Questions from Senator Tillis

1. My subcommittee has spent this year holding a series of hearings on the Digital Millennium Copyright Act (DMCA) and the safe harbors that internet service providers enjoy when it comes to copyright infringement by their users. Technology company witnesses said they can't identify infringements by their users – but Twitter is very good at identifying and censoring or taking down or flagging content from conservative voices. Don't you think that if Twitter can track and take down political content that it also has the ability to take down content that infringes the IP of hard-working American creators?

Twitter responds to copyright complaints submitted under the DMCA pursuant to our copyright policy. Section 512 of the DMCA outlines the statutory requirements necessary for formally reporting copyright infringement, as well as providing instructions on how an affected party can appeal a removal by submitting a compliant counter-notice. Twitter will respond to reports of alleged copyright infringement, such as allegations concerning the unauthorized use of a copyrighted image as a profile or header photo, allegations concerning the unauthorized use of a copyrighted video or image uploaded through our media hosting services, or Tweets containing links to allegedly infringing materials.

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.

2. Can you tell me about the number of human moderators you have at Twitter, and how many are working on copyright infringement?

Twitter uses a combination of machine learning and human review to adjudicate reports of violations of the Twitter Rules and make determinations on whether the activity violates our rules. With hundreds of millions of Tweets sent every day, we recognize that engaging in content moderation at scale requires increased use of machine learning and algorithms to surface and take action against violations of our rules. Thus, we continue to invest in efforts to automate content moderation. As a result of these investments, more than 50 percent of Tweets we take action on for abuse are now proactively surfaced using technology, rather than relying on reports to Twitter.

3. How does Twitter determine that it is appropriate to take a proactive step to counter, address, block, or remove content? Are the considerations different for political content and copyright-infringing content?

Whether Twitter takes enforcement action is dependent on whether an individual is in violation of the Twitter Rules. The Twitter Rules include a range of policies, including policies that prohibit abuse, harassment, hateful conduct, and child sexual exploitation, among many others. The Twitter Rules also include a copyright policy. Under this policy, Twitter will respond to reports of alleged

copyright infringement, such as allegations concerning the unauthorized use of a copyrighted image as a profile or header photo, allegations concerning the unauthorized use of a copyrighted video or image uploaded through our media hosting services, or Tweets containing links to allegedly infringing materials.

4. I know that Twitter has developed significant technological tools to guard against foreign interference in elections or to identify and label content that you deem politically misleading, inaccurate, or dispute. What types of technological tools do you use to identify and address content that infringes copyright?

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.

In an effort to be as transparent as possible regarding the removal or restriction of access to Tweeted content, we clearly mark withheld Tweets and media to indicate to viewers when content has been withheld in response to a complaint from a rightsholder. We also send a redacted copy of each copyright complaint that we process to [Lumen](#), where they are posted to a public-facing website, with all personal information removed.

One important consideration as we contemplate voluntary agreements and technological measures to address this problem 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.

5. Can you give me an update on Twitter's ongoing issues with RIAA and the steps you are taken to be more proactive in taking down the content of musicians and other artists that is being pirated on your site?

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. A full accounting of actions taken to protect copyrighted material on our platform can be found in our [Transparency Report.](#) For 2019, the total copyright takedown notices we received worldwide from all rightsholders was less than 285,000; those notices resulted in 1.1 million pieces of media removed in 2019 worldwide.

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.

6. We're now continuing the DMCA reform process with draft legislation that I want stakeholders from all perspectives to help me refine so we can build consensus. Do I have your commitment that Twitter will participate constructively and in good faith?

Yes, we look forward to engaging with you on constructive dialogue on this issue.

7. It is not uncommon for legitimate media outlets to make corrections or update stories that they publish and post on Twitter. Does your platform place a notification on posts for stories that have been corrected? If not, why not?

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. 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.

For now, we have focused these labeling efforts on areas where we believe there is the greatest risk of harm, including COVID-19, civic integrity, and manipulated media. However, we continue to explore additional ways in which we can provide additional context to people who use Twitter, in order to address potential harms associated with misinformation.

8. You claim you are not a publisher. At the same time you substitute your judgment for what can be published on your website for the editorial decision of traditional news outlets. The main example of this is the recent censoring of the New York Post story on Hunter Biden. How are you not engaging in editorial decision making by prohibiting a traditional news outlet from publishing a story on your site?

We issued the Distribution of Hacked Materials Policy in advance of the U.S. 2018 midterm elections to discourage and mitigate harms associated with hacks and unauthorized exposure of private information. Pursuant to these policies, on October 14, 2020, we took action on Tweets related to two articles published by the *New York Post* that, based on preliminary information, linked to content we determined to be in violation of our policies, including the Distribution of Hacked Materials Policy. Following our enforcement actions, we received significant feedback — both positive and negative — on the action.

After reviewing the feedback, we made changes within 24 hours to the policy to address concerns that there could be unintended consequences to journalists, whistleblowers, and others in ways that are contrary to Twitter's purpose of serving the public conversation. We also noted publicly that the only enforcement action available under the Distribution of Hacked Materials Policy was removal, which was no longer in alignment with new product capabilities, such as a label, that provide people with additional context.

Under Section 230 of the Communications Decency Act, services like Twitter are not considered publishers or speakers for hosting content by third parties. This protection allows us to act responsibly to promote healthy conversations by taking action against misinformation, abuse, harm and illegal activity that makes its way onto Twitter, and has been critical to promoting free

expression online. Eliminating Section 230 or prescribing reactionary government speech mandates will neither address concerns nor align with the First Amendment. Indeed, such actions could have the opposite effect, likely resulting in increased removal of speech, the proliferation of frivolous lawsuits, and severe limitations on our collective ability to address harmful content and protect people online. In addition, complex regulatory schemes or carveouts that favor already dominant market participants could stifle innovation and competition, ultimately giving consumers less choice online.

9. Has your company or any subsidiary entered into an agreement, either written or not, with a government to either limit access to certain information, or to not limit government officials posts or place warnings on those posts?

No, we have not entered into separate agreements with governments that would exempt them from complying with the Twitter Rules or enforce these rules differently with regards to specific government accounts.

Notwithstanding this, many countries, including the United States, have laws that may apply to Tweets and Twitter account content. In our continuing effort to make our services available to people everywhere, if we receive a valid and properly scoped request from an authorized entity, it may be necessary to withhold access to certain content in a particular country from time to time. Such withholdings will be limited to the specific jurisdiction that has issued the valid legal demand or where the content has been found to violate local laws.Upon receipt of requests to withhold content, we will promptly notify affected users unless we are prohibited from doing so (e.g., if we receive a court order under seal). We also clearly indicate within the product when content has been withheld and publish requests to withhold content on Lumen—unless, similar to our practice of notifying users, we are prohibited from doing so. The Transparency Report discloses information on government requests to remove content under local laws pursuant to our "Country Withheld Content" policy.

10. I asked if Twitter used a platform similar to the Tasks platform identified by Facebook. It was hard to hear your response but you did say that Twitter uses a similar platform.
>   **- What is the name of this platform?**
>   **- How does it work?**
>   **- What is it primarily used for?**
>   **- Does Twitter collect metadata from this platform?**
>   **- Does any**

Twitter uses a variety of methods and software to assist with businesses functions, including project management. For example, Twitter uses the software Jira to help with project management and issue tracking. Among numerous tasks, the software can help track issues logged in a range of contexts, including questions from employees to human resources or computer issues flagged to the IT Department.

11. Do you coordinate with any other company or outside group when you make decisions about content moderation?

Twitter does not coordinate with other entities when making content moderation decisions. However, we have partnerships with government agencies, nonprofits, and industry peers to facilitate information sharing to inform our policy and enforcement decisions.

For example, the National Center for Missing & Exploited Children, whose mission is to help find missing children, reduce child sexual exploitation, and prevent child victimization, is an important partner for Twitter and our industry peers. When we are made aware of content depicting or promoting child sexual exploitation, including links to images or content or third party sites where this content can be accessed, the material is removed without further notice and reported to NCMEC. While our general practice is to notify Twitter users when their content is reported to third-parties or law enforcement, we do not notify users when the reported content includes child sexual exploitation material. Furthermore, we participate in NCMEC's hash sharing database for industry and non-governmental organizations which consists of image and video hashes of known child sexual abuse material.

We also partner with nonprofits dedicated to child protection across the globe. In addition to our important relationship with NCMEC, Twitter is an active member of the Technology Coalition. This industry-led non-profit organization strives to eradicate child sexual exploitation by mentoring emerging or established companies, sharing trends and best-practices across industry, and facilitating technological solutions across the ecosystem. The Technology Coalition serves as an effective model because it gives companies the flexibility to create, test, and iterate across our diverse products and models.

12. Does Twitter receive any information from any other company or entity other than Twitter about posts and content moderation decisions?

Twitter has numerous partnerships that we rely on to better inform policy and decision making. In addition to the partnerships described above, Twitter is part of the Global Internet Forum to Counter Terrorism, which brings together industry, government, civil society, and academia to share information and collaborate to counter terrorist or extremist content online. Through the GIFCT, we have assembled a shared industry database of "hashes" or digital "fingerprints" for violent terrorist propaganda that spans more than 100,000 hashes. The database allows a company that discovers terrorist content on one of its sites to create a digital fingerprint and share it with the other companies in the forum, who can then use those hashes to identify such content on their services or platforms, review against their respective policies and individual rules, and remove matching content as appropriate or block extremist content before it is posted.

We also began to work with a small group of companies to test a new collaborative system to share URLs. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective, recognizing the resource constraints of smaller companies.

In order to safeguard the conversation regarding the 2020 U.S. election, we also have partnerships with leaders in civic tech, industry, and governments organizations, such as the National Association of Secretaries of State, National Association of State Election Directors, Department of Homeland Security, Federal Bureau of Investigation, Department of Justice, Office of the Director of National Intelligence, and elections officials across the country. We have also developed partnerships with news organizations, civil society, and others, which have been instrumental in informing policies and helping to identify potential threats regarding the integrity of the election conversation occurring on Twitter.

13. What procedure or policies does Twitter have in place to ensure content moderation is done in a objective manner?

Twitter does not use political viewpoints, perspectives, or party affiliation to make decisions, and we have taken several steps to ensure objective content moderation. For example, the Twitter Rules themselves are objective and not rooted in a particular ideology; the rules are focused on preventing harm and safeguarding the public conversation. In addition, we take a behavior-first approach to content moderation enforcement, meaning we look at how accounts behave before we look at the content they are posting. Moreover, we have invested in advancing procedural fairness, to facilitate impartial decision-making and provide avenues for individuals to appeal decisions if there has been a mistake.

14. In response to Senator Ernst you mentioned a tool for tracking all content moderation decisions.
- **What is the name of this tool?**
- **What information does it track?**
- **Who has access to this tool?**
- **Does it track who makes a content moderation decision?**
- **Do you collect metadata on this platform? Are you willing to share trends based on this metadata with Congress?**
- **Are Republicans subject to content moderation more than Democrats based on the information you have from this platform?**
- **You indicated there are safeguards in this system. Please outline and describe each step in the content moderation process, including the process for appeals.**

Twitter does not use political viewpoints, perspectives, or party affiliation to make content moderation decisions. We apply the Twitter Rules impartially. Twitter uses a variety of methods and software to assist with businesses functions, including project management. For example, Twitter uses the software Jira to help with project management and issue tracking, including in some cases where there may be potential violations of Twitter Rules. We work to be transparent about our enforcement, and have provided information about enforcement trends in the Twitter Transparency Center.

With regards to appeals, if an account was suspended or locked in error, an individual can appeal. First, the individual must log in to the account that is suspended and file an appeal. The individual must describe the nature of the appeal and provide an explanation of why the account is not in violation of the Twitter Rules. Twitter employees will typically engage with the account holder via email to resolve the appeal. Mistakes in enforcement — made either by a human or algorithm —

are inevitable, and why we strive to make appeals easier. We recognize that enhancing procedural fairness, including through a straightforward appeals process is a critical part of building consumer trust and we look forward to working with the Committee on this issue.

15. Why do you only place warnings on some stories published by traditional news sources and not go one step farther and provide more context on all posts from traditional news sources when the text in the post or the title of the post may not fully tell an accurate story?

Our policies focus on labeling misinformation in the contexts where there is the greatest risk of harm, including COVID-19, civic integrity, and manipulated media. However, we continue to explore additional ways in which we can provide additional context to people who use Twitter, in order to address potential harms associated with misinformation.

Questions from Senator Blackburn

1. During the November 3, 2020 election (before and after this date), did Twitter maintain any informal or formal lists of U.S. public officials who were specifically targeted for special monitoring of their Twitter posts?

During the election period, we used a combination of human and automated mechanisms to enforce our policies. For example, we reviewed Tweets reported as potential violations by the public, civil society partners, or government agencies. We also used automated systems to detect suspicious behaviors or identify potential violations of our rules. Twitter's enforcement teams prioritized the review of Tweets from the accounts of each of the presidential candidates and their campaigns and reviewed each to ensure compliance with our terms of service, beginning two weeks prior to election day.

2. During the November 3, 2020 election (before and after this date), did Twitter's enforcement teams prioritize the review of Tweets from the official and campaign accounts of any U.S. Senators for compliance with Twitter's terms of service?

As noted in the response to Question 1, we used a variety of mechanisms to enforce our policies during the election period. All people who use Twitter are governed by The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service, which collectively make up the "Twitter User Agreement." Our enforcement teams receive reports of Tweets and other media to review against our Terms of Service through a variety of channels, including through in-app reporting from the people who use our service to reports from government partners, political stakeholders, the press, and civil society.

3. In advance of the January 5, 2021 Georgia Senate run-off elections, are Twitter's enforcement teams prioritizing the review of Tweets from the official and campaign accounts of any U.S. Senators for compliance with Twitter's terms of service?

Twitter's work to safeguard the conversation regarding the 2020 U.S. election is ongoing; we continue to work to safeguard the conversation around the 2021 Georgia Senate run-off elections.

4. In advance of the January 5, 2021 Georgia Senate run-off elections, are Twitter's enforcement teams prioritizing the review of Tweets from the official and campaign accounts of any candidates for U.S. Senate for compliance with Twitter's terms of service?

Twitter's work to safeguard the conversation regarding the 2020 U.S. election is ongoing; we continue to work to safeguard the conversation around the 2021 Georgia Senate run-off elections.

5. Does Twitter's enforcement team prioritize the review of Tweets from any reporters from One America News for compliance with Twitter's terms of service?

All people who use Twitter are governed by The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service, which collectively make up the "Twitter User Agreement." Our enforcement teams receive reports of Tweets and other media to review against our Terms of

Service through a variety of channels, including through in-app reporting from the people who use our service to reports from government partners, political stakeholders, the press, and civil society.

6. Does Twitter's enforcement team prioritize the review of Tweets from any reporters from Fox News for compliance with Twitter's terms of service?

As noted in the response to Question 1, we used a variety of mechanisms to enforce our policies during the election period. All people who use Twitter are governed by The Twitter Rules and all incorporated policies, Privacy Policy, and Terms of Service, which collectively make up the "Twitter User Agreement." Our enforcement teams receive reports of Tweets and other media to review against our Terms of Service through a variety of channels, including through in-app reporting from the people who use our service to reports from government partners, political stakeholders, the press, and civil society.

7. Twitter placed a blue elections flag label on a November 5, 2020 clip from Fox News' Hannity Show quoting Rep. Jim Jordan. The flag stated, "This claim about election fraud is disputed." Upon clicking the notice, the user is directed to a page which reads, "Voter fraud of any kind is exceedingly rare in the US, election experts confirm." In the flagged clip, Congressman Jim Jordan discusses Justice Samuel Alito's decision ordering Pennsylvania counties to comply with a state directive to separate late ballots received after Election Day. Please explain why Twitter's election fraud label was imposed on a Tweet that discussed a Supreme Court decision ordering Pennsylvania officials to follow election laws.

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. 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. In the case referenced, we labeled the Tweet to provide individuals with additional context regarding the claims being made.

8. Does Twitter maintain a list of users that have a history of posting "misleading" or "false" speech as determined by Twitter's criteria?

When determining whether to take enforcement action, we may consider a number of factors, including whether the account holder has a history of violating our policies. If an account holder repeatedly violates our Rules then the penalty for violating our policies will increase. This includes requiring violators to remove the Tweet(s) and taking additional actions like verifying account ownership and/or temporarily limiting their ability to Tweet for a set period of time.

9. Does Twitter maintain a list of users with links to Chinese state-sponsored actors that have a history of posting "misleading" or "false" speech as determined by Twitter's criteria?

In August 2019, Twitter disclosed 936 accounts originating from within the People's Republic of China (PRC). These accounts were deliberately and specifically attempting to sow political discord in Hong Kong, including undermining the legitimacy and political positions of the protest

movement on the ground. Based on our intensive investigations, we have reliable evidence to support that this is a coordinated state-backed operation. Specifically, we identified large clusters of accounts behaving in a coordinated manner to amplify messages related to the Hong Kong protests.

As Twitter is blocked in PRC, many of these accounts accessed Twitter using VPNs. However, some accounts accessed Twitter from specific unblocked IP addresses originating in mainland China. The accounts represent the most active portions of this campaign; a larger, spammy network of approximately 200,000 accounts — many created following our initial suspensions — were proactively suspended before they were substantially active on the service. These accounts were disclosed in Twitter's comprehensive public archive of Tweets and media associated with suspected state-backed information operations on Twitter.

10. Based on previous cybersecurity breaches, has Twitter suspected in the past that Chinese state-sponsored actors have targeted Twitter accounts affiliated with Chinese dissidents for hacking or other internet attacks?

We have well-established relationships with law enforcement agencies active in this arena, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of State's Global Engagement Center. We look forward to continued cooperation with federal partners on the threats posed by state-backed information operations, because in certain circumstances only they have access to information critical to our joint efforts to stop bad faith actors.

11. Does Twitter's enforcement team prioritize the review of Tweets from any accounts affiliated with China's state propaganda outlets, which include the People's Daily, China Daily, China News Service, the Global Times, and CCTV, for compliance with Twitter's terms of service?

An important part of our work is providing people with context so that they can make informed decisions about the content they see on Twitter. Accordingly, in August 2020, we made the decision to add labels to the Twitter accounts of key government officials, with a focus on senior officials who are the voices of the state abroad, and accounts belonging to state-affiliated media entities, their editors-in-chief, and senior staff. We believe providing these labels are an important step, so that when people see an account discussing geopolitical issues from another country, they have context about its affiliation and who it represents. We are applying labels to the accounts that represent the five permanent members of the UN Security Council: China, France, Russian Federation, the United Kingdom, and the United States before expanding to a wider range of countries. Additionally, we do not permit news media entities controlled by state authorities to purchase advertisements. This policy extends to individuals reporting on behalf of or who are directly affiliated with such entities.

12. Of the $3.59 billion in revenue Twitter generated in 2019, how much ad revenue did Twitter generate from relationships with Chinese businesses or government agencies?

The Twitter annual report contains additional information about the company's revenue streams.

13. How much ad revenue did Twitter generate from Huawei for marketing the launch of its tablet, the Huawei Ascend Mate 7? For reference, this marketing campaign is found on

Twitter's "Success Stories" page:
https://marketing.twitter.com/en_apac/success-stories/huaweis-new-tablet-goes-global-with-promoted-tweets.

The Twitter annual report contains additional information about the company's revenue streams. As a global platform, Twitter provides advertising services to companies around the world. All ads must comply with Twitter policies, which among other things, prohibit ads from state media or that promote unacceptable business practices.

14. How much ad revenue did Twitter generate from Huawei for marketing the launch of its phone, the Huawei Mate 30 Pro 5G? For reference, this marketing campaign is found on Twitter's "Success Stories" page:
https://marketing.twitter.com/en_gb/success-stories/huawei-mate-30-launch.

The Twitter annual report contains additional information about the company's revenue streams. All ads must comply with Twitter policies, which among other things, prohibit ads from state media As a global platform, Twitter provides advertising services to companies around the world. All ads must comply with Twitter policies, which among other things, prohibit ads from state media or that promote unacceptable business practices.

15. How much ad revenue did Twitter generate from Huawei for marketing the @HuaweiMobile brand in advance of the GSMA Mobile World Congress, a global tech conference? For reference, this marketing campaign is found on Twitter's "Success Stories" page:
https://marketing.twitter.com/en_apac/success-stories/how-huawei-became-the-most-talked-about-brand-at-a-global-event.

The Twitter annual report contains additional information about the company's revenue streams. As a global platform, Twitter provides advertising services to companies around the world. All ads must comply with Twitter policies, which among other things, prohibit ads from state media or that promote unacceptable business practices.

1. The following questions examine how Twitter views its activities in moderating and directing its platform:

> **a. When Twitter hosts, unaltered, the material of a third party – as in the form of a basic Tweet – is Twitter acting as publisher?**

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case.

Section 230's protection has been critical to preserving free expression online, as well as promoting innovation. Looking forward, we encourage Congress to work with industry and civil society to build upon Section 230's foundation, whether it be through additions to Section 230, industry-wide self-regulation best practices, or a new legislative framework. We believe that the best way to address concerns with content moderation is to require the publication of moderation processes and practices, a straightforward process to appeal decisions, and best efforts around algorithmic choice, while protecting the privacy of individuals. We look forward to working with the committee to achieve these goals.

> **b. When Twitter blocks a Tweet, is it acting as a publisher?**

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case. Twitter does not block Tweets. In cases where an account violates the Twitter Rules, it may be subject to a range of enforcement actions, which can include account suspension or removal of a Tweet.

> **c. When Twitter intentionally limits the reach of a Tweet, is it acting as a publisher?**

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case. In cases where an individual violates the Twitter Rules, they may be subject to various enforcement actions, including actions designed to provide additional context to individuals about the Tweet. For example, in some circumstances where we do not remove content which violates the civic integrity policy, we may take other actions, including applying a label and/or warning message to the content where it appears in the Twitter product; showing a warning to people before they share or like the content; reducing the visibility of the

content on Twitter and/or prevent it from being recommended; and/or providing a link to additional explanations or clarifications, such as in a Twitter Moment or relevant Twitter policies.

d. When Twitter covers the face of a Tweet with a warning or other label written by Twitter, requiring the user to click through in order to access the content of that Tweet, is Twitter acting a publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case.

In cases where an individual violates the Twitter Rules, they may be subject to various enforcement actions. In certain cases involving a violation of the Twitter Rules by 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.

e. When Twitter labels a Tweet, instructing the user that the information in the Tweet is subject to controversy or is in dispute, is Twitter acting as a publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case.

In cases where an individual violates the Twitter Rules, they may be subject to various enforcement actions. 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.

f. When Twitter labels a Tweet, informing the user that Twitter has determined that the information in the Tweet is untrue, is Twitter acting as a publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case.

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.

g. When Twitter makes a judgment regarding the content of a tweet, and accordingly reduces or limits its dissemination, is Twitter acting as a publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case. In cases where an individual violates the Twitter Rules, they may be subject to various enforcement actions.

h. When Twitter conditions usage of its platform upon not sharing information, is Twitter acting as publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case. When individuals use Twitter, they agree to comply with the Twitter Rules, which are designed to prevent harm. For example, the Twitter Rules prohibit the promotion of terrorism or violent extremism.

i. When Twitter categorizes and organizes Tweets, deciding on their order or presence in a "Trending" list, or employs similar tools to suggest content, is Twitter acting as a publisher?

Under subsection (c)(1), Section 230 of the Communications Decency Act provides that neither providers nor the people who use our service are to "be treated as the publisher or speaker of any information provided by another information content provider." Whether a plaintiff's claim treats an entity as a publisher for purposes of Section 230 is a determination that a court would make based on the specific details of a case.

The goal of Trends is to provide individuals who use Twitter topics that are popular now, so they can discover emerging topics of discussion. Trends are, by default, tailored to an individual based on who they follow, their interest, and other factors. However, individuals can choose to see Trends that are not tailored for them by selecting a specific location, which identify popular topics among people in a specific geographic location.

2. Twitter recently limited the distribution and dissemination of reporting from the *New York Post* regarding Joe Biden, his family's business dealings, and his campaign.

a. **Has Twitter ever acted to limit the distribution of:**

    i. **News stories regarding President Trump's personal financial information, including but not limited to confidential tax records?**

The Twitter Rules prohibit posting other people's private information (such as home phone number and address) without their express authorization and permission. They also prohibit threatening to expose private information or incentivizing others to do so. News stories referencing the fact that a disclosure has been made, which does not directly share private information, would not fall under this prohibition. In addition, our Distribution of Hacked Materials policy prohibits distribution of hacked materials, but does not prohibit reporting on such materials.

    ii. **The disclosures of Edward Snowden, including but not limited to the illegally obtained material for which he is currently under federal investigation?**

Our Distribution of Hacked Materials was not in place at the time of the Snowden disclosures. On October 23, 2020, we revised the Distribution of Hacked Materials to state that we will no longer remove hacked content unless it is directly shared by hackers or groups directly associated with a hack.

    iii. **Stories regarding First Lady Trump, and personal conversations recorded without her knowledge or consent?**

The Twitter Rules prohibit posting other people's private information (such as home phone number and address) without their express authorization and permission. They also prohibit threatening to expose private information or incentivizing others to do so. News stories referencing the fact that a disclosure has been made, which does not directly share private information, would not fall under this prohibition.

    b. If the answer to the previous questions is "no," please explain, with specificity, the disparate treatment between these news items. If the answer is "yes," please state with specificity what actions were taken to limit distribution.

Twitter is constantly refining its rules and enforcement practices to better safeguard the public conversation and respond to the feedback from the public. For example, following the application of our Distribution of Hacked Materials Policy against specific content shared by @NYPost, we received significant feedback — both positive and negative — on the action.. After reviewing the feedback, we made changes within 24 hours to the policy to address concerns that there could be unintended consequences to journalists, whistleblowers and others in ways that are contrary to Twitter's purpose of serving the public conversation. We also noted publicly that the only enforcement action available under the Distribution of Hacked Materials Policy was removal, which was no longer in alignment with new product capabilities, such as the application of a label, that provides people with additional context.

3. Despite the unprecedented steps Twitter has taken to define for the average user which information is "false," and which information is to be relied upon, you were unable to answer

questions on the subject of voter fraud in your testimony before the Committee. Please answer the following questions regarding the public discussion of voter fraud:

    a. Does voter fraud exist?

Twitter does not tell people what information is true or false as it relates to civic processes. Our intention is to connect the dots of conflicting statements and show the information in dispute so people can judge for themselves.

    b. Are any of the executives at Twitter, including but not limited to those who are making content moderation policy decisions, experts in voter fraud?

We seek to employ individuals with a range of backgrounds, expertise, and skill sets.

    c. What sources of authority does Twitter rely upon in making determinations regarding the truthfulness of claims of voter fraud?

Twitter's enforcement of its civic integrity policy does not tell people what information is true or false. Our intention is to connect the dots of conflicting statements and show the information in dispute so people can judge for themselves. 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.

    d. Would the following statement violate Twitter's election information policies: "Absentee ballots remain the largest source of potential voter fraud"?

When making enforcement decisions, Twitter considers a range of behaviors, including the content of the Tweet itself, but also platform behavior. Thus, it is difficult to provide an answer to such hypothetical scenarios.

    e. Would the following statement violate Twitter's election information policies: "Voter fraud is particularly possible where third-party organizations, candidates, and political party activists are involved in handling absentee ballots"?

When making enforcement decisions, Twitter considers a range of factors. This can include the content of the Tweet itself, but also platform behavior. Thus, it is difficult to provide an answer to such hypothetical scenarios.

    f. In making determinations regarding the truth or verifiability of voter fraud claims, does Twitter employ corporate values, beliefs, priorities, or opinions when deciding what content is removed, republished, moderated, labelled, or otherwise promoted or demoted? What are those values?

The Twitter Rules are objectively enforced, without regards to political viewpoints, party affiliation, political ideology, or personal beliefs. As outlined in our testimony, in order to build trust, we fully support efforts to enhance transparency, procedural fairness, privacy, and algorithmic choice.

       g. Does Twitter make these determinations on a viewpoint-neutral basis?

In developing and enforcing our rules for the service we seek to be impartial. As part of this, Twitter supports and has invested in advancing procedural fairness to ensure decisions are made impartially and individuals have mechanisms to appeal in cases where they believe a mistake has been made.

4. The day before you testified before the Committee, you and I spoke on the telephone and you told me that Twitter was committed to transparency with regard to its content moderation policies and enforcement. Accordingly, the following questions relate to Twitter's enforcement of its content moderation policies. For this question and its subparts, please construe "content moderation policies" broadly, including decisions regarding the position or order in which content is displayed, the position or order in which users or content appear in searches, whether users or content are promoted or demoted, and all other modifications of content, such as flagging, qualifying, labelling, and denoting.

       a. In your enforcement of content moderation policies:

              i. How many times has Twitter blocked, flagged, censored, limited the reach of, or otherwise affected the tweets, posts, or content of Republican candidates for office between January 1, 2016 and November 24, 2020? Please include enforcements that were later reversed or recognized to be in error.

Twitter does not track the political affiliation of the people who use our service. Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices.

              ii. How many times has Twitter blocked, flagged, censored, limited the reach of, or otherwise affected the tweets, posts, or content of Democratic candidates for office between January 1, 2016 and November 24, 2020? Please include enforcements that were later reversed or recognized to be in error.

Twitter does not track the political affiliation of the people who use our service. Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices.

              iii. How many times has Twitter blocked, flagged, censored, limited the reach of, or otherwise affected the tweets, posts, or content of Republican elected officials between January 1, 2016 and November 24,

2020? Please include enforcements that were later reversed or recognized to be in error.

Twitter does not track the political affiliation of the people who use our service. Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices.

      iv. **How many times has Twitter blocked, flagged, censored, limited the reach of, or otherwise affected the tweets, posts, or content of Democratic elected officials between January 1, 2016 and November 24, 2020? Please include enforcements that were later reversed or recognized to be in error.**

Twitter does not track the political affiliation of the people who use our service. Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices.

    b. For each instance marked and counted in the previous question, please provide the name of the account affected, the content of the material affected, and the specific reason for the enforcement action.

Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices.

    c. In your enforcement of election information content moderation policies, are decisions made and executed by humans, or by algorithm with preset code?

Twitter uses a combination of automation and human review to adjudicate reports of violations and make determinations on whether activity violates our rules.

      i. **Has anyone been tasked with keeping track of which content or users are affected by these policies?**

We have made efforts to be as transparent as possible about our enforcement decisions. Because Twitter is a public platform, enforcement decisions are often apparent to the public. In some contexts, we have also taken steps to provide additional datasets to researchers. For example, in [October 2018](#), we published the first comprehensive archive of Tweets and media associated with known state-backed information operations on Twitter. This one of a kind resource, used by researchers, journalists and experts around the world, now spans operations across 15 countries, including more than nine terabytes of media and 200 million Tweets.

      ii. **If not, has anyone been tasked with keeping track of other information of this type, in other contexts?**

We have made efforts to be as transparent as possible about our enforcement decisions, providing information about enforcement through the Twitter Transparency Center. Apart from this, we fully support efforts to increase transparency around content moderation. We believe that companies like

Twitter should publish their moderation process, and should be transparent about how cases are reported and reviewed, how decisions are made, and what tools are used to enforce.

          iii. If so, please explain the difference in the maintenance of records.

          d. In drafting Twitter's election information content moderation policies and enforcing those same policies with regard to the 2020 elections, did Twitter collaborate with, confer with, or defer to any outside individuals or organizations? If so, please list the individuals and organizations and state the nature of their relationship with Twitter.

As part of our civic integrity efforts, we have developed partnerships that allowed us to share information, gather input from experts, and better gain context on how misinformation was being spread and impacting the public conversation. These partnerships included leaders in civic tech, our peers, federal, state, and local governments organizations (e.g., National Association of Secretaries of State, National Association of State Election Directors, Department of Homeland Security, Federal Bureau of Investigation, Department of Justice, Office of the Director of National Intelligence, and elections officials across the country), news organizations, and civil society, among others.

          i. Does Twitter take account of the political, philosophical, or ideological orientation or reputation of those sources with which it cooperates in executing its election information content moderation policies?

Twitter does not use political viewpoints, perspectives, or party affiliation to make any decisions.

          e. The following questions relate to the individuals with supervisory authority who are responsible for the formulation and implementation of content moderation policies.

          i. Among those individuals with supervisory authority who make substantive decisions regarding content moderation policy, how many:

          1. Self-identify or are registered as Democrats?

Twitter does not ask employees to disclose their political affiliation.

          2. Self-identify or are registered as Republicans?

Twitter does not ask employees to disclose their political affiliation.

          3. Would identify themselves as "liberal?"

Twitter does not ask employees to disclose their political affiliation.

          4. Would identify themselves as "conservative?"

Twitter does not ask employees to disclose their political affiliation.

5. Have donated to:

a. The Democratic Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

b. A candidate running for office as a Democrat?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

c. A cause primarily affiliated with or supported by the Democratic Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

d. A cause primarily affiliated with or supported by liberal interest groups?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

e. A political action committee primarily advocating for the Democratic Party, Democratic candidates or office-holders, or causes primarily supported by the Democratic Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

f. The Republican Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

g. A candidate running for office as a Republican?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

h. A cause primarily affiliated with or supported by the Republican Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

       i. A cause primarily affiliated with or supported by conservative interest groups?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

       j. A political action committee primarily advocating for the Republican Party, Republican candidates or office-holders, or causes primarily supported by the Republican Party?

Twitter does not ask employees to publicly disclose political contributions, apart from disclosures required under existing laws.

      6. Worked on or volunteered for a Democratic campaign?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

      7. Worked on or volunteered for a Republican campaign?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

      8. Worked on, interned for, or volunteered for a Democratic legislator, State or federal?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

      9. Worked on, interned for, or volunteered for a Republican legislator, State or federal?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

      10. Worked on or interned for a Democratic administration or candidate?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

11. Worked on or interned for a Republican administration or candidate?

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology.

ii. Following a Constitution Subcommittee in April 2019 in which a Twitter executive testified, I submitted questions similar to those in 4.d.1 above. Twitter responded by refusing to answer the questions asked and instead stating that it "does not use political ideology as a factor in its hiring decisions."

1. Does Twitter believe that it can adopt and enforce viewpoint neutral content moderation rules if the individuals developing and enforcing those rules overwhelmingly identify as Democrats and progressives?

Yes. Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology. To help ensure that decisions are made objectively, we have invested in developing fair processes.

2. When hiring individuals to make substantive decisions regarding content moderation policies or to enforce those policies, does Twitter consider any personal characteristics or demographic information, including but not limited to race, ethnicity, sex, gender, or sexual orientation, to ensure that Twitter's content moderation policies and enforcement actions reflect diverse views?

We believe that being a diverse and inclusive company is key to serving the public conversation. By 2025 we have set a goal to have at least 25% of our overall U.S. workforce be underrepresented minorities. We consistently share our progress with the public every quarter. Our most recent report reflects U.S. ethnicity for all roles and levels at 41.4% White, 28.4% Asian, 6.3% Black, 5.2% Latinx, 3.7 Multiracial, and less than 1% Indigenious.

a. If the answer is "yes", please identify the characteristics or demographic information Twitter considers and explain how that characteristic promotes view-point neutrality but ensuring diverse political ideologies does not.

As a company, Twitter is focused on advancing the principle of procedural fairness in our decision-making across the board. Procedural fairness at Twitter means we ensure that all decisions are made without using political viewpoints, party affiliation, or political ideology, whether related to automatically ranking content on our service or how we develop or enforce the Twitter Rules. Our Twitter Rules are not based on ideology or a particular set of beliefs. We believe strongly in being impartial, and we strive to enforce our Twitter Rules fairly.

In addition, we strive to give people an easy, clear way to appeal decisions we make that they think are not right. Mistakes in enforcement — made either by a human or algorithm — are inevitable, and why we strive to make appeals easier. We believe that all companies should be required to provide a straightforward process to appeal decisions. This makes certain people can let us know when we do not get it right, so that we can fix any mistakes and make our processes better in the future.

> **f. A top Twitter executive—Vijaya Gadde—has been praised as the leader of the "resistance" and the "lead architect of the policy approach that led Twitter to clamp down on . . . President Donald Trump."[1]**
>
>> **i. Does this description of the top policy executive at Twitter reflect the orientation and priorities of the election information content moderation team as a whole?**

Decisions at Twitter are made without regard to political viewpoints, party affiliation, or political ideology. We focus our content moderation efforts to reduce harm and facilitate the public conversation.

>> **ii. Is Ms. Gadde ultimately responsible for what is determined to be "true" and "untrue" regarding claims made by users on Twitter?**

Twitter does not tell people what information is true or false as it relates to information concerning COVID-19, civic integrity, or manipulated media. Our intention is to connect the dots of conflicting statements and show the information in dispute so people can judge for themselves. 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.

>> **iii. If she is not ultimately responsible, who is?**

No one individual makes content decisions at Twitter. As Twitter's Chief Executive Officer Jack Dorsey Tweeted in May 2020: "Fact check: there is someone ultimately accountable for our actions as a company, and that's me." We fully support efforts to enhance transparency and procedural fairness, so that individuals have trust in our company and the platforms they use. We also support efforts to increase competition and innovation, so that individuals have choices if a platform fails to earn their trust.

---

[1]https://www.politico.com/news/magazine/2020/10/28/twitter-vijaya-gadde-free-speech-policies-technology-social-media-429221

<u>**Questions from Senator Grassley**</u>

1. How do your users hold you and your companies accountable for your content moderation practices? Congress has the benefit of bringing you before this Committee to answer questions. But when one of my constituents thinks his or her speech was wrongfully moderated or fact-checked, why should they have any faith that their objection will even be heard?

As a company, Twitter is focused on advancing the principle of procedural fairness in our decision-making across the board. We strive to give people an easy, clear way to appeal decisions we make that they think are not right. Mistakes in enforcement — made either by a human or algorithm — are inevitable, and why we strive to make appeals easier. We believe that all companies should be required to provide a straightforward process to appeal decisions. This makes certain people can let us know when we do not get it right, so that we can fix any mistakes and make our processes better in the future. We fully support efforts to increase procedural fairness and look forward to working with your office on this issue.

2. Do you agree that users should be entitled to due process when content they post is taken down or moderated? And do you agree that there could be more transparency by your companies in explaining why certain speech or content is moderated?

We agree that increased transparency and procedural fairness are key to address concerns with content moderation. We believe that all companies should be required to provide a straightforward process to appeal decisions. In addition, we believe that all companies should publish their content moderation process, and be transparent about how cases are reported and reviewed, how decisions are made, and what tools are used to enforce rules.

1. Does Twitter have a policy prohibiting its employees from coordinating content moderation decisions with outside companies such as Facebook or Google, where such moderation is not strictly required by law?

Twitter does not coordinate its content moderation decisions with outside entities. However, Twitter has numerous partnerships that we rely on to better inform decision making and facilitate information sharing. For example, we share information in three critical areas: combatting child sexual exploitation, prohibiting terrorism and violent extremism, and safeguarding the conversation about the U.S. election.

The National Center for Missing & Exploited Children is a nonprofit whose mission is to help find missing children, reduce child sexual exploitation, and prevent child victimization. NCMEC is an important partner for Twitter and our industry peers. When we are made aware of content depicting or promoting child sexual exploitation, including links to images or content or third party sites where this content can be accessed, the material is removed without further notice and reported to NCMEC. We also partner with nonprofits dedicated to child protection across the globe. In addition to our important relationship with NCMEC, Twitter is an active member of the Technology Coalition. This industry-led non-profit organization strives to eradicate child sexual exploitation by mentoring emerging or established companies, sharing trends and best-practices across industry, and facilitating technological solutions across the ecosystem.

In addition, Twitter is part of the Global Internet Forum to Counter Terrorism, which brings together industry, government, civil society, and academia to share information and collaborate to counter terrorist or extremist content online. Through the GIFCT, we have assembled a shared industry database of "hashes" or digital "fingerprints" for violent terrorist propaganda that spans more than 100,000 hashes. The database allows a company that discovers terrorist content on one of its sites to create a digital fingerprint and share it with the other companies in the forum, who can then use those hashes to identify such content on their services or platforms, review against their respective policies and individual rules, and remove matching content as appropriate or block extremist content before it is posted.

We have also begun to work with a small group of companies to test a new collaborative system to share URLs. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective, recognizing the resource constraints of smaller companies.

Furthermore, in order to safeguard the conversation regarding the 2020 U.S. election, we have critical partnerships with leaders in civic tech, industry, and government organizations, such as the National Association of Secretaries of State, National Association of State Election Directors, Department of Homeland Security, Federal Bureau of Investigation, Department of Justice, Office of the Director of National Intelligence, and elections officials across the country. We also have

partnerships with news organizations, civil society, and others, which have been instrumental in informing policies and helping to identify potential threats regarding the integrity of the election conversation occurring on Twitter.

2. Please provide a list of all hashtags, URLs, and real-world individuals that Twitter has banned across its platform.

Our biannual [Twitter Transparency Center](#) highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices. In addition, in 2018, we published the first comprehensive archive of Tweets and media associated with known state-backed information operations on Twitter. This one of a kind resource, used by researchers, journalists and experts around the world, now spans operations across 15 countries, including more than nine terabytes of media and 200 million Tweets.

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.

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 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 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.

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.

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.

<u>**Questions from Senator Hirono**</u>

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.