

Questions from Senator Blumenthal

1. During your testimony, you stated that Twitter will take steps to combat rampant disinformation on social media in Spanish-speaking sites.

- a. Who is in charge of setting Spanish-language content policy and making high-level decisions about enforcement, including about posts that come from high-ranking U.S. officials and the president?**
- b. How many U.S.-based Spanish language content moderators does Twitter have and what is the proportion of U.S.-based Spanish-language moderators to U.S.-based Spanish-language Twitter users? How does that compare to U.S.-based English-language moderators vs. U.S.-based English-language users?**

Twitter has taken numerous steps to combat election misinformation and provide voting information to Spanish-speaking communities. In advance of the election, when [searching](#) for key terms related to voter registration, individuals who used Twitter received prompts in English and Spanish pointing them to official resources. In addition, we trained voter education nonprofits and government partners on how to use Twitter tools and create content targeted at Spanish-language audiences. As part of this, we have worked with @USAGovEspanol, @NALEO, @HispanicFed, @MALDEF, @LULAC, and @WeAreUnidos to promote voter education and misinformation pre-bunking targeted at Spanish-speaking communities. In addition to this, during the election period, our content moderators routinely took enforcement action on Tweets that violated our rules.

In the lead up to the Georgia run-off, we will continue to monitor the platform for election misinformation targeted at Spanish-speaking communities and proactively debunk misinformation through Twitter Moments which provide additional context to the conversation. In addition, we have an Elections Hub available for all people in the US with Twitter Moments, available in English and Spanish, that provides information about voter registration and early voting for the Georgia runoff.

2. We all have an interest in setting a baseline standard for privacy protections. Some in the tech industry want preemption, but there are significant concerns over whether the standard we set will be enforced. Would you be open to a national consumer privacy law that includes some form of private enforcement?

Twitter supports passage of a strong national federal privacy law, which includes robust enforcement. Our team is happy to provide feedback on specific enforcement proposals.

3. Section 230 shuts the courthouse door on victims seeking justice. Even when tech companies are aware of abuse and crime – even when they amplify and promote that conduct – they are immune.

- a. Should an online platform that actively facilitates discriminatory advertising be fully accountable under our civil right laws?**

- b. Should Section 230 serve as a shield for online platforms against protections for children, civil rights, and consumer protection where platforms enable the promotion of illegal content or conduct?**
- c. How should Section 230 be reformed to ensure that it encourages platforms to be good Samaritans—not bad actors?**

The Communications Decency Act already exempts federal criminal activity from liability protections, including illegal conduct that violates civil rights, consumer protection, or childhood sexual exploitation laws. As explained in more detail in our written testimony, we do not believe that the solution to concerns raised about content moderation is to eliminate or create carveouts to existing Section 230 liability protections. Such changes can counterproductively harm speech and safety. Instead, we believe the solution to addressing content moderation challenges should be focused on enhancing transparency, procedural fairness, privacy, and algorithmic choice.

- 4. In December, Twitter announced “Bluesky,” a project to develop an open standard for social networks. The long-term goal of Bluesky is to foster interoperability between social networks. In the Microsoft antitrust case, as well as AOL’s merger with Time Warner, enforcers required firms to adopt open standards to enable competition.**
 - a. Do you agree that antitrust enforcers should consider requiring that dominant tech platforms offer interoperability and open standards to smaller competitors?**

Twitter supports interoperability, and making sure the people understand and have transparency into how their data is used and can control who it is shared with and when, including between tech companies.

- b. Recently, Senators Warner, Hawley, and I introduced the ACCESS Act, which would require dominant platforms to offer interoperability with competitors. What recommendations would you provide to build on the ACCESS Act?**

Our teams are happy to work with you in more depth on the ACCESS Act. Twitter appreciates efforts to enable data transfer projects to increase interoperability. Frameworks like Bluesky will help us meet that objective too by creating common standards for the industry.

- 5. During your testimony, you stated that Twitter was on heightened alert for misinformation and disinformation in the lead-up to and immediately following the 2020 election, and took additional steps to label misinformation and disinformation to protect the integrity of the 2020 election.**
 - a. How, if at all, will Twitter’s content moderation policies and practices differ from the period surrounding the 2020 election in the lead-up to the Georgia runoff?**
 - b. How, if at all, will Twitter’s content moderation policies and practices differ after the Georgia runoff?**

In the lead up to the 2020 U.S. election, we made significant enhancements to our policies to protect the integrity of the conversation occurring on Twitter regarding the election. The civic integrity policy and others aimed at safeguarding the public conversation remain in effect and will continue to be enforced in the lead up to and after the Georgia runoff.

For example, this year, we updated our civic integrity policy to more comprehensively enforce labeling or removing of false and misleading information. The updated policy covers the following activities:

- False or misleading information about how to participate in an election or civic process;
- Content intended to intimidate or dissuade people from participating;
- Misrepresentation about affiliation (for ex., a candidate or political party);
- Content that causes confusion about laws and regulations of a civic process, or officials and institutions executing those civic processes;
- Disputes of claims that could undermine public confidence in the election (e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results); and
- Content that misleads about outcomes (e.g., claiming victory before results are in, inciting unlawful conduct to prevent the procedural or practical implementation of election results).

The civic integrity policy augmented and enhanced other important rules aimed at preventing interference with the election. Twitter banned all political advertising in 2019, only allowing some cause-based advertising for non-partisan civic engagement, in line with our belief that the reach of political speech should be earned, not bought. Additionally, we adopted rules prohibiting deceptively shared synthetic or manipulated media, sometimes referred to as “deep fakes,” that may lead to serious offline harm; and labeling deceptive or synthetic media to provide additional context. Moreover, we have rules prohibiting platform manipulation, impersonation, hateful conduct, ban evasion, and attributed activity, among other harmful activities. We have also labeled specific government and state-media accounts from UN P-5 nation states, and plan to expand this effort in the near future.

6. Since Joe Biden was declared the President-elect, Twitter has scaled its content moderation. However, President Trump routinely flouts Twitter’s policies, hourly seeking to delegitimize the election. There is a real threat of violence, and these unfounded allegations are corrosive to our democracy.

- a. Under what conditions would you return to preventing a viewer from seeing the President’s misinformation about the election results unless the user affirmatively clicks “view” on a warning label?**
- b. Under what conditions would you return to preventing a user from commenting on or retweeting the President’s misinformation about the election results?**

In October 2020, we clarified our civic integrity policy to provide more information about our efforts to safeguard the public conversation against false claims of victory in the 2020 U.S. election. Applying warnings to premature claims of victory or victory claims that differed from official sources was always intended to be a temporary measure designed to guard against claims of

victory when the election outcome was still being determined and the risk of harm was most acute. Once the race was called by official sources and the outcome was widely disseminated, we determined that the risk associated with false claims of victory in the Presidential race significantly decreased and that warnings were no longer necessary to safeguard the public conversation.

- 7. This year, Russia and Iran engaged in aggressive campaigns to suppress the vote and undermine candidates throughout the election.**
 - a. Is it correct that Twitter received warnings from law enforcement about the threat of hack-and-leak operations during the election?**
 - b. Would you have taken the same steps as you did with the *New York Post*, if it were *The Guardian* and about one of the Trump children?**

In 2018, we created a policy to prevent Twitter from being used to spread hacked materials. This policy was informed from the activity we saw on the service during the 2016 elections by Russian-state actors. Additionally, we were warned by government partners generally about the threats that hack and leak operations could pose to the integrity of the conversation regarding the 2020 U.S. election.

It was against this backdrop that we enforced our Hacked Materials Policy against very specific content shared by the @NYPost. Under this policy, people on Twitter were blocked from sharing certain links from the @NYPost, publicly or privately, as those specific articles contained the source materials themselves. References to the contents of the materials or discussion about the materials were not restricted under the policy. After hearing feedback from journalists and others, we quickly updated our policy to limit its scope to only cover the removal of materials shared by hackers directly. Our initial decision was not impacted by political ideology or the party affiliation of the individuals who had information disclosed by the materials. It was informed by our Hacked Materials Policy at the time, and the warnings we had received about potential hack-and-leak operations.

- 8. Your counterpart, Facebook, has updated its policies to prohibit dehumanization based on race, ethnicity, and immigration status; whereas Twitter's does not clearly cover dehumanizing content targeting people based on race. How does Twitter account for differences in its policies on hate speech?**

Twitter recently expanded its hateful conduct policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin. More information about the expansion can be found [here](#).

Questions from Senator Booker

1. **Social media platforms, including Twitter, have a responsibility to stem the flow of election misinformation on their platforms. I believe it is possible for platforms like Twitter to ensure Americans' freedom to speak out while protecting the legitimacy of our democratic process and the public's safety.**
 - a. **Has Twitter considered implementing viral circuit breakers as proposed by Professor Ellen Goodman and the Center for American Progress,¹ where social media platforms would design a pause in the algorithmic amplification of fast-growing content about the election until content moderators can conduct an effective review for accuracy? Do you think this would be an effective tool in combatting the flow of misinformation on social media?**
 - b. **Has Twitter considered instituting a short delay on content from specific high-reach accounts to allow for human review, just as live network TV institutes a short delay to prevent unacceptable content from airing? Do you think this would be an effective tool in combatting the flow of misinformation on social media?**
 - c. **Will Twitter commit to hiding false and misleading content that baselessly delegitimizes our democratic process—content designed to sow doubt and division— behind a click-through warning label? Will Twitter commit to ensuring that its algorithm does not amplify such content?**

Twitter has taken numerous steps to combat the spread of misinformation. We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue. We will only add descriptive text that is reflective of the existing public conversation to let people determine their own viewpoints. In addition, we will reduce the visibility of labeled Tweets, meaning we will not amplify the Tweets on a number of surfaces across Twitter. We also alert people with a warning in cases where they seek to share a Tweet that has been labeled for misinformation, and in some cases disable engagement altogether. This has helped us to combat the potential spread of misinformation on the platform.

While we do not currently institute a short delay on content from high-reach accounts and have not instituted viral circuit breakers, we continue to study and refine our approach to addressing harms associated with misinformation. We look forward to continuing the conversation with your office about additional steps we can take to address harmful misinformation.

2. **Do you agree that social media platforms like yours have a responsibility to continue enforcing their enhanced election-related rules for user-generated content during the period between Election Day and the presidential inauguration on January 20?**

In the lead up to the 2020 U.S. election, we made significant enhancements to our policies to protect the integrity of the conversation occurring on Twitter regarding the election. The civic

integrity policy and others aimed at safeguarding the public conversation remain in effect and will continue to be enforced in the lead up and after the Georgia runoff.

Specifically, this year, we updated our civic integrity policy to more comprehensively enforce labeling or removing of false and misleading information. The updated policy covers the following activities:

- False or misleading information about how to participate in an election or civic process;
- Content intended to intimidate or dissuade people from participating;
- Misrepresentation about affiliation (for ex., a candidate or political party);
- Content that causes confusion about laws and regulations of a civic process, or officials and institutions executing those civic processes;
- Disputes of claims that could undermine public confidence in the election (e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results); and
- Content that misleads about outcomes (e.g., claiming victory before results are in, inciting unlawful conduct to prevent the procedural or practical implementation of election results).

The civic integrity policy augmented and enhanced other important rules aimed at preventing interference with the election. Twitter banned all political advertising in 2019, only allowing some cause-based advertising for non-partisan civic engagement, in line with our belief that the reach of political speech should be earned, not bought. Additionally, we adopted rules prohibiting deceptively shared synthetic or manipulated media, sometimes referred to as “deep fakes,” that may lead to serious offline harm; and labeling deceptive or synthetic media to provide additional context. Moreover, we have rules prohibiting platform manipulation, impersonation, hateful conduct, ban evasion, and attributed activity, among other harmful activities. We have also labeled specific government and state-media accounts from UN P-5 nation states, and plan to expand this effort in the near future.

3. What steps have you taken to modify Twitter’s algorithms to ensure that blatantly false election disinformation posted by election officials that receives high levels of interaction isn’t amplified?

In cases where a label or interstitial is applied, we take steps to reduce the visibility of Tweets, meaning we will not amplify the Tweets on a number of surfaces across Twitter. We may also remove the ability for people to retweet or like the Tweet.

¹ Adam Conner & Erin Simpson, *Results Not Found: Addressing Social Media’s Threat to Democratic Legitimacy and Public Safety After Election Day*, CTR. FOR AM. PROGRESS (Oct. 23, 2020), <https://www.americanprogress.org/issues/technology-policy/reports/2020/10/23/492232/results-not-found-addressing-social-medias-threat-democratic-legitimacy-public-safety-election-day>.; ² Donald J. Trump (@realDonaldTrump), TWITTER (Nov. 12, 2020, 11:34 A.M.), <https://twitter.com/realdonaldtrump/status/1326926226888544256>.; ³ *Joint Statement from Elections Infrastructure Government Coordinating Council & the Election Infrastructure Sector Coordinating Executive Committees*, CYBERSECURITY & INFRASTRUCTURE SEC. AGENCY (Nov. 12, 2020), <https://www.cisa.gov/news/2020/11/12/joint-statement-elections-infrastructure-government-coordinating-council-election>.

4. If an elected official repeatedly and flagrantly violates Twitter’s policies, at what point would Twitter impose a more severe enforcement penalty beyond labeling or hiding individual tweets?

We assess reported Tweets from world leaders against the Twitter Rules, which are designed to ensure people can participate in the public conversation freely and safely. We take enforcement action for any account on our service that involves the promotion of terrorism; clear and direct threats of violence against an individual; posting of private information; posting or sharing intimate photos or videos of someone produced or distributed without their content; engaging in behaviors related to child sexual exploitation; engaging in violations of the copyright policy, and encouraging or promoting self-harm. Direct interactions with fellow public figures, comments on political issues of the day, or foreign policy saber-rattling on economic or military issues are generally not in violation of these Twitter Rules.

In other cases involving a violation of the Twitter Rules, we will err on the side of leaving the content up if there is [a clear public interest](#) in doing so. In such cases, we may place the violative content behind a warning notice that provides context about the violation and allows people to click through should they wish to see the content.

5. What is the most significant enforcement action Twitter has taken against an American elected official’s account for violating your policies?

Twitter has taken a range of enforcement actions against the accounts of American elected officials, including removing, labeling, limiting the visibility of Tweets, or placing warning labels on content.

6. What is the most significant enforcement action Twitter has taken against a foreign leader’s account for violating your policies?

Twitter has taken a range of enforcement actions against the accounts of foreign leaders, including removing, labeling, or placing warning labels on content.

7. Has Twitter undertaken any studies about how hiding a tweet behind a warning label—as opposed to just adding a label beneath a tweet—limits the spread of disinformation? Has Twitter studied what content fills the void when these steps are taken? And if so, what has Twitter found? If not, do you plan to initiate such studies, or will you make a commitment to do so?

We are continuously analyzing our enforcement policies – including the impact of specific types of enforcement actions – so that we can learn and improve. We are currently analyzing the effectiveness of actions we took during the 2020 U.S. election and look forward to sharing our findings with the Committee. An initial analysis of the impact of some of our efforts around the U.S. 2020 election is [here](#).

8. President Trump is spreading dangerous misinformation about our electoral process on your platforms right now. What specific lessons have you learned since Election Day? And what concrete steps has Twitter taken to enhance its enforcement policies

regarding election disinformation since Election Day?

Our efforts to safeguard the conversation on Twitter regarding the 2020 U.S. election are ongoing and we continue to apply labels, warnings, and additional restrictions to Tweets that included potentially misleading information about the election. We continue to assess the impact of our enforcement actions, but an initial examination of our efforts from October 27th to November 11th has found:

- Approximately 300,000 Tweets have been labeled under our Civic Integrity Policy for content that was disputed and potentially misleading. These represent 0.2% of all US election-related Tweets sent during this time period;
- 456 of those Tweets were also covered by a warning message and had engagement features limited (Tweets could be Quote Tweeted but not Retweeted, replied to, or liked);
- Approximately 74% of the people who viewed those Tweets saw them after we applied a label or warning message; and
- There was an estimated 29% decrease in Quote Tweets of these labeled Tweets due in part to a prompt that warned people prior to sharing.

9. At noon on January 20, Donald Trump will no longer be President of the United States. If he continues to spread election misinformation in the future, will Twitter treat Donald Trump’s tweets differently—as an ex-President—from how the platform does now?

We assess reported Tweets from world leaders against the Twitter Rules, which are designed to ensure people can participate in the public conversation freely and safely. We take enforcement action for any account on our service that involves the promotion of terrorism; clear and direct threats of violence against an individual; posting of private information; posting or sharing intimate photos or videos of someone produced or distributed without their consent; engaging in behaviors related to child sexual exploitation; engaging in violations of the copyright policy, and encouraging or promoting self-harm. Direct interactions with fellow public figures, comments on political issues of the day, or foreign policy saber-rattling on economic or military issues are generally not in violation of these Twitter Rules.

In cases involving a world leader, we will err on the side of leaving the content up if there is a clear public interest in doing so. In such cases, we may place the violative content behind a warning notice that provides context about the violation and allows people to click through should they wish to see the content. Twitter’s world leader policy no longer applies when the account in question is no longer a world leader.

10. On November 10, President Trump issued a baseless tweet falsely claiming that an election technology company had “DELETED” millions of his votes and had “SWITCHED” hundreds of thousands more.² In fact, a group of federal and state officials responsible for election cybersecurity issued a statement debunking President Trump’s claims. “There is no evidence that any voting system deleted or lost votes, changed votes, or was in any way compromised,” they wrote, adding that “[t]he November 3rd election was the most secure in American history.”³

- a. **Is Twitter aware of any evidence to suggest that any election security company deleted millions of Trump votes nationwide?**
- b. **Do you think baseless claims about election fraud are harmful to our democracy?**
- c. **When President Trump posted similarly outrageous tweets during the week of the election, Twitter had hidden those tweets behind a warning label, which had the effect of reducing how many users were exposed to this false and misleading information. Why was this more recent tweet by President Trump—spreading outrageous falsehoods about the 2020 election—treated differently?**

Twitter does not have additional information related to the election security company referenced, beyond what has been publicly reported and shared by government sources. As the response to Question 2 details, in 2020, we updated our civic integrity policy to better safeguard the public conversation around critical civic processes, like the election and census. This policy permits us to take action in cases where individuals make claims that could undermine public confidence in the election, including unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.

With regards to the specific Tweet referenced in (c), it was posted on November 12, 2020 and [labeled](#) pursuant to Twitter's civic integrity policy.

11. On November 15, President Trump tweeted, "I WON THE ELECTION!"⁴ This blatant election misinformation was liked and retweeted hundreds of thousands of times. This and other similar tweets by President Trump were false declarations of victory aimed at undermining the integrity of our electoral process. Why did Twitter decide not to hide this disinformation behind a warning label, as it did for some earlier tweets?

In October 2020, we clarified our civic integrity policy to provide more information about our efforts to safeguard the public conversation against false claims of victory in the 2020 U.S. election. Applying labels and warnings to premature claims of victory or victory claims that differed from official sources was always intended to be a temporary measure designed to guard against claims of victory when the election outcome was still being determined and the risk of harm was most acute. Once the race was called by official sources and the outcome was widely disseminated, we determined that the risk associated with false claims of victory in the Presidential race significantly decreased and that warnings were no longer necessary to safeguard the public conversation.

Questions from Senator Coons

- 1. Some have criticized social media platforms for evaluating new products and business unit performance based on engagement-oriented metrics that in their view, reward misinformation and harmful content.**
 - a. What metrics does Twitter use when evaluating a new product or the performance of its business units?**
 - b. How does Twitter ensure that its metrics and incentive structures do not reward the development of products that contribute to misinformation or other harmful content?**

At Twitter, we prioritize healthy public conversation through our product, policies, and enforcement. The health principles that guide our work include decreasing potential for likely harm; harmful bias and incentives; and reliance on content removal. Our principles also push us to increase diverse perspectives and public accountability. These principles connect to everything for us — from our decision to ban all political ads, to our policy around public-interest notices, and even a product test that allows people to choose who can reply to their Tweets.

- 2. Does Twitter calculate, as Facebook now does, the prevalence of hate speech on its platform?**
 - a. If not, please explain why not and how Twitter believes it can effectively address hate speech without this information.**
 - b. If so, please explain and state whether that information will be made public.**

Twitter is a uniquely open service, which is used by academics across the world as a data source to study important issues, including hate online. [Published papers](#) have recognised that “Twitter provides a unique big data source for public health researchers” while Carl Miller highlighted that “Twitter is the only major service to make public conversation data available via an API, for the purposes of study.” The accessibility of Twitter’s APIs allows researchers to pursue their own areas of focus, measuring whatever is relevant to their field of study. Enabling independent research is a different way to inform the public debate and something we believe strongly in.

In addition to this, we have numerous policies aimed at combating hateful conduct on our platform. Our [Hateful Conduct Policy](#) prohibits references to violent events or types of violence where protected categories were the primary victims, or attempts to deny or diminish such events. We recently [expanded](#) this policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin. In addition, our [Glorification of Violence Policy](#) prohibits content that glorifies or praises historical acts of violence and genocide, including the Holocaust. And, our [Violent Organizations Policy](#) prohibits violent organizations, including extremist groups, from using Twitter to promote violence. Under this policy, we have removed 204 groups, half of which had links to white supremacy, and permanently suspended 1.7 million unique accounts for the promotion of terrorism.

- 3. You indicated in your testimony that you would not conduct an independent civil rights audit, as Facebook has done, in light of your consultation with civil rights groups and the transparency reports you issue. Please provide more detail about the specific groups that you consult with, the nature of the consultations, and why you**

believe that consultation is an adequate substitute for an independent audit of the company that would permit direct access and analysis of Twitter’s internal policies, products, and actions.

We agree that third-party feedback and metrics can be valuable resources to inform our work. Our focus is not only assessment but building a framework both internally and externally to make substantive change over time. To that end, we have established a global, cross-functional group with C-suite representation to inform and evaluate our work related to civil rights. In addition, several national organizations that represent the interests and defense of civil rights, serve in advisory roles on our Trust and Safety Council. These partnerships inform how we shape and enforce our policies. Prospectively, we will continue to engage groups in the U.S. that are focused on civil rights as we advance our work.

- 4. You testified, “We wanted to scope our approach to start to focus on the highest severity of harm. We focused on three areas: manipulated media... civic integrity... and public health, specifically around COVID.” The Fourth National Climate Assessment concluded that climate change “affects human health by altering exposures to heat waves, floods, droughts, and other extreme events; vector-, food- and waterborne infectious diseases; changes in the quality and safety of air, food, and water; and stresses to mental health and well-being.”¹ Given the immediate public health risks posed by climate change, does Twitter intend to develop a climate change misinformation policy? If not, please explain why not. If so, please provide details including timing.**

At this time, Twitter does not take enforcement action against climate change misinformation. However, we continue to evaluate our policies and if additional safeguards are necessary to better serve the needs of the people who use Twitter.

- 5. The *Guardian* reported in February 2020 that a draft study by Brown University researchers suggests “a substantial impact of mechanized bots in amplifying denialist messages about climate change.”² What steps has Twitter taken to address the amplification of climate denialism by bots?**

Twitter has various policies to address potential harms associated with automated activity. For example, our [Platform Manipulation and Spam Policy](#) prohibits engaging in bulk, aggressive, or deceptive activity that misleads and/or disrupts individuals’ experience on Twitter. This policy is intended to address a range of behavior, including commercially motivated spam, inauthentic engagements, coordinated harmful activity, and coordinated activity that attempts to artificially influence conversations through the use of multiple accounts, fake accounts, automation and/or scripting.

- 6. In the coming months, it is likely that extensive new information about COVID-19**

¹ Kristi L. Ebi et al. “Human Health,” in *Impacts, Risks, and Adaptation in the United States: Fourth National Climate Assessment, Volume II*, U.S. Global Change Research Program (2018), available at <https://nca2018.globalchange.gov/chapter/14/>.

² See Oliver Milman, “Revealed: Quarter of all Tweets About Climate Crisis Produced by Bots,” *The Guardian* (Feb. 21, 2020), available at <https://www.theguardian.com/technology/2020/feb/21/climate-tweets-twitter-bots-analysis> (Feb. 21, 2020).

vaccine candidates will become available. Unfortunately, misinformation about vaccines abounds, and the World Health Organization named resulting vaccine hesitancy one of the top ten threats to global health in 2019.³ In addition, a recent study found that social media users exposed to content on certain vaccines were more likely to grow misinformed over time than were consumers of traditional media.⁴

- a. Is Twitter proactively engaged in planning efforts to address misinformation about emerging COVID-19 vaccines on its platforms?**
- b. If so, how does Twitter plan to assess the accuracy of information about these vaccines?**
- c. Has Twitter partnered with (or will Twitter partner with) fact checkers with relevant training and expertise to address misinformation about COVID-19 vaccines?**
- d. How will Twitter handle vaccine-related content deemed valid when posted but which more recent guidance or consensus suggests is misleading or inaccurate?**
- e. How will Twitter engage public health, immunology, and other related experts to identify and contextualize content that is incomplete or misleading?**

The public conversation occurring on Twitter is critically important during this unprecedented public health emergency. With a critical mass of expert organizations, official government accounts, health professionals, and epidemiologists on our service, our goal is to elevate and amplify authoritative health information as far as possible. To address this global pandemic, on March 16, 2020, we announced [new enforcement guidance](#), broadening our definition of harm to address, specifically, content related to COVID-19 that goes directly against guidance from authoritative sources of global and local public health information. We require individuals to remove violative Tweets in a variety of contexts with the goal of preventing offline harm. Additionally, we are currently engaged in an effort launched by the Office of the U.S. Chief Technology Officer under President Trump in which we are coordinating with our industry peers to provide timely, credible information about COVID-19 via our respective platforms. This working group also seeks to address misinformation by sharing emerging trends and best practices.

In addition, in December 2020, we updated our policy approach to misleading information about COVID-19. Beginning December 21, we may require people to remove Tweets which advance harmful false or misleading narratives about COVID-19 vaccinations, including:

- False claims that suggest immunizations and vaccines are used to intentionally cause harm to or control populations, including statements about vaccines that invoke a deliberate conspiracy;
- False claims which have been widely debunked about the adverse impacts or effects of receiving vaccinations; or
- False claims that COVID-19 is not real or not serious, and therefore that vaccinations are unnecessary.

³ See Sarah Geoghegan et al., “Vaccine Safety: Myths and Misinformation,” *Frontiers in Microbiology* (Mar. 2020), available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7090020/>.

⁴ Dominik Andrezej Stecula et al., “How Trust in Experts and Media Use Affect Acceptance of Common Anti-Vaccination Claims,” *Misinformation Review* (Jan. 2020).

Starting in early 2021, we may label or place a warning on Tweets that advance unsubstantiated rumors, disputed claims, as well as incomplete or out-of-context information about vaccines. Tweets that are labeled under this expanded guidance may link to authoritative public health information or [the Twitter Rules](#) to provide people with additional context and authoritative information about COVID-19. We will enforce this policy in close consultation with local, national and global public health authorities around the world, and will strive to be iterative and transparent in our approach.

- 7. In the lead-up to the election, reporting and company representatives have referenced efforts by Twitter to alter content distribution algorithms in order to restrict the reach of harmful content. The scope and extent of such efforts, and to what degree they are temporary or permanent, is unclear. Please elaborate on the following:**
 - a. Any systems put in place, whether automated or manual, to alter content distribution or recommendation algorithms;**
 - b. Any associated changes in content moderation practices;**
 - c. Specific definitions of content areas to which changes were applied;**
 - d. An evaluation of the results of such changes, including any metrics you can provide for each week during October and November (or a statement that such evaluation will be included in a subsequent post-mortem analysis).**

Twitter made several policy and product changes in an effort to safeguard the public conversation occurring on Twitter regarding the 2020 U.S. election. On the policy side, notably, we expanded and enforced our Civic Integrity Policy, which prohibits manipulating or interfering in elections or other civic processes. This includes posting or sharing content that may suppress participation or mislead people about when, where, or how to participate in a civic process. Under this policy, Twitter labeled and reduced the visibility of Tweets containing false or misleading information about civic processes in order to provide additional context.

During the period from October 27 to November 11, 2020, we labeled approximately 300,000 Tweets under our [Civic Integrity Policy](#) for content that was disputed and potentially misleading. These represent 0.2% of all U.S. election-related Tweets sent during this time period. Approximately 450 of those Tweets were also covered by a warning message and had engagement features limited, including Tweets could be Quote Tweeted but not Retweeted, replied to or liked. Approximately 74% of the people who viewed those Tweets saw them after we applied a label or warning message. We saw an estimated 29% decrease in Quote Tweets of these labeled Tweets due in part to a prompt that warned people prior to sharing.

In addition to policy changes, we implemented [significant product changes](#) intended to increase context and encourage more thoughtful consideration before Tweets are amplified. For example, we encouraged people to add their own commentary when amplifying content by prompting Quote Tweets instead of Retweets. This change introduced some friction, and gave people an extra moment to consider why and what they were adding to the conversation. The change slowed the spread of misleading information by virtue of an overall reduction in the amount of sharing on the service. We observed a 23% decrease in Retweets and a 26% increase in Quote Tweets, but on a net basis the overall number of Retweets and Quote Tweets combined decreased by 20%.

In addition, we stopped providing “liked by” and “followed by” Tweet recommendations from accounts you do not follow in the Home Timeline and through notifications. While we had initially hoped that this would reduce the potential for misleading information to spread on our service, we did not observe a statistically significant difference in misinformation prevalence as a result of this change (nor any meaningful reduction in abuse reports). Instead, we found that pausing these recommendations prevented many people from discovering new conversations and accounts to follow, and we have since reverted the change. We are continuing to assess the impact of these product and policy changes and look forward to sharing further information with the Committee.

8. Twitter has affirmed its interest in working with outside researchers to improve conversations on the platform, but academics and activists have expressed frustration over the roadblocks they have encountered in getting research projects off the ground. Twitter has acknowledged that “[a] number of factors, including internal employee changes and unforeseen complexity in establishing partnerships, have contributed to delays and uncertainty.”⁵ What progress, if any, has Twitter made on removing obstacles for outside researchers seeking to study the health of discourse on Twitter, and what research projects, if any, are currently under way?

Twitter is a uniquely open service, which is used by academics across the world as a data source to study important issues, including hate online. Published papers have recognised that “Twitter provides a unique big data source for public health researchers” while Carl Miller highlighted that “Twitter is the only major service to make public conversation data available via an API, for the purposes of study.” Enabling independent research is a different way to inform the public debate and something we believe strongly in. For example, earlier this year we made available at no cost to researchers a [dedicated API-endpoint](#) of Covid-19 Tweets, enabling a wide range of research into the pandemic.

Moreover, to further support the work of academics, we have been testing an Academic Research product track, which will launch next year. The Academic Research product track will offer qualified academic researchers a significantly higher monthly Tweet volume cap for free, as well as endpoints and enhanced features to get more precise and complete data for analyzing the public conversation.

In addition to accessing public Tweets, we have also taken the decision to make the data relating to state-linked information operations that we have removed available to researchers and the public. Beginning [in 2018](#), we now make available a comprehensive archive of Tweets and media associated with known state-backed information operations that we have removed. This unique archive, the only source of its kind made available by any company, is used by researchers, journalists and experts around the world and now spans operations across 15 countries, including more than nine terabytes of media and 200 million Tweets.

9. While important questions about the potential addictive properties of social media remain open for further research, the association between social media use and mental

⁵ See Deepa Seetharaman, “Jack Dorsey’s Push to Clean Up Twitter Stalls, Researchers Say,” *Wall Street Journal* (Mar. 15, 2020), available at <https://www.wsj.com/articles/jack-dorseys-push-to-clean-up-twitter-stalls-researchers-say-11584264600>.

health disorders, especially among youth, raises urgent questions.⁶

- a. What research has Twitter conducted internally on the mental health impacts of social media use?
- b. What resources has Twitter made available for independent, external study of the potential implications of its platform for users' mental health?
- c. Has Twitter integrated the findings of any such studies into its product design?

We continue to follow this issue, including academic research, closely. Studies have produced a range of insights and findings that do not suggest that these questions are as clear cut as some discussion may suggest, with some finding that the impact of digital technology on mental health and wellbeing is positive.

Our Trust & Safety Council has a dedicated advisory group on [Suicide Prevention and Mental Health](#), while we partner with organisations around the world to better understand these issues, in addition to supporting our partners in their work providing support and raising understanding. In addition to this, earlier this summer, we worked with various mental health partners across the globe to raise [awareness](#) and encourage honest conversation around the emotional challenges we are experiencing together, amid the unprecedented COVID-19 crisis. Since then, we've expanded our work with NGOs to identify, connect, and engage vulnerable people across the world. In particular, we've continued to engage suicide prevention organizations and counseling services to ensure that people on Twitter feel safe and have access to support when they need it most.

10. The 1998 Digital Millennium Copyright Act (DMCA) established a “notice and takedown” system for policing copyright infringement on the internet, but a recent Copyright Office report concluded that “Congress’ original intended balance has been tilted askew” and that takedown notices have not remedied the widespread problem of digital piracy.⁷ I am concerned about the impact of this ongoing piracy on our nation’s creative community.

- a. What responsibility should platforms like Twitter shoulder in proactively identifying, removing, and blocking infringing content?
- b. What steps has Twitter taken to protect copyright owners beyond processing DMCA takedown notices?
- c. Does Twitter ever monetize copyrighted content that is posted without the permission of the copyright owner?
- d. I am concerned by reports that Twitter charges copyright owners for access to the tools necessary to search for infringing content on the platform and limits the number of infringing posts that can be identified for removal. How do you respond?

⁶ See, e.g., Cleveland Clinic, “Is It Possible to Become Addicted to Social Media?” (Mar. 1, 2019), available at <https://health.clevelandclinic.org/is-it-possible-to-become-addicted-to-social-media/>; Elroy Boers et al., “Association of Screen Time and Depression in Adolescence,” *JAMA Pediatrics* (2019), available at https://jamanetwork.com/journals/jamapediatrics/article-abstract/2737909?guestAccessKey=f3fe2ed6-1fb3-44cc-a9a8-a38bd0463942&utm_content=weekly_highlights&utm_term=081019&utm_source=silverchair&utm_campaign=jama_network&cmp=1&utm_medium=email.

⁷ United States Copyright Office, *Section 512 of Title 17: A Report of the Register of Copyrights* (May 21, 2020), available at <https://www.copyright.gov/policy/section512/section-512-full-report.pdf>.

- e. **As further noted in the recent Copyright Office report, “Congress’ vision of broad, open, cross-industry standards-setting for the creation of standard technical measures has not come to pass.” Why do you think that is, and do you have any hope that future voluntary standardization of technical measures will combat digital piracy effectively?**

Twitter responds to all legitimate copyright complaints as laid out in Section 512 of the DMCA. Furthermore, Twitter does not allow for full-length music streaming as some of our competitors do. Twitter is unique among our industry peers and remains a text-first service for the creation of content. When media is shared on Twitter, it is often in service of commentary or criticism to drive real-time, public conversations and debate. For example, when our customers post snippets of third-party content (videos, photos and GIFs), they are frequently doing so in an exercise of political and newsworthy speech, and for the purpose of commentary and criticism (not to share and watch pirated content). As you can see in our [transparency reports](#), the number of takedown notices Twitter receives annually is a fraction of what other platforms receive and reflects the fact that allegedly infringing material is a small fraction of the total Tweets uploaded to Twitter.

Twitter relies on rightsholders, who are in the best position to know whether use of their content is infringing or not, to notify us of infringing material on Twitter. Twitter’s response to copyright complaints may include the removal or restriction of access to allegedly infringing material. If we remove or restrict access to content in response to a copyright complaint, Twitter will make a good-faith effort to contact the affected account holder with information concerning the removal or restriction of access, including a full copy of the complaint, along with instructions for filing a counter-notice.

We are actively engaged with the RIAA and many other rightsholder bodies regarding copyright concerns they have. We dedicate significant resources to quickly respond to takedown notices, and we work with rightsholders and representative bodies to address specific concerns if and when they arise.

One important consideration as we contemplate voluntary agreements and technological measures is that in certain circumstances, heightened specificity can hurt small businesses that are launching new apps and platforms. Small companies will not necessarily have the resources to afford expensive third-party technical solutions. Additionally, we are always concerned that voluntary agreements will further entrench the market dominant players.

Questions from Senator Durbin

1. **On October 28, *The New York Times* published an article entitled “Evidence of anti-conservative bias by platforms remains anecdotal.” The article says: “Conservatives have said for years that online social media platforms censor their views. But their evidence is largely anecdotal, and conservative accounts frequently perform extremely well online.”**
 - a. **Do you agree that evidence of online social media platforms censoring conservative views is “largely anecdotal”?**

Yes. Twitter enforces our policies impartially, without regards to political viewpoints or affiliation. Political bias in the enforcement of our policies or other related decision-making has not been credibly confirmed by any studies of which we are aware.

- b. **Do you agree that “conservative accounts frequently perform extremely well online”?**

Conservative voices have a strong presence on Twitter. For example, President Trump was the most mentioned person in the U.S. last year with more than 300 million mentions. Additionally, the hashtag #MAGA was the third most popular hashtag with more than 60 million Tweets.

2. **In their sobering book “How Democracies Die,” authors Steven Levitsky and Daniel Ziblatt make the following observation:**

“Under President Trump, America has been defining political deviancy down. The president’s routine use of personal insult, bullying, lying, and cheating has, inevitably, helped to normalize such practices. Trump’s tweets may trigger outrage from the media, Democrats, and some Republicans, but the effectiveness of their responses is limited by the sheer quantity of violations. As [Senator Daniel Patrick Moynihan] observed [in 1993], in the face of widespread deviance, we become overwhelmed and then desensitized. We grow accustomed to what we previously thought to be scandalous. Furthermore, Trump’s deviance has been tolerated by the Republican Party, which has helped make it acceptable to much of the Republican electorate.”

Mr. Dorsey, I know that Twitter was not conceived as a medium for desensitizing Americans to political deviancy. But President Trump’s tweets have had that effect, with serious consequences to our democratic institutions. What is your reaction to the authors’ discussion of the normalization of personal insults, bullying, lying, and cheating that has been accelerated through President Trump’s tweets?

We assess reported Tweets from world leaders, including President Trump, against the Twitter Rules. In response to violations of the Twitter Rules, we have taken action on a variety of Tweets posted by President Trump, including labeling Tweets or placing them behind an interstitial and limiting amplification. Importantly, we believe there is a value in keeping the content available on

our service. There is a public interest in enabling the people to be informed and engage directly with their elected leaders.

While Twitter has a responsibility to safeguard the integrity of the public conversation, we recognize that we are only one part of the broader ecosystem that impacts the broader public discourse. The internet has lowered traditional media barriers to entry for all voices, allowing for unprecedented discourse and community building across the political and socio-economic spectrum.. We are happy to work with Congress on efforts to increase civic resilience to better safeguard against harmful misinformation and other concerning behavior.

3. Has President Trump ever posted or retweeted tweets that contain objectively false information during his presidency? If so, has Twitter ever identified such tweets as containing objectively false information?

Twitter does not make determinations about whether information on the platform is true or false. We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. Consistent with this feedback from our customers, we have expanded our enforcement options to allow us to label misinformation related to manipulated media, COVID-19, and civic integrity. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue.

4. Twitter has been trying to address election disinformation, sometimes by adding small disclaimers at the bottom of false or misleading tweets, sometimes by putting disclaimer screens so users have to click past the disclaimer in order to view the tweet. For example, recently President Trump tweeted, “Most fraudulent Election in history!” At the bottom of this tweet, Twitter posted a small disclaimer, saying, “This claim about election fraud is disputed.” When President Trump makes an objectively false claim like this, shouldn’t Twitter say that it’s false, not disputed?

We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. In addition, we recognize that Twitter is only one part of the broader ecosystem that is necessary to build civic resilience and combat harmful misinformation. We encourage Congress to engage in efforts to promote civic resilience more broadly and are happy to work with you on such efforts.

5. It is not widely known that Twitter allows users to post graphic pornographic content. While Twitter does not allow children under the age of 13 to sign up for an account, a child does not need an account to access content posted on Twitter, including pornographic content.

- a. What measures is Twitter taking to ensure that children are protected from exposure to graphic pornographic content on your service, especially at a time when millions of American children are spending many hours online every day due to remote learning?**

Our Twitter Rules and Twitter Media Policy limit the types of content that may be shared on Twitter and describe requirements for users who choose to share potentially sensitive content on Twitter. In addition, we also provide people, irrespective of their age, a variety of settings to control their experience, including their safety and content experience on the platform.

Every account holder has the choice of whether they will see a warning for sensitive content or not. When an individual on Twitter has this setting enabled, people who visit a specific profile may see a message that the account may include potentially sensitive content and inquire if the individual wants to view it. This setting enables individuals on Twitter to control their own experience and protects them from seeing sensitive content without first having made a choice to click through the warning, or to never see warnings.

b. Will Twitter consider reaching out to local school systems to encourage them to block access to Twitter on school-issued devices and other online learning platforms, especially for elementary school students?

We believe strongly in protecting children on Twitter. As part of our work, we comply with the Children's Online Privacy Protection Act, or COPPA and provide a range of tools for individuals to control whether or not they see sensitive content on Twitter. We encourage schools to use these tools in cases where they are concerned about children accessing sensitive content and will continue to work with your office on addressing this concern.

Questions from Senator Hirono

1. In your opening statement, you expressed your belief that “companies like Twitter should publish their moderation process.”

- a. What is preventing Twitter from publishing its content moderation process right now?**
- b. Does Twitter commit to publishing its content moderation process? If so, when? If not, why not?**

We have made significant efforts to be transparent about our content moderation rules and enforcement. Our Twitter Rules are publicly available, and we strive to ensure that these rules are easily understood. In addition, an important component of our transparency efforts is the Twitter Transparency Center. This year, we expanded our biannual transparency report site to become a comprehensive Twitter Transparency Center. Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This site includes data visualizations making it easier to compare trends over time and more information for the individuals who use Twitter, academics, researchers, civil society groups and others who study what we do to understand bigger societal issues.

The Transparency Center includes data on enforcement actions under the Twitter Rules that requires the removal of specific Tweets or to suspend accounts. The Center also includes sections covering information requests, removal requests, copyright notices, trademark notices, email security, platform manipulation, and state-backed information operations. We continue working on additional ways we can be transparent about our enforcement practices, and support efforts by this committee to enhance platform transparency.

2. Section 230(c)(1) of the Communications Decency Act currently grants platforms like yours broad immunity for content posted by third parties, even if platforms have knowledge of the content, promote the content, or profit off the content. This immunity applies regardless of platform’s size, resources, or efforts to moderate content.

- a. Do you believe that all internet platforms should receive the same degree of immunity under Section 230(c)(1) regardless of their size and resources?**
- b. Do you believe that all internet platforms should receive the same degree of immunity under Section 230(c)(1) regardless of whether, and to what extent, they moderate content?**
- c. Would you support legislation that required platforms to earn their immunity under Section 230(c)(1) by conditioning immunity on meeting a minimum standard of care?**

Section 230 is the Internet’s most important law for free speech and safety. Weakening Section 230 protections will remove critical speech from the Internet. We must ensure that all voices can be heard, and we continue to make improvements to our service so that everyone feels safe participating in the public conversation—whether they are speaking or simply listening. The protections offered by Section 230 help us achieve this important objective. Eroding the foundation of Section 230 could collapse how we communicate on the Internet, leaving only a small number of giant and well-funded technology companies. We should also be mindful that undermining Section 230 will result in far more removal of online speech and impose severe limitations on our collective ability to address harmful content and protect people online.

As explained in more detail in our written testimony, we do not believe that the solution to concerns raised about content moderation is to eliminate Section 230 liability protections. Instead, we believe the solution should be focused on enhancing transparency, procedural fairness, privacy, and algorithmic choice, which can be achieved through additions to Section 230, industry-wide self-regulation best practices, or additional legislative frameworks.

3. I have heard from copyright holders that Twitter has imposed hurdles to the efficient removal of infringing content on its platform, including by limiting the number of legitimate Digital Millennium Copyright Act notices that can be sent through its webform and by withholding from copyright holders free search capabilities that would help them enforce their rights.

Do you commit to working with copyright holders to remove these hurdles as other platforms already have?

Twitter relies on rightsholders, who are in the best position to know whether use of their content is infringing or not, to notify us of infringing material on Twitter. Twitter's response to copyright complaints may include the removal or restriction of access to allegedly infringing material. If we remove or restrict access to content in response to a copyright complaint, Twitter will make a good-faith effort to contact the affected account holder with information concerning the removal or restriction of access, including a full copy of the complaint, along with instructions for filing a counter-notice.

Twitter responds to all legitimate copyright complaints as laid out in Section 512 of the DMCA. Furthermore, Twitter does not allow for full-length music streaming as some of our competitors do. Twitter is unique among our industry peers and remains a text-first service for the creation of content. When media is shared on Twitter, it is often in service of commentary or criticism to drive real-time, public conversations and debate. For example, when our customers post snippets of third-party content (videos, photos and GIFs), they are frequently doing so in an exercise of political and newsworthy speech, and for the purpose of commentary and criticism. As you can see in our [transparency reports](#), the number of takedown notices Twitter receives annually is a fraction of what other platforms receive and reflects the fact that allegedly infringing material is a small fraction of the total Tweets uploaded to Twitter.

We are actively engaged with the RIAA and many other rightsholder bodies regarding copyright concerns they have. We dedicate significant resources to quickly respond to takedown notices, and we work with rightsholders and representative bodies to address specific concerns if and when they arise.

Questions from Senator Leahy

1. **You committed to me during your testimony that Twitter will conduct a post-mortem analysis of election misinformation and disinformation spread on your platform and make it available for the public to review.**
 - a. **What is the scope of this review, including what timeframe will it cover?**
 - b. **When do you anticipate the review will be complete?**
 - c. **You testified your review would be made available to the public so academics and researchers can review it. Do you intend to place any limitations on the availability of any data made public compared to your internal review?**

Our work to safeguard the conversation on Twitter regarding the 2020 U.S. election is ongoing, and we continue to assess the measures we have taken for effectiveness. We have already engaged in an unprecedented level of transparency around our election work, including routinely publicly releasing datasets of information that can reliably be linked to state backed actors. The archive of these datasets, which includes information released in October 2020, is the only type of its kind in industry.

Apart from these measures, we have already [released](#) initial analysis regarding our efforts to safeguard the conversation occurring on Twitter about the 2020 U.S. election. We are in the process of conducting additional analysis on actions taken throughout the entire election period and plan to release additional information to the public early next year.

2. **The label used by Twitter to describe the President’s election-related tweets states that his claims are disputed, or that other sources called the election differently. In fact, many of these tweets are demonstrably, factually false. To this day, the President is absurdly claiming he won the election. When we know a tweet on your platform will be seen by millions and is objectively false, why not simply label it as “false?”**

We have heard from the people who use Twitter that we should not determine the truthfulness of Tweets and we should provide context to help people make up their own minds in cases where the substance of a Tweet is disputed. Consistent with this feedback from our customers, we have expanded our enforcement options to allow us to label misinformation related to manipulated media, COVID-19, and civic integrity. When we label Tweets, we link to Twitter conversation that shows three things for context: (1) factual statements; (2) counterpoint opinions and perspectives; and (3) ongoing public conversation around the issue.

3. **Will Twitter maintain and continue building on its Civic Integrity Policy to ensure that its platform does not amplify misinformation in political elections and other civic matters?**

We will continue to assess and build on our efforts to safeguard the public conversation around elections and other civic processes.

4. What is Twitter doing to ensure that Live Videos are not utilized as a loophole to evade the platform's Civic Integrity Policy?

Our policies apply to all content posted on Twitter, including live videos. When there is a violation of these policies, we can apply a range of enforcement options, including removal, limiting visibility, and providing additional context. For example, Twitter removed live video of a speech from President Trump on November 5, 2020, due to violations of our Civic Integrity Policy.

5. Misinformation campaigns targeting communities of color are a very dangerous tool that contribute to voter suppression. A report by the Senate Intelligence Committee noted that Russian interference in the 2016 election targeted African Americans in our country with disinformation campaigns. In October, the *Washington Post* reported that Twitter had deleted fake accounts posing as Black Americans. One of the fake accounts that was suspended had garnered 24,000 followers and its most popular tweet was like 75,000 times. What steps is Twitter taking to improve the speed at which fake accounts are found and to prevent the spread of misinformation campaigns targeting communities of color on the platform?

As platform manipulation tactics evolve, we are continuously updating and expanding our rules to better reflect what types of inauthentic activity violate our guidelines. We continue to develop and acquire sophisticated detection tools and systems to combat malicious automation on our service.

Individuals are not permitted to use Twitter in a manner intended to artificially amplify, suppress information, or engage in behavior that manipulates or disrupts other people's experience on the service. We do not allow spam or platform manipulation, such as bulk, aggressive, or deceptive activity that misleads others and disrupts their experience on Twitter. We also prohibit the creation or use of fake accounts.

We also know that certain groups and individuals engage in persistent, organized efforts to manipulate and interfere with the conversation on Twitter. Therefore, when we are able to reliably attribute an account on Twitter to an entity known to violate the Twitter Rules, we will remove additional accounts associated with that entity. For instance, if we are able to identify activity associated with the Russian Internet Research Agency, all accounts tied to that entity will be removed, regardless of the content they share. We likewise will remove accounts that deliberately mimic or are intended to replace accounts we have previously suspended for violating our rules. These steps allow us to take more aggressive action against known malicious actors.

6. Misinformation spreads in many languages on social media platforms. In October 2020, *The New York Times* reported that misinformation in America thrives in both English and Spanish. In October 2020, a study published in *The American Journal of Tropical Medicine and Hygiene*, detailed how COVID-19 misinformation has circulated in 25 different languages across at least 87 countries. What are you doing to prevent misinformation from spreading in languages other than English and alerting users about this on the platforms? How prepared is Twitter to handle misinformation in Spanish specifically?

Twitter has taken numerous steps to address misinformation in Spanish that violates our policies. We have content moderators with Spanish language capability who can enforce our policies. In addition, we partner with a range of government and civil society organizations on how to use Twitter tools to facilitate the flow of credible information to Spanish speaking communities. For example, as part of our efforts to safeguard the public conversation around the election when [searching](#) for key terms related to voter registration, individuals who used Twitter received prompts in English and Spanish pointing them to official resources. In addition, we trained voter education nonprofits and government partners on how to use Twitter tools and create content targeted at Spanish-language audiences. As part of this, we have worked with @USAGovEspanol, @NALEO, @HispanicFed, @MALDEF, @LULAC, and @WeAreUnidos to promote voter education and misinformation pre-bunking targeted at Spanish-speaking communities. In addition to this, during the election period, our content moderators routinely took enforcement action on Tweets that violated our rules.

7. Will you commit to take additional steps based on lessons learned during the 2020 presidential election to prevent the spread of misinformation during the upcoming U.S. Senate elections in Georgia? What steps are you planning, if any, at this stage?

Our work to safeguard the conversation occurring on Twitter regarding the 2020 U.S. election is ongoing. In the lead up to the 2020 elections, we made significant enhancements to our policies to protect the integrity of the election. The civic integrity policy and others aimed at safeguarding the public conversation remain in effect and will continue to be enforced around the Georgia runoff.