The Honorable Jan Schakowsky (D-IL)

1. Please provide our committee with a list within 30 days of all the Universities, think tanks, NGOs, publications, journalists, activists, and activist groups that Twitter currently supports financially or has provided more than $50,000 in funding to in the past five years. This should include all grants or other payments paid through a third party that ultimately produced reports, studies, editorials, or other publicly available content.

   RESPONSE: Twitter is committed to corporate responsibility, sustainability, and philanthropy. We strive to unite philanthropy with our business objectives and be good stewards in the communities where we work and live. Our 2020 Global Impact Report, details this work.

2. Does Twitter share data on disinformation with independent scholars or researchers? If so, what information is shared, with whom is it shared, and what is the process for sharing it?

   RESPONSE: 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. Twitter makes data available via an API for the purposes of study. Moreover, to support the work of academics, we have launched a dedicated Academic Research product track, which offers 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, 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.

   We have also decided 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.

3. Clarify who is responsible for Spanish-language content moderation in the U.S. Please indicate which specific team is responsible for that work and share the title of the manager who most directly carries that responsibility. In parallel, please also indicate the title of the manager who is responsible for efforts to counter Spanish-language disinformation in the U.S. and the name of the relevant team working on this issue. In both instances, please identify where these teams and managers sit in the organizational chart and to whom they report.

   a. Furthermore, clarify whether Spanish-language content in the U.S. is a specifically identified and assigned responsibility. Do the teams and managers involved have specific responsibility for Spanish-language content in the U.S. or are they responsible for all Spanish-language content on the platform globally?

RESPONSE: Twitter’s Trust & Safety Team is responsible for creating and overseeing content policies, while our dedicated Twitter Service team leads our enforcement work. We have content moderators with Spanish language capability, including training on culturally relevant context, who can enforce the Twitter Rules. In addition, we partner with a range of government and civil society organizations to ensure we are specifically addressing the needs of Spanish-speaking communities in the United States.

For example, as part of work around COVID-19, Twitter has worked to ensure that Spanish-speaking communities have access to credible information. We launched a home timeline prompt in 16 countries in seven languages, including English and Spanish in the US, to elevate reliable, up-to-date information about COVID-19 vaccinations to the top of people’s timelines. The prompts were available April 26 to May 3, 2021, aligning with World Immunization Week, and linked to localized vaccine information from public health organizations and experts. In addition, our enforcement teams have taken action on misinformation in Spanish that violates our COVID-19 policy.

Similarly, 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 worked with @USAGovEspanol, @NALEO, @HispanicFed, @MALDEF, @LULAC, and @WeAreUnidos to promote voter education and misinformation pre-bunking targeted at Spanish-speaking communities.

4. Phrases and symbols associated with minority communities are systematically more likely to be flagged as hate speech on social media platforms, regardless of how benign
the content is. The Sikh community has noticed the targeted removal of benign posts that reference Sikh religious artifacts, political speech, and community organizing. How will your platform create channels of recourse that will prevent the structural and systematic suppression of content from minority communities?

RESPONSE: If there are specific examples on Twitter of potential improper enforcement of symbols associated with minority communities, please provide additional information to our team for analysis.

5. Children and young adults who are members of minority communities are plagued by uniquely impactful hate speech and cyberbullying on social media platforms. While more than 15% of youth social media users experience cyber bullying, people of color receive disproportionately more vitriolic messages that can have substantial impact on emotional and cognitive development. How has Twitter sought to provide recourse to individuals facing online harassment at the intersection of cyberbullying and racial/ethnic hate speech?

RESPONSE: Twitter has taken a variety of steps to combat abusive behavior and hateful conduct on our service, which we recognize often disproportionately impacts people of color. While we still have work to do in this area, we have taken significant steps to ensure that all individuals feel safe when using our service. Our **Abusive Behavior** policy prohibits targeted harassment of someone, or incite others to do so. In addition, our **Hateful Conduct Policy** prohibits promotion of violence against people on the basis of protected categories. This includes wishing or calling for serious harm against a person or group or people, or targeting individuals with the intent to harass with references to violent events or types of violence where protected categories were the primary victims. In December 2020, we expanded this policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin. Moreover, our **Glorification of Violence Policy** prohibits content that glorifies or praises historical acts of violence and genocide.

In addition to combatting hateful and abusive content, we champion media literacy among our younger generations by providing critical information and skills about how to navigate an increasingly complicated media environment. We have a global partnership with UNESCO on media and information literacy where our users and our product benefit from their expertise. We also work in collaboration with the National Association for Media Literacy, a coalition of educators and advocates, on how media and information literacy can be a defense — particularly in our young people — against disinformation and the spread of harmful content.

6. How often do policy compliance standards change on Twitter, and who is involved in creating these policies? What is the role of international policy teams in determining global policies? How do international policies impact the policies of other nations? What safeguards do you have to prevent the limitations from one country, such as India, from impacting users based in the United States?
RESPONSE: The global Twitter Trust & Safety team develops the policies governing our service. This team also consults with our Trust & Safety Council, which includes safety experts from across the globe. We are continually working to update, refine, and improve our enforcement and our policies, informed by in-depth research around trends in online behavior and feedback from the people who use Twitter. Our content moderation policies, encapsulated in the Twitter Rules, apply globally. In some cases, we may take action on content that is illegal in a particular country, but is not illegal in other countries and does not violate our global rules. Specifically, if we receive a valid legal demand or content violates local law, in some cases we may be required to withhold access to certain content in a particular country. To protect the public interest, we publish these requests to the Lumen Database and in aggregate in the Twitter Transparency Report.

7. Can you explain what exactly you have done to ensure Twitter’s tools — algorithms, recommendation engines, and targeting tools — are not amplifying conspiracy theories and disinformation, connecting people to dangerous content, or recommending hate groups or purveyors of disinformation and conspiracy theories to people? For example, can you provide detailed answers on some of the Capitol riot suspects’ use history, to include the following:

a. Were any of the conspiracy theorists or other purveyors of electoral misinformation and Stop the Steal activity recommended to them as people to follow? Were their feeds curated to show more Stop the Steal and other conspiracy theory tweets than authoritative sources?

RESPONSE: Prior to the attack on the Capitol, Twitter had suspended several groups reportedly involved in the attack, including the Proud Boys, Oathkeepers, and Three Percenters, under our Violent Organizations Policy. In addition, in July 2020, we took action to deamplify QAnon content, including by preventing URLs associated with QAnon from being shared on Twitter, working to deamplify QAnon content from our search feature, no longer recommending Tweets with QAnon content, and blocking QAnon topics and terms from appearing in the Trending section. Following the attack, we began permanently suspending accounts that were primarily dedicated to sharing QAnon content, resulting in the removal of over 140,000 accounts.

Long-term, we are continuing to innovate to identify new ways to address misinformation more broadly. For example, in January, we launched the “Birdwatch” pilot, a community-based approach to misinformation. Birdwatch is expected to broaden the range of voices involved in tackling misinformation, and streamline the real-time feedback people already add to Tweets. We hope that engaging diverse communities here will help address current deficits in trust for all. We expect data related to Birdwatch will be publicly available at Birdwatch Guide, including the algorithm codes that power it.
8. How will Twitter resist the increasing demands for targeted censorship of minority communities by the Indian Government on its platform in India?

RESPONSE: Twitter is committed to protecting democratic values and human rights in India, including freedom of expression, and working constructively with the Government of India. During the unrest that occurred in Delhi during the Farmers Protest on January 26, 2021, and its aftermath, we took enforcement actions to protect the conversation happening on the service. As part of this, took action on hundreds of accounts that violated the Twitter Rules, particularly inciting violence, abuse, wishes of harm, and threats that could trigger the risk of offline harm. Despite requests from the government to remove content, we did not take action on accounts that consist of news media entities, journalists, activists, and politicians. To do so, we believe, would have violated their fundamental right to free expression.

As we receive more requests to withhold content from the Government of India within the country, we will continue to engage openly and constructively based on our overarching commitment to serve the public conversation during the global health crisis posed by COVID. We notify the account holder directly so they are aware that we have received a legal order pertaining to the account by sending a message to the email address associated with the account(s), if available, unless we are prohibited from doing so. We also clearly indicate on Twitter when content has been withheld and publish requests to withhold content on Lumen database — a level of transparency that is unique in the industry, unless, similar to our practice of notifying users, we are prohibited from doing so. Globally there is a concerning trend towards regulatory action that could result in increased instances of censorship and internet shutdowns, threatening the Open Internet. This is an issue that is far bigger than any one company and requires a concerted effort by all stakeholders to ensure this trend does not accelerate.

9. Is there an identified level of anti-Democratic regulations that would make it untenable for Twitter to continue its operations within a country?

RESPONSE: Defending and respecting the user’s voice is one of our core values at Twitter, which is grounded in our commitment to freedom of expression, privacy, and user safety. Our core mission is defending open, public conversation. We have a designated human rights function at Twitter that seeks to embed a human rights lens across our processes, product, and policies. We regularly assess the regulatory environment in countries in which we operate to ensure that we are able to fulfill our core mission and protect our employees and the people who use Twitter.

10. On several occasions you have touted a decentralized version of Twitter, such as “Bluesky.” How do you envision interventions for disinformation taking place on a distributed version of Twitter, if at all, and what business model, if any, would Twitter
contemplate for such a version? How far along is it, and how open is it in its conception, whether in code or in participation by other software developers and organizations?

RESPONSE: Bluesky is an independent community of open source architects, engineers, and designers, whose goal is to develop open and decentralized standards for social media. This community has already completed an initial review of the ecosystem around protocols for social media to aid this effort. Currently, we are focused on hiring the project lead, who will be responsible for taking this work forward, independent of Twitter. We will continue to provide financial support, and in time hope others will also do so. The hope is that Bluesky can serve as a repository for Twitter and other companies to contribute to and access open standards that promote healthy conversation and ultimately provide individuals greater choice.

11. Does Twitter believe its labels and other restrictions on Trump and other tweets that shared election disinformation were effective? How exactly do you measure that effectiveness? Why, or why not? What criteria was used to decide on these sanctions and who applied them?

RESPONSE: Twitter has conducted research aimed at identifying the needs of people who use Twitter, including how to address potential misinformation on the service and how to ensure that people have access to credible information to inform their viewpoints. A key piece of feedback we got from the people who use our service is that we should not determine the truthfulness of Tweets; 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, we continue to explore ways to expand our enforcement options beyond a binary choice of leaving content up or taking it down. Consistent with these efforts, 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.

We continue to assess our efforts surrounding the 2020 election. A preliminary review of Tweets from October 27, 2020 to November 11, 2020 found that 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. Approximately 74% of the people who viewed those Tweets saw them after we applied a label or warning message.

In addition, we are exploring other ways to promote healthy conversation on our service. For example, we are rolling out prompts that encourage people to reconsider a potentially harmful or offensive reply. Based on an initial pilot test last year, we found that after such prompts, 34% of people revised their initial reply or decided to not send their reply at all.
and after being prompted once, people composed, on average, 11% fewer offensive replies in the future.

12. Following the deadly siege on the US Capitol on January 6th, you introduced a detailed strike system specifically for civic integrity policy. Has Twitter applied this new policy since its creation? And do you intend to expand the strike system to other problem areas, such as COVID-19 misinformation?

**RESPONSE:** Our civic integrity policy’s strike system remains in effect, and we have adopted a similar approach to our COVID-19 misinformation policy. Repeat violations of the COVID-19 misinformation policy and the civic integrity policy are enforced against on the basis of the number of strikes an accounts had accrued for violations of this policy. Specifically:

- 1 strike: No account-level action
- 2 strikes: 12-hour account lock
- 3 strikes: 12-hour account lock
- 4 strikes: 7-day account lock
- 5 strikes: Permanent suspension

13. Unlike other online platforms, Twitter does not have an explicit policy to categorically ban or address Holocaust denial for purposes of content rules. As you reiterated in your testimony, Holocaust denial does not necessarily violate Twitter’s “hateful conduct” policy against “the dehumanization of a group of people based on their religion, caste, age, disability, serious disease, national origin, race, or ethnicity” or its misinformation policy. Will Twitter adopt an explicit policy against Holocaust denial or add Holocaust denial expressly to other content removal policies? If so, when? If not, why not?

**RESPONSE:** Twitter’s Hateful Conduct Policy prohibits a wide range of behavior, including making references to violent events or types of violence where protected categories of people were the primary victims, or attempts to deny or diminish such events. Twitter also has a robust policy against glorification of violence and takes action against content that glorifies or praises historical acts of violence and genocide, including the Holocaust. We are committed to expanding our approach in this area and will have more to share on this soon.

14. Little to no information has been released by Twitter regarding why or how #Sikh and #Sikhism were blocked multiple times throughout 2020. What procedures does your platform follow to remove or block hashtags?

**RESPONSE:** Our understanding is that industry peers may have blocked or removed these hashtags in 2020, although we are not aware of specific actions that Twitter has taken related to this content. If there are specific examples on Twitter of potential
improper enforcement on #Sikh or #Sikhism being blocked, please provide additional information to our team for analysis.

   a. Were automated systems responsible for the initial takedowns of these hashtags or was human decision-making involved in these processes?

**RESPONSE:** Our teams are not aware of these takedowns on Twitter, and we welcome additional information on this issue.

15. Your testimony cited Twitter’s success at removing COVID-19 and vaccine-related disinformation. At the same time, white supremacist content, racist, and other hate content is running rampant on the platform. Is there anything inherently more difficult about policing white supremacist, racist, and other hate content compared to health misinformation and disinformation? Why?

**RESPONSE:** Hateful conduct has no place on Twitter. Under our [Hateful Conduct](https://help.twitter.com/en/rules-and-policies/hateful-conduct) policy, hateful imagery is not permitted within live video, account bio, profile or header images. We consider hateful imagery to be logos, symbols, or images whose purpose is to promote hostility and malice against others based on their race, religion, disability, sexual orientation, gender identity or ethnicity/national origin. Additionally, sending an individual unsolicited hateful imagery is a violation of our [Abusive Behavior](https://help.twitter.com/en/rules-and-policies/abusive-behavior) policy. Our [Violent Organizations Policy](https://help.twitter.com/en/rules-and-policies/violent-organizations) also prohibits violent organizations, including violent extremist groups, from using Twitter. Under this policy, we have removed over 200 groups, half of which have links to white supremacy.

**The Honorable Bobby L. Rush (D-IL)**

1. Social media platforms have been used to suppress the ability of many users to exercise their civil rights both online and offline. What systems have been implemented on your platforms to prevent this from happening? Do you have any data as to how successful these systems are?

**RESPONSE:** Twitter has taken several steps to ensure that our products and policies safeguard civil rights. We have a Civil Rights Executive Working Group that is dedicated to bringing civil rights considerations into various parts of business, training, and policy decisions. In addition, we work with key civil society groups and organizations both on an ad hoc basis and through our Trust & Safety Council. These partnerships shape the development of our products and policies. We also believe it is critical that our efforts take into account the views and feedback of those who use Twitter, which is why we seek input from the public on our efforts—including, most recently, around our world leaders policy.

We have taken significant steps to ensure that all individuals, including those from vulnerable communities, feel safe when using our service. Our [Abusive Behavior](https://help.twitter.com/en/rules-and-policies/abusive-behavior) policy prohibits targeted harassment of someone, or incite others to do so. In addition, our
Hateful Conduct Policy prohibits promotion of violence against people on the basis of protected categories. This includes wishing or calling for serious harm against a person or group or people, or targeting individuals with the intent to harass with references to violent events or types of violence where protected categories were the primary victims. In December 2020, we expanded this policy to prohibit language that dehumanizes people on the basis of race, ethnicity, or national origin. Moreover, our Glorification of Violence Policy prohibits content that glorifies or praises historical acts of violence and genocide.

The Honorable Anna G. Eshoo (D-CA)

1. You testified that the tweet I referenced by Robert F. Kennedy, Jr. that suggests that Hank Aaron died from the COVID-19 vaccine does not violate your policies. Multiple fact-checking organizations debunked this claim. The tweet seems to clearly violate your COVID-19 misinformation policy.

   a. Does this tweet violate your policies? Why or why not?

      RESPONSE: This Tweet was sent prior to the expansion of our COVID-19 enforcement policy, which included labeling of Tweets. We do not apply our policies retroactively, and thus did not take enforcement action on this content.

   b. If so, why have you not removed the tweet?

      RESPONSE: We do not apply our policies retroactively, and thus did not take enforcement action on the Tweet.

   c. Have you taken any actions to reduce the amplification of this tweet?

      RESPONSE: We have taken no action to reduce the amplification of the Tweet.

   d. How many people have seen this tweet (and related retweets or quote tweets)?

      RESPONSE: This tweet has received approximately 9,100 retweets.

   e. Do you ban repeat offenders of your COVID-19 misinformation policy? If so, under what circumstances?

      RESPONSE: Repeated violations of the COVID-19 misinformation policy are enforced against on the basis of the number of strikes an account has accrued for violations of this policy:

      ● 1 strike: No account-level action
      ● 2 strikes: 12-hour account lock
      ● 3 strikes: 12-hour account lock
● 4 strikes: 7-day account lock
● 5 or more strikes: Permanent suspension

If we determine that an account is primarily dedicated to Tweeting or promoting a particular misleading narrative (or set of narratives) about COVID-19, this would also be grounds for suspension.

f. Have you taken any adverse actions against tweets from Robert F. Kennedy, Jr. or against his account?

**RESPONSE:** We have taken enforcement action against this account for violations of our COVID-19 misinformation policy.

2. Leading up to the November elections, Twitter prompted users to quote tweets rather than simply retweet, removed recommendations in users’ timelines and notifications, and added context to “Trends” in the “For You” part of its platform.

a. What was the impact of each of these changes on reducing the spread of misinformation?

**RESPONSE:** In the weeks leading up to and during election week, we implemented significant product changes intended to increase context and encourage more thoughtful consideration before Tweets are amplified. As part of this effort, 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. As part of our initial analysis, 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 short, this change may have slowed the spread of misleading information by virtue of an overall reduction in the amount of sharing on the service.

b. For any of the changes that reduced misinformation, why did Twitter reverse the change?

**RESPONSE:** The product changes implemented in advance of the election were intended to be temporary measures. On December 16, 2020, we halted prompting of Quote Tweets, and reenabled standard Retweet behavior. We hoped this change would encourage thoughtful amplification and also increase the likelihood that people would add their own thoughts, reactions, and perspectives to the conversation. However, we observed that prompting Quote Tweets didn’t appear to increase context: 45% of additional Quote Tweets included just a single word and 70% contained less than 25 characters. You can find additional information about our product changes in advance of the election and their impact [here](#).
3. For election disinformation and COVID-19 misinformation, Twitter partially relies on labeling tweets. Some are skeptical that labels are effective in debunking false or misleading claims.

   a. What percentage of users engage (e.g., read or click on) the additional information or links?

   RESPONSE: Engagement rates with labels vary significantly. We are looking for opportunities to make the labels themselves more informative and visible, and further direct people to additional information linked from these labels.

   b. Has Twitter tested the efficacy of labeling misinformation in improving users’ understandings of issues? If so, what is the result of this testing?

   RESPONSE: We have conducted research aimed at identifying the needs of people who use Twitter, including how to address potential misinformation on the service and how to ensure that people have access to credible information to inform their viewpoints. A key piece of feedback we got from the people who use our service is that we should not determine the truthfulness of Tweets; 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, we continue to explore ways to expand our enforcement options beyond a binary choice of leaving content up or taking it down. Instead, we are looking at ways to ensure that people can access additional information to inform their viewpoints.

   c. How do labels impact user engagement (e.g., likes and retweets) with labeled tweets?

   RESPONSE: When we label content, we typically reduce the visibility of the Tweets, meaning we will not amplify the Tweets on a number of surfaces across Twitter. We also typically alert people with a warning in cases where they seek to engage with a Tweet that has been labeled for misinformation, and in some cases disable engagement altogether.

4. On March 24th attorneys general from 12 states urged you to do more to combat COVID-19 misinformation. What changes will you make in response to the request?

   RESPONSE: We have policies in place to address vaccine misinformation and have invested in efforts to increase access to credible information on our service. On March 18, 2020, we moved quickly to implement a COVID-19 misinformation policy for content that goes directly against guidance from authoritative sources of global and local public health information. Subsequently, in December, we expanded our COVID-19 misleading information policy to allow the removal of Tweets that advance harmful, false, or misleading narratives related to COVID-19 vaccines in particular. We are currently prioritizing the removal of the most harmful misleading accounts and information and labeling Tweets that contain potentially misleading information about
COVID-19 vaccines.

In addition to combating harmful misinformation reactively, Twitter has prioritized ensuring that individuals who use our service have access to authoritative and credible information about COVID-19 and vaccines. We have invested heavily in product features and partnerships for this critical work, which have been described in more detail on our website. For example:

- We launched a home timeline prompt in 16 countries, including in the United States, and in seven languages, to elevate reliable, up-to-date information about COVID-19 vaccinations to the top of people’s timelines. The prompts, available April 26 to May 3, 2021, and strategically aligned with World Immunization Week, link to localized vaccine information from public health organizations and experts.
- In January 2020, we launched a dedicated search prompt feature to ensure that when individuals search for key terms related to COVID-19, they are directed to credible, authoritative content. In the United States, people are directed to the dedicated website on coronavirus administered by the Centers for Disease Control and Prevention (CDC).
- We created a COVID-19 information hub to surface Tweets containing credible and relevant information. We also have created national and state-level Events, which consist of a collection of relevant Tweets from accounts of government officials, media partners, and public health experts.
- We have partnered with the Ad Council to support their “It’s Up To You” campaign, whose goal is to address vaccine hesitancy among key racial and ethnic groups.
- We have committed Ads for Good credits to nonprofit organizations, including those that serve communities that have been disproportionately impacted by the pandemic, to ensure they can build campaigns to fact-check and get reputable health information to the widest possible audiences.

5. A New York Times article titled “How Misinformation ‘Superspreaders’ Seed False Election Theories” (11/23/2020) found that a small number of users were responsible for spreading election disinformation. How many of these accounts have you banned from your platform?

**RESPONSE:** All of the accounts referenced in the article were extensively reviewed by Twitter, and in many cases, actioned during the election period. Additionally, this year, we expanded our enforcement of our civic integrity policy to address repeat violators. For severe or repeated violations of our civic integrity policy, accounts will be permanently suspended. Repeated violations of this policy are enforced against on the basis of the number of strikes an account has accrued for violations of this policy:

- 1 strike: No account-level action
- 2 strikes: 12-hour account lock
6. The Center for Countering Digital Hate recently published a report titled “The Disinformation Dozen” that identifies 12 individuals responsible for 65% of all anti-vaccine content. For each of the 12 individuals, please identify the following:

a. Have you taken adverse actions against any content the individual’s account(s) posted? If so, how many times?

**RESPONSE:** Our team has examined the accounts referenced in the “Disinformation Dozen” report issued by the Center for Countering Digital Hate. We have taken enforcement action against six of the accounts for violating our COVID-19 misinformation policy, including two permanent account suspensions. Notably, several of the Tweets related to the report predate expansion of our COVID-19 policy and were thus not actioned since we do not typically apply rule violations retroactively. In addition, the aforementioned report did not distinguish between harmful vaccine misinformation that contradicts credible public health information, which is prohibited under our policy, and negative vaccine sentiment that is a matter of opinion.

b. Have you banned or taken other adverse actions against the individual’s account(s)?

**RESPONSE:** Please see the response to (a).

7. Have you observed a rise in hate directed at the Asian American and Pacific Islander (AAPI) on Twitter in the prior year relative to the same 12-month period a year earlier?

**RESPONSE:** Please see below.

a. If so, please share any quantifications of this change.

**RESPONSE:** Ensuring that individuals who use Twitter feel safe using our service is an executive level priority and we have several policies designed to address anti-AAPI hate. Our Hateful Conduct Policy prohibits promotion of violence against people on the basis of protected categories. This includes wishing or calling for serious harm against a person or group or people, or targeting individuals with the intent to harass with references to violent events or types of violence where protected categories were the primary victims. This policy was expanded this year 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. And, our Violent Organizations Policy prohibits violent organizations, including
violent extremist groups, from using Twitter to promote violence. Under this policy, we have removed over 200 groups, half of which have links to white supremacy.

b. What have you done to combat anti-AAPI hate on Twitter?

**RESPONSE:** In addition to these policies cited in (a), we have taken several other actions to combat ant-AAPI hate, including:

- Protecting and supporting the health of the public conversation by activating and expanding partnerships with credible and authoritative leaders in the AAPI community.
- Working with third-party groups to identify academics, activists and their organizations from the Asian community for verification.
- Providing monetary grants to the Asian Americans Advancing Justice, the OCA - Asian Pacific American Advocates, Act to Change and the AAPI GoFundMe community fund.
- Empowering our employees to support through our Allyship Right Now #StandforAsians campaign by offering educational resources and donation matching at 3:1.
- Donating premium product enhancements such as the #StopAsianHate emoji and on-platform takeover inventory, known as First-View, to trusted AAPI advocacy partners.
- Encouraging solidarity and support for the Asian community across peer companies.

8. What disinformation campaigns did you track in the lead up to the November 2020 elections that targeted people of color, women, LGBTQ+, veteran, or older American users?

a. How did you respond to these campaigns?

**RESPONSE:** We continue to fully assess our work and the conversation that occurred on Twitter during the 2020 election period. During the election period, we took action to address efforts aimed at disrupting the election from foreign actors, including Russia and Iran. For example, in October, we disclosed that we removed approximately 130 accounts originating in Iran that were attempting to disrupt the public conversation around the first U.S. 2020 Presidential debate. Following investigation, we suspended a total of 238 accounts operating from Iran for various violations of our platform manipulation policies.

Tweets that we have removed and that are associated with known state-backed information operations on Twitter are included in a public archive, to facilitate research by the government and researchers. This one of a kind resource now spans operations across 15 countries, including more than nine terabytes of
media and 200 million Tweets. Many of these disclosures engaged a wide range of themes, including the targeting of specific communities within the U.S.

In addition, during the election period, we also saw efforts to undermine the US election from domestic actors. We enforced our civic integrity, which covers efforts to undermine election processes and outcomes. In addition, we invested heavily in efforts to facilitate access to credible information surrounding the election. An initial assessment of our work around the immediate election period is here.

b. Did you notify the users they had been targeted and share corrective information?

RESPONSE: A key part of our efforts around the 2020 election involved getting ahead of potentially misleading information by facilitating the ability of individuals to access information from official and credible sources. As part of this effort, prior to the November election, we showed everyone on Twitter in the U.S. a series of pre-bunk prompts. These prompts, which were seen 389 million times, appeared in people’s home timelines and in Search, and reminded people that election results were likely to be delayed, and that voting by mail is safe and legitimate. We also had search prompts pointing people to official sources when they searched for key terms related to voter registration, in addition to an Election Hub that directed people to election information from official sources.

9. What categories of information related to the January 6th attacks have you shared with law enforcement agencies?

RESPONSE: Following the Capitol attack, we worked closely with law enforcement and government partners, including the Federal Bureau of Investigation, Department of Homeland Security, the Capitol Police and other agencies. In the event’s immediate aftermath, we prioritized processing requests related to the attack, including emergency disclosure requests, subpoenas, court orders, and preservation requests. We have continued to comply as efficiently as possible with lawful requests to disclose information.

a. Have you shared any of this information with researchers?

RESPONSE: Twitter is a uniquely open service, which is used by academics across the world as a data source to study important issues. Twitter makes data available via an API for the purposes of study. Moreover, to support the work of academics, we have launched an Academic Research product track, which offers 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.

b. Have you permanently deleted any of this information?
RESPONSE: We recognize that law enforcement’s efforts are ongoing and, in some cases, still in their early stages. We continue to comply with lawful requests to preserve information.

c. Will you commit to retaining any such information until all legitimate law enforcement and research requests are fulfilled?

RESPONSE: Twitter complies with lawful requests to preserve information.

10. What did you do to combat voter suppression targeted at Black Americans in the recent elections?

RESPONSE: In the lead up to the 2020 elections, we made significant enhancements to our policies to protect the integrity of the election. Most notably, we updated our civic integrity policy to more comprehensively enforce labeling or removing of false and misleading information. The updated policy covers 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.

11. If a user’s post is designed to meet the letter of your policies but clearly infringes on the spirit of your misinformation policies, how do you treat the content and the content creator?

RESPONSE: We take action on conduct that violates the Twitter Rules. Details on our enforcement approach can be found here. People come to Twitter to stay informed, and they want credible information to help them do so. Currently, we have misinformation policies that address civic integrity, manipulated media, and COVID-19, as these are the areas where we believe there is the greatest risk of harm. We continually work to improve
these policies and adapt them to changing behaviors and tactics as they emerge on Twitter.

However, we continue to innovate to try to address more broadly the issue of misinformation. We also want to broaden the range of voices that are part of tackling this problem, and we believe a community-driven approach can help. That’s why we recently introduced Birdwatch, a pilot in the US of a new community-driven approach to help address misleading information on Twitter. Birdwatch allows people to identify information in Tweets they believe is misleading and write notes that provide informative context. We believe this approach has the potential to allow us to respond quickly when misleading information spreads, adding context that people trust and find valuable. Eventually we aim to make notes visible directly on Tweets for the global Twitter audience, when there is consensus from a broad and diverse set of contributors. More information on Birdwatch can be found here.

12. Spanish, Cantonese, Mandarin, Tagalog, Vietnamese, French and French Creole, Korean, German, Arabic, and Russian are the most spoken languages in the U.S. What are you doing to combat misinformation in these languages?

**RESPONSE:** Twitter is a global company that serves the needs of individuals who speak a variety of languages around the world. We take a behavior-first approach to content moderation, meaning we look at how accounts behave before we review the content they are posting. In addition, individuals who enforce our misinformation policy are fluent in a variety of languages spoken around the world, including many of the ones cited above. As it relates to our COVID-19 efforts, we have worked to increase access to credible information in a variety of languages. As part of this, we launched a home timeline prompt in 16 countries, including in the United States, and in seven languages, to elevate reliable, up-to-date information about COVID-19 vaccinations to the top of people’s timelines.

13. How often are you personally involved in content moderation decisions?

**RESPONSE:** Enforcement decisions at Twitter are not based on any one individual’s opinion. Rather, we have invested in procedures to ensure that the Twitter Rules are enforced impartially and that any mistakes can be promptly rectified. Although I do not make all decisions at Twitter, as CEO, I take ultimate responsibility for the decisions made by the company.

14. How many employees and contractors at your company are dedicated to content moderation?

**RESPONSE:** As of 1 May 2021, we currently employ more than 2,200 full time employees and contractors across the globe who are directly involved in enforcing our rules on Twitter, all of whom work in different ways to improve the health of the platform. This number does not include support-related employees that augment our overall content moderation work, nor subject matter specialists or those working on
safety product and technology investments. We believe that technology, coupled with proportional growth of our review teams is generating the biggest impact in improving the health of the public conversation.

a. How does this number compare to one year and four years prior?

**RESPONSE:** We continue to invest in employees and new technology to enforce the Twitter Rules and create healthy conversation on the service.

b. To what degree are your content moderation team members?

**RESPONSE:** Twitter’s Trust & Safety Team is responsible for creating and overseeing content policies, while our Twitter Service team leads our dedicated enforcement work. We use 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.

15. How many employees and contractors does your company employ for selling advertising?

**RESPONSE:** At Twitter, some advertisers create and manage their ad campaigns through an online self-service portal, while others have account management teams who serve as points of contact and advise on how the advertisers can best manage their ad campaigns on Twitter. We have various teams that support these functions across self-serve and direct sales and support teams.

16. To what degree are your product teams and executives incented – through bonuses, salary changes, or non-monetary incentives – to increase user engagement? Please describe any such incentives.

**RESPONSE:** Our product teams and executives are not incentivized based purely on user engagement; rather the incentive structure is designed to promote healthy conversation on the platform. To give you more insights as to how our business goals are aligned with promoting healthy conversations, please see the transcript from our most recent Analyst Day, noting the primary audience was Twitter’s advertisers and investors.

17. Have you studied the effectiveness of targeted or behavioral advertising relative to contextual advertising?

**RESPONSE:** On December 14, we announced partnerships with Integral Ad Science and DoubleVerify, with the goal of delivering independent transparency to advertisers regarding the context in which their ads serve. We are currently working on additional research to gain more knowledge and insight around the impact of adjacency on Twitter ads.
a. What is the difference between targeted or behavioral advertising and contextual advertising with respect to user engagement and time spent on your platform?

**RESPONSE:** Twitter doesn’t directly sell any goods or services on our platform. Twitter’s mission is to serve the public conversation. To enable that mission, we enable people to see the most relevant content to them, whether organic content or advertising content. Ensuring that people see the most relevant content for them is an important feature of our service. Thus, the information we rely on to power our advertising experience is generally the same as the information we rely on to enable people to see the most relevant content for them.

Advertising revenue allows us to support and improve our services. We use the information described in our Privacy Policy to help make our advertising more relevant to our users, to measure its effectiveness, and to help recognize devices for those individuals on our platform to serve ads on and off of Twitter.

In addition, Twitter adheres to the Digital Advertising Alliance (DAA) Self-Regulatory Principles for Online Behavioral Advertising (also referred to as “interest-based advertising”) and respects the DAA’s consumer choice tool for you to opt out of interest-based advertising. If an individual does not want Twitter to show interest-based ads on and off of Twitter, there are ways to turn off this feature. In addition, our ads policies prohibit advertisers from targeting ads based on categories that we consider sensitive or are prohibited by law, such as race, religion, politics, sex life, or health.

b. Have you partnered with independent researchers to study this? If so, what did they find?

**RESPONSE:** See the response above to question 17.

18. As you have in the past, will you commit to continue promoting authoritative information to Twitter users about Open Enrollment?

**RESPONSE:** Yes.

**The Honorable Peter Welch (D-VT)**

Twitter users can find a massive amount of music on the platform. In 2020, music creators noticed over 2 million infringements for Twitter to take down. And there was likely more scattered throughout the Twitter platform.

1. At a time when the pandemic has closed venues and these same creators have struggled to change their business operations, please explain why Twitter is charging copyright owners to use its tools to locate instances of online theft on your platform? The “free” version has limited functionality and the burden should be shared between the creator and the platform to protect copyright.
RESPONSE: Twitter takes copyright seriously and complies with the requirements of the DMCA. We also take steps to be transparent about copyright enforcement. For example, in the last two years, as our Transparency Report shows, we received 535,000 takedown notices, resulting in the removal of 4.5 million pieces of content. These figures incorporate takedown notices for all Tweets, not just videos, and the notices submitted by all copyright owners, not only those from the music industry. Copyright removals represent a tiny fraction of the total content uploaded to our service every day.

Twitter has worked directly with rightsholders to address Twitter-specific issues with infringing content appearing on our services. Among other efforts, we have voluntarily created a live content matching system in partnership with Audible Magic to detect and stop unauthorized copies of live broadcasts. Our help center for live matching is publicly available online.

Additionally, Twitter makes APIs available for any use case that complies with our developer agreements and policies. Rightsholders have access to Twitter’s standard API, for free. In addition to providing access, we work with rightsholders to help them make more efficient use of the free APIs to maximize the data available and achieve their goals of searching for and identifying content on the platform. Customers determine what their data and product functionality needs are and whether they want to use the standard API for free or premium/enterprise products. More information on access to Twitter’s API can be found here:

The Honorable Yvette D. Clarke (D-NY)

We all know the music and live entertainment industries have been decimated by the pandemic. In 2020 alone, as creators have tried to stay afloat, the music industry has notified Twitter of over 2 million instances of infringement of their works on your platform. Yet not only has Twitter refused to negotiate a license to use musical works, but the platform also charges creators upwards of $100,000 to use Twitter’s search tools to find stolen versions of their works.

1. Why is Twitter charging copyright owners to use search tools to locate stolen works on your platform?

RESPONSE: Twitter makes APIs available for any use case that complies with our developer agreements and policies. Rightsholders have access to Twitter’s standard API at no cost. In addition to providing access, we work with rightsholders to help them make more efficient use of the free APIs to maximize the data available and achieve their goals of searching for and identifying relevant content on Twitter. Customers determine what their data and product functionality needs are and determine whether they want to use the standard, free API or premium/enterprise products. More information on access to Twitter’s API can be found here: https://help.twitter.com/en/rules-and-policies/twitter-api
2. The free version of Twitter’s search API limits things like the number of search queries, number of results, and the search terms that can be used. Why does a creator need to pay not to have you throttle their ability to search for theft of their works on your platform?

**RESPONSE:** See the response to question 1. In addition to free API access, we have undertaken extensive efforts to comply with DMCA requirements and have worked collaboratively with rightsholders to address their concerns about copyright infringement on the platform through a number of voluntary efforts that go well beyond DMCA requirements. These efforts include:

- Scaling our copyright enforcement teams to be staffed globally, providing coverage 24 hours a day, seven days a week;
- Working to ensure that we remove content in response to valid claims as expeditiously as possible, nearly always within hours of initial reports, if not minutes;
- Offering automated content matching for live broadcasts; and
- Investing in automated processing of copyright removals in response to DMCA takedown notices from trusted reporters.

While these are just a few of the recent investments we have made, we have a continued commitment to understanding rightsholder needs and understanding improvements to our voluntary measures.

3. Mr. Dorsey, what is your company doing to address the amplification and discrimination in ad-targeting?

**RESPONSE:** Twitter fully complies with existing civil rights laws. We do not allow targeting of paid ads based on race, religion, sexual orientation, political affiliation, or other sensitive categories. In addition, advertisers must contact Twitter for prior approval of ads related to:

- Offers of housing to rent, sell, or buy, including certain related services including realtor services, home insurance, and mortgage brokerage services
- Offers of credit; or
- Offers of paid or unpaid employment including internships

The above advertisers are also prohibited from targeting ads on the basis of age, language, disability, familial status, civil or marital status, sex, race, color, national origin, religion, receipt of income from public assistance programme, exercise of rights under Consumer Credit Protection Act, or precise location (at the zip code level or more precise). Apart from these protections, we continue to test and innovate ways to measure our systems to protect against discrimination in the delivery of ads.
1. Does Twitter have an explicit policy against Holocaust denial? If not, will Twitter commit to using existing policy against hateful conduct and misinformation to deploy content moderation tools against Holocaust denial?

**RESPONSE:** Twitter’s Hateful Conduct Policy prohibits a wide range of behavior, including making references to violent events or types of violence where protected categories of people were the primary victims, or attempts to deny or diminish such events. Twitter also has a robust policy against glorification of violence and takes action against content that glorifies or praises historical acts of violence and genocide, including the Holocaust. We are committed to expanding our approach in this area and will have more to share on this soon.

The Honorable Robin Kelly (D-IL)

1. At the hearing you said that you do not profit off harmful misinformation, conspiracy theories, and violent content.

   a. How do you manage to avoid collecting revenue from ads served on misinformation content?

   **RESPONSE:** Our mission is to increase healthy public conversation, and we evaluate the incentives created by our product and adjust for the mission. As part of our commitment to this, we have a variety of tools designed to promote health conversation, including, for example, nudges that prompt individuals to read articles before sharing them. In addition, we continue to innovate new ways to address misinformation, including through our new pilot Birdwatch, a new community driven approach to help address misinformation. As it relates to ads, Twitter differs from many other services to the extent that we do not serve ads on content; rather, promoted tweets sit independent of other content on the service, with clear separation.

   In addition to this, we have invested in several initiatives to expand brand safety. On December 14, we announced partnerships with Integral Ad Science and DoubleVerify, with the goal of delivering independent transparency to advertisers regarding the context in which their ads serve. We are currently working on additional research to gain more knowledge and insight around the impact of adjacency on Twitter ads. In addition, Twitter is an active and founding member of the Global Alliance for Responsible Media (GARM). In conjunction with our ongoing efforts to better align our brand safety solutions with GARM’s principles, we support DoubleVerify and Integral Ad Science in their efforts to adapt their current categorization frameworks to better map to the harmful monetized content categories identified by the GARM, the 4A’s APB, and the ANA.
b. Are you claiming that your company has not received any payments from ads that displayed before or adjacent to harmful content and was later removed for violating your policies?

**RESPONSE:** See the response to subpart (a).

2. Mr. Dorsey, you mentioned your support for Bluesky to create an open-source protocol for social media companies to use.

   a. Do you believe that government standards around disparate impact would be helpful to this initiative?

   **RESPONSE:** We encourage partnerships between the government and the private sector to promote responsible use of machine learning algorithms, including efforts to prevent bias. Through our Responsible Machine Learning Initiative, we are conducting in-depth analysis and studies to assess the existence of potential harms in the algorithms we use. As part of this, in the coming months, we plan to make available analysis on:

   - A gender and racial bias analysis of our image cropping (saliency) algorithm;
   - A fairness assessment of our Home timeline recommendations across racial subgroups; and
   - An analysis of content recommendations for different political ideologies across seven countries.

   b. When do you anticipate Twitter switching to a decentralized social networking model?

   **RESPONSE:** The Bluesky team has already created an initial review of the ecosystem around protocols for social media to aid their efforts. Bluesky will eventually allow Twitter and other companies to contribute to and access open recommendation algorithms that promote healthy conversation and ultimately provide individuals greater choice. This effort is emergent, complex, and unprecedented, and therefore it will take time. However, we are excited by its potential and will continue to provide the necessary exploratory resources to push this project forward.

   While Bluesky is in its nascent stages, we hope to someday be a client of the standard it creates. In the meantime, we continue to invest in innovative approaches to address misinformation and ensure responsible use of machine learning algorithms.
1. In October, the ADL issued a study showing a dramatic spike in anti-Asian sentiment and conspiracy theories about the spread of coronavirus in conversations on Twitter in the days immediately following former-President Trump’s first tweet about his positive COVID-19 diagnosis. On and off social media, Mr. Trump and many others have made inflammatory remarks regarding the COVID-19 virus and blaming China and the AAPI community. As we have seen in the spike in AAPI hate crimes and the horrific murder to several AAPI women in Atlanta, this rhetoric leads to violence.

   a. Did Mr. Trump’s comments violate your community standards? Why was he not removed from the platform at that time?

   **RESPONSE:** We have a variety of policies designed to address hateful conduct, including such conduct targeted at the AAPI community. Our [Hateful Conduct Policy](#) prohibits promotion of violence against people on the basis of protected categories. This includes wishing or calling for serious harm against a person or group or people, or targeting individuals with the intent to harass with references to violent events or types of violence where protected categories were the primary victims. This policy was expanded this year 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. And, our [Violent Organizations Policy](#) prohibits violent organizations, including violent extremist groups, from using Twitter to promote violence.

   These policies, in conjunction with our COVID-19 misinformation policy, prohibit content improperly suggesting that Asian Americans are responsible for the coronavirus or which encourage violence against AAPI communities. President Trump was permanently suspended from the platform in January 2021. In recent months, there have been increased questions about how we should address rule violations by world leaders, and whether we are striking the right balance between mitigating harm and ensuring access to information that is in the public interest. As a result, we are re-examining our approach to world leaders and have solicited feedback from the public. We are simultaneously consulting with a range of human rights experts, civil society organizations, and academics worldwide whose feedback will be reflected in forthcoming revisions to the policy framework.

2. There have been some successful partnerships between big tech companies like yours and law enforcement. However, the sales of illicit live animals and their parts and products continue to grow in online marketplaces and communities of violence promoting illegal animal abuse content continue to be prevalent problems. Some online crime experts have raised concerns that the widespread company policies of removing illicit wildlife for sale
may be hindering law enforcement efforts by prematurely alerting criminals who then open new accounts under different usernames. It also poses the problem of deleting public evidence that could help authorities catch traffickers. Company policies also differ between platforms, creating an inconsistent landscape for both companies and law enforcement to work in.

a. How could Congressional mandates assist your company in building stronger and more comprehensive partnerships with law enforcement?

RESPONSE: Information sharing and collaboration are critical to Twitter’s success in combating illegal conduct on our platform. We have well-established relationships with federal government agencies, including the Federal Bureau of Investigation, Department of Homeland Security, National Counterterrorism Center, and state and local law enforcement agencies. We encourage continued investments by the federal, state, and local governments to share relevant threat information with online platforms.

The Honorable Gus M. Bilirakis (R-FL)

During the hearing, a New York Times article entitled, “Square, Jack Dorsey’s Pay Service, Is Withholding Money Merchants Say They Need.” was entered into the record. Mr. Dorsey, your company, withheld between 20 and 30 percent of each transaction from companies struggling to stay afloat. Your reasons for withholding these funds were “to protect against risky transactions or customers who demanded their money back.” But many of the affected businesses provided documents to the New York Times they had not had any returns or risk flags.

1. Mr. Dorsey, why was money withheld from companies that had no returns or risky transactions?

2. Mr. Dorsey, have you returned the money to the rightful companies? If no, when will you return the money to the companies?

Square is better positioned to answer Questions 1 and 2.

3. Mr. Dorsey, some of the businesses complained via your platforms, but were blocked. What specific policies did they break to warrant being blocked? Did you block them to prevent criticism of your “Square” financial property from circulating online?

RESPONSE: Blocking does not prevent tweets from circulating online or being seen by other users. Instead, it is a tool to ensure the account managers do not receive repeated abusive tweets or content from spam accounts on their feed or in their messages inbox. There are rare cases where accounts engage in abusive messages and spam towards the Square Twitter account.
4. Mr. Dorsey, your sensitive media policy defines adult content as “any consensually produced and distributed media that is pornographic or intended to cause sexual arousal.” It states, “You can share graphic violence and consensually produced adult content within your Tweets, provided that you mark this media as sensitive.” With this in mind, let’s talk about the dark side of user-generated adult content: sexual exploitation. Republicans and Democrats may have different beliefs about the impact of hate speech and misinformation, but hopefully we can all agree that the impact of child sexual abuse material and non-consensual nudity is in a class all by itself.

While the reasons may differ, a strong bipartisan consensus exists that you’re not doing a good job of content moderation. Your policies for adult content do ban child sexual abuse material and non-consensual nudity, but is your enforcement of these policies actually effective? If you’re struggling with content moderation in general, then why would you magically succeed at stopping online sexual exploitation?

**RESPONSE:** Twitter has a zero tolerance policy for childhood sexual exploitation, and we have invested heavily in partnerships to combat this harmful content. For example, 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, through our proactive detection or reports, 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.

5. The tech industry can often be a metrics-driven and data-driven world. Frequently, you’re trying to improve metrics that measure engagement: video watch time for YouTube, the number of daily active users on Facebook, or the number of daily tweets on Twitter. However, sometimes you can increase engagement by pushing both sides into their echo chambers. This echo chamber effect leads to less civil discourse and more political polarization, and it also allows misinformation to more easily spread on both sides. With respect to the problem of echo chambers, what metrics has your company developed which can measure this problem?

**RESPONSE:** We have conducted research aimed at identifying the needs of people who use Twitter, including how to address potential misinformation on the platform and how
to ensure that people have access to credible information to inform their viewpoints. A
key piece of feedback we got from the people who use our service is that we should not
determine the truthfulness of Tweets; 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, we continue to explore ways to expand our enforcement options beyond a
binary choice of leaving content up or taking it down. Instead, we are looking at ways to
ensure that people can access additional information to inform their viewpoints.

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 may 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 continue to explore additional improvements to our service that we can make
to ensure individuals have access to credible information and diverse viewpoints.

Notwithstanding these efforts, we recognize that Twitter is only part of a larger
ecosystem that impacts the online discourse. For example, a Knight Foundation study
found that “evidence for echo chambers is actually strongest in offline social networks,
which can increase exposure to like-minded views and information and amplify partisan
messages.” The study also found that “several studies have found evidence for offline
echo chambers that are as strong or stronger than those documented online.”

**The Honorable Gary Palmer (R-AL)**

There have been a lot of discussions about Russian and Chinese attempts to influence our
elections and political discussions. I think we need to realize that these disinformation campaigns
are not only aimed at us, but also our allies. For example, Radio Free Asia highlighted that China
mobilized 300,000 cyber operatives to target our ally Taiwan’s 2018 and 2020 elections. A 2016
Harvard study found that a group allied with the Chinese Communist Party (CCP) produced 488
million “fake” social media posts a year to distract other internet users from news and online
discussions painting the Communist Party in a negative light.

1. What is your company doing outside of the U.S. to protect our allies like Taiwan from the
nefarious activities of the CCP?

**RESPONSE:** Combatting attempts by state-backed actors to manipulate the conversation
on Twitter remains a top priority for the company, and we continue to invest heavily in
our detection, disruption, and transparency efforts related to state-backed information
operations. Our goal is to remove bad faith actors and to advance public understanding of
these critical topics.

In addition to Twitter Rules aimed at preventing platform manipulation, we have invested
in efforts to safeguard the public conversation on Twitter from state-backed actors. 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.

Additionally, we have worked to provide 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 initially applied these 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. This year, we expanded these labels to accounts from Group of Seven (G7) countries, and to a majority of countries that Twitter has attributed to state-linked information operations.

Furthermore, we have worked to better enable the public, lawmakers, and researchers to better understand the threat of state-backed information operations, data is essential. In October 2018, we published the first comprehensive archive of Tweets and media associated with known state-backed information operations on Twitter. We now routinely update this archive to add additional datasets when we disrupt state-backed information operations. This one of a kind resource now spans operations across 15 countries, including more than nine terabytes of media and 200 million Tweets.

2. What actions are you taking to keep individuals associated with the CCP from using your platform to attack the U.S. and our allies?

RESPONSE: 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, including government actors, 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. 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.

An important part of our work is also 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 initially 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, and this year expanded these efforts to cover accounts from Group of Seven (G7) countries and a majority of countries that Twitter has attributed to state-linked information operations. In addition, 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.

3. Do you pledge to work with our allies to push back against Russian and Chinese disinformation campaigns?

**RESPONSE:** As explained in more detail in Question 1, combatting attempts by state-backed actors to interfere in conversations on Twitter remains a top priority for the company. As part of this work, we have significantly deepened our partnership with industry peers, establishing formal processes for information sharing with both large and small companies and a regular cadence of discussion about shared threats. In addition to this, we have well-established relationships with federal government agencies, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of Homeland Security to address these threats. We welcome further collaboration to prevent bad faith state actors from disrupting the public conversation on Twitter.

**The Honorable Michael C. Burgess, M.D. (R-TX)**

According to Subsection (f)(3) of Section 230 of the Communications Decency Act, an “information content provider means any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.”
Based on this almost anyone can be a content provider. One of my biggest concerns over the years has been the accessibility of Terms of Service and whether they have been applied equally to all users – users who must accept them in order to access your services.

1. On your platforms, if you limit distribution of user created content or include additional information to accompany a user’s content or post, are you then a content creator since you were “responsible…in part, for the…development of information”?

**RESPONSE:** 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.

2. Should Congress be considering an update to the definition of an “information content provider”?

**RESPONSE:** We believe that the fundamentals of online speech, which are addressed through Section 230, remain; however, we should build upon Section 230 to reflect the realities of the modern digital age. Sound proposals can advance free speech and online safety while guaranteeing that competition is protected. Unwittingly entrenching the dominant companies through regulation only they can comply with—and by extension harming American innovation and global competition—should be avoided at all costs. Instead, we encourage Congress to look at ways to enhance transparency, procedural fairness, algorithmic choice, and privacy.

3. What would your service look like without Section 230?

**RESPONSE:** Twitter, like so many online services Americans depend on daily, would not be recognizable and could not have served as the same forum for the expression of diverse views had Section 230 not existed. 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.

4. What about when you were first starting out?

**RESPONSE:** Section 230 has undoubtedly contributed to the development and success of services like Twitter and the U.S. technology sector more broadly. It has enabled new
companies—small ones seeded with an idea—to build and compete with established companies globally. 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. Section 230 enables new companies to promote free expression, while also taking steps to mitigate online harms. Eliminating Section 230 would make it harder for new and smaller companies to innovate to compete with larger platforms and innovate to address the challenges of today.