.* In Nebraska, Facebook contacted the governor's office "to learn more about Nebraska's social distancing restrictions, and the governor's staff provided already

study undertaken by the human rights group Avaaz examined the dissemination of over 100 pieces of misinformation content about the virus that were rated false and/or misleading by independent fact-checkers and that could cause public harm.[30] Avaaz's review found that "millions of the platform's users are still being put at risk," and that "the pieces of [false and/or misleading] content [sampled by Avaaz] were shared over 1.7 million times on Facebook, and viewed an estimated 117 million times.[31] For example, according to Avaaz, "a harmful misinformation post that claimed that one way to rid the body of the virus is to … gargle with water, salt or vinegar was shared over 31,000 times before eventually being taken down after Avaaz flagged this content for action by Facebook."[32]Avaaz found that, notwithstanding Facebook's policy described above, 41 percent of virus-related misinformation content remains on the platform without warning labels, including 65% of content that had been flagged by Facebook's own fact-checking program.[33] Beyond failing to apply warning labels to content that its fact-checking partners deemed to be misleading, Facebook suffers from delays in the implementation of its policies. In this age of instant digital news, unless imminently harmful medical disinformation is rapidly curbed, it runs the risk of hastening the spread of the virus. And yet according to Avaaz, it "can take up to 22 days for the platform to downgrade [false and/or misleading content related to the virus] and issue warning labels." The lag is even more severe in the case of non-English content, where "over half (51%) of non-English misinformation content had no warning labels."[34] Fortunately, Facebook seems willing to change course, as evidenced by its willingness to institute a retroactive alert system, whereby each user exposed to harmful misinformation will be notified and provided with

---

publicly available information about Nebraska's 10-person limit and directed health measures." *Id.* Facebook is apparently reaching out to governments on these matters because "[u]nless government prohibits the event during this time, we allow it to be organized on Facebook." *Id.* Yet, at least in the case of Nebraska, Facebook's effort seems to be a somewhat fumbling one, with Facebook employees reaching out to state officials to learn about information that is already readily available. Facebook's hesitance to remove posts related to protests likely stems from a deeper worry about becoming the online policeman on the question of the constitutional right to assemble. *Id.*

[30] *How Facebook can Flatten the Curve of the Coronavirus Infodemic,* Avaaz (Apr. 15, 2020), https://secure.avaaz.org/campaign/en/facebook_coronavirus_misinformation/. For the full text of the report, see https://avaazimages.avaaz.org/facebook_coronavirus_misinformation.pdf.

[31] *Id.*

[32] *How Facebook can Flatten the Curve of the Coronavirus Infodemic,* Avaaz (Apr. 15, 2020). "2,611 clones" of the same false post "remain on the platform with over 92,246 interactions. Most of these cloned posts have no warning labels from Facebook."

[33] *Id.* (emphasis in original).

[34] *Id.*

accurate information. (Avaaz indicated that members of Facebook's misinformation team made such a commitment in a conversation with Avaaz staff in April 2020.[35])

Facebook's intent to combat medical misinformation in the context of the pandemic is commendable, but the failures in the timely implementation of its new policies are highly problematic given the degree of the public health risk. While Facebook cannot reasonably be expected to identify and curb every piece of misinformation on the pandemic on its platform, it must commit to staffing up and improving the implementation of its measures to counter medical misinformation. Facebook's success rate in curbing harmful misinformation might never be perfect, but its present approach has glaring flaws that must be remedied.

Twitter's Response to Medical Misinformation

Twitter's response to medical misinformation in the context of the pandemic – like its response to false political ads, as discussed below – has been more forceful than that undertaken by Facebook. According to this response, Twitter will remove harmful posts containing medical misinformation that "could pose a threat to people's health or well-being," will label other posts containing medical misinformation, and will also direct users to truthful and accurate information about the pandemic.

Under a recent update to Twitter's rules, tweets that cause harm (including in the context of the pandemic) will be removed. Accordingly, Twitter will remove from its platform tweets along the lines of "social distancing is not effective" or "the news about washing hands is propaganda."[36] An important component of Twitter's effort includes broadening its definition of harmful tweets, so as to more proactively target and remove content that expressly contradicts the most up-to-date guidance from authoritative health sources (such as the Center for Disease Control and the World Health Organization).[37] Relevant portions of Twitter's currently applicable set of rules regarding its handling of pandemic-related misinformation is set out in the margin.[38]

---

[35] *See supra* note x.

[36] Jack Morse, *Twitter Steps up Enforcement in the Face of Coronavirus Misinformation*, MASHABLE (Mar. 18, 2020), https://mashable.com/article/twitter-cracks-down-coronavirus-misinformation/.

[37] @vijaya and @Derelia (Vijaya Gadde and Matt Derella), *An Update on Our Continuity Strategy During COVID-19*, TWITTER BLOG (Mar. 16, 2020), https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html.

[38] Steps we're taking

...

Broadening our definition of harm to address content that goes directly against guidance from authoritative sources of global and local public health information. Rather than reports, we will enforce this in close coordination with trusted partners, including public health authorities and governments, and continue to use and consult with information from those sources when reviewing content.

- We'll continue to prioritize removing content when it has a clear call to action that could directly pose a risk to people's health or well-being, but we want to make it clear that we will not be able to take enforcement action on every Tweet that contains incomplete or disputed information about COVID-19. This is not meant to limit good faith discussion or expressing hope about ongoing studies related to potential medical interventions that show promise.
- … We will continue to use both technology and our teams to help us identify and stop spammy behavior and accounts.
- We may also apply the public interest notice in cases where world leaders violate the COVID-19 guidelines.

Under this guidance, we will require people to remove tweets that include:

- Denial of global or local health authority recommendations to decrease someone's likelihood of exposure to COVID-19 with the intent to influence people into acting against recommended guidance, such as: "social distancing is not effective", or actively encouraging people to not socially distance themselves in areas known to be impacted by COVID-19 where such measures have been recommended by the relevant authorities.
- Description of alleged cures for. COVID-19, which are not immediately harmful but are known to be ineffective, are not applicable to the COVID-19 context, or are being shared with the intent to mislead others, even if made in jest, such as "coronavirus is not heat-resistant - walking outside is enough to disinfect you" or "use aromatherapy and essential oils to cure COVID-19."
- Description of harmful treatments or protection measures which are known to be ineffective, do not apply to COVID-19, or are being shared out of context to mislead people, even if made in jest, such as "drinking bleach and ingesting colloidal silver will cure COVID-19."
- Denial of established scientific facts about transmission during the incubation period or transmission guidance from global and local health authorities, such as "COVID-19 does not infect children because we haven't seen any cases of children being sick."
- Specific claims around COVID-19 information that intends to manipulate people into certain behavior for the gain of a third party with a call to action within the claim, such as "coronavirus is a fraud and not real - go out and patronize your local bar!!" or "the news about washing your hands is propaganda for soap companies, stop washing your hands".
- Specific and unverified claims that incite people to action and cause widespread panic, social unrest or large-scale disorder, such as "The National Guard just announced that no more shipments of food will be arriving for 2 months - run to the grocery store ASAP and buy everything!"

Twitter states that it is working with "trusted partners" (including public health authorities and governments) to both identify and remove harmful medical misinformation

---

- Specific and unverified claims made by people impersonating a government or health official or organization such as a parody account of an Italian health official stating that the country's quarantine is over.
- Propagating false or misleading information around COVID-19 diagnostic criteria or procedures such as "if you can hold your breath for 10 seconds, you do not have coronavirus."
- False or misleading claims on how to differentiate between COVID-19 and a different disease, and if that information attempts to definitively diagnose someone, such as "if you have a wet cough, it's not coronavirus - but a dry cough is" or "you'll feel like you're drowning in snot if you have coronavirus - it's not a normal runny nose."
- Claims that specific groups, nationalities are never susceptible to COVID-19, such as "people with dark skin are immune to COVID-19 due to melanin production" or "reading the Quran will make an individual immune to COVID-19."
- Claims that specific groups, nationalities are more susceptible to COVID-19, such as "avoid businesses owned by Chinese people as they are more likely to have COVID-19."

...
Instituting a global content severity triage system so we are prioritizing the potential rule violations that present the biggest risk of harm and reducing the burden on people to report them.
Executing daily quality assurance checks on our content enforcement processes to ensure we're agile in responding to this rapidly evolving, global disease outbreak.
Engaging with our partners around the world to ensure escalation paths remain open and urgent cases can be brought to our attention.
Continuing to review the Twitter Rules in the context of COVID-19 and considering ways in which they may need to evolve to account for new behaviors.
As we've said on many occasions, our approach to protecting the public conversation is never static. That's particularly relevant in these unprecedented times. We intend to review our thinking daily and will ensure we're sharing updates here on any new clarifications to our rules or major changes to how we're enforcing them.
Finally, we're encouraged that our service is being used around the world to provide free, authoritative health information, and to ensure that everyone has access to the conversations they need to protect themselves and their families. For more, our dedicated COVID-19 Event page has the latest facts right at the top of your timeline. @vijaya and @Derelia (Vijaya Gadde and Matt Derella), *An Update on Our Continuity Strategy During COVID-19*, TWITTER BLOG (Mar. 16, 2020), https://blog.twitter.com/en_us/topics/company/2020/An-update-on-our-continuity-strategy-during-COVID-19.html.

content.[39] The range of content that Twitter states it will remove is broad. Twitter is focused on tweets that contain "denial of global or local health authority recommendations," those that "describe alleged cures for COVID-19, which are not immediately harmful but are known to be ineffective," those that "describe harmful treatments or protection measures which are known to be ineffective," and those that "deny established scientific facts about transmission during the incubation period."[40]

Twitter is also targeting tweets that go beyond medical misinformation and appear to encourage societal unrest. For example, the company states that it will target and remove tweets that contain "specific and unverified claims that incite people to action and cause widespread panic, social unrest, or large-scale social disorder," as well as tweets that contain "specific and unverified claims made by people impersonating a government or health official or organization," such as a parody account of an Italian health official stating that the country's quarantine is over.[41]

In May 2020, Twitter announced a policy of placing warning labels on tweets containing misinformation related to COVID-19, including tweets that are issued by world leaders. According to Twitter's head of site integrity Yoel Roth, such warning labels will apply to anyone sharing misleading information that meets the requirements of Twitter's policy, and no exceptions will be made for the tweets of world leaders.[42]

Pursuant to its policy of removing COVID-19 related medical misinformation, including from world leaders, in August 2020 Twitter removed a tweet from the Trump Campaign in which Trump claimed that children being "almost immune" to the virus. Twitter suspended the account's tweeting privileges until the post was deleted, citing its rules on COVID-19.[43]

In addition to removing and labeling tweets in its attempts to restrict false and misleading medical information in the context of the pandemic, Twitter is also pursuing other means, including by restricting the functionality of Twitter accounts that spread such information. For example, on July 28, Twitter penalized Donald Trump, Jr., for posting misinformation in the form of a Breitbart video showing a group of doctors making misleading and false claims about the COVID-19 pandemic. See below. In the video, a group of people dressed in white lab coats who call themselves "America's Frontline Doctors" staged a press conference in front of the U.S. Supreme Court making claims such

---

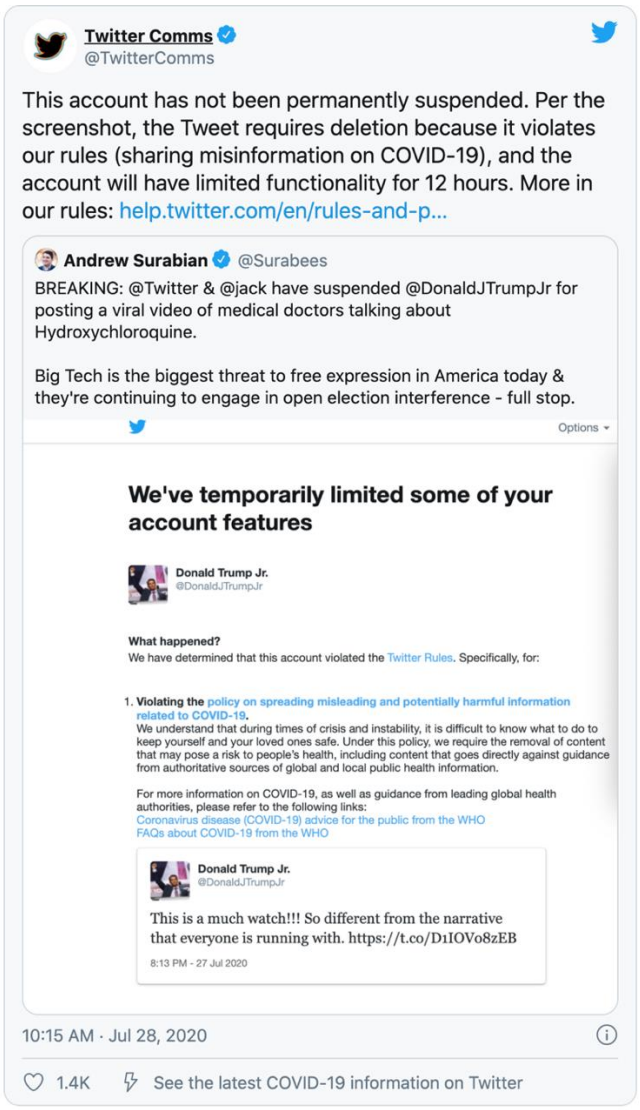
[39] *Id.*
[40] *Id.*
[41] *Id.*
[42] Yoel Roth (@yoyoel), TWITTER (May 11, 2020, 3:10 PM), https://twitter.com/yoyoel/status/1259923758522855426.
[43] Kelly, *supra* note x. Facebook deleted the same post. *Id.*

as that hydroxychloroquine is "a cure for Covid" and "you don't need a mask" to slow the spread of coronavirus.[44] Twitter order Trump Jr. to delete this misleading tweet, added a note to its trending topics warning about the potential dangers of hydroxychloroquine, and in addition, took measures to limit Trump Jr.'s account functionality for 12 hours.[45] (Facebook and YouTube also removed the offending video, but not before it had been viewed millions of times.[46])

[44] *See* Sam Shead, *Facebook, Twitter and YouTube Pull 'False' Coronavirus Video After It Goes Viral*, CNBC (July 28, 2020), https://www.cnbc.com/2020/07/28/facebook-twitter-youtube-pull-false-coronavirus-video-after-it-goes-viral.html.

[45] *See* Rachel Lerman, Katie Shepherd, and Taylor Telford, *Twitter Penalizes Donald Trump Jr. for Posting Hydroxychloroquine Misinformation Amid Coronavirus Pandemic*, WASH. POST (July 28, 2020), https://www.washingtonpost.com/nation/2020/07/28/trump-coronavirus-misinformation-twitter/.

[46] *See* Shead, supra n. 45.



[47]

In addition to implementing systems to remove false and misleading medical information, Twitter is also prominently making available truthful and accurate information about COVID-19 through its Know The Facts search prompt. In early 2020, Twitter expanded its #KnowTheFacts program – which it had earlier put in place to help the public find credible information on immunization and vaccine health – to surface and highlight credible information on the virus and to ensure that when Twitter users access the platform to search for information about the virus, they are first met with credible, authoritative information from reliable sources. A further component of Twitter's #KnowTheFacts program limits auto-suggest results that may direct Twitter users to

[47] Twitter Comms (@TwitterComms), TWITTER (July 28, 2020, 10:15 AM), https://twitter.com/TwitterComms/status/1288115957005578246.

misinformation on Twitter.[48] And, similar to Facebook, Twitter has created a specific webpage dedicated to providing the latest authoritative information on the pandemic. This resource – Twitter's "COVID-19 Event Page" – provides an aggregation of credible news updates on the pandemic, curated with content from verified sources like The New York Times, Associated Press, and Reuters, as well as other public health sources such as the Center for Disease Control, which provide relevant virus-related updates.[49]

Twitter's aggressive approach to medical misinformation on its platform appears to be effective, so far. According to Twitter's reporting on the implementation of its policies regarding medical misinformation in the context of the pandemic, it has removed thousands of tweets containing misleading and potentially harmful content and has challenged over 1.5 million accounts that were targeting discussions around COVID-19 with spammy or manipulative behaviors.[50]

Google/YouTube's Response to Medical Misinformation

Google's approach to searches that seek information on COVID-19 is similarly proactive. Whereas the company has for most of its history deferred solely to its complex algorithms to serve up search results without human intervention, Google now has taken the approach of having searches related to coronavirus trigger a type of "SOS alert," and

---

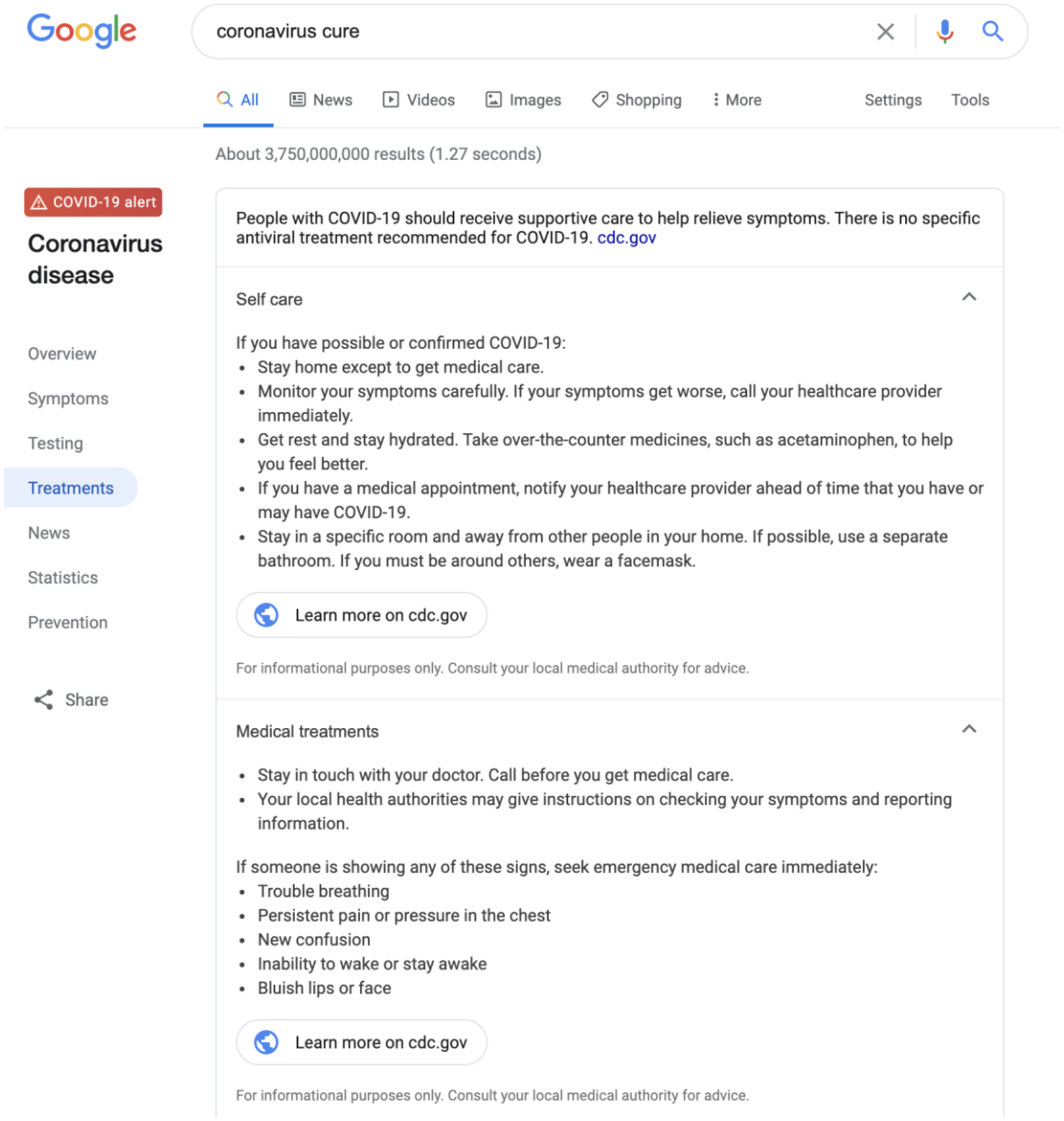
[48] *See* @Twitter, *Coronavirus: Staying Safe and Informed on Twitter*, TWITTER BLOG (Apr. 3, 2020), https://blog.twitter.com/en_us/topics/company/2020/COVID-19.html, Launch of a new dedicated #KnowTheFacts search prompt:

> As the global conversation continues around the spread of COVID-19, we want to share the work we're doing to surface the right information, to promote constructive engagement, and to highlight credible information on this emerging issue. We've seen tens of millions of Tweets on this topic in the past four weeks and that trend looks set to continue. Given the rapidly evolving nature of the issue and the growing international response, we've launched a new dedicated search prompt to ensure that when you come to the service for information about the #coronavirus, you're met with credible, authoritative information first. In addition, we're halting any auto-suggest results that are likely to direct individuals to noncredible content on Twitter. This is an expansion of our Know The Facts prompt, which we specifically put in place for the public to find clear, credible information on immunization and vaccination health. *Id.*

[49] *See* Jack Morse, *Twitter Steps up Enforcement in the Face of Coronavirus Misinformation*, MASHABLE (Mar. 18, 2020), https://mashable.com/article/twitter-cracks-down-coronavirus-misinformation/; see also COVID-19 Events Page, Twitter, available at https://twitter.com/i/events/1219057585707315201. Most tweets and updates are from verified news accounts such as the New York Times, Associated Press, and Reuters, as well as other public health sources such as the CDC and similar international entities.

[50] See text accompanying notes x – y.

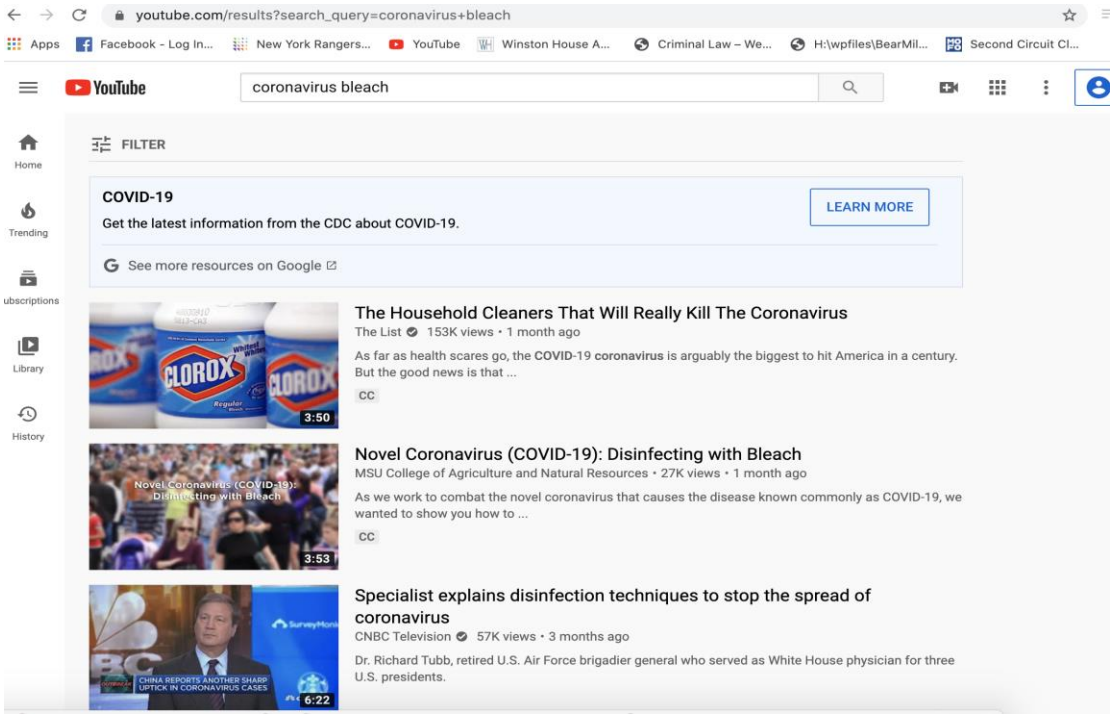
resulting in prominent displays of news from mainstream publications including National Public Radio, followed by information from the U.S. Centers for Disease Control and the World Health Organization."[51]



[Google search page in response to searching for coronavirus cure, as of July 6, 2020]

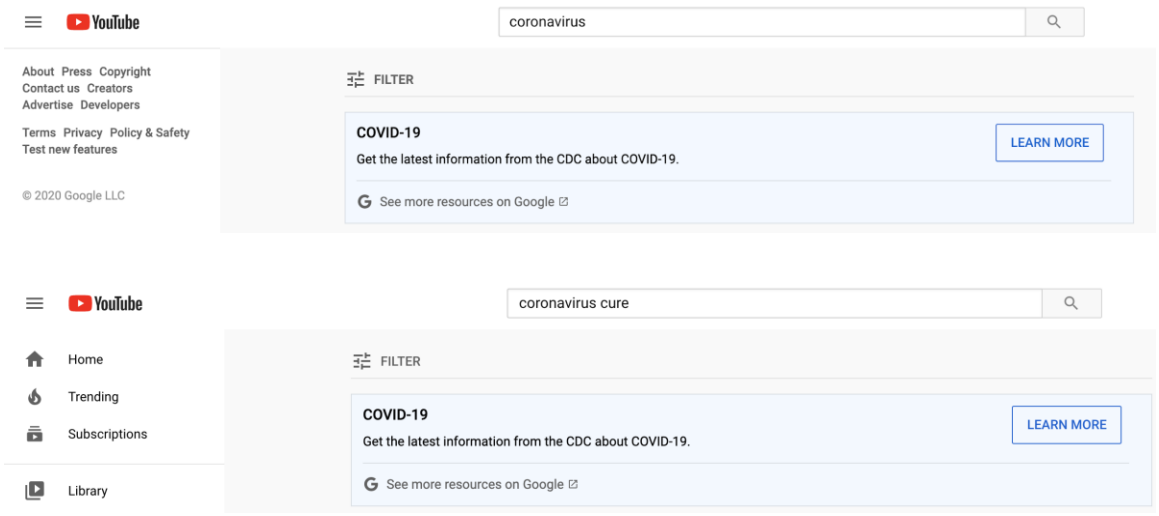
[51] Mark Bergen and Garret De Vynyck, *Google Scrubs Coronavirus Misinformation on Search, YouTube*, BLOOMBERG (Mar. 10, 2020), https://www.bloomberg.com/news/articles/2020-03-10/dr-google-scrubs-coronavirus-misinformation-on-search-youtube.

In addition, YouTube has modified its Terms of Service to prohibit any content that directly contradicts advice from the WHO.[52] In an update to its monetization policy, YouTube announced that it will prohibit videos that seek to capitalize on coronavirus-related conspiracies, and has directed users instead to videos "debunking" the conspiracies.[53]

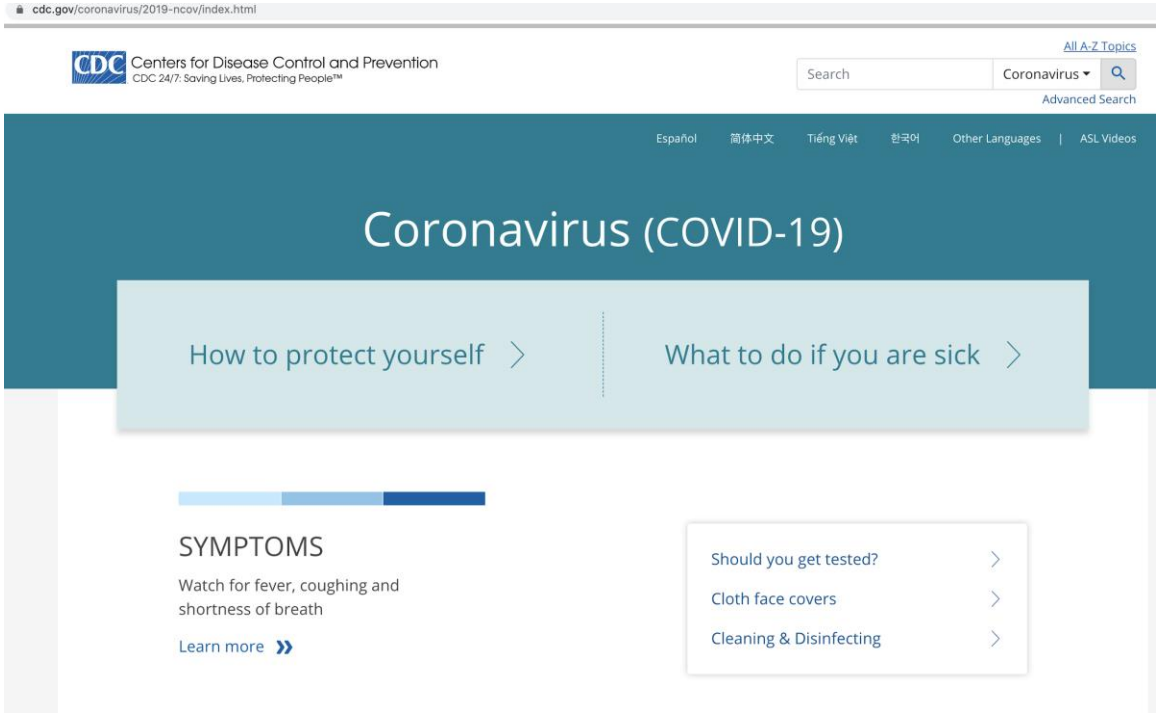


YouTube, however, like Facebook, has run into difficulties in countering medical misinformation on its platform, especially in regions where fact-checking is not as readily achievable or practical as in the United States. This is partly a result of the nature of the medium itself. YouTube videos typically involve some creative elements such that the question of whether the resulting content is true or false becomes more complex and nuanced. YouTube has recently adopted the same approach as Google search, providing a banner at the top of searches for terms such as "coronavirus" or "coronavirus cure," that provide a link to the CDC's official page (see below).

[52] *Coronavirus: YouTube Bans 'Medically Unsubstantiated' Content*, BBC NEWS (Apr. 22, 2020), https://www.bbc.com/news/technology-52388586.
[53] *Id.*



The above page provides a link to the official CDC page on COVID-19:



II. PLATFORMS' EFFORTS TO ADDRESS POLITICAL MISINFORMATION: MEASURES TO COMBAT FALSE AND MISLEADING POLITICAL SPEECH AND MICROTARGETING OF POLITICAL ADS

Introduction and Misinformation in the 2016 Election

Today's online information ecosystem continues to be a forum for political and election-related misinformation, as it was four years ago in the context of the 2016 election, Misinformation and intentionally false disinformation[54] on the Internet is particularly problematic given that the Internet is a dominant (if not *the* dominant) source of information in the political sphere, with two-thirds of Americans identifying internet sources as their leading sources of information in connection with the 2016 U.S. presidential election.[55] In addition, misinformation spread via social media is much more susceptible to being spread faster and farther than is truthful information. According to a recent study published in Science,[56] false news -- and in particular, false political news -- is spread farther, faster, deeper, and more broadly than the truth -- with the top 1% of false news cascades diffused to between 1,000 and 100,000 people (whereas the truth rarely diffused to more than 1,000 people) and with false news diffusing faster than the truth.[57] The authors of the study in Science investigated the differential diffusion of all of the verified true and false news stories distributed on Twitter from 2006 to 2017, composed of approximately 126,000 stories tweeted by approximately 3 million people more than 4.5 million times.[58] They observed that "[f]alsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects

---

[54] In this article, I will use the term "misinformation" to refer to false information regardless of whether the speaker of that information had an intent to mislead, and I use the term "disinformation" to refer to intentionally false information, where the speaker of that information had an intent to mislead. *See, e.g.*, Valerie Strauss, *Word of the Year: Misinformation. Here's Why*., WASH. POST (Dec. 10, 2018), https://www.washingtonpost.com/education/2018/12/10/word-year-misinformation-heres-why/.

[55] *See* Honest Ads Act, S. 1989, 115th Cong. § 3(10) (2017); PEW RES. CTR., ELECTION 2016: CAMPAIGNS AS A DIRECT SOURCE OF NEWS 28 (2016). https://assets.pewresearch.org/wp-content/uploads/sites/13/2016/07/PJ_2016.07.18_election-2016_FINAL.pdf.

[56] Soroush Vosoughi, Deb Roy, and Sinan Aral, *The Spread of True and False News Online*, SCIENCE, 09 Mar 2018, Vol. 359, Issue 6380, pp. 1146-1151.

[57] *Id.*

[58] The authors of the study in Science classified news as "true" or "false" using information from six independent fact-checking organizations that exhibited 95 to 98% agreement on the classifications. *Id.*

were more pronounced for false political news" than for false news concerning other subjects, such as natural disasters, science, urban legends, or financial information.[59]

False and misleading political content on social media platforms—especially on Facebook—played a significant role in influencing members of the electorate leading up to the 2016 election. More than one quarter of voting age adults visited a false news website in the final weeks of the 2016 campaign.[60] Indeed, in the months leading up to the election, the top twenty fake news stories had more "engagements" (which includes shares, reactions, and comments) on Facebook (with 8.7 million engagements) than the twenty top hard news stories (with 7.3 million engagements).[61] In the final three months of the U.S. presidential campaign, the top performing fake election news stories on Facebook generated more engagements than the top stories from major news outlets such as *The New York Times*, *The Washington Post*, *Huffington Post*, and *NBC News*,[62] and material generated by Russian operatives reached a hundred and twenty-six million American

[59] Interesting, the authors further observe that "[c]ontrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that false news spreads more than the truth because humans, not robots, are more likely to spread it." Id. But that is not to diminish the role that bots played in Russian interference in the 2016 election. Foreign interference in our 2016 presidential elections was clearly exacerbated by the use of automation in the form of bots, trolls, and fake accounts and by the use of microtargeted political advertisements to amplify disinformation, manipulate public discourse, exacerbate political and social divisions, and deceive voters on a mass scale, especially via Twitter platform, in a manner that was targeted to members of the U.S. electorate, especially in swing states. Natasha Bertrand, *Twitter Users Spreading Fake News Targeted Swing States in the Run-Up to Election Day*, BUS. INSIDER (Sept. 28, 2017), https://www.businessinsider.com/fake-news-and-propaganda-targeted-swing-states-before-election-2017-9. Russian bots, for example, were responsible for 30 to 40% of election related tweets directed to the swing states of Pennsylvania, Michigan, and Wisconsin, as well as to the battleground states of Ohio, Missouri, Florida, North Carolina, and Colorado, during the 2016 presidential elections. Trump won the Electoral College because some eighty thousand votes went his way in Wisconsin, Michigan, and Pennsylvania. *See e.g.*, Kathleen Hall Jamieson, CYBERWAR: HOW RUSSIAN HACKERS AND TROLLS HELPED ELECT A PRESIDENT 67 (2018).

[60] *See* Danielle Kurtzleben, *Did Fake News on Facebook Help Elect Trump? Here's What We Know*, NPR (Apr. 11, 2018), https://www.npr.org/2018/04/11/601323233/6-facts-we-know-about-fake-news-in-the-2016-election.

[61] *Id.*

[62] *See* Craig Silverman, *This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook*, BUZZFEED NEWS (Nov. 16, 2016, 5:15 PM), https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook#.emA15rzd0.

Facebook users.[63] Twitter was a primary target of Russia's false news and misinformation offensives during the 2016 elections, as a St. Petersburg-based troll factory known as the Internet Research Agency used Twitter as a vehicle to create fake accounts to exacerbate political and social tensions in the United States and to mislead U.S. voters.[64] The Internet Research Agency controlled more than 3,000 Twitter accounts during the 2016 U.S. elections, and another 50,000 automated accounts were connected to the Russian government.[65]

Political Speech and Political Advertising on Social Media Platforms Today

Political advertising on social media platforms is big business–and, as of this writing–big and still largely *unregulated* business in the United States. The total amount spent on digital political advertising in the U.S. is expected to reach $2.9 billion in 2020 (an increase of over 100% from 2016), with Google and Facebook capturing the vast majority of digital political advertising.[66] Because of the power of such ads in influencing our democratic processes, the method of dissemination -- including the use of microtargeting to target specific, narrow segments of the electorate -- and the substance of such ads -- including the issue of false and misleading information in such ads -- have been subject to intense scrutiny, as discussed below.

Because political advertising has the potential to affect our democratic processes in powerful ways, it has traditionally been subject to a host of government regulations, including transparency regulations, disclosure and public file regulations, and prohibitions on foreign participation, as I discuss below. Yet, as of this writing, such government regulations only apply to traditional media and generally do not apply to online mediums. This is despite the fact that Facebook's user base of 204 million American users, for example, is ten times larger than the subscriber base of the largest cable and satellite providers and despite the fact that over one billion dollars was spent on online advertising in 2016 – and many billions more will be spent on digital advertising in the context of the 2020 elections. Although the federal Honest Ads Act was introduced as an attempt to

---

[63] *See* Jane Mayer, *How Russia Helped Swing the Election for Trump*, THE NEW YORKER (Oct. 1, 2018), https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump/.

[64] *See* Craig Timberg & Elizabeth Dwoskin, *Twitter is Sweeping Out Fake Accounts Like Never Before, Putting User Growth at Risk*, WASH. POST (July 6, 2018), https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk.

[65] *Id.*

[66] Emily Glazer, *Facebook Weighs Steps to Curb Narrowly Targeted Ads*, WALL STREET J. (Nov. 21, 2019), https://www.wsj.com/articles/facebook-discussing-potential-changes-to-political-ad-policy-11574352887.

remedy this regulatory gap and to extend this host of regulations on political advertising to social media platforms, as of this writing, the Act has not been enacted into law. In addition, several state legislative attempts to regulate political advertising online have been subject to successful First Amendment challenges, as discussed below. Accordingly, the social media giants -- Facebook, Twitter, and Google -- have been left to their own devices to determine whether and how to regulate political advertising -- and how to regulate the microtargeting of political ads -- on their platforms.

Below in Part A, I briefly survey the current state of the regulation of political advertising applicable to traditional mediums of expression. In Part B, I examine the proposed federal Honest Ads Act, as well as similar state versions of such legislation. In Part C, I turn to the special problems of microtargeting of political advertising on social media. I then analyze the steps that the major social media platforms have taken -- and have declined to take -- to address the problems caused by false political speech on their forums and by the microtargeting of political ads in particular.

A. Federal Regulation of Political Advertising Applicable to Traditional Media

Various federal statutes, Federal Election Commission rules, and Federal Communications Commission rules currently impose transparency requirements on political advertisements disseminated by broadcast, cable, and satellite providers, and also impose requirements on these providers prohibiting foreign participation in U.S. elections. First, FEC regulations impose transparency requirements on political advertisements disseminated via non social media: Any public communication made by a political committee — including communications that do not expressly advocate the election or defeat of a clearly identified federal candidate or solicit a contribution — must display a disclaimer. Disclaimers must also appear on political committees' internet websites, and in certain email communications. All public communications that expressly advocate the election or defeat of a clearly identified candidate, electioneering communications, and all public communications that solicit any contribution require a disclaimer, regardless of who has paid for them. Public communications include electioneering communications and any other form of general public political advertisement, including communications made using the following media: broadcast, cable or satellite; newspaper or magazine; outdoor advertising facility; mass mailing (more than 500 substantially similar mailings within 30 days); phone bank (more than 500 substantially similar calls within 30 days); and communications placed for a fee on another person's website.[67]

[67] *See* FEDERAL ELECTION COMMISSION, ADVERTISING AND DISCLAIMERS, https://www.fec.gov/help-candidates-and-committees/making-disbursements/advertising/ (last accessed Jul. 19, 2020).

Second, what is known as the "Foreign Participation Ban" prohibits foreign nationals from attempting to influence elections through donations, expenditures, or other things of value.[68] Existing regulations applicable to broadcast, cable, and satellite platforms include a broad prohibition on the involvement of foreign nationals with elections in the United States, under which foreign nationals are prohibited from making any contribution, donation, or expenditure in connection with any federal, state, or local election; making any contribution or donation to any committee or organization of any national, state, or local political party; or making any disbursement for an electioneering communication.

Third, the Bipartisan Campaign Reform Act (BCRA), which applies to traditional media, imposes disclosure and public file requirements, aimed at informing the electorate about the source of election related advertisements, and these provisions have been upheld by the Supreme Court.[69] BCRA § 311 requires that televised "electioneering communications" funded by anyone other than a candidate must include a statement clearly indicating who was responsible for the ad (in the form of "[_____] is responsible for the content of this advertising"), along with the name and address (or web address) of the person who funded the ad. In addition, BCRA requires that anyone who spent more than $10,000 on electioneering communications within a calendar year file a detailed statement with the FEC, providing his or her name, amount of expenditure, and the name of the election to which the communication was directed, among other details. In upholding BCRA's disclosure and public file requirements against a First Amendment challenge by Citizens United, the Supreme Court explained that these provisions "provid[e] the electorate with information" and "insure that voters are fully informed about the person or group who is speaking . . . so that people will be able to evaluate the arguments to which they are being subjected." The Court concluded that these requirements were a less restrictive alternative compared to other, more extensive regulations of political speech, that "the public has an interest in knowing who is speaking about a candidate shortly before an election," and that this "informational interest alone is sufficient to justify application of [the Act] to these ads."

None of the above regulations currently apply to political advertising on social media.

B. The Proposed Honest Ads Act

[68] 52 U.S.C. § 30121(a)(1)(A).
[69] See text accompanying notes x – y.

As discussed above, various federal statutes and Federal Election Commission rules[70] currently impose transparency requirements on political advertisements disseminated by broadcast, cable, and satellite providers, and also impose requirements on these providers prohibiting foreign participation in U.S. elections; yet, social media platforms like Google, Facebook, and Twitter are currently not subject to analogous regulations. Such social media platforms have been largely immune from Federal Election Commission and related regulations that have long been applicable to other sources of news and information in our political information ecosystem that mandate transparency and accountability requirements.[71]

The Honest Ads Act, introduced in October 2017 by Senators Mark Warner (D-Virginia), Amy Klobuchar (D-Minnesota) and the late John McCain (R-Arizona), seeks to remedy this regulatory disparity. The Act attempts to address some of the problems created by foreign interference in U.S. elections in the online arena by imposing transparency regulations on online political advertisements and by requiring that online platforms enforce the longstanding ban on foreign participation in United States elections.[72] Although, as discussed infra, social media platforms like Twitter, Google, and Facebook are undertaking substantial measures themselves to address such problems, such measures may be revisited or revoked by the platforms at any time; therefore, government regulation in the form of the Honest Ads Act is still an important tool for addressing these problems, and indeed one that is welcomed by the platforms.[73]

The Honest Ads Act seeks to address problems in the online marketplace of ideas by extending three sets of requirements that have long been imposed on communications platforms to online platforms: (1) the expansion of disclosure requirements applicable to political advertisements; (2) the expansion of public file requirements; and (3) the expansion of the obligation to undertake reasonable efforts to limit foreign interference in U.S. elections.

---

[70] *See* Fredreka Schouten, *Federal Regulators Approve Narrow Facebook Ad Disclosure*, USA TODAY (Dec. 15, 2017, 11:42 AM), https://www.usatoday.com/story/news/politics/2017/12/14/federal-regulators-weigh-whether-unmask-online-political-ad-buyers/951425001.

[71] See text accompanying notes x – y.

[72] *See* S. 1989, 115th Cong. §§ 3–4 (2018).

[73] Both Facebook and Twitter have come out in support of the Honest Ads Act. *See* Aimee Picchi, *Facebook: What Is the Honest Ads Act?*, CBS NEWS (Apr. 11, 2018), https://www.cbsnews.com/news/facebook-hearings-what-is-the-honest-ads-act; Twitter Public Policy (@Policy), TWITTER (Apr. 10, 2018, 8:54 AM), https://twitter.com/Policy/status/983734917015199744.

First, the Honest Ads Act extends the disclosure obligations governing political advertisements that print, broadcast, and cable advertisements must meet to online platforms.[74] The Federal Election Campaign Act of 1971 requires that certain political ads in print, broadcast, and cable disclose who has paid for the advertisement.[75] This requirement currently does not extend to paid Internet or digital advertisements. Under the Honest Ads Act, the Federal Election Campaign Act's definition of "electioneering communication" would be expanded to include *online* paid political advertisements.[76] Existing federal law also imposes disclosure requirements on "public communications" that expressly advocate for a candidate's election or defeat, are paid for or authorized by a candidate, solicit a political contribution, or are by a political committee. The Honest Ads Act would update the definition of "public communication" as well, to ensure that disclosure obligations applicable to these types of advertisements extend to the online environment.[77]

Second, under the Honest Ads Act, large digital platforms (those with more than fifty million unique monthly visitors) would be required to maintain publicly available records of political advertisements by a purchaser whose aggregate requests to purchase political advertisements on that platform exceed $500 within the past year.[78] Such records must include a digital copy of the political advertisement, as well as a description of the target audience, the ad rate, the name of the candidate or office that the ad was supporting, and the contact information of the purchaser of the ad.[79] Like the FCC's broadcast file rules, the Act would apply to ads made by, for, or about political candidates, about elections, and about "national legislative issues of public importance."[80]

Third, the Honest Ads Act would mandate that all advertising platforms—including online platforms — make reasonable efforts to comply with the "foreign participation ban."[81] This longstanding ban prohibits foreign nationals from attempting to influence elections through donations, expenditures, or other things of value.[82] Existing regulations applicable to broadcast, cable, and satellite platforms include a broad prohibition on the involvement of foreign nationals with elections in the United States, under which foreign

---

[74] S. 1989, 115th Cong. § 2 (2018) ("The purpose of this Act is to enhance the integrity of American democracy and national security by improving disclosure requirements for online political advertisements . . . .").
[75] 52 U.S.C. § 30120.
[76] S. 1989, 115th Cong. § 6 (2018).
[77] *Id.* § 5.
[78] *Id.* § 8(a)(1)(A).
[79] *Id.* § 8(a)(2).
[80] *Id.* § 8(a)(4).
[81] *See id.* § 9.
[82] 52 U.S.C. § 30121(a)(1)(A).

nationals are prohibited from making any contribution, donation, or expenditure in connection with any federal, state, or local election; making any contribution or donation to any committee or organization of any national, state, or local political party; or making any disbursement for an electioneering communication. The Honest Ads Act would extend these prohibitions to the online environment as well.

The Honest Ads Act was reintroduced to the 116th Congress on May 7, 2019 in the Senate and May 8, 2019 in the House.[83] For several months, it sat before the respective Committees on Rules and Administration[84] and House Administration.[85] In October, the Honest Ads Act was incorporated[86] into the Stopping Harmful Interference in Elections for a Lasting Democracy Act,[87] passed in the House,[88] and it presently sits again before the Rules and Administration Committee.[89]

State Measures to Regulate Political Ads on Social Media

In the absence of federal legislation, several states have attempted to regulate political advertising online, but these attempts have been subject to legal challenges (which have been successful, as of this writing). Like the Honest Ads Act, Maryland's Online Electioneering Transparency and Accountability Act[90] establishes disclosure requirements

---

[83] Zach Montellaro, *The Honest Ads Act Returns*, POLITICO (May 9, 2019), https://www.politico.com/newsletters/morning-score/2019/05/09/the-honest-ads-act-returns-615586.

[84] *S.1356- Honest Ads Act, Actions*, CONGRESS.GOV, https://www.congress.gov/bill/116th-congress/senate-bill/1356/all-actions.

[85] *H.R.2592- Honest Ads Act, Actions*, CONGRESS.GOV, https://www.congress.gov/bill/116th-congress/house-bill/2592/all-actions.

[86] Comm. on H. Admin., *Markup of H.R. 4617, Stopping Harmful Interference in Elections for a Lasting Democracy Act*, 116th Cong. (2019) https://cha.house.gov/committee-activity/markups/markup-hr-4617-stopping-harmful-interference-elections-lasting-democracy.

[87] Not to be confused with the Showing How Isolationism Effects Long-Term Development Act: https://www.congress.gov/bill/116th-congress/house-bill/7005?q=%7B%22search%22%3A%5B%22SHIELD+Act%22%5D%7D&s=2&r=1.

[88] *H.R.4617- SHIELD Act, Actions*, CONGRESS.GOV, https://www.congress.gov/bill/116th-congress/house-bill/4617/actions?q=%7B%22search%22%3A%5B%22H.R.+4617%22%5D%7D&r=1&s=1.

[89] *S.2669- SHIELD Act, Actions*, https://www.congress.gov/bill/116th-congress/senate-bill/2669/actions?q=%7B%22search%22%3A%5B%22Stopping+Harmful+Interference+in+Elections+for+a+Lasting+Democracy%22%5D%7D&r=1&s=3.

[90] Md. Code Ann. Elec. Law § 13-405 (2018).

for "qualifying paid digital communications disseminated through the online platform."[91] Under the Maryland Act, upon notice by the purchaser, the publisher must identify the amount paid, and as applicable, the purchaser's name and controlling interest.[92] A number of media outlets, including the Washington Post, brought a challenge to this law in the U.S. District Court for the District of Maryland, alleging that the Maryland Act infringed their First Amendment rights of free speech and free press.[93] The court, while crediting Maryland's interest in legislating to prevent misinformation online such as was prevalent in the 2016 election,[94] nonetheless temporarily enjoined the statute's enforcement on the grounds that the Act was subject to strict scrutiny and therefore likely unconstitutional. Although the court held that the state's interest "in preventing foreign governments and their nationals from interfering in their elections" was "compelling,"[95] it nonetheless found that the Act was both unconstitutionally over- and underinclusive: "[The Act] regulates substantially more speech than it needs to, while, at the same time, neglecting to regulate the primary tools that foreign operatives exploited to pernicious effect in the 2016 election."[96]

Washington State has also sought to regulate social media companies' political advertisements. In April 2020, Washington State sued Facebook, alleging violations of the state's campaign finance laws. This marks the second action taken by Washington State against Facebook for similar violations in the past two years: in December 2018, Facebook paid a $200,000 penalty for failure to maintain legally required information under a Washington State law governing political ads.[97] Washington State requires commercial advertisers in all mediums (including social media) to maintain records identifying the sponsors of political ads in both state and local elections, and it further requires that these

---

[91] Md. Code Ann. Elec. Law § 13-405(b)(1) (2018).
[92] *See* Md. Code Ann. Elec. Law § 13-405(b)(6) (2018).
[93] *See Washington Post v. McManus*, 335 F.Supp.3d 272 (D. Md. 2019) *aff'd*, No. 19-1132, 2019 WL 6647336 (4th Cir. Dec. 6, 2019).
[94] *Washington Post*, 335 F.Supp.3d at 282 ("It was in February 2018, amid the uproar over Russian meddling in American political affairs, that Maryland legislators – noting an absence of any federal statutory or regulatory activity aimed at thwarting foreign interference in its elections – resolved to act and introduced the bill at issue here.").
[95] *Id.*, at 300 ("To characterize the State of Maryland's interest in addressing this threat as anything less than compelling would be a profound error; on the contrary, the Maryland legislature should be commended, not criticized, for attempting to address this threat to the fairness of its electoral process.").
[96] *Id.*
[97] Washington State Office of the Attorney General, *AG Ferguson Sues Facebook for Repeatedly Violating Washington Campaign Finance Law* (Apr. 14, 2020), https://www.atg.wa.gov/news/news-releases/ag-ferguson-sues-facebook-repeatedly-violating-washington-campaign-finance-law.

records be made readily available for public inspection.[98] The present case alleges that Facebook hosted hundreds of political ads in violation of state law since the time the company announced it would stop accepting Washington state political ads.[99] Although Facebook does store some information about its ads in an online, publicly accessible database (Facebook's Ads Library, discussed infra[100]), according to the Washington State action, Facebook failed to disclose additional information required under applicable state law, such as the address of the person who sponsored the advertising, the name of that person, the precise cost and dates of payment, and the method of payment.[101]

In sum, unlike political advertising on traditional media, political advertising on social media is currently largely unregulated.

C. The Special Problems Caused by Microtargeting of Political Ads

Microtargeting of ads on social media platforms is a practice that generally allows advertisers to limit their messaging to narrow slices or subsets of individuals, by exploiting the vast trove of social data about individuals' online behavior and preferences that has been collected by social media platforms.[102] Microtargeting of ads in general stands in sharp contrast to the broadcasting of ads in mediums like major metropolitan newspapers, radio and television, through which advertisers provide content to a large and mostly homogenous audience (e.g., to all readers of The Washington Post). In contrast, microtargeting delivers ad content to very specific subgroups (e.g., readers who shop at Whole Foods who are between the ages of 25 and 49, and who have watched a certain video on YouTube) or even to specific, listed individuals (by using tools such as Facebook's Custom Audiences).[103] The practice of microtargeting employs and capitalizes on the social data -- such as an individual's likes, dislikes, interests, preferences, behaviors and viewing and purchasing habits -- collected by social media platforms about their users and made available to advertisers to enable advertisers to segment individuals into small groups so as to more accurately and narrowly target advertising to them. Facebook, for

[98] RCW 42.17A, https://app.leg.wa.gov/RCW/default.aspx?cite=42.17A.060. "Timely," here means within 24 hours of the ad's publication. See supra, n. 56.
[99] *AG Ferguson Sues Facebook for Repeatedly Violating Washington Campaign Finance Law*, *supra* note 53.
[100] See text accompanying notes x – y.
[101] *Id.*
[102] See text accompanying notes x – y.
[103] *See, e.g.*, Dipayan Ghosh, *What is Microtargeting and What Is It Doing in Our Politics?*, MOZILLA: INTERNET CITIZEN (Oct. 4, 2018), https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/.

example, reportedly tracks a list of over 1,100 attributes on each of its users spanning users' demographic, behavioral, and interest categories.[104]

The practice of microtargeting enables advertisers to capitalize on the comprehensive social data about individuals collected by social media platforms to design and disseminate content that advertisers predict will be the most effective and relevant with respect to the targeted segment of individuals. For example, an advertiser might limit the scope of an ad's distribution to single men between 25 and 35 who live in apartments and "like" the Washington Nationals.[105] While businesses derive certain benefits from the microtargeting of ads in nonpolitical contexts, microtargeting of ads in the political context can pose serious problems for the democratic process and for the marketplace of ideas model that underlies our First Amendment model of freedom of speech.[106] Unlike political advertising on mass media like broadcast television or radio – in which large national or regional audiences are exposed to the same political advertisement – by employing narrowcasted microtargeted ads on social media, a political advertiser can craft a specific ad to a much narrower intended audience and *only* that specific audience, thereby essentially preventing others from accessing and scrutinizing the content of the ad.

As described by Facebook's former Chief Security Officer Alex Stamos, the chief benefit of political micro-targeting is that allows political advertisers to deploy "messages that are extremely finely targeted to a very small number of people."[107] By microtargeting political ads, a campaign can make different – and even contradictory – appeals to voters in Michigan and voters in New York or Atlanta. As such, extensively deployed microtargeting of political ads – which is by definition immune from the check of broad public scrutiny – increases the possibility that a politician might lie with impunity. As Stamos explains, "if you allow people to show an ad to just 100 folks and then you run tens

---

[104] *See* Till Speicher et al., *Potential for Discrimination in Online Targeted Advertising*, 81 PROC. MACHINE LEARNING RES. 1, 5, 7 (2018) ("For each user in the US, Facebook tracks a list of over 1,100 binary attributes spanning demographic, behavioral and interest categories that we refer to as curated attributes. Additionally, Facebook tracks users' interests in entities such as websites, apps, and services as well as topics ranging from food preferences (e.g., pizza) to niche interests (e.g., space exploration).").

[105] Ellen L. Weintraub, *Don't Abolish Political Ads on Social Media. Stop Microtargeting.*, WASH. POST (Nov. 1, 2019), https://www.washingtonpost.com/opinions/2019/11/01/dont-abolish-political-ads-social-media-stop-microtargeting/.

[106] *See, e.g.*, Dawn C. Nunziato, *The Marketplace of Ideas Online*, 94 NOTRE DAME L. REV. 1519 (2019).

[107] Peter Kafka, *Facebook's Political Ad Problem, Explained By an Expert*, VOX (Dec. 10, 2019), https://www.vox.com/recode/2019/12/10/20996869/facebook-political-ads-targeting-alex-stamos-interview-open-sourced.

of thousands of ads, it makes it extremely difficult for your political opponent and the print media to call you out."[108] Such microtargeting of political ads also exacerbates problems of balkanization, in which the messages that individuals are receiving are so disparate as to dissolve the larger communities of interest that otherwise ostensibly bind the country as a nation.[109]

The Internet Research Agency – the notorious agent of Russian disinformation during the 2016 election cycle – was able to spend pennies on the dollar (or ruble) compared to U.S. presidential campaigns by deploying powerful microtargeted political ads on social media. With its use of microtargeted political ads, the Agency was able to powerfully leverage its influence to interfere with U.S. elections. While the Trump and

---

[108] *Id.*

[109] Craig Timberg, *Critics Say Facebook's Powerful Ad Tools May Imperil Democracy. But Politicians Love Them.*, WASH. POST (Dec. 9, 2019), https://www.washingtonpost.com/technology/2019/12/09/critics-say-facebooks-powerful-ad-tools-may-imperil-democracy-politicians-love-them/. *See also* Isaac Stanley-Becker, *Facebook's Ad Tools Subsidize Partisanship, Research Shows. And Campaigns May Not Even Know It.*, WASH. POST (Dec. 10, 2019), https://www.washingtonpost.com/technology/2019/12/10/facebooks-ad-delivery-system-drives-partisanship-even-if-campaigns-dont-want-it-new-research-shows/ (Eli Pariser, the Internet activist who coined the term "filter bubble" explaining that serving "users with information that aligns with their existing worldview ... 'fragments political discourse.'"). A recent study in fact showed that the very mechanism of Facebook's ad delivery increases partisanship. Muhammad Ali, et al, *Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging*, ARXIV 13 (Dec. 17, 2019), https://arxiv.org/pdf/1912.04255.pdf (see "Discussion"). The authors isolated the role that Facebook's perception of an ad's content plays in determining the audience that receives it by creating a generic, non-partisan ad with a call to register to vote that linked to a generic domain. They then "configured [their] web server to deliver a different response for requests for these pages based on the IP address of the requestor." If the requestor were identified as Facebook, it would be served "a copy of the HTML from the official Trump campaign website, the official Sanders campaign website, or a generic voting information website." All other requestors were simply "redirected to the generic voting information website." *Id.* at 7 (See 4.3 "Isolating role of content"). The ads therefore appeared identical to users, but misled Facebook's algorithm to associate them with different political content. The authors found that even after selecting a target audience, Facebook will prefer delivering the ad to those who it predicts will identify with its message. "Counterintuitively, advertisers who target broad audiences may end up ceding platforms even more influence over which users ultimately see which ads." *Id.* at 1. Beyond the "ad creation and targeting phase, where the advertiser selects their desired audience," the actual delivery of the ad further discriminates among possible recipients. The selection is "rooted in the desire to show relevant ads to users" and, the study notes, "can lead to dramatic skew in delivery along gender and racial lines, even when the advertisers aims to reach gender and race-balanced audiences." *Id.* at 1-2.

Clinton campaigns spent a combined $81 million on pre-election Facebook ads,[110] for example, the Internet Research Agency was able to sew tremendous discord by spending only $46,000.[111] This miniscule amount of spending took advantage of the powerful ability to target custom audiences by inferring interests from social media users' social data. The Internet Research Agency used the microtargeting tools developed by leading technology companies -- including Facebook's advertising customization tools -- to target specific audiences that they believed would be particularly susceptible to false and misleading election-related information. In particular, Russian operatives used Facebook's Custom Audiences[112] tool to display specific ads and messages to voters who had visited the operatives' fake social media sites -- and used this microtargeting technique to sew division among voters – specifically to suppress black voter turnout.[113] Facebook's Custom Audiences tool allows advertisers (including, in this case, the Russian operatives) to input into Facebook's system a specific list of users they wish to target. While such technological tools have long been used by corporate America to deliver advertising to target audiences, Facebook and other social media platforms were taken by surprise by the use of such tools for purposes of interference in the U.S. elections. As The Washington Post explains:

> [Russian operatives' microtargeted political ads] focused on such hot-button issues as illegal immigration, African American political activism and the rising prominence of Muslims in the United States. The Russian operatives then used a Facebook "retargeting" tool, called Custom Audiences, to send specific ads and messages to voters who had visited those sites....One such ad featured photographs of an armed black woman "dry firing" a rifle — pulling the trigger of the weapon without a bullet in the chamber. Investigators believe the advertisement may have been

---

[110] Josh Constine, *Trump and Clinton Spent $81M on US Election Facebook Ads, Russian Agency $46k*, TECHCRUNCH (Nov. 1, 2017), https://techcrunch.com/2017/11/01/russian-facebook-ad-spend/.

[111] Figure given in testimony: *Hearing on Social Media Influence in the 2016 United States Elections, Before the S. Select Comm. on Intelligence*, 115th Cong. (2017) https://www.intelligence.senate.gov/hearings/open-hearing-social-media-influence-2016-us-elections#; *See also* Nicholas Thompson, *A Facebook Executive Apologizes to His Company—and to Robert Mueller*, WIRED (Feb. 19, 2018), https://www.wired.com/story/facebook-executive-rob-goldman-apologizes-to-company-and-robert-mueller/; Kevin Roose, *On Russia, Facebook Sends a Message It Wishes It Hadn't*, N.Y. TIMES (Feb. 19, 2018), https://www.nytimes.com/2018/02/19/technology/russia-facebook-trump.html for Facebook VP of ads Rob Goldman's ham-fisted reaction to the difference.

[112] *About Website Custom Audiences*, FACEBOOK FOR BUS., https://www.facebook.com/business/help/610516375684216 (Last accessed Jul. 19, 2020).

[113] Spencer Overton, *State Power to Regulate Social Media Companies to Prevent Voter Suppression*, 53 U.C. DAVIS L. REV. 1793 (2020).

> designed to encourage African American militancy and, at the same time, to stoke fears within white communities...[114]

Russian operatives used other Facebook tools in addition to Custom Audiences to target groups by demographics, geography, gender, and interests. As Clinton Watts, a fellow at the Foreign Policy Research Institute, explains, "This means that any American who knowingly or unknowingly clicked on a Russian news site may have been targeted through Facebook's advertising systems to become an agent of influence — a potentially sympathetic American who could spread Russian propaganda with other Americans."[115] Accordingly, "every successful click [provided the Russian operatives with] more data that they can use to retarget [thereby speeding up] the influence dramatically."[116] Targeted Facebook users were then shown ads featuring divisive topics that the Russians wanted to promote in their Facebook news feeds, which displayed the ads alongside messages from friends and family members.

Professor Spencer Overton explains in greater detail how the Russian Internet Research Agency used Facebook's complex ad targeted tools to microtarget political ads to African Americans in order to suppress the black vote in the 2016 election.[117] Overton explains that African American audiences accounted for over 38% of U.S.-focused ads purchased by the Internet Research Agency, which created social media accounts that falsely claimed they were African American-operated and urged African Americans to "boycott the election." Overton writes:

> Facebook's "Ad Manager" allows an advertiser to select from a series of 52,000 targeting attributes, including demographics/ethnic affinity (e.g., African American), issue interests (e.g., "Malcolm X" or the "Civil Rights Movement"), and Facebook engagement (e.g., liked a particular post). About 73% of the [Internet Research Agency's] ads used interest-based targeting, and most of the interest-based targeting focused on African American communities and interests like Martin Luther King, Jr., Nelson Mandela, Malcolm X and Muhammad Ali. Facebook develops these profiles by collecting vast amounts of data on its two billion users, including posts, likes, clicks, profile information, zip codes (including user preferences such as "likes" and comments), and by utilizing predictive modeling techniques to make inferences. This microtargeting is also enhanced by real-time re-targeting algorithms, a constant loop between users' voluntary choices (e.g.,

---

[114] Adam Entous, Craig Timberg, and Elizabeth Dwoskin, *Russian Facebook Ads Show Black Woman Firing a Rifle, Amid Efforts to Stoke Racial Strife*, WASH. POST (Oct. 2, 2017), https://www.washingtonpost.com/business/technology/russian-facebook-ads-showed-a-black-woman-firing-a-rifle-amid-efforts-to-stoke-racial-strife/2017/10/02/e4e78312-a785-11e7-b3aa-c0e2e1d41e38_story.html.

[115] *Id.*

[116] *Id.*

[117] Spencer Overton, *State Power to Regulate Social Media Companies to Prevent Voter Suppression*, 53 U.C. Davis L. Rev. 1793 (2020).

> liking) and the machine's feedback on their choices. Another targeting tool — Facebook's Lookalike Audience — allows advertisers to ask Facebook to create target audiences that are demographically similar to (i.e., "look like") another clearly defined audience — and by doing so Facebook "clones" audiences …[In addition,] Facebook's Custom Audience function requires that an advertiser give Facebook personally identifiable information (e.g., voter records, email addresses) of the precise people the advertiser wants to target, and Facebook uses that data to identify corresponding social media accounts.[118]

In short, using Facebook's powerful microtargeting tools, Russian operatives were able to target African-American members of our electorate, sew division, and – among other problems – suppress the black vote.

The Internet Research Agency was not alone in its masterful deployment of microtargeted political ads in the 2016 presidential election. The Trump Campaign, for example, also targeted black Americans in specific neighborhoods in an effort to decrease voter participation.[119] The benefit, as then-Trump digital media director Brad Parscale described it, was that "only the people we want to see it, see it."[120] Parscale claimed that the use of microtargeted political ads on Facebook and Twitter enabled the Trump Campaign to be one hundred to two hundred times more effective in targeting members of the electorate than the Hillary for President Campaign.[121] Whether or not Parscale's particular claim is true, research shows that political microtargeting indeed "had a significant effect in persuading undecided voters to support Trump and in persuading Republic supporters to turn out on polling day."[122] Specifically, researchers found that political advertising on Facebook made voters five to ten percent more likely to vote, and that "targeted Facebook campaigning increased the probability that a previously non-

---

[118] *Id.* (internal quotation marks omitted).
[119] Joshua Green and Sasha Issenberg, *Inside the Trump Bunker, With Days to Go*, BLOOMBERG (Oct. 27, 2016), https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go.
[120] *Id.* https://www.bloomberg.com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go
[121] Brad Parscale (@parscale), TWITTER (Feb. 24, 2018, 4:46 PM), https://twitter.com/parscale/status/967516077956755457.
[122] University of Warwick, *Targeted Facebook Ads Shown to be Highly Effective in the 2016 Presidential Election*, PHYS.ORG (Oct. 25, 2018), https://phys.org/news/2018-10-facebook-ads-shown-highly-effective.html.

aligned voter would vote for Donald Trump by at least five percent" if they were a regular Facebook user.[123]

Problems Caused by Microtargeting Political Ads

The microtargeting of political ads, compared to the dissemination of political ads via traditional media outlets, is problematic for a number of reasons from a free speech perspective – and this is not even considering the problems caused by the weaponization of microtargeting by Russian operatives interfering in our elections, sewing division, and suppressing the black vote. First, political ads disseminated via traditional media are subject to a host of federal regulations requiring transparency, disclosure, limitations on foreign interference, etc., as discussed above,[124] whereas ads disseminated via social media are not. Second, political ads disseminated via traditional media are subject to broad exposure and broad public scrutiny – which are necessary for the truth-facilitating features of the marketplace of ideas mechanisms to function. Microtargeted ads, on the other hand – which are essentially the "online equivalent of whispering millions of different messages into zillions of different ears for maximum effect and with minimum scrutiny"[125] – are not similarly subject to broad exposure or broad public scrutiny. Third, and relatedly, microtargeted ads on social media are more likely to be susceptible to the spread of misinformation. As politics and technology expert Dipayan Ghosh explains: "[Microtargeting of political ads facilitates] organic shares and reshares of content pushed by unpaid users who appreciate what they see ... and wish to spread it around their networks. This results in free content consumption for the political campaign….[and this] viral spread of unpaid or organic content ... further encourages the success of misinformation campaigns."[126]

In short, the microtargeting of political ads disseminated via social media is especially pernicious because it is not subject to regulatory scrutiny, not subject to meaningful widespread public scrutiny, and because – as discussed above – false claims in

[123] Federica Liberini, et al, *Politics in the Facebook Era: Evidence from the 2016 Presidential Election*, 5 (Centre for Competitive Advantage in the Global Economy, Working Paper No. 389, Oct. 2018), https://warwick.ac.uk/fac/soc/economics/research/centres/cage/manage/publications/389-2018_redoano.pdf.

[124] See text accompanying notes x – y.

[125] Kara Swisher, *Google Changed Its Political Ad Policy. Will Facebook Be Next?*, N.Y. TIMES (Nov. 22, 2019), https://www.nytimes.com/2019/11/22/opinion/google-political-ads.html.

[126] Dipayan Ghosh, *What is Microtargeting and What Is It Doing in Our Politics?*, MOZILLA: INTERNET CITIZEN (Oct. 4, 2018), https://blog.mozilla.org/internetcitizen/2018/10/04/microtargeting-dipayan-ghosh/.

such political ads are likely to be spread farther, faster, deeper, and more broadly than true claims in political ads.[127]

Measures to Address False Political Advertising and Microtargeting by Social Media Platforms

As of this writing, despite a heightened awareness of the problems caused by microtargeted political advertising and by false political ads, such problems have yet to be effectively addressed via regulation or legislation (at least in the United States). Instead, political advertising on social media– and the regulation of false political speech and microtargeting in particular – is subject to an ad hoc patchwork of voluntary, piecemeal measures recently adopted by the social media platforms themselves. Some of the social media platforms – notably Twitter – are adopting rigorous measures to combating such problems, while others – notably Facebook – have adopted more of a hand-off approach, at least with respect to political ads that constitute "direct speech by politicians."[128] Below I examine the measures undertaken by the social media platforms to address problems caused by the microtargeting of political ads and by false and misleading political ads.

Twitter's Regulation of Political Ads

Of the three major social media platforms, Twitter has taken the most aggressive stance with respect to false and misleading political ads – by banning political ads altogether. In October 2019, Twitter CEO Jack Dorsey announced that the platform would ban all political advertising.[129] The decision places Twitter in stark contrast with Facebook, which allows political ads and exempts politicians' political ads from its fact-checking program[130] and whose CEO Mark Zuckerberg had stridently defended his company's laissez-faire attitude towards political content moderation on the grounds that this approach upholds the ideal of free expression.[131] By contrast, Dorsey distinguished Twitter's new policy by explaining that it is not about free expression, but rather about politicians "paying for reach."[132]

---

[127] See text accompanying notes x – y.
[128] See text accompanying notes x – y.
[129] Jack Dorsey (@jack), TWITTER (Oct. 30, 2019, 4:05 PM), https://twitter.com/jack/status/1189634360472829952.
[130] Nick Clegg, *Facebook, Elections and Political Speech*, FACEBOOK NEWSROOM (Sept. 24, 2019), https://about.fb.com/news/2019/09/elections-and-political-speech/.
[131] Cecilia Kang and Mike Isaac, Defiant Zuckerberg Says Facebook Won't Police Political Speech, N.Y. TIMES (Oct. 21, 2019), https://www.nytimes.com/2019/10/17/business/zuckerberg-facebook-free-speech.html.
[132] Jack Dorsey (@jack), TWITTER (Oct. 20, 2019, 4:05 PM), https://twitter.com/jack/status/1189634377057067008.

Twitter published its policy for implementing its political advertising ban on November 11, 2019, a little less than a year before the 2020 presidential election. Twitter defines political content as that which "references a candidate, political party, elected or appointed government official, election, referendum, ballot measure, legislation, regulation, directive, or judicial outcome."[133] Ads that reference the above, including by "appeals for votes, solicitations of financial support, and advocacy for or against any of the above-listed types of political content" are prohibited. PACs, SuperPACs, candidates, political parties, and elected or appointed government officials are also banned from advertising on Twitter.[134] There are, however, some exemptions. Advertisers that Twitter deems to be news publishers may reference political content so long as the reference does not amount to advocacy.[135]

Twitter's Legal, Policy and Trust & Safety Lead Vijaya Gadde also identified an exemption for "cause-based ads"[136] – ads that "educate, raise awareness, and/or call for people to take action in connection with civic engagement, economic growth, environmental stewardship, or social equity causes."[137] Political organizations, candidates, and politicians may not use such ads, but other groups may.[138] Among other restrictions,[139] caused-based ads may not be micro-targeted. Twitter's allowance of caused-based ads is an apparent response to initial criticism of Twitter's policy. Many users reacted to Twitter's announcement by requesting more precise definitions, including questions about what

---

[133] *Political Content*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content.html (last accessed July 19, 2020).
[134] *Political Content FAQs*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/political-content/political-content-faqs.html (last accessed July 19, 2020).
[135] *Political Content*, *supra* note 90. Such publishers must have a minimum of 100,000 monthly unique visitors in the United States. They must also have a searchable archive, may not be primarily user-generated or aggregated content, and must not be dedicated to a single issue. *How to Get Exempted As a News Publisher from the Political Content Policy*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/ads-content-policies/political-content/news-exemption.html (last accessed July 19, 2020).
[136] Vijaya Gadde (@vijaya), TWITTER (Nov. 15, 2019, 1:30 PM), https://twitter.com/vijaya/status/1195408747926917120.
[137] *Cause-Based Advertising Policy*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/restricted-content-policies/cause-based-advertising.html (last accessed July 19, 2020).
[138] @vijaya *supra* note 93. https://twitter.com/vijaya/status/1195408747926917120.
[139] Restrictions include certification for caused-based advertisers. *Cause-Based Advertiser Certification*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/ads-content-policies/cause-based-advertising/cause-based-certification.html (last accessed July 19, 2020).

constitutes a "political" ad[140] and what constitutes an "ad."[141] As yet, it is unclear. Twitter states that for-profit organizations may place "cause ads" if they do not "have the primary goal of driving political, judicial, legislative, or regulatory outcomes" and are "tied to the organization's publicly stated values, principles, and or/beliefs."[142] However, it is not clear at this time how Twitter will interpret the "primary goal" language in its policy.[143]

In addition to prohibiting political ads on its platform, Twitter recently announced measures to combat misinformation in the form of manipulated media like deepfakes and shallow fakes.[144] On February 4, 2020, Twitter announced its new policy on "synthetic and manipulated media,"[145] which provides: "You may not deceptively share synthetic or manipulated media that are likely to cause harm. In addition, we may label Tweets

---

[140] https://twitter.com/aaronhuertas/status/1189672683400761344

[141] Brad Koenig (@MavsLaker), TWITTER (Oct. 31, 2019, 4:19 PM), https://twitter.com/MavsLaker/status/1190000411559780358.

[142] *Cause-Based Advertising FAQs*, TWITTER: BUS., https://business.twitter.com/en/help/ads-policies/ads-content-policies/cause-based-advertising/faqs.html (last accessed July 19, 2020).

[143] Some have commented that Sierra Club could promote their causes but not single out politicians or legislation, or that a group could run a gun violence awareness ad but not call for a ban on assault weapons as that would imply a legislative outcome. Sheila Dang and Paresh Dave, *Twitter Tightens Bans on Political Ads and Causes Ahead of 2020 U.S. Election*, REUTERS (Nov. 15, 2019), https://www.reuters.com/article/us-twitter-politics-adban-idUSKBN1XP224 Others have observed the challenges Twitter can expect to face in distinguishing between causes and political outcomes, for instance is an ad about universal healthcare a cause or is it about a related bill and how would that be determined? Id. Still others have observed that, if Twitter's misinformation policy is not integrated with its cause-based ads policy, Twitter could still permit inaccurate, but "softer" talking points that don't rise to the level of lobbying, e.g. an anti-minimum wage ad would not be permitted but an inaccurate ad about how a minimum wage law bankrupted a town could conceivably be permitted. Emily Stewart, *Twitter Is Walking into a Minefield with Its Political Ads Ban*, VOX: RECODE (Nov. 15, 2019), https://www.vox.com/recode/2019/11/15/20966908/twitter-political-ad-ban-policies-issue-ads-jack-dorsey.

[144] A "deepfake" is defined as a "product of artificial intelligence or machine learning, including deep learning techniques (e.g., a technical deepfake), that merges, combines, replaces, and/or superimposes content onto a video, creating a video that appears authentic." See, e.g., Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIFORNIA LAW REVIEW 1753 (2019); Richard L. Hasen, Deep Fakes, Bots, and Siloed Justices: American Election Law in a Post Truth World, U. C. Irvine Legal Studies Research Paper 2019.

[145] See Yoel Roth & Ashita Achuthan, *Building rules in public: Our approach to synthetic & manipulated media*, TWITTER BLOG (Feb. 4, 2020), https://blog.twitter.com/en_us/topics/company/2020/new-approach-to-synthetic-and-manipulated-media.html.

containing synthetic and manipulated media to help people understand their authenticity and to provide additional context."[146] Pursuant to this rule, Twitter will label content that is deceptively altered or fabricated, and will remove content if it impacts public safety or is likely to cause serious harm.[147] Twitter has already shown, on five separate occasions,

---

[146] *Synthetic and Manipulated Media Policy,* TWITTER HELP CENTER, https://help.twitter.com/en/rules-and-policies/manipulated-media (last visited July 19, 2020).

[147] *Id.* Relevant portions of Twitter's synthetic and manipulated media policy are as follows: You may not deceptively promote synthetic or manipulated media that are likely to cause harm. In addition, we may label Tweets containing synthetic and manipulated media to help people understand their authenticity and to provide additional context. [W]e may label Tweets that include media (videos, audio, and images) that have been deceptively altered or fabricated. In addition, you may not share deceptively altered media on Twitter in ways that mislead or deceive people about the media's authenticity where threats to physical safety or other serious harm may result.….In order for content to be labeled or removed under this policy, we must have reason to believe that media, or the context in which media are presented, are significantly and deceptively altered or manipulated. …In assessing whether media have been significantly and deceptively altered or fabricated, some of the factors we consider include:

- whether the content has been substantially edited in a manner that fundamentally alters its composition, sequence, timing, or framing;
- any visual or auditory information (such as new video frames, overdubbed audio, or modified subtitles) that has been added or removed; and
- whether media depicting a real person have been fabricated or simulated

We are most likely to take action (either labeling or removal…) on more significant forms of alteration, such as wholly synthetic audio or video or content that has been doctored (spliced and reordered, slowed down) to change its meaning. Subtler forms of manipulated media, such as isolative editing, omission of context, or presentation with false context, may be labeled or removed on a case-by-case basis. …

In order to determine if media have been significantly and deceptively altered or fabricated, we may use our own technology or receive reports through partnerships with third parties. In situations where we are unable to reliably determine if media have been altered or fabricated, we may not take action to label or remove them.

…We also consider whether the context in which media are shared could result in confusion or misunderstanding or suggests a deliberate intent to deceive people about the nature or origin of the content, for example by falsely claiming that it depicts reality. We assess the context provided alongside media to see whether it makes clear that the media have been altered or fabricated. Some of the types of context we assess in order to make this determination include:

that it will place warnings on posts from the President that violate its policies, such as its policies on abusive behavior and on misinformation, including manipulated media.[148]

- The text of the Tweet accompanying or within media
- Metadata associated with media
- Information on the profile of the account sharing media
- Websites linked in the Tweet, or in the profile of the account sharing media

…

Tweets that share synthetic and manipulated media are subject to removal under this policy if they are likely to cause serious harm. Some specific harms we consider include:

- Threats to the physical safety of a person or group
- Risk of mass violence or widespread civil unrest
- Threats to the privacy or ability of a person or group to freely express themselves or participate in civic events…

We also consider the time frame within which the content may be likely to impact public safety or cause serious harm, and are more likely to remove content under this policy if we find that immediate harms are likely to result from the content's presence on Twitter….In most cases, if we have reason to believe that media shared in a Tweet have been significantly and deceptively altered or fabricated, we will provide additional context on Tweets sharing the media where they appear on Twitter. This means we may:

- Apply a label to the content where it appears in the Twitter product;
- Show a warning to people before they share or like the content;
- Reduce the visibility of the content on Twitter and/or prevent it from being recommended; and/or
- Provide a link to additional explanations or clarifications, such as in a Twitter Moment or landing page.

In most cases, we will take all of the above actions on Tweets we label.

Media that meet all three of the criteria defined above—i.e. that are synthetic or manipulated, shared in a deceptive manner, and is likely to cause harm—may not be shared on Twitter and are subject to removal. Accounts engaging in repeated or severe violations of this policy may be permanently suspended.

https://help.twitter.com/en/rules-and-policies/manipulated-media#:~:text=Overview,are%20likely%20to%20cause%20harm.&text=In%20addition%2C%20you%20may%20not,other%20serious%20harm%20may%20result.

[148] Twitter's first warning labels on Tweets from the President involved unsubstantiated claims about mail-in ballots being fraudulent, glorifying violence/use of force, and a manipulated video (discussed further below). Elizabeth Dwoskin, *Twitter's Decision to Label Trump's Tweets Was Two Years in the Making*, WASH. POST (May 29, 2020), https://www.washingtonpost.com/technology/2020/05/29/inside-twitter-trump-label/. As of the time of this writing, Twitter most recently affixed a warning label to a second Tweet from the President promoting use of force against protestors, citing its policy

In the first case of Twitter applying its new policy on disinformation through deliberately altered content, Twitter labeled as "manipulated media" an edited video featuring presidential candidate Joe Biden in which Biden appeared to be endorsing President Trump for re-election in 2020, which was tweeted by White House social media director Dan Scavino and retweeted by the President.[149] The video had been edited so as to mislead viewers into believing that Biden was actually endorsing Trump.

In short, Twitter's absolute ban on political ads and its restrictions on manipulated media constitute strong and likely effective measures toward addressing the problems of false and misleading political speech. Some skeptics of the ban, however, have pointed out that the ban will not affect "organic" content or messages from politicians that are shared or retweeted by supporters, and that it could encourage the use of "bots" or paid

---

regarding "the presence of a threat of harm against an identifiable group." Rachel Lerman, *Twitter Slaps Another Warning Label on Trump Tweet About Force*, Wash. Po (June 23, 2020), https://www.washingtonpost.com/technology/2020/06/23/twitter-slaps-another-warning-label-trump-tweet-about-force/. Facebook left the post up without a warning. *Id.*

[149] *See* Ivan Mehta, *Trump's Retweet with doctored Biden Video Earns Twitter's First 'Manipulated Media' Label*, THE NEXT WEB (March 9, 2020), https://thenextweb.com/twitter/2020/03/09/trumps-tweet-with-doctored-biden-video-earns-twitters-first-manipulated-media-label/.

users to amplify the tweets.[150] In addition, it remains to be seen whether Twitter's carve-out for caused-based ads will provide sufficient opportunities for important speech on topics of civic and social activism.

Facebook's Regulation of Falsity in Political Ads

Facebook is taking a number of steps to combat misinformation in general on its platform.[151] The company has adopted extensive measure to combat publicly-available misinformation, including by partnering with independent third-party fact checkers to evaluate posts, providing counterspeech in the form of "Related Articles"/"Additional Reporting on This" on topics similar to false or misleading posts, and limiting the distribution of posts from content providers who repeatedly share false news and eliminating their ability to profit.[152] These extensive measures to combat misinformation and false content on Facebook are applicable to political content and political ads, but are not applicable to posts that are considered "direct speech by a politician." Thus, under Facebook's currently applicable fact-checking policies, political speech and the content of political ads are subject to fact-checking – except if such content constitutes "direct speech by a politician." This exception for politicians' content has come under substantial scrutiny in recent months, especially given the highly controversial posts of President Trump. Before examining this controversial exception to Facebook's general fact-checking policy for public posts on its platform, I first examine the company's generally-applicable policy itself.

Facebook's General Fact-Checking Policy for Publicly-Available Posts – Excluding the Posts of Politicians

Facebook is continuing to expand the partnership that it began in December 2016 with fact-checkers to evaluate publicly-available content posted on its platform.[153] Through its fact-checking initiatives, Facebook is working with select independent third-party fact checkers, which are certified through the non-partisan International Fact-

[150] See https://www.financialexpress.com/industry/technology/twitter-exempts-some-cause-based-messages-from-political-ad-ban/
https://techcrunch.com/2019/10/15/twitter-world-leaders-break-rules/

[151] *See* Tessa Lyons, *Hard Questions: What's Facebook's Strategy for Stopping False News?*, FACEBOOK NEWSROOM (May 23, 2018), https://newsroom.fb.com/news/2018/05/hard-questions-false-news.

[152] *See* Hunt Allcott et al., *Trends in the Diffusion of Misinformation on Social Media* app. at 4 (Stan. Inst. for Econ. Pol'y Res., Working Paper No. 18-029, 2018), http://web.stanford.edu/~gentzkow/research/fake-news-trends-appx.pdf (listing in Table 1 all of Facebook's efforts to combat false news).

[153] *See* Lyons, supra note x.

Checking Network.[154] In the United States, the certified fact-checking organizations with whom Facebook works are the Associated Press, factcheck.org, Lead Stories, Check Your Fact, Science Feedback, and PolitiFact.[155] (Notably, Facebook had added *The Weekly Standard* to these ranks for a period of time in an attempt to respond to critics who claimed that its fact-checking program was politically biased, but this publication is now defunct.)

Facebook has expanded its fact-checking initiative to include the fact checking of all public, newsworthy Facebook posts, including links, articles, photos, and videos.[156] The fact-checking process on Facebook also applies to political advertisements -- except if those advertisements (or other posts) constitute "direct speech made by an elected official."[157] As Facebook explains, "Posts and ads from politicians are generally not subjected to fact-checking [including] politicians at every level. This means candidates running for office, current office holders - and, by extension, many of their cabinet appointees - along with political parties and their leaders." This conspicuous exception to Facebook's fact-checking process has major ramifications for the political process, and has subjected Facebook to substantial criticism in recent months. Below, I first examine Facebook's fact-checking process generally, and then turn to the exception to this process for posts made by elected officials (including their political advertisements).

Facebook's fact-checking process can be initiated by Facebook users by flagging a post as being potentially false. Subject to the exception for direct speech by politicians, any public, newsworthy post (including text, photos, and videos) can be flagged for fact-checking, either by a user, by an outside journalist, or – as is most commonly the case – by Facebook's machine learning algorithms. For a user to flag a post as potentially false, a

---

[154] *See id. and* https://ifcncodeofprinciples.poynter.org/signatories

[155] *See* Mike Ananny, *Checking in with the Facebook Fact-Checking Partnership*, COLUM. JOURNALISM REV. (Apr. 4, 2018), https://www.cjr.org/tow_center/facebook-fact-checking-partnerships.php; *see also How Are Third-Party Fact-Checkers Selected?*, FACEBOOK HELP CENT., https://www.facebook.com/help/1599660546745980?helpref=faq_content (last visited Sept. 29, 2018); *Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK HELP CENT., https://www.facebook.com/help/publisher/182222309230722 (last accessed July 19, 2020).

[156] *See* Antonia Woodford, *Expanding Fact-Checking to Photos and Videos*, FACEBOOK NEWSROOM (Sept. 13, 2018), https://newsroom.fb.com/news/2018/09/expanding-fact-checking.

[157] "If a claim is made directly by a politician on their Page, in an ad or on their website, it is considered direct speech and ineligible for our third party fact checking program — even if the substance of that claim has been debunked elsewhere." *Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK HELP CENT., https://www.facebook.com/help/publisher/182222309230722 (last accessed July 19, 2020).

user clicks ". . ." next to the post he or she wishes to flag as false, then clicks "Report post," then clicks "It's a false news story," then clicks "Mark this post as false news."[158]

Once a post is flagged by a user as a potential false news story, it is submitted for evaluation to a third-party independent fact-checker.[159] While the process of evaluating posts in the past was triggered only by user flagging, Facebook now incorporates other ways of triggering such evaluation, including by providing its independent fact checkers with the authority to proactively identify posts to review[160] as well as by using machine learning to identify potentially false posts.[161] For each piece of content up for review, a fact checker has the option of providing one of eight different ratings: false, mixture, false headline, true, not eligible (if, for example, the post is not verifiable, opinion, etc.), satire, opinion, or prank generator.[162]

Once a third-party fact-checker has determined that a post is false, Facebook then initiates several steps. First, Facebook deprioritizes false posts in users' News Feeds, *i.e.*, the constantly updating list of stories in the middle of a user's home page (including status updates, photos, videos, links, app activity, and likes)—such that future views of each false post will be reduced by an average of eighty percent.[163] Second, Facebook commissions a fact-checker to write a "Related Article" or "Additional Reporting on This" setting forth truthful information about the subject of the false post and the reasons why the fact-checker rated the post as false.[164] Such content is then displayed in conjunction with the false post

---

[158] *See How Do I Mark a Post as False News?*, FACEBOOK HELP CENT., https://www.facebook.com/help/572838089565953?helpref=faq_content (last visited Sept. 29, 2018). Alternatively, a user can click ". . ." next to a post, then click "Find Support or Report Post," then select "False News."

[159] See Lyons, supra note x ("[W]hen people on Facebook submit feedback about a story being false or comment on an article expressing disbelief, these are signals that a story should be reviewed.").

[160] *See id.* ("Independent third-party fact-checkers review the stories, rate their accuracy, and write an article explaining the facts behind their rating.").

[161] *See* Dan Zigmond, *Machine Learning, Fact-Checkers and the Fight Against False News*, FACEBOOK NEWSROOM (Apr. 8, 2018), https://about.fb.com/news/2018/04/inside-feed-misinformation-zigmond.

[162] *Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK HELP CENT., https://www.facebook.com/help/publisher/182222309230722 (last accessed July 19, 2020).

[163] *See id.*; see also Tessa Lyons, *Increasing Our Efforts to Fight False News*, FACEBOOK NEWSROOM (June 21, 2018), https://newsroom.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/.

[164] *See* Tessa Lyons, *Replacing Disputed Flags With Related Articles*, FACEBOOK NEWSROOM (Dec. 20, 2017), https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation.

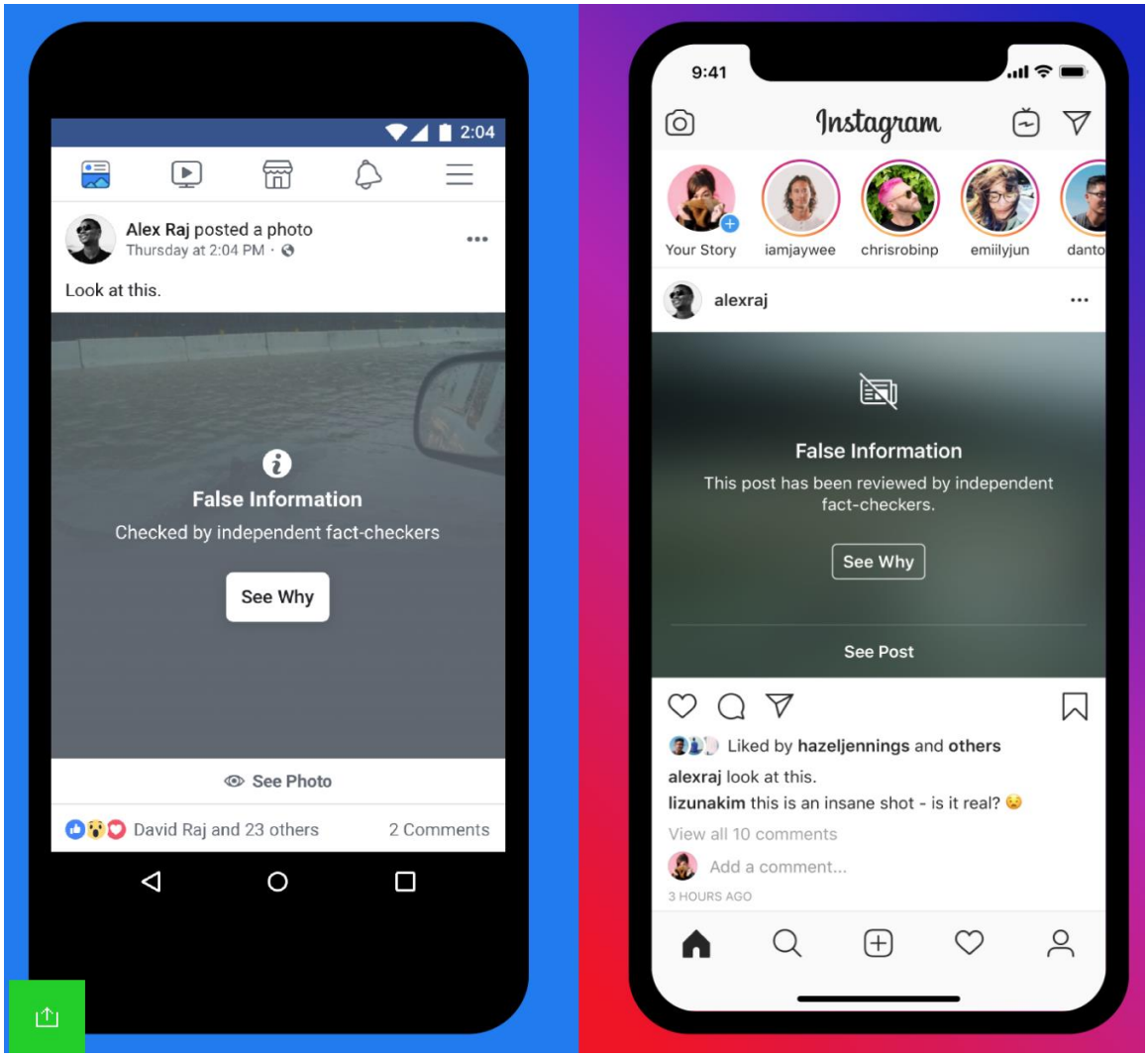
on the same subject.[165] While Facebook formerly flagged false news sites with a "Disputed" flag, the company changed its approach in response to research suggesting that such flags may actually entrench beliefs in the disputed posts. Facebook now provides "Related Articles"/"Additional Reporting on This" in conjunction with false news stories (which apparently does not result in similar entrenchment).[166] In addition, users who attempt to share the false post will be notified that the post has been disputed and will be informed of the availability of a "Related Article"/"Additional Reporting on This," as will users who earlier shared the false post,[167] as in the example below (setting forth Facebook and Instagram's flags).[168]

[165] *See id.*; *see also* Geoffrey A. Fowler, *I Fell for Facebook Fake News. Here's Why Millions of You Did Too.*, WASH. POST (Oct. 18, 2018), https://www.washingtonpost.com/technology/2018/10/18/i-fell-facebook-fake-news-heres-why-millions-you-did-too/ (describing steps undertaken by Facebook to respond to fake video, including posting "Additional Reporting on This," with links to reports from fact-checking organizations); supra Lyons, note 14; Sara Su, *New Test with Related Articles*, FACEBOOK NEWSROOM (Apr. 25, 2017), https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles.

[166] See Lyons, supra note x (explaining that "[a]cademic research on correcting misinformation has shown that putting a strong image, like a red flag, next to an article may actually entrench deeply held beliefs . . . [but that] Related Articles, by contrast, are simply designed to give more context, which our research has shown is a more effective way to help people get to the facts. . . . [W]e've found that when we show Related Articles next to a false news story, it leads to fewer shares than when the Disputed Flag is shown.").

[167] See Lyons, supra note x.

[168] Elle Hunt, *'Disputed by Multiple Fact-Checkers': Facebook Rolls Out New Alert to Combat Fake News*, THE GUARDIAN (Mar. 21, 2017), https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news.



In addition, As Facebook explains: "When fact-checkers write articles with more information about a story, you'll see them in Related Articles immediately below the story in your News Feed."[169] Facebook also provides its users who are about to share posts that have been debunked by a fact-checker by alerting them to additional reporting. Facebook users are provided with the following notice, in an attempt to keep such false posts from going viral: "Before you share this content, you might want to know that there's additional reporting on this from [list of fact-checkers that have debunked the post]."[170] As an

[169] *How is Facebook Addressing False News Through Third-Party Fact-Checkers*, FACEBOOK HELP CENT., https://www.facebook.com/help/1952307158131536 (last visited July 21, 2020).
[170] *See Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK BUS., https://www.facebook.com/help/publisher/182222309230722 (last accessed Dec. 12, 2019).

example, Facebook users who attempted to share the infamous video of Nancy Pelosi, which was doctored to make it appear that she was drunk and slurring her words, were alerted to the fabrication.[171] In addition, Facebook will now post more prominent fact-checking labels as interstitials warnings atop photos and videos on Facebook (and Instagram) that were fact-checked as false, as in the examples below.

Third, content providers—*i.e.*, Facebook pages and domains—that repeatedly publish and/or share false posts will have their ability to monetize and advertise reduced and ultimately disabled by Facebook unless and until they issue corrections or successfully dispute fact-checkers' determination that their posts are false.[172]

Facebook's Policies Regarding False Political Ads

With respect to false political ads, Facebook's policy is complex. Although Facebook has implemented extensive measures with respect to false posts generally (described above), this false news policy does not apply to "direct speech" by politicians. Accordingly, Facebook's general false news policy -- composed of the fact-checking process described above -- has an exception for "direct speech" by politicians, such that direct speech by politicians is not run through Facebook's external fact checking process. Facebook provides the following justification for this exception to its fact-checking policy:

> We rely on third-party fact-checkers to help reduce the spread of false news and other types of viral misinformation, like memes or manipulated photos and videos. We don't believe, however, that it's an appropriate role for us to referee political debates and prevent a politician's speech from reaching its audience and being subject to public debate and scrutiny. ...This means that we will not send organic content or ads from politicians to our third-party fact-checking partners for review.[173]

---

[171] Cecilia Kang, *Nancy Pelosi Criticizes Facebook for Handling of Altered Videos*, N.Y. TIMES (May 29, 2019), https://www.nytimes.com/2019/05/29/technology/facebook-pelosi-video ("People who see the video in feed, try to share it from feed, or already shared it are alerted that it's false.").

[172] *See* Satwik Shukla & Tessa Lyons, *Blocking Ads from Pages That Repeatedly Share False News*, FACEBOOK NEWSROOM (Aug. 28, 2017), https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news; https://www.facebook.com/business/help/182222309230722

[173] Nick Clegg, *Facebook, Elections and Political Speech*, FACEBOOK NEWSROOM (Sep. 24, 2019), https://about.fb.com/news/2019/09/elections-and-political-speech/. Facebook will, however, subject the posts of political action committees and political advocacy groups to its fact-checking process. Facebook has explained that while it will not fact check political ads from candidates, it does evaluate the accuracy of political ads from

> Posts and ads that constitute "direct speech" from current "politicians" at any/every level and their appointees -- i.e., the politician's own claim or statement -- are not subjected to fact-checking -- even if the substance of the claim has been debunked elsewhere.[174]

Facebook's decision not to submit direct speech from current politicians to fact-checking is apparently grounded in the belief that such political speech is already subject to sufficient scrutiny among the polity and the free press and should not be subject to further scrutiny by Facebook's fact-checkers. Facebook further justifies its policies as follows: "In a democracy, people should decide what is credible, not tech companies...That's why-–like other internet platforms and broadcasters– we don't fact check ads from politicians."[175] The company further justifies its decision by adverting to the importance of political ads to challengers and local candidates: "Given the sensitivity around political ads, we have considered whether we should ban them altogether...But political ads are important for local candidates, up-and-coming challengers, and advocacy groups that use our platform to reach voters and their communities."[176]

As a result, political speech and political ads made by politicians themselves – posts and campaign ads by politicians – operate in a separate system on Facebook. While ordinary users who publicly post false content are subject to a range of actions and consequences –including possibly outright ban from Facebook[177] – elected officials are exempt from such consequences.

Facebook's decision not to screen for or remove false ads by politicians came into sharp focus in October 2019, when President Donald Trump's reelection campaign began running an ad that was proven to be false[178] about former Vice President Joe Biden on

---

political advocacy groups or political action committees. *See* David Klepper, *Facebook Clarifies Zuckerberg Remarks on False Political Ads*, AP NEWS (Oct. 24, 2019), https://apnews.com/64fe06acd28145f5913d6f815bec36a2.

[174] *Fact-Checking on Facebook: What Publishers Should Know*, FACEBOOK BUS., https://www.facebook.com/help/publisher/182222309230722 (last accessed Dec. 12, 2019) (emphasis added).

[175] David Klepper, *Facebook Clarifies Zuckerberg Remarks on False Political Ads*, AP NEWS (Oct. 24, 2019), https://apnews.com/64fe06acd28145f5913d6f815bec36a2.

[176] David Klepper, *Facebook Clarifies Zuckerberg Remarks on False Political Ads*, AP NEWS (Oct. 24, 2019), https://apnews.com/64fe06acd28145f5913d6f815bec36a2.

[177] See text accompanying notes x – y.

[178] Amy Sherman, *Donald Trump Ad Misleads About Joe Biden, Ukraine, and the Prosecutor*, POLITIFACT (Oct. 11, 2019),

Facebook. The Trump Campaign released a 30-second video ad accusing Biden of promising Ukraine money in exchange for firing a prosecutor investigating a company with ties to Biden's son, Hunter Biden.[179] The video ad falsely claimed that Joe Biden offered Ukraine $1 billion in aid if Ukraine pushed out the official investigating a company tied to Hunter Biden. The Biden campaign asked Facebook to take down the ad, but Facebook refused. In justifying its refusal, Facebook's head of global elections policy Katie Harbath explained: "Our approach is grounded in Facebook's fundamental belief in free expression, respect for the democratic process, and the belief that, in mature democracies with a free press, political speech is already arguably the most scrutinized speech there is." Accordingly, the false Trump Campaign ad on Biden remained available on Facebook, garnering at least 4.6 million views.[180]

Former presidential candidate Senator Elizabeth Warren -- who has a history of locking horns with Facebook and with big tech in general[181] -- took particular aim at Facebook's policy towards political ads by placing an intentionally false ad on the platform

---

https://www.politifact.com/factchecks/2019/oct/11/donald-trump/trump-ad-misleads-about-biden-ukraine-and-prosecut/.

[179] Michael M. Grynbaum & Tiffany Hsu, *CNN Rejects 2 Trump Campaign Ads, Citing Inaccuracies*, N.Y. TIMES (Oct. 3, 2019), https://www.nytimes.com/2019/10/03/business/media/cnn-trump-campaign-ad.html.

[180] Jeremy B. Merrill, *While Everyone Was Looking at Facebook, Trump's False Biden Ad Appeared More Often on YouTube*, QUARTZ (Nov. 1, 2019), https://qz.com/1739780/trumps-biden-ad-appeared-more-often-on-youtube-than-on-facebook/.

[181] *See* Elizabeth Warren, (@TeamWarren), *Here's How We Can Break Up Big Tech*, MEDIUM (Mar. 8, 2019), https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c. After announcing her ambition to break up big tech companies, Warren took out ads on Facebook that denounced Facebook itself as well as Amazon and Google for their "vast power over our economy and our democracy." Cristiano Lima, *Facebook Backtracks After Removing Warren Ads Calling for Facebook Breakup*, POLITICO (Mar. 11, 2019), https://www.politico.com/story/2019/03/11/facebook-removes-elizabeth-warren-ads-1216757. Facebook initially removed the ads, apparently because they contained an unauthorized reproduction of Facebook's logo, but soon after, the company reversed course and restored them "in the interest of allowing robust debate." Isaac Stanley-Becker and Tony Romm, *Facebook Deletes, and Then Restores, Elizabeth Warren's Ads Criticizing the Platform, Drawing Her Rebuke*, WASH. POST (Mar. 12, 2019), https://www.washingtonpost.com/nation/2019/03/12/facebook-deletes-then-restores-elizabeth-warrens-ads-criticizing-platform-drawing-her-rebuke/; Elizabeth Warren, *Elizabeth's Plan: Break up the Big Tech Companies*, FACEBOOK (Mar. 8, 2019), https://www.facebook.com/ElizabethWarren/videos/396777104233421/. Warren meanwhile warned of the danger of a "social media marketplace" that is "dominated by a single censor." Elizabeth Warren (@ewarren), TWITTER (Mar. 11, 2019, 7:59 PM), https://twitter.com/ewarren/status/1105256905058979841.

in October 2019. Warren's ad declared that "Mark Zuckerberg and Facebook just endorsed Donald Trump for re-election."[182] She explained that the ad was a test to see "just how far" Facebook's policy went and accused Facebook of becoming a "disinformation-for-profit machine."[183] Adhering to its policy of refusing to fact-check direct speech by politicians, Facebook declined to remove Warren's intentionally (and provocatively) false ad, stating "if Senator Warren wants to say things she knows to be untrue, we believe Facebook should not be in the position of censoring that speech."[184]

At least one of Facebook's fact-checking partners expressed dissatisfaction with the company's policy of refusing to fact-check direct speech by politicians. Lead Stories – one of Facebook's six fact-checking organizations in the United States – has argued that Facebook should modify its policy and should scrutinize ads from politicians and recommended that such a fact-checking process should be reviewed by a new nonpartisan blue-ribbon panel.[185] This fact-checking partner reportedly proposed these changes to the policy at Facebook's November 2019 fact-checking partner summit. Alan Duke, Lead Stories' editor-in chief, contended that there is "an urgent need for a fair method to identify egregiously false political ads in 2020" because "too many people are too fast to fall for disinformation."[186] However, Facebook declined to make any changes in its fact-checking policy in response.

In addition, Facebook employees recently rose up in strong opposition to Facebook's policy exempting politicians' (and especially President Trump's) posts from fact-checking (and from other of the company's content policies as well, including those prohibiting threats of imminent violence). The particular flashpoint most recently at issue involved violent speech, not misinformation, in the form of Donald Trump's May 2020 post following the murder of George Floyd and the ensuing demonstrations. Trump

---

[182] Brian Fung, *Elizabeth Warren Targets Facebook's Ad Policy -- with a Facebook Ad*, CNN (Oct. 12, 2019), https://www.cnn.com/2019/10/11/politics/elizabeth-warren-facebook-ad/index.html.
[183] Elizabeth Warren (@ewarren), TWITTER (Oct. 12, 2019, 10:01 AM), https://twitter.com/ewarren/status/1183019880867680256.
[184] Brian Fung, *Elizabeth Warren Targets Facebook's Ad Policy -- with a Facebook Ad*, CNN (Oct. 12, 2019), https://www.cnn.com/2019/10/11/politics/elizabeth-warren-facebook-ad/index.html.
[185] *See* Donie O'Sullivan, *A Facebook Fact-Checker Will Propose a Possible Solution to the Company's False Ad Debacle*, CNN (Oct. 31, 2019), https://www.cnn.com/2019/10/31/tech/facebook-fact-checking-political-ads/index.html.
[186] Donie O'Sullivan, *A Facebook Fact-Checker Will Propose a Possible Solution to the Company's False Ad Debacle*, CNN (Oct. 31, 2019), https://www.cnn.com/2019/10/31/tech/facebook-fact-checking-political-ads/index.html.

threatened to deploy the military in Minneapolis to "bring the City under control" and infamously stated "when the looting starts, the shooting starts."[187]

President Trump made this post across several platforms. While Twitter appended a notice to the President's post explaining that the post violated the platform's rules against glorifying violence and requiring users to click through the notice to view the tweet (see below),

[187] Megan Rose Dickey and Taylor Hatmaker, *Facebook Employees Stage Virtual Walkout in Protest of Company's Stance on Trump Posts*, TECHCRUNCH (June 1, 2020), https://techcrunch.com/2020/06/01/facebook-employees-stage-virtual-walkout-in-protest-of-companys-stance-on-trump-posts/ (screenshot included).

Facebook took no action.[188] Facebook's CEO Mark Zuckerberg explained that he was personally appalled by the President's tweet, but felt that Facebook's institutional role was to "enable as much expression as possible unless it will cause imminent risk of specific harms or dangers spelled out in [Facebook's] clear policies." Zuckerberg explained further that "we read it as a warning about state action, and we think people need to know if the government is planning to deploy force."[189] Some of Facebook's employees, however, were extremely dissatisfied by the company's response, resulting in "intense debate" on Facebook's internal employee messaging system about the company's laissez-faire policies regarding politicians' posts.[190] In response, Zuckerberg hosted an internal town-hall to explain his and the company's rationale for inaction.[191] Facebook ultimately retreated from its non-interventionist stance towards Donald Trump and his campaign, at least with respect to its hate speech content regulation, as it removed a Trump Campaign page ad because of its use of a symbol of hate.[192] However, many companies felt Facebook

---

[188] Brian Stelter and Donie O'Sullivan, *Trump Tweets Threat That 'Looting' Will Lead to 'Shooting.' Twitter Put a Warning Label on It*, CNN (May 29, 2020), https://www.cnn.com/2020/05/29/tech/trump-twitter-minneapolis/index.html.

[189] Mark Zuckerberg, FACEBOOK (May 29, 2020, 4:19 PM), https://www.facebook.com/zuck/posts/10111961824369871.

[190] Rachel Siegel and Elizabeth Dwoskin, *Facebook Employees Blast Zuckerberg's Hands-Off Response to Trump Posts as Protests Grip Nation*, WASH. POST (June 1, 2020), https://www.washingtonpost.com/business/2020/06/01/facebook-zuckerberg-donation-trump/.

[191] Elizabeth Dwoskin and Nitasha Tiku, *Facebook Employees Said They Were 'Caught in an Abusive Relationship' with Trump as Internal Debates Raged*, WASH. POST (June 5, 2020), https://www.washingtonpost.com/technology/2020/06/05/facebook-zuckerberg-trump/.

[192] Days later when a Trump-affiliated campaign page posted an advertisement denouncing "dangerous MOBS of far-left groups … causing absolute mayhem" accompanied by an image of a downward facing red triangle, Facebook deactivated those ads because the image was the same symbol used by the Nazis to denote political prisoners in its concentration camps. Facebook representatives stated that the ad violated a policy against using a "banned hate group's symbols" outside of a condemnatory context or as an object for discussion. Isaac Stanley-Becker, *Facebook Removes Trump Ads with Symbol Once Used by Nazis to Designate Political Prisoners*, WASH. POST (June 18, 2020), https://www.washingtonpost.com/politics/2020/06/18/trump-campaign-runs-ads-with-marking-once-used-by-nazis-designate-political-prisoners/. Zuckerberg has also since announced that Facebook will begin labeling "newsworthy content." Occasionally, he explains, "we leave up content that would otherwise violate our policies if the public interest value outweighs the risk of harm." Now, Facebook will append a notification that the content violates Facebook's policy but remains so that people can engage with and discuss it. Mark Zuckerberg, FACEBOOK (June 26, 2020, 11:25 AM), https://www.facebook.com/zuck/posts/10112048980882521. Facebook will also further restrict content that can be included in paid advertisements. Ads that claim people from "a specific race, ethnicity, national origin, religious affiliation, caste, sexual orientation,

still had not gone far enough, and joined a growing advertising boycott to pressure the platform to take more aggressive action against the hate speech and misinformation being spread by political figures such as President Trump.[193] Facebook responded by announcing that it would remove posts from political leaders that incited violence or attempted to suppress voting, and affix labels on posts violating its hate speech prohibitions.[194]

Facebook's decision to exempt speech by politicians from its fact-checking and other content regulation policies also drew sharp criticism recently from civil rights experts, who conducted an extensive, independent two-year civil rights audit of Facebook's content regulation policies and their implementation.[195] The experts' concerns were magnified by Facebook's response to President Trump's posts regarding recent civil rights protests and mail-in ballots in the context of the pandemic. The civil rights experts expressed strong criticisms of the company's policies and exemption of Trump's posts from its content regulation policies and voiced particular concern about the ramifications of this exemption for our political process:

> We have grave concerns that the combination of the company's decision to exempt politicians from fact-checking and the precedents set by its recent

---

gender identity or immigration status are a threat to the physical safety, health or survival of others" are now prohibited when they were not before, and Facebook also intends to "better protect immigrants, migrants, refugees and asylum seekers from ads suggesting these groups are inferior or expressing contempt, dismissal or disgust directed at them." *Id.*

[193] "Marketers are expressing unease with how [Facebook] handles misinformation and hate speech, including its permissive approach to problematic posts by President Trump." *All the Companies Quitting Facebook*, N.Y. TIMES: DEALBOOK NEWSLETTER (July 7, 2020), https://www.nytimes.com/2020/06/29/business/dealbook/facebook-boycott-ads.html. Boycott lists can be found here: Tiffany Hsu and Gillian Friedman, *CVS, Dunkin', Lego: The Brands Pulling Ads from Facebook over Hate Speech*, N.Y. TIMES (July 7. 2020), https://www.nytimes.com/2020/06/26/business/media/Facebook-advertising-boycott.html; and here: Allen Kim and Brian Fung, *Facebook Boycott: View the List of Companies Pulling Ads*, CNN (July 2, 2020), https://www.cnn.com/2020/06/28/business/facebook-ad-boycott-list/index.html.

[194] Craig Timberg and Elizabeth Dwoskin, *Silicon Valley is Getting Tougher on Trump and His Supporters over Hate Speech and Disinformation,* Wash. Post (July 10, 2020), https://www.washingtonpost.com/technology/2020/07/10/hate-speech-trump-tech/. Twitch recently suspended President Trump's account and Reddit closed a long-controversial forum named after the President (this same forum helped to popularize the dangerous Pizzagate false conspiracy theory). Reddit's action may have been in response to "employee" concerns as well, as it came after an open letter written by hundreds of volunteer moderators chastised Reddit's leadership for the proliferation of hateful speech, calling it the company's "most glaring problem." Id.

[195] *See Facebook's Civil Rights Audit – Final Report* (July 8, 2020), available here: https://muslimadvocates.org/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

> decisions on President Trump's posts, leaves the door open for the platform to be used by other politicians to interfere with voting. If politicians are free to mislead people about official voting methods (by labeling ballots illegal or making other misleading statements that go unchecked, for example) and are allowed to use not-so-subtle dog whistles with impunity to incite violence against groups advocating for racial justice, this does not bode well for the hostile voting environment that can be facilitated by Facebook in the United States. We are concerned that politicians, and any other user for that matter, will capitalize on the policy gaps made apparent by the president's posts and target particular communities to suppress the votes of groups based on their race or other characteristics. With only months left before a major election, this is deeply troublesome as misinformation, sowing racial division and calls for violence near elections can do great damage to our democracy.[196]

Facebook's Refusal to Regulate the Microtargeting of Political Ads

Facebook, as discussed above, has been the primary social media platform facilitating the microtargeting of political ads to members of the public, allowing political advertisers to have access to the vast trove of social data that it collects on its users, to serve up ads to users with great precision and with no public scrutiny. In addition to Facebook's reticence to adopt measures applicable to false political ads by politicians, the company has also been flatly unwilling to regulate or modify its practice of allowing for the microtargeting of political ads on its platform, despite calls for the company to do so. Although Facebook as of late 2019 was reportedly considering increasing the minimum number of people who can be targeted in political ads on its platform from 100 to a few thousand, as of this date, Facebook has not made any changes to its policy allowing for the microtargeting of political ads.[197]

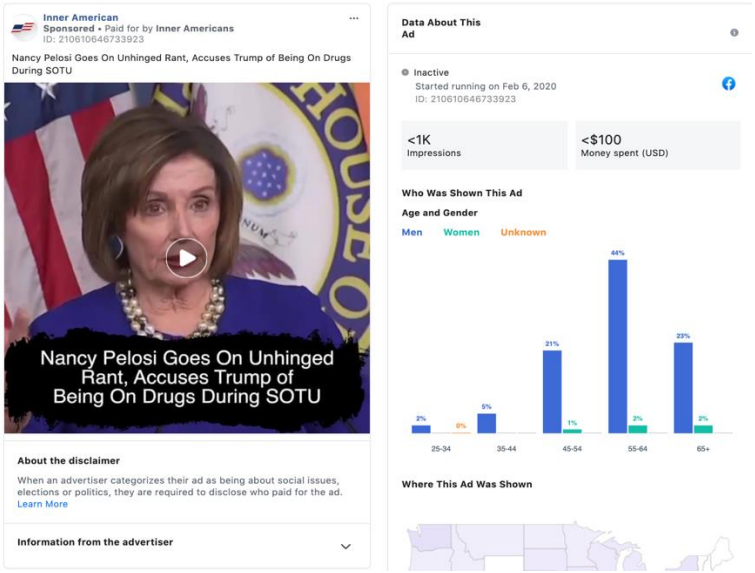
Facebook's Transparency and Disclosure Requirements Regarding Political/Electioneering Advertisements

---

[196] *Id.*

[197] *See* Associated Press, *Facebook Refuses to Restrict Untruthful Political Ads and Micro-Targeting*, THE GUARDIAN (Jan. 9, 2020), https://www.theguardian.com/technology/2020/jan/09/facebook-political-ads-micro-targeting-us-election; Emily Glazer, *Facebook Weighs Steps to Curb Narrowly Targeted Political Ads*, WALL STREET J. (Nov. 21, 2019), https://www.wsj.com/articles/facebook-discussing-potential-changes-to-political-ad-policy-11574352887?mod=followfacebook. The company has, however, adopted a policy that will allow users to opt out of political ads. https://thehill.com/policy/technology/477486-facebook-will-still-allow-misinformation-micro-targeting-under-new-ad-rules

Facebook has recently implemented a Political Advertising Policy that requires, first, that every election-related and issue advertisement made available on Facebook to users in the United States be clearly labeled as a "Political Ad" and include a "Paid for by" disclosure, with the name of the individual or organization who paid for the advertisement at the top of the advertisement.[198] Second, under the Policy, Facebook has committed to collecting and maintaining a publicly available archive of political advertisements as part of its Ad Library, which provides information regarding the campaign budget associated with each individual ad and how many people saw it, including their age, location, and gender.[199] See example below.



And Facebook has recently updated its Ad Library's functionality in an effort to increase transparency and provide more useful data to researchers, advocates, and the public generally – including by permitting users to search for and filter ads based on the estimated audience size – which enable researchers to identify and study micro-targeted ads.[200] Finally, under the Policy, Facebook will prohibit foreign entities from purchasing political ads directed at U.S. audiences. Facebook implements this last prohibition by mailing to prospective political advertisers a postcard to a U.S. address in order to verify U.S. residency. If a prospective purchaser of a political ad is not verified under this process, it will not be able to purchase a political ad on Facebook. Commenting on the recently

---

[198] *See* Rob Goldman & Alex Himel, *Making Ads and Pages More Transparent*, FACEBOOK NEWSROOM (Apr. 6, 2018), https://newsroom.fb.com/news/2018/04/transparent-ads-and-pages/.

[199] Since 2018, Facebook has maintained a library of ads about social issues, elections or politics that ran on the platform. These ads are either classified as being about social issues, elections or politics or the advertisers self-declare that the ads require a "Paid for by" disclaimer. *See, e.g., Facebook's Civil Rights Audit – Final Report*, July 8, 2020; Rob Leathern, *Shining a Light on Ads with Political Content*, FACEBOOK NEWSROOM (May 24, 2018), https://newsroom.fb.com/news/2018/05/ads-with-political-content/.

[200] *See Facebook's Civil Rights Audit – Final Report*, July 8, 2020.

implemented Political Advertising Policy, Facebook's CEO Mark Zuckerberg explained, "These changes won't fix everything, but they will make it a lot harder for anyone to do what the Russians did during the 2016 election and use fake accounts and pages to run ads."[201] Facebook's recently implemented measures imposing disclosure requirements on political ads and limiting foreign entities from purchasing political ads go beyond those that are encompassed in the proposed Honest Ads Act, discussed above, and may at least be moderately successful in preventing the type of foreign interference in U.S. elections that occurred in 2016.

Google's Measures to Address Microtargeting of Political Ads

Google recently amended its rules governing the practice of microtargeting of political advertisements. While Google maintains that it has never offered "granular microtargeting" of election ads, in November 2019, Google officially amended its rules to restrict microtargeting so that political advertisers can only target ads based on three characteristics: an individual's age, gender, and general location (defined by postal code).[202] Political advertisers can also use contextual targeting, which enables them to serve users with ads according to the content that users are accessing.[203] Google claims this approach aligns it with industry practice in television, radio and print media.[204] Google's policy on microtargeting took effect in the European Union at the end of 2019, and became effective worldwide (including in the United States) in January 2020.[205]

---

[201] Josh Constine, *Facebook and Instagram Launch US Political Ad Labeling and Archive*, TECHCRUNCH (May 24, 2018), https://techcrunch.com/2018/05/24/facebook-political-ad-archive/.

[202] Scott Spencer, *An Update on Our Political Ads Policy*, GOOGLE BLOG: THE KEYWORD (Nov. 20, 2019), https://blog.google/technology/ads/update-our-political-ads-policy/.

[203] *See id.* Regarding its newly-announced policy, Google explains:

In the U.S., we have offered basic political targeting capabilities to verified advertisers, such as serving ads based on public voter records and general political affiliations (left-leaning, right-leaning, and independent). While we've never offered granular microtargeting of election ads, we believe there's more we can do to further promote increased visibility of election ads. That's why we're limiting election ads audience targeting to the following general categories: age, gender, and general location (postal code level). Political advertisers can . . . continue to do contextual targeting, such as serving ads to people reading or watching a story about, say, the economy. Of course, we recognize that robust political dialogue is an important part of democracy, and no one can sensibly adjudicate every political claim, counterclaim, and insinuation. So we expect that the number of political ads on which we take action will be very limited—but we will continue to do so for clear violations. *Id.*

[204] Scott Spencer, *An Update on Our Political Ads Policy*, GOOGLE BLOG: THE KEYWORD (Nov. 20, 2019), https://blog.google/technology/ads/update-our-political-ads-policy/.

[205] Rachel Sandler, *Google Limits Microtargeting for Paid Political Ads*, Forbes (Nov. 20, 2019), https://www.forbes.com/sites/rachelsandler/2019/11/20/google-limits-microtargeting-for-paid-political-ads/#55c667fd51ec.

Accordingly, under Google's rules, only the following characteristics may be used to target election ads: geographic location (but not radius around a location), age, gender, and contextual targeting options such as ad placements, topics, keywords against sites, apps, pages and videos. All other types of targeting are not allowed for use in election ads, including the use of Google's powerful Audience Targeting products,[206] Remarketing,[207] Customer Match,[208] and Geographic Radius Targeting.[209] Google's microtargeting policy applies to ads shown to users of Google's search engine and YouTube, as well as display advertisements sold by Google that appear on other websites. In an email to political campaigns, Google outlined these new rules, explaining that election ads will no longer be allowed to target what are called "affinity audiences" that look like other groups that campaigns might want to target. Further, political campaigns can no longer upload their own lists of people to whom to show ads. In addition, Google will prohibit what is known as "remarketing," the process of serving ads to people who have previously taken an action like visiting a campaign's website.

Google's microtargeting policy prevents political advertisers from taking advantage of some of Google's most sophisticated targeting tools, upon which it has built its dominant market position.[210] The most granular of those targeting tools are "custom affinity" audiences, an offering that has allowed advertisers to create tailor-made audiences by targeting individual interests and lifestyles as defined by keyword phrases. Google's sophisticated targeting tools also have allowed advertisers to target or exclude according to demographic data such as age, gender, household income, homeownership, and the like.[211] General advertisers may also target users who have previously interacted with their

[206] *About Audience Targeting*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/2497941 (last accessed: July 21, 2020).
[207] *About Remarketing*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/2453998 (last accessed: July 21, 2020).
[208] *About Customer Match*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/6379332 (last accessed: July 21, 2020).
[209] *Target Ads to Geographic Locations*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/1722043 (last accessed: July 21, 2020).
[210] According to this WSJ report citing "research firm eMarketer": 37% of the $130 billion U.S. digital ad market. Patience Haggin and Kara Dapena, *Google's Ad Dominance Explained in Three Charts*, WALL STREET J. (June 17, 2019), https://www.wsj.com/articles/why-googles-advertising-dominance-is-drawing-antitrust-scrutiny-11560763800.
[211] *About Demographic Targeting*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/2580383?hl=en&ref_topic=3122881 (last accessed July 21, 2020).

site[212] or by submitting previously collected customer data to re-engage with the same group or expand to similar audiences.[213] These sophisticated targeting tools are now unavailable to political advertisers.

One of the greatest challenges Google faces in implementing its policy restricting the use of microtargeting by political advertisers is how to meaningfully and accurately define political/election advertising. With respect to the United States, Google currently defines election ads as those that feature:

- A current officeholder or candidate for an elected federal office (including federal offices such as that of the President or Vice President of the United States, members of the United States House of Representatives or United States Senate).
- A current officeholder or candidate for a state-level elected office, such as Governor, Secretary of State, or member of a state legislature.
- A federal or state level political party.
- A state-level ballot measure, initiative, or proposition that has qualified for the ballot in its state.[214]

Yet, few election ads as they are popularly understood are likely to be so specific. For example, "issue ads" funded by Super-PACs may not specifically "advocate the election or defeat of a clearly identified federal candidate,"[215] yet such outside spending makes up the vast majority of political advertising.[216] Thus, Google's definition of election ads may turn out to be substantially underinclusive and ineffective.

Google has also implemented a host of procedural requirements for political advertisers. Advertisers who wish to purchase and run election ads[217] or use political

---

[212] *Reach People Who Visited Your Site Or App*, GOOGLE ADS HELP, https://support.google.com/google-ads/topic/3122874?hl=en&ref_topic=3121935 (last accessed July 21, 2020).
[213] *About Customer Match*, GOOGLE ADS HELP, https://support.google.com/google-ads/answer/6379332?hl=en&ref_topic=6296507 (last accessed July 21, 2020).
[214] "Disclosure Requirements for Election Advertising", *Political Content*, GOOGLE ADVERT. POLICIES HELP, https://support.google.com/adspolicy/answer/6014595#701 (last accessed July 21, 2020).
[215] FEDERAL ELECTION COMMISSION, ADVERT. AND DISCLAIMERS, https://www.fec.gov/help-candidates-and-committees/making-disbursements/advertising/ (last accessed Jul. 19, 2020).
[216] *2020 Outside Spending, by Race*, OPENSECRETS.ORG, https://www.opensecrets.org/outsidespending/summ.php?disp=R (last visited June 21, 2020).
[217] "Election Ads", *Political Content*, GOOGLE ADVERT. POLICIES HELP, https://support.google.com/adspolicy/answer/6014595 (last accessed July 21, 2020).

affiliation in personalized advertising[218] in the United States must go through a verification process, which is required for all ad formats/extensions, and all personalized ads features. Political advertisers must provide a Federal Election Commission (FEC) ID and either an Employer Identification Number (EIN) (for organizations) or Social Security Number (for individuals). Google collects such data and makes available a transparency report on political ad spending by each advertiser/campaign. The transparency report lists top advertisers and the amount of political ad spending by each advertiser. A recent transparency report (as of June 6, 2020) provides this list of top political ad spending since May 31, 2018:[219]

Advertiser	Total ad spend
MIKE BLOOMBERG 2020 INC	$62,263,700
TRUMP MAKE AMERICA GREAT AGAIN COMMITTEE	$23,277,300
BIDEN FOR PRESIDENT	$12,045,800
DONALD J. TRUMP FOR PRESIDENT, INC.	$9,081,000
TOM STEYER 2020	$8,854,400
BERNIE 2020	$8,709,900
PETE FOR AMERICA, INC.	$7,183,000
SENATE LEADERSHIP FUND	$5,023,200
CONSERVATIVE BUZZ LLC	$4,959,600
WARREN FOR PRESIDENT, INC.	$4,872,200
NRCC	$4,480,100
CONGRESSIONAL LEADERSHIP FUND	$4,203,200
PRIORITIES USA	$4,023,800
Need to Impeach	$4,015,100
NRSC	$3,582,000

Google's Regulation of Falsity and Misleading Content in Political Ads

Google also recently revised its rules about truth in advertising to prohibit ads with "demonstrably false claims that could significantly undermine participation or trust" in elections.[220] Google has stated, however, that by reframing these truth-in-advertising rules,

[218] "Political Affiliation in Personalized Advertising", *Political Content*, GOOGLE ADVERT. POLICIES HELP, https://support.google.com/adspolicy/answer/143465?#533 (last accessed July 21, 2020).
[219] *Political Advertising in the United States*, GOOGLE TRANSPARENCY REPORT, https://transparencyreport.google.com/political-ads/region/US?hl=en (last accessed July 6, 2020).
[220] Scott Spencer, *An Update on Our Political Ads Policy*, GOOGLE BLOG: THE KEYWORD (Nov. 20, 2019), https://www.blog.google/technology/ads/update-our-political-ads-

it does not intend to appoint itself as the arbiter of truth in politics. Google explains that since "no one can sensibly adjudicate every political claim, counterclaim, and insinuation," it will focus its efforts on claims that are something more than generic falsehood or exaggeration. It will not take comprehensive action against every misleading political ad but will do so for "clear violations." That line will likely be difficult to define and maintain. In its announcement, Google gives the example of "deep fakes" as the type of content that it will now remove. These are addressed by Google's policy prohibiting "manipulating media to deceive, defraud, or mislead others." The example the company provides is "deceptively doctoring media related to politics, social issues, or matters of public concern."[221] Google has also released an open-source database containing 3,000 manipulated videos in order to help identify and target deepfakes.[222]

It is as yet unclear what falls within the category of demonstrably false political ads according to Google,[223] but a few examples provide some guidance. When YouTube CEO Susan Wojcicki was asked whether YouTube would remove President Trump's advertisement (which he placed on Facebook) falsely accusing Joe Biden of corruptly sheltering his son from a Ukrainian investigation through bribery, Wojcicki explained that this ad "would not be a violation of our policies" because "politicians are always accusing their opponents of lying."[224] On the other hand, Wojcicki cited the (now infamous) video that showed Nancy Pelosi speaking at an artificially reduced rate, which made Pelosi

---

policy/; *Misrepresentation*, GOOGLE ADVERT. POLICIES HELP, https://support.google.com/adspolicy/answer/6020955?hl=en (last accessed July 21, 2020).

[221] *Misrepresentation*, GOOGLE ADVERT. POLICIES HELP, https://support.google.com/adspolicy/answer/6020955?hl=en (last accessed July 21, 2020).

[222] Karen Hao, *Google Has Released a Giant Database of Deepfakes to Help Fight Deepfakes*, MIT TECH. REV. (Sept. 25, 2019), https://www.technologyreview.com/f/614426/google-has-released-a-giant-database-of-deepfakes-to-help-fight-deepfakes/; see also Nick Dufour and Andrew Gully, *Contributing Data to Deepfake Detection Research*, GOOGLE AI BLOG (Sept. 24, 2019), https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html.

[223] Google and YouTube have removed over 300 Trump ads in the last half of 2019, but the archive in which removed ads are listed does not indicate why specific ads were removed. Shachar Bar-On & Natalie Jimenez Peel, *300+ Trump Ads Taken Down by Google, YouTube*, CBS NEWS: 60 MINUTES OVERTIME (Dec. 1, 2019), https://www.cbsnews.com/news/300-trump-ads-taken-down-by-google-youtube-60-minutes-2019-12-01/.

[224] Lesley Stahl, *How Does YouTube Handle the Site's Misinformation, Conspiracy Theories, and Hate?*, CBS NEWS: 60 MINUTES (Dec. 1, 2019), https://www.cbsnews.com/news/is-youtube-doing-enough-to-fight-hate-speech-and-conspiracy-theories-60-minutes-2019-12-01/.

appear to be drunk. Wojcicki noted that that video was removed "very fast" because "it's not okay to have technically manipulated content that would be misleading."[225]

With respect to manipulated media in particular, YouTube has adopted specifically applicable policies.[226] Its deceptive practices policies state that "[C]ontent that has been technically manipulated or doctored in a way that misleads users (beyond clips taken out of context) and may pose a serious risk of egregious harm" are prohibited and will be removed from YouTube.[227] YouTube has further stated that it will remove content that attempts to mislead people about the voting process or any other false information relating to elections.[228] YouTube also recently created an Intelligence Desk to help review technically-manipulated content and take proactive approaches to mitigate the spread of such content,[229] and the company has also changed its recommendations systems to prevent people from viewing misinformation on its site.[230]

III. Analysis and Assessment of Platforms' Measures to Combat Medical and Political Misinformation

The efforts undertaken by the major social media platforms' measures to address medical and political misinformation are not without their problems. These efforts, however, are generally consistent with First Amendment substantive and procedural values, are trending in the right direction, and are by and large welcomed by the American public. The platforms' efforts are not subject to First Amendment scrutiny, since the platforms are not state actors.[231] On the contrary, the platforms enjoy great discretion with respect to the choices they make regarding content regulation on their platforms, thanks to Section 230 of the Communications Decency Act (at least for now).[232] That said, the

---

[225] *Id.*
[226] *How YouTube Supports Elections*, YouTube Official Blog (Feb. 3, 2020), https://youtube.googleblog.com/2020/02/how-youtube-supports-elections.html.
[227] *Id.*
[228] *Id.*
[229] *Id.*
[230] *Id.*
[231] See, e.g., Dawn C. Nunziato, Virtual Freedom: Net Neutrality and Free Speech in the Internet Age (2009).
[232] The Communications Decency Act Section 230 prohibits any attempt to hold social media platforms liable for hosting harmful speech or for taking steps to remove harmful speech. 47 U.S.C. § 230(c) (2019). Section 230(c)(1) of the Act provides that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider." Courts have consistently interpreted this provision to immunize social media platforms from liability for hosting a variety of categories of harmful speech, including causes of action such as defamation, negligence, gross negligence, nuisance, sending threatening messages, and even statutory violations of the Fair Housing Act and related anti-discrimination violations. In addition, the "good Samaritan" provision of Section 230 immunizes platforms from liability for undertaking measures to screen or block content

measures that the platforms have undertaken to combat misinformation have been largely consistent with First Amendment substantive and procedural values.

First, the platforms' most interventionist efforts with respect to false medical misinformation and false/misleading statements of fact in the health and medical context are consistent with First Amendment substantive values, in which lesser protection is accorded for false and misleading statements of fact (especially in the medical field). While the marketplace of ideas theory (and its default response of counterspeech as a remedy for bad speech) accords broad protection to good and bad *ideas*, it does not accord the same broad protections to good and bad claims or assertions of *fact*. The Supreme Court in embracing the marketplace of ideas theory has made clear that there is no such thing as a false *idea*—that all *ideas* are protected—but that false statements of *fact* are not similarly immune from regulation. While the Court has sometimes recognized the minimal potential contributions to the marketplace of ideas made by harmless lies[233] or some false statements

---

on their platforms, providing that platforms cannot "be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene . . . excessively violent, harassing, or otherwise objectionable . . . ." 47 U.S.C. § 230(c)(2)(A). President Trump has recently taken aim at Section 230. See

https://www.whitehouse.gov/presidential-actions/executive-order-preventing-online-censorship/

https://thehill.com/policy/technology/499871-trump-to-order-review-of-law-protecting-social-media-from-responsibility

https://www.techdirt.com/articles/20200714/23061044903/house-government-appropriations-bill-would-bar-ftc-fcc-doing-anything-related-to-trumps-inane-anti-230-executive-order.shtml

[233] *See United States v. Alvarez*, 567 U.S. 709, 732 (2012). In *United States v. Alvarez*, 567 U.S. 709 (2012), the Supreme Court, in a 6-3 decision, struck down a portion of the Stolen Valor Act, a federal law that criminalized the making of false statements about having a military medal. The Act made it a misdemeanor to falsely represent oneself as having received any U.S. military decoration or medal and provided for prison terms up to six months (and up to one year if the subject of such lies was the Medal of Honor). In a challenge brought by Xavier Alvarez, who was convicted under the Act for publicly lying about receiving the Congressional Medal of Honor, the Court struck down the Stolen Valor Act on First Amendment grounds. Justice Kennedy, writing for a plurality, held that harmless false statements are not, by the sole reason of their falsity, excluded from First Amendment protection. *See also* Breyer, J., concurring in judgment, arguing that when Alvarez posed as a military medal recipient, this was a seemingly harmless lie, since this did not hurt anyone and was a lie that could be easily remedies by counterspeech – i.e., if a list of medal recipients were made available on the Internet.

of fact,[234] it has also emphasized that the First Amendment does not stand in the way of regulating intentionally false or misleading harmful assertions of fact,[235] especially in the medical context. Indeed, with regard to false and misleading statements of fact regarding medical treatments, cures, medicine, etc., the Food and Drug Administration and the Federal Trade Commission have extensive authority, consistent with the First Amendment, to prohibit false and misleading claims. The FDA and the FTC are empowered to prohibit the false or misleading branding, advertising, marketing, and/or sale of products – including products that claim to be cures or treatments for COVID-19 – and these agencies have recently cracked down on online purveyors of such products.[236] Thus, it is not inconsistent with First Amendment values for the social media platforms to undertake

---

[234] *See New York Times v. Sullivan*, 376 U.S. 254 (1964).

[235] *See Gertz v. Welch*, 418 U.S. 323, 340 (1974): "[T]here is no constitutional value in false statements of fact. Neither the intentional lie nor the careless error materially advances society's interest in uninhibited, robust, and wide-open debate on public issues." (citations omitted).

[236]The FDA has authority to regulate purveyors of such products on the grounds that they "misleadingly" represent that their products as safe and effective for the treatment or prevention of COVID-19, and that the products are therefore illegal unapproved and misbranded products under Section 502 of the Food Drug & Cosmetic Act (the "FD&C Act"). *See* Alexandra Sternlicht, *The FTC Has Sent Cease-And-Desist Letters to Over 150 Companies Who Claim to Have COVID-19 Cures*, FORBES (July 9, 2020), https://www.forbes.com/sites/alexandrasternlicht/2020/07/09/the-ftc-has-sent-cease-and-desist-letters-to-over-150-companies-who-claim-to-have-COVID-19-cures/#34ef5282722e (FDA has sent warning letters to over 150 companies who claim to have COVID-19 cures); Meagan Flynn, *Leader of Fake Church Peddling Bleach as COVID-19 Cure Sought Trump's Support. Instead, He Got Federal Charges*., WASH. POST (July 9, 2020), https://www.washingtonpost.com/nation/2020/07/09/fake-coronavirus-cure-bleach/ (criminal charges for conspiracy to defraud the United States and deliver misbranded drugs brought against fake Florida church that claims to have COVID-19 cures). The FDA has authority to regulate purveyors of such products on the grounds that they "misleadingly" represent that their products as safe and effective for the treatment or prevention of COVID-19, and that the products are therefore illegal unapproved and misbranded products under Section 502 of the Food Drug & Cosmetic Act (the "FD&C Act"). In addition, by marketing products as intended to mitigate, prevent, treat, diagnose or cure COVID-19 in people, the products were deemed "drugs" as defined under the FD&C Act, *see* 21 USC § 321(g)(1), and – absent advance review and approval by the FDA – these products were deemed "unapproved new drugs" sold in violation of Section 505(a), 301(a) and (d) of the FD&C Act. In addition, under the Federal Trade Commission Act (the "FTC Act"), 15 U.S.C. 41 et seq., it is unlawful to advertise that a product can prevent, treat, or cure human disease unless the purveyor possesses competent and reliable scientific evidence (including, for example, well-controlled human clinical studies) substantiating that the claims are true at the time they are made. Accordingly, to make or exaggerate such claims without scientific evidence sufficient to substantiate them violates the FTC Act.

measures to combat false and misleading statements of fact, especially in the area of medical and health related information.[237]

In addition, the platforms' efforts to remove content likely to incite violence or great public harm is consistent with the emergency exception in First Amendment jurisprudence, as originally articulated by Holmes and Brandeis[238] and as recognized by the Court in its incitement jurisprudence in *Brandenburg v. Ohio* and progeny. Content that is created or shared with the purpose of immediately contributing to or exacerbating violence or physical harm is generally subject to regulation under the First Amendment's incitement jurisprudence, under which the government is permitted to regulate "advocacy of … law violation . . . where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action."[239]

Further, the platforms' efforts to label less harmful false and misleading medical information and to develop and refer users to accurate information revolves primarily around providing *counterspeech* instead of implementing censorship as a remedy. This is consistent with First Amendment substantive values and with the marketplace of ideas theory of the First Amendment, according to which – ever since the formative years of modern First Amendment jurisprudence – the accepted response to bad speech is not censorship but more (better) speech. [240] As Justice Brandeis explained in his oft-quoted concurrence in *Whitney v. California,* joined by Justice Holmes: "If there be time to expose through discussion the falsehood and fallacies [of speech], to avert the evil by the process of education, the remedy to be applied is more speech, not enforced silence."[241] According to the marketplace theory of the First Amendment, ideas should generally be allowed to compete freely in the marketplace unfettered by government restrictions (absent emergency conditions). The default remedy for harmful ideas in the marketplace of speech is not censorship, but counterspeech, which operates by allowing those who are exposed to bad speech to be exposed to good speech as a counterweight. The platforms' efforts to respond to false and misleading medical and political information by labeling them as such and to refer users to accurate information is consistent with this counterspeech approach in First Amendment jurisprudence. In addition, the platforms' efforts in regulating misinformation in political speech and political advertising contribute toward "producing an informed public capable of conducting its own affairs" and facilitating the preconditions necessary for citizens to engage in the task of democratic self-government,[242] which are also foundational goals of our First Amendment jurisprudence.

---

[237] Id.
[238] See text accompanying notes x – y.
[239] *Brandenburg v. Ohio,* 395 U.S. 444, 447 (1969).
[240] *Abrams,* 250 U.S. at 630. As Holmes explained in his Abrams dissent, "[o]nly the emergency that makes it immediately dangerous to leave the correction of evil counsels to time warrants making an exception to the sweeping command, 'Congress shall make no law . . . abridging the freedom of speech.'" *Id.* at 630–31.
[241] *Whitney v. California,* 274 U.S. 357, 375–77 (1927) (Brandeis, J., concurring) (emphasis added).
[242] *Red Lion Broad. Co.,* 395 U.S. at 392.

The platforms' efforts are also generally consistent with First Amendment *procedural* values and with principles of due process generally.[243] The First Amendment's protections for freedom of expression not only embody a substantive dimension of which categories of speech to protect; they also embody procedural dimensions, imported from the Due Process Clause, which require that "sensitive tools" be implemented by decisionmakers in restricting speech.[244] As free speech theorist Henry Monaghan explains, "procedural guarantees play an equally large role in protecting freedom of speech; indeed, they assume an importance fully as great as the validity of the substantive rule of law to be applied."[245] Accordingly, First Amendment jurisprudence incorporates a powerful "body of procedural law which defines the manner in which [decisionmakers] must evaluate and resolve [free speech] claims — [establishing] a First Amendment due process."[246] This jurisprudence embodies "a comprehensive system of procedural safeguards designed to obviate the dangers of a censorship system."[247] Consistent with these procedural safeguards embodied in First Amendment jurisprudence, social media platforms should impose speech restrictions on medical and political misinformation in a clear, neutral, and transparent manner, such that speakers are adequately and clearly informed of the platforms' rules regarding speech, speakers are specifically informed of the reasons why their speech was restricted (removed or labeled), such decisions are made consistently by impartial decisionmakers, and speakers have an opportunity to be heard/to appeal any such speech restrictions. In general, the platforms have provided clear notice to users of their (evolving) terms of service regarding medical and political misinformation and have provided users with clear notice when implementing speech removal or labeling decisions. For example, as discussed above, when Twitter restricted Donald Trump, Jr.'s posts embodying false claims about and unproven cures for COVID-19 on the grounds that the post violated Twitter's rules regarding medical misinformation,[248] it did so in the context of providing clear prior notice of what speech was restricted and a process to appeal Twitter's decisions,[249] and it also provided notice to Trump, Jr., of the specific reason why his speech was restricted. See below.

---

[243] *See*, e.g., Dawn Carla Nunziato, How (Not) To Censor: Procedural First Amendment Values and Internet Censorship Worldwide, 42 Geo. J. Int'l L. 1123 (2014); Dawn Carla Nunziato, Forget About It? Harmonizing European and American Protections for Privacy, Free Speech, and Due Process, Privacy and Power (Cambridge University Press 2017).

[244] *Bantam Books v. Sullivan*, 372 U.S. 58, 66 (1963).

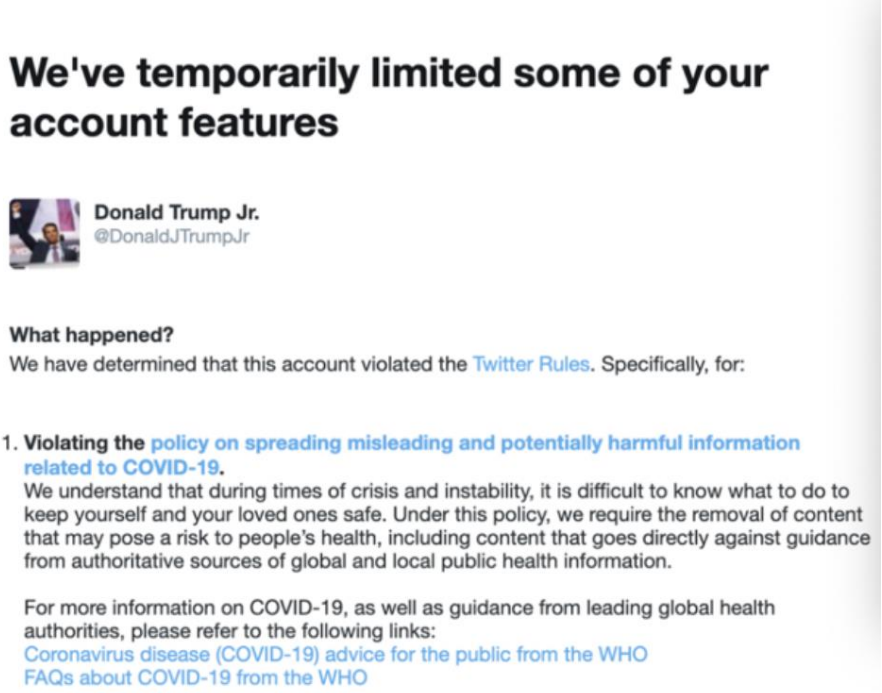
[245] Henry Monaghan, *First Amendment "Due Process,"* 83 HARV. L. REV. 518 (1970) (internal quotations omitted).

[246] *Id.*

[247] *Id*. (internal quotations omitted).

[248] See text accompanying notes x – y.

[249] See https://help.twitter.com/forms/general?subtopic=suspended (setting forth the procedural for users to appeal severe violations of Twitter's rules resulting in suspending and/or blocked accounts).



We've temporarily limited some of your account features

Donald Trump Jr.
@DonaldJTrumpJr

What happened?
We have determined that this account violated the Twitter Rules. Specifically, for:

1. Violating the policy on spreading misleading and potentially harmful information related to COVID-19.
We understand that during times of crisis and instability, it is difficult to know what to do to keep yourself and your loved ones safe. Under this policy, we require the removal of content that may pose a risk to people's health, including content that goes directly against guidance from authoritative sources of global and local public health information.

For more information on COVID-19, as well as guidance from leading global health authorities, please refer to the following links:
Coronavirus disease (COVID-19) advice for the public from the WHO
FAQs about COVID-19 from the WHO

In short, the extensive measures undertaken by the major social media platforms to respond to false and misleading misinformation in the medical and political contexts are generally consistent with our First Amendment substantive and procedural values.

In addition, recent studies have shown that the efforts undertaken by the major social media platforms' measures to address political and medical misinformation have been moderately successful. As Hunt Allcott and his co-authors report in their article *Trends in the Diffusion of Misinformation on Social Media*, based on their study of "trends in the diffusion of content from 570 fake news websites and 10,240 fake news stories on Facebook and Twitter between January 2015 and July 2018," while "[u]ser interactions with false content rose steadily on . . . Facebook . . . through the end of 2016," since then, "interactions with false content have fallen sharply."[250] The authors of the study find that "user interaction with known false news sites has declined by 50 percent since the 2016 election."[251] Based on these findings, the authors conclude that "efforts by Facebook following the 2016 election to limit the diffusion of misinformation [namely, the "suite of policy and algorithmic changes made by Facebook following the 2016 election"[252]] may have had a meaningful impact."[253]

[250]Hunt Allcott, Matthew Gentzkow & Chuan Yu, *Trends in the Diffusion of Misinformation on Social Media* 1 (Stanford Institute for Economic Policy Research, Working Paper No. 18-029, 2018), https://web.stanford.edu/~gentzkow/research/fake-news-trends.pdf.
[251]Allcott, Gentzkow & Yu, supra note x, at 5.
[252] *Id.* at 6.
[253] *Id.* at 3.

Further, the labeling of content as false or misleading on social media platforms has been shown to be effective in limiting the dissemination of such false or misleading content. According to a recent study, social media users were about 50% less likely to share false stories if the stories had been labeled as false. With no labels being used at all, participants considered sharing 29.8 percent of false stories in the sample, but that figure dropped to 16.1 percent of false stories that had a label attached.[254] In addition, the labeling of posts as false led to improved accuracy in social media users' beliefs. Researchers found, in an exhaustive series of surveys across more than 10,000 participants on a wide range of topics, that 60% of respondents gave accurate answers when presented with a fact-check/correction, while only 32% expressed accurate beliefs when they were not presented with a fact-check/correction.[255]

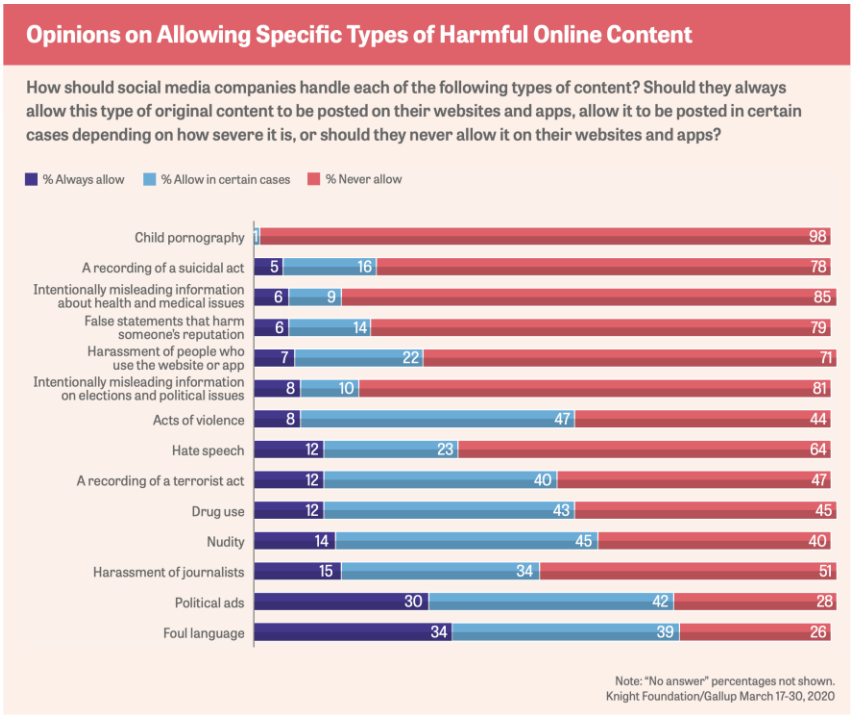
Finally, there is broad public support among Americans for social media platforms' continuing to take a meaningful role in combating political and medical misinformation on their platforms. A March 2020 Knight Foundation/Gallup Poll found that the vast majority of Americans surveyed (81%) supported the removal of intentionally misleading information on elections or other political issues, and an even greater majority of Americans surveyed (85%) supported social media companies' removal of intentionally misleading health information.[256]

---

[254] *See* Peter Dizikes, *The Catch to Putting Warning Labels on Fake News*, MIT NEWS (Mar. 2, 2020), http://news.mit.edu/2020/warning-labels-fake-news-trustworthy-0303.

[255] Lee Drutman, *Fact-Checking Misinformation Can Work. But It Might Not Be Enough.*, FIVETHIRTYEIGHT (June 4, 2020), https://fivethirtyeight.com/features/why-twitters-fact-check-of-trump-might-not-be-enough-to-combat-misinformation. The political scientists conducting the surveys, Ethan Porter and Thomas J. Wood, found that the most effective fact-checks shared four characteristics: they were from a highly credible source, they offered a new frame for the issue rather than merely calling the misinformation "wrong," they didn't directly challenge a worldview or identity, and they happened before a false narrative could gain traction. Id.

[256] *Free Expression, Harmful Speech and Censorship in a Digital World,* Knight Foundation and Gallup (2020), https://knightfoundation.org/wp-content/uploads/2020/06/KnightFoundation_Panel6-Techlash2_rprt_061220-v2_es-1.pdf.



CONCLUSION

Social media platforms are playing an ever-expanding role in shaping the contours of information ecosystem today, as these platforms have shouldered the burden of ensuring that the public is informed – and not misinformed – about matters affecting our democratic institutions in the context of our elections, as well as about matters affecting our very health and lives in the context of the pandemic. The platforms are attempting to address these serious problems alone, in the absence of federal or state regulation or guidance in the United States. While the platforms' intervention in the online marketplace of ideas is not without its problems, this Article has argued that this intervention is by and large a salutary development and is one that has brought about improvements in the online information ecosystem. Social media companies have been generally inspired by First Amendment free speech values – both substantive and procedural – to protect a vibrant marketplace of ideas online while imposing limited, moderately effective checks on harmful false and misleading speech, with complex systems of removal, fact-checking, and labeling, and by serving up prominent information from independent fact-checkers and trusted authorities to counter medical and political misinformation. In the absence of effective regulatory measures in the United States to combat medical and political misinformation online, social media companies should be commended for their efforts thus far and should continue to develop and deploy even more successful measures to combat such misinformation online.