OPEN TECHNOLOGY INSTITUTE

NEW AMERICA

June 24, 2020

The Honorable Jan Schakowsky
Chair, Subcommittee on Consumer
Protection & Commerce
House Committee on Energy and Commerce
2125 Rayburn House Office Building
Washington, DC 20515

The Honorable Cathy McMorris Rodgers
Ranking Member, Subcommittee on Consumer
Protection & Commerce
House Committee on Energy and Commerce
2125 Rayburn House Office Building
Washington, DC 20515

The Honorable Michael F. Doyle
Chair, Subcommittee on
Communications and Technology
House Committee on Energy and Commerce
2125 Rayburn House Office Building
Washington, DC 20515

The Honorable Robert E. Latta
Ranking Member, Subcommittee on
Communications and Technology
House Committee on Energy and Commerce
2125 Rayburn House Office Building
Washington, DC 20515

Dear Representatives Schakowsky, Doyle, McMorris Rodgers, and Latta:

New America's Open Technology Institute (OTI) appreciates the opportunity to submit a statement for the record for the joint hearing entitled, "A Country in Crisis: How Disinformation Online is Dividing the Nation," being held by the Subcommittee on Consumer Protection and Commerce and the Subcommittee on Communications and Technology. OTI works at the intersection of technology and policy to ensure that every community has equitable access to digital technologies that are open and secure, and their benefits. We support and defend the right to privacy and freedom of expression, and press internet platforms to provide greater transparency and accountability around their operations, technologies, and impacts. On June 1, OTI released a report which outlines how eight internet platforms have responded to the spread of COVID-19 misinformation and disinformation and which makes recommendations on how internet platforms can provide greater transparency around these efforts, and how policymakers can encourage greater accountability from these platforms.[1]

The recent wave of protests across the nation calling for racial justice and equity in the United States, as well as the spread of the novel coronavirus across the world have demonstrated how misinformation and disinformation online can yield harmful real-life, offline consequences. Specifically, protests have become fertile grounds for disinformation with fake accounts propping up to spread false information and spread exaggerated accounts to paint protestors in a negative light.[2] Misinformation and disinformation around COVID-19 have disproportionately impacted communities of color and already marginalized

---

[1] Spandana Singh and Koustubh "K.J." Bagchi, *How Internet Platforms Are Combating Disinformation and Misinformation in the Age of COVID-19*, June 1, 2020, https://www.newamerica.org/oti/reports/how-internet-platforms-are-combating-disinformation-and-misinformation-age-covid-19.

[2] *Morning Edition*, "'None Of This Is True': Protests Become Fertile Ground for Online Disinformation," performed by Martin Austermuhle and Miles Parks, aired June 1, 2020, on NPR, https://www.npr.org/2020/06/01/867137863/none-of-this-is-true-protests-become-fertile-ground-for-online-disinformation.

groups. For example, an article widely shared on Facebook earlier this year claimed that people of color may be immune to COVID-19, as they have melanin.[3] Additionally, a Pew study indicated that the false belief that the coronavirus was created in a lab was more prevalent among Hispanic and Black people than among whites.[4] The same study concluded that educational attainment is also a factor in being susceptible to misinformation.[5] Specifically, those with a bachelor's degree or more education were less likely than those with a high school diploma or less education to say the coronavirus was created in a lab. Finally, a team of BBC reporters examined the impact of misinformation on a global scale and tracked coronavirus misinformation. Their review found links to assaults, arsons and deaths.[6]

Participation in civic engagement has also become impacted by the proliferation of misinformation and disinformation. As primary and general elections continue to be held throughout the country, disinformation related to voter suppression and electoral issues more broadly is also on the rise, posing significant threats to political discourse and democracy in the United States. For example, an organization named Judicial Watch made online claims that the U.S. Census Bureau has been hiring "serious criminals" including sex offenders "to enter the homes of unsuspecting Americans to gather statistics" and help conduct the 2020 Census.[7] Another misleading claim around the Census stated that information provided on the Census would be shared with the government—a strong concern for many marginalized communities.[8] The scope and potential impact of Census misinformation and disinformation campaigns has become so influential, the U.S. Census Bureau established its own Trust & Safety team to combat these efforts.[9]

## How Internet Platforms Respond to Misinformation and Disinformation Online

Since the 2016 U.S. presidential elections, internet platforms have come to recognize the severe influence and impacts that online misinformation and disinformation can have on the strength and nature of democracy and discourse. In response to the growing presence of organized hate groups and harmful content online, most internet platforms have developed and instituted content policies which prohibit

---

[3] Tom Kertscher, "Melanin Doesn't Protect Against Coronavirus," PolitiFact, last modified March 10, 2020, https://www.politifact.com/factchecks/2020/mar/10/facebook-posts/melanin-doesnt-protect-against-coronavirus/.

[4] Katherine Schaeffer, "Nearly Three-In-Ten Americans Believe COVID-19 Was Made In A Lab," Pew Research Center, last modified April 8, 2020, https://www.pewresearch.org/fact-tank/2020/04/08/nearly-three-in-ten-americans-believe-covid-19-was-made-in-a-lab/.

[5] Schaeffer, "Nearly Three-In-Ten," Pew Research Center.

[6] Marianna Spring, "Coronavirus: The Human Cost of Virus Misinformation," *BBC*, May 27, 2020, https://www.bbc.com/news/stories-52731624.

[7] Judicial Watch, "Census Bureau Still Hiring Felons, Child Sex Offenders Pass Background Check - Judicial Watch," Facebook, December 29, 2019, https://www.facebook.com/JudicialWatch/posts/10157456842846943.

[8] *Weekend Edition Sunday*, "Combating Fear And Disinformation Regarding The 2020 Census," hosted by Lulu Garcia-Navarro, performed by Jeri Green, aired March 8, 2020, on NPR, https://www.npr.org/2020/03/08/813384375/combating-fear-and-disinformation-regarding-the-2020-census.

[9] Zack Schwartz, "Census Partners With Social Media Platforms, Community Organizations, the Public to Stop Spread of False Information," United States Census Bureau, last modified February 10, 2020, https://www.census.gov/library/stories/2020/02/putting-2020-census-rumors-to-rest.html.

violence-inciting speech and hate speech.[10] These categories of speech are not only harmful, they can also support ongoing voter suppression and disinformation efforts.[11] However, outside the context of violence-inciting hate speech, internet platforms have adopted a broad range of approaches to address these issues which include removing misleading content, reducing the spread or distribution of such content, and promoting accurate and verified information on topics such as the COVID-19 pandemic and voting.

Facebook, for example, outlines in its Community Standards that it does not remove false news from the platform. Instead, the company partners with a series of independent fact-checking organizations to evaluate different pieces of content. If content is deemed misleading, the company will reduce the spread and distribution of the content on the platform so that this content appears lower in a user's News Feed.[12] In addition, Pages, domains, and Groups that continuously share misinformation will also be demoted on the platform, and such Pages may see restrictions around their ability to advertise and monetize.[13] However, Facebook has instituted some exceptions to this approach.

Facebook has taken a more active approach during the COVID-19 pandemic, and has begun removing misleading content related to the pandemic that can pose "imminent physical harm" to users.[14] This is a notable divergence from Facebook's policies on other forms of health misinformation, such as misleading information about vaccines, which outlined that such content would be demoted rather than removed.[15] The company has also stated that it will append warning labels to certain debunked posts related to the pandemic, and it will also alert users if they have engaged with or viewed harmful misleading content related to COVID-19 that has been debunked by its fact-checking partners. The company will also connect users who have liked, reacted to, or commented on these posts to additional, legitimate information sources.[16]

Similarly, Facebook has modified its regular approach with regard to moderation of voting related information, and has stated that it will remove content that misrepresents the dates, locations, times, methods, and qualifications for voting or voter registration, as well as content that includes threats of violence related to voting, voter registration, or the outcome of an election.[17] However, critics,

---

[10] Spandana Singh, *The Transparency Report Tracking Tool: How Internet Platforms Are Reporting on the Enforcement of Their Content Rules*, June 17th, 2020, https://www.newamerica.org/oti/reports/transparency-report-tracking-tool/.

[11] Rosen et al., "Helping to Protect," Facebook Newsroom.

[12] Facebook, "Community Standards: False News," Facebook, https://www.facebook.com/communitystandards/false_news.

[13] Guy Rosen et al., "Helping to Protect the 2020 US Elections," Facebook Newsroom, last modified October 21, 2019, https://about.fb.com/news/2019/10/update-on-election-integrity-efforts/.

[14] Mark Zuckerberg, "I want to share an update on the work we're doing to connect people with accurate information and limit the spread of misinformation about Covid-19.," Facebook, April 16, 2020, 9:01 am, https://www.facebook.com/zuck/posts/10111806366438811

[15] Stephie Grob Plante, "Facebook Says It Will Crack Down On Health Misinformation," Vox, last modified July 8, 2019, https://www.vox.com/the-goods/2019/7/8/20686737/facebook-groups-news-feed-anti-vaxx-health.

[16] Singh and Bagchi, *How Internet*.

[17] Rosen et al., "Helping to Protect," Facebook Newsroom.

particularly civil rights organizations, have pressed the company to do more, as these efforts have not been sufficient to undermine the spread of voting-related disinformation.[18]

In addition, when it comes to political candidates, elected individuals, and other government officials, Facebook has refused to play an intervening role. As a result, the company does not typically remove content posted by such individuals, even if the content violates their content policies. The company has shared that there are exceptions to this approach when it comes to advertising,[19] and when the content causes imminent risk of harm or danger that are covered by Facebook's policies.[20]

Twitter's approach to misinformation and disinformation differs from Facebook's in some important ways. The company has implemented a Civic Integrity policy that prohibits individuals from using the platform to engage in manipulative behaviors related to elections and other civic processes such as censuses and major ballot initiatives. The company's Civic Integrity policy penalizes individuals for sharing misleading information around when, where, or how to participate in a civic process, and for attempting to suppress participation, including through intimidation, in these processes more broadly.[21] In response to violations of this policy, the company will determine the severity of the violation and may subsequently remove the Tweets in question, and temporarily or permanently suspend accounts.[22]

In addition, Twitter has adopted a public interest exception policy for cases in which elected world leaders and other government officials violate some of their content policies, including policies related to COVID-19 content guidelines and policies related to promoting violence.[23] In these cases, the company determines whether there is public interest value in keeping the content on the service, such as enabling the public to know that these leaders are publishing misinformation. Then, instead of removing the content, Twitter will place the content behind a notice that provides context about the violation, but still allow users to view the content if they wish to see it.[24] This approach provides a middle course through which the platform takes action against posts that would otherwise violate their policies, but that recognizes that users have a right to access information, including from world leaders and other

---

[18] "Letter Urging Facebook to Protect Civil Rights," The Leadership Conference on Civil and Human Rights, last modified October 21, 2019, https://civilrights.org/resource/letter-urging-facebook-to-protect-civil-rights/.

[19] Adi Robertson, "Facebook Says It Won't Remove Politicians' Posts For Breaking Its Rules," *The Verge*, September 24, 2019, https://www.theverge.com/2019/9/24/20882401/facebook-political-posts-fact-checking-nick-clegg-speech .

[20] Mark Zuckerberg, "This has been an incredibly tough week after a string of tough weeks.," Facebook, May 29, 2020, 7:19 PM, https://www.facebook.com/zuck/posts/this-has-been-an-incredibly-tough-week-after-a-string-of-tough-weeks-the-killing/10111961824369871/.

[21] Twitter, "Civic Integrity Policy," Twitter Help Center, https://help.twitter.com/en/rules-and-policies/election-integrity-policy.

[22] Twitter, "Civic Integrity," Twitter Help Center.

[23] Twitter, "About public-interest exceptions on Twitter," Twitter Help Center, https://help.twitter.com/en/rules-and-policies/public-interest.

[24] "World Leaders On Twitter: Principles & Approach," Twitter Blog, last modified October 15, 2019, https://blog.twitter.com/en_us/topics/company/2019/worldleaders2019.html.
Twitter, "About Public-Interest Exceptions on Twitter," Twitter Help Center, https://help.twitter.com/en/rules-and-policies/public-interest.

government officials, as well as a right to know what their leaders are saying. This is especially during a crisis period such as this one.

Online platforms are a major outlet for information and as a result companies should institute such a public interest exception and notice policy, as Twitter currently does. However, companies should institute this policy responsibly. If a leader's content violates the platform's content policies, in most cases, it should be left up with a clear notice that explains why the content has been left up. In addition, the content should be fact-checked and the companies should provide additional information to users in the notice detailing whether the post contains misleading information. However, if the content posted by these leaders poses imminent harm, then platforms should remove this content just as they would for content from anyone else, as it can have significant offline consequences. In addition, this policy should be implemented consistently.

**Pushing Internet Platforms For More Transparency and Accountability**

Although many internet platforms have developed and instituted content policies that aim to address the expanding spread of misinformation and disinformation on their services, the effectiveness of these policies vary, as do their enforcement.

Currently, most internet platforms provide little transparency and accountability around how they create and implement their policies related to misinformation and disinformation, and what the impact of their efforts to combat these types of content are. For example, every quarter Facebook publishes a Community Standards Enforcement Report which outlines the scope and scale of its content moderation efforts on Facebook and Instagram. The report provides quantitative data on how prevalent certain types of content are on the platforms (e.g. child nudity and sexual exploitation of children, fake accounts, violent and graphic content) and includes data on how much of this content Facebook was able to proactively detect through its automated tools, and how much of the violating content in each category the company took action on.[25] Although the company has a number of policies outlining enforcement actions on content containing misinformation or disinformation, the company does not publish any data in its report which explains the results of these enforcement actions. This makes it difficult to understand the efficacy of the platforms' response to combat misinformation and disinformation. In the context of the pandemic, Facebook has shared some data related to the number of posts containing COVID-19 misinformation that the company has appended warning labels to.[26] However, this does not go far enough in painting an accurate and granular picture of how the company has responded to the spread of misinformation during this time.

---

[25] Facebook, "Community Standards Enforcement Report," Facebook, https://transparency.facebook.com/community-standards-enforcement.

[26] Guy Rosen, "An Update on Our Work to Keep People Informed and Limit Misinformation About COVID-19," Facebook Newsroom, last modified April 16, 2020, https://about.fb.com/news/2020/04/covid-19-misinfo-update/.

Similarly, Twitter does not provide information related to the moderation of misinformation and disinformation on its platform in its Twitter Rules Enforcement Report either.[27] In its Community Guidelines Enforcement Report, YouTube does provide some data on the number of videos, channels, and comments that have been removed for containing misleading information.[28] However, these data points are combined with data points related to spam and scams, thus making it difficult to understand how impactful the company's responses are. As elections are held throughout the year, transparency around corporate efforts to combat other types of disinformation, including campaigns related to voter suppression and voting will be of the utmost importance.

In addition, as outlined above, many internet platforms have policies which prohibit violence-inciting and hate speech content on their services. Most platforms, including Facebook, Twitter, and YouTube, publish data on their content moderation of these types of content in their transparency reports. However, the companies fail to provide adequate transparency around their use of automated tools when moderating such content.[29]

Automated tools are often unable to make accurate determinations when subjective definitions or contextual information is needed. Hate speech, for example, is not an easily defined category of content, and as a result automated tools have been found to be less accurate when reviewing these types of content. This can result in erroneous removals of content, which can negatively impact communities of color and already vulnerable groups.[30] If companies do not offer users with robust appeals processes, this leaves users with no method for remedy or redress.[31] For example, in 2017, Ijeoma Uluo, a Black activist and writer, spoke out after Facebook suspended her account after she posted screenshots of racist comments and threads she had received. Uluo's story is one of many stories about people of color, particularly Black people, who have had their posts removed and accounts suspended for speaking out against racism. As a result, civil rights groups have called on companies to improve their "racially biased" content moderation systems,[32] which are built off of human review and automated review processes. One key method for improvement cited by these experts was an effective appeals process.[33]

As OTI recommends in our report, going forward, there are a number of mechanisms and approaches that

---

[27] Twitter, "Twitter Rules Enforcement Report," Twitter Transparency Report, https://transparency.twitter.com/en/twitter-rules-enforcement.html.

[28] YouTube, "YouTube Community Guidelines Enforcement," Google Transparency Report, https://transparencyreport.google.com/youtube-policy/removals.

[29] Spandana Singh, "The Transparency Report Tracking Tool: How Internet Platforms Are Reporting on the Enforcement of Their Content Rules," New America's Open Technology Institute, last modified June 17, 2020, https://www.newamerica.org/oti/reports/transparency-report-tracking-tool/.

[30] Spandana Singh, *Everything in Moderation: An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User-Generated Content*, July 22, 2019, https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/.

[31] Singh, *Everything in Moderation*.

[32] Sam Levin, "Civil rights groups urge Facebook to fix 'racially biased' moderation system," *The Guardian,* January 18, 2017, https://www.theguardian.com/technology/2017/jan/18/facebook-moderation-racial-bias-black-lives-matter..

[33] Guynn, "Facebook Apologizes".

internet platforms can implement in order to provide greater transparency and accountability around their efforts to combat misinformation and disinformation on their services. In particular, companies should:

- Remove or reduce the spread of content that has been fact-checked and deemed to contain misinformation or disinformation.
- Publish a detailed description of all misinformation and disinformation-related content policies online, including examples of how these policies are enforced. Companies should also provide public notice if these policies change and should include an archive of past policies.
- Provide adequate notice to users who have engaged with misleading content in the past and direct them to authoritative sources of information
- Give users the opportunity to appeal moderation decisions. Given that many companies rely on automated tools to detect and remove content at scale, they should enable users to appeal moderation decisions which have resulted in the removal or suspension of their content and accounts. This appeals process should be timely and should enable users to provide additional information on the case and have their case reviewed by a new reviewer. Users who flag content and accounts should also have access to an appeals process.
- Institute a public interest exception and notice policy that enables companies to leave content that has been posted by world leaders and elected and other government officials on their services, even if the content has been fact-checked and deemed to contain misinformation. In such instances, the company should append a label to the content that provides additional context, including notice that the content has been fact-checked and contains misleading information. Companies should also direct users viewing such content to authoritative sources of information. However, where companies determine that content posted by such officials could result in imminent harm, they should not apply this public interest exception and notice policy, and should instead promptly remove the content as they would with any other user.
- Publish comprehensive transparency reports which include data on the scope and scale of their content moderation efforts related to misinformation and disinformation. These reports should break data points down based on the category of misinformation (e.g. health misinformation, voting-related misinformation), and should also outline the range of enforcement actions a company has taken against content (e.g. removing the content, demoting the content, labeling the content, etc.). As companies institute and expand appeals processes, the company should also publish data outlining the number of appeals requests the company received for each category of content, and the amount of content the company reinstated as a result of appeals. Companies should also provide data on the amount of content companies proactively recognized they erroneously moderated and subsequently reinstated. These data points provide valuable data on the effectiveness and accuracy of company moderation efforts.
- Partner with reputable fact-checking organizations and authoritative entities such as the WHO, CDC, and U.S. Census Bureau to verify or refute information circulated on these platforms, and promote reliable and legitimate sources of information.

In addition, although the U.S. government is limited in the extent to which it can direct platforms on how to decide what content to permit on their sites, the U.S. government can take certain steps to improve

accountability mechanisms from platforms and to support efforts to combat the spread of misinformation and disinformation. In particular:

- Policymakers should enact rules to require greater transparency from online platforms, including regular reporting regarding their content moderation and algorithmic-decision making processes.
- Government agencies and representatives should ensure that they are disseminating verified information and are not contributing to the spread of unproven or debunked information. These entities should also help debunk misleading claims and information using their own online accounts.
- Given the increase of misinformation and disinformation-fuelled discrimination, policymakers should clarify that all offline anti-discrimination statutes apply in the digital environment. Congress and state legislatures should also enact appropriate legislation where necessary in order to fill gaps or clarify the applicability of such laws.

**Conclusion:**

As the COVID-19 pandemic continues to upend our society and economy, and as the 2020 presidential election rapidly approaches, it is vital that the public is armed with accurate and reliable information. Misinformation and disinformation have the power to uproot democracy and democratic values, and internet platforms must do more to combat the spread of this harmful content. These companies must also provide greater transparency and accountability around these efforts, in order to ensure that they are safeguarding user rights and acting in the best interest of the public.

We are happy to discuss any of the above points further with you or your staff. Thank you again for the opportunity to provide this submission for the hearing record.

Sincerely,

Koustubh "K.J." Bagchi
Senior Policy Counsel

Spandana Singh
Policy Analyst