Responses to Questions for the Record
U.S. House Committee on Energy and Commerce
Mr. Jack Dorsey, Chief Executive Officer, Twitter

The Honorable Greg Walden

1. Mr. Dorsey, you testified that Twitter uses thousands of "signals" to determine and decide what to show, downrank, or filter. Please outline the specific behavioral signals that were implicated in the search auto-suggest issue. Additionally, please outline the specific signals that Twitter currently uses and how those signals determine what content to show, downrank, or filter.

Twitter also uses a range of behavioral signals to determine how Tweets are organized and presented in the home timeline, conversations, and search based on relevance. Twitter relies on behavioral signals—such as how accounts behave and react to one another—to identify content that detracts from a healthy public conversation, such as spam and abuse. Unless we have determined that a Tweet violates Twitter policies, it will remain on the platform, and is available in our product. Where we have identified a Tweet as potentially detracting from healthy conversation (*e.g.*, as potentially abusive), it will often only be available to view if you click on "Show more replies" or choose to see everything in your search setting.

Some examples of behavioral signals we use, in combination with each other and a range of other signals, to help identify this type of content include: an account with no confirmed email address, simultaneous registration for multiple accounts, accounts that repeatedly Tweet and mention accounts that do not follow them, or behavior that might indicate a coordinated attack. Twitter is also examining how accounts are connected to those that violate our rules and how they interact with each other. The accuracy and usefulness of the algorithms developed from these behavioral signals will continue to improve over time.

These behavioral signals are an important factor in how Twitter organizes and presents content in communal areas like conversation and search. Our primary goal is to ensure that relevant content and Tweets contributing to healthy conversation will appear first in conversations and search. Because our service operates in dozens of languages and hundreds of cultural contexts around the globe, we have found that behavior is a strong signal that helps us identify bad faith actors on our platform. The behavioral ranking that Twitter utilizes does not consider in any way political views or ideology. It focuses solely on the behavior of all accounts. Twitter is always working to improve our behavior-based ranking models such that their breadth and accuracy will improve over time. We use thousands of behavioral signals in our behavior-based ranking models—this ensures that no one signal drives the ranking outcomes and protects against malicious attempts to manipulate our ranking systems.

2. Mr. Dorsey, you testified that "about 600,000 accounts" were impacted by the autosuggest search issue. With respect to these accounts, please provide the following:

a. The number of Members of Congress' accounts that were impacted. Please provide a complete list of all the names of Congressional members, or their accounts, that were affected.

b. The number of Republican Members of Congress' accounts that were impacted.

c. The number of Democratic Members of Congress' accounts that were impacted.

d. The number of other Federal, State or local government accounts that were impacted.

e. Please provide a complete list of all the names of the remaining non-Congressional accounts.

In July 2018, we publicly acknowledged that some accounts were not being auto-suggested even when people were searching for their specific name. Our technology was using a decision-making criteria that considered the behavior of people following these accounts. Specifically, the followers of these accounts had a higher proportion of abusive behavior on the platform.

This issue impacted 600,000 accounts across the globe. The vast majority of impacted accounts were not political in nature. The issue impacted 53 accounts of politicians in the U.S., representing 0.00883 percent of total affected accounts. This subset of affected accounts includes 10 official accounts of Republican Members of Congress. The remainder of impacted political accounts relate to campaign activity and affected candidates across the political spectrum.

To be clear, this issue was limited solely to the search auto-suggestion function. The accounts, their Tweets, and all surrounding conversation about those accounts remained displayed in search results. Once identified, this issue was promptly resolved within 24 hours.

An analysis of accounts for Members of Congress that were affected by this search issue demonstrates that there was no negative effect on the growth of their follower counts. To the contrary, follower counts of impacted Members of Congress increased.

Twitter will not reveal the names of the impacted accounts due to privacy and security concerns, but we can disclose information concerning the accounts of those who have publicly discussed they were impacted.

We have enclosed information concerning follower count data displayed in graph format for the official accounts of Congressmen Matt Gaetz, Jim Jordan, Mark Meadows, and Devin Nunes. The graphs contain publicly available data on the follower counts for the three month period surrounding the incident, which occurred in late July. The other impacted members of Congress also saw their followers increase during this time period.

3. Mr. Dorsey, you testified that Twitter needs to do a better job with your Terms of Service to make them more understandable and approachable. Please explain the steps Twitter is taking or intends to take to rework the Twitter Terms of Service to accomplish this?

Twitter's Terms of Service ("Terms") govern an individual's access to and use of our services, including our various websites, SMS, APIs, email notifications, applications, buttons, widgets, ads, commerce services, and our other covered services. The Terms also cover any information, text, links, graphics, photos, audio, videos, or other materials or arrangements of materials uploaded, downloaded or appearing on the Services (collectively referred to as "Content").

By using the Services an individual who uses the platform agrees to be bound by these Terms. Twitter may revise these Terms from time to time. The changes will not be retroactive, and the most current version of the Terms, which will always be at twitter.com/tos, will govern our relationship with the individual on the platform. We will try to notify individuals on our platform of material revisions, for example via a service notification or an email to the email associated with an account. By continuing to access or use the Services after those revisions become effective, the individual agrees to be bound by the revised Terms.

Making a policy change requires in-depth research around trends in online behavior, developing language that sets expectations around what's allowed, and reviewer guidelines that can be enforced across millions of Tweets. Once drafted, we gather feedback from our teams and Trust & Safety Council. We gather input from around the world so that we can consider diverse, global perspectives around the changing nature of online speech, including how our rules are applied and interpreted in different cultural and social contexts. We then test the proposed rule with samples of potentially abusive Tweets to measure the policy effectiveness and once we determine it meets our expectations, build and operationalize product changes to support the update. Finally, we train our global review teams, update the Twitter Rules, and start enforcing it. In 2019, as in prior years, we intend to review the Twitter Terms of Service to determine how to make the rules that govern our platform more approachable and accessible.

Twitter recently updated our Privacy Policy to include callouts, graphics, and animations designed to enable people to better understand the data we receive, how it is used, and when it is shared.

4. Mr. Dorsey, how does Twitter determine whether the information included in Twitter's Terms of Service is understandable and approachable?

a. How does Twitter define understandable with respect to Twitter' s Terms of Service?

b. How does twitter define approachable with respect to Twitter's Terms of Services?

The Twitter Rules (along with all incorporated policies), Privacy Policy, and Terms of Service collectively make up the "Twitter User Agreement" that governs an individual's access to and use of Twitter's services. We have the Twitter Rules in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency.

Our policies and enforcement options evolve continuously to address emerging behaviors online and we sometimes come across instances where someone is reported for an incident that took place prior to that behavior being prohibited. In those instances, we will generally require the individual to delete the Tweet that violates the new rules but we won't generally take other enforcement action against them (e.g. suspension). This is reflective of the fact that the Twitter Rules are a living document. We continue to expand and update both them and our enforcement options to respond to the changing contours of online conversation. This is how we make Twitter better for everyone.

We are continually working to update, refine, and improve both our enforcement and our policies, informed by in-depth research around trends in online behavior both on and off Twitter, feedback from the people who use Twitter, and input from a number of external entities, including members of our Trust & Safety Council.

We believe we have to rely on a straight-forward, principled approach and focus on the long term goal of understanding - not just in terms of the service itself - but in terms of the role we play in society and our wider responsibility to foster and better serve a healthy public conversation.

5. Mr. Dorsey, does Twitter ever review, use, or consider data or information about a user unrelated to Twitter to make decisions about content posted by a Twitter user?

a. If yes, what data or information does Twitter consider and why?

Yes, but only in rare circumstances. A good example of this is our approach to violent extremist groups. We take pride in Twitter being a platform where a diverse range of opinions can be held and discussed, but we will not tolerate groups or individuals associated with them who engage in and promote violence against civilians both on and off the platform. Accounts affiliated with groups in which violence is a component of advancing their cause risk having a chilling effect on opponents and bystanders. The violence that such groups promote could also have dangerous consequences offline, jeopardizing their targets' physical safety.

We prohibit the use of Twitter's services by violent extremist groups – i.e., identified groups subscribing to the use of violence as a means to advance their cause, whether political, religious, or social. Groups that are prohibited on our services are those that identify as such or engage in activity — both on and off the platform — that promotes violence. An individual on Twitter may not affiliate with organizations that – whether by their own statements or activity both on and off the platform – use or promote violence against civilians to further their causes.

We consider violent extremist groups to be those which identify through their stated purpose, publications, or actions, as an extremist group; have engaged in, or currently engage in, violence (and/or the promotion of violence) as a means to further their cause; and target civilians in their acts (and/or promotion) of violence.

6. Mr. Dorsey, you testified that Twitter's verification program is not where you'd like it to be and that it needs a "serious reboot." When was the verification program shut down? Please explain the current problems and inadequacies with the verification program and what steps Twitter is taking or intends to take to address this.

Verification was meant to authenticate identity and voice but it is interpreted as an endorsement or an indicator of importance. We gave verified accounts visual prominence on the service which deepened this perception. We should have addressed this earlier but did not prioritize the work as we should have. This perception became worse when we opened up verification for public submissions and verified people who we in no way endorse.

Twitter recognizes that we have created this confusion and need to resolve it. Beginning in November 2017, we have paused all general verifications while we work to resolve this issue.

Twitter is currently working on a new authentication and verification program. In the meantime, we are not accepting any public submissions for verification and have introduced new guidelines for the program.

7. Mr. Dorsey, in Twitter's July 26, 2018 blog post "Setting the record straight on shadow banning," company officials indicated:

> **"For the most part, *we believe the issue had more to do with how other people were interacting with these representatives' accounts than the accounts themselves* (see bullet #3 above). There are communities that try to boost each other's presence on the platform through coordinated engagement. *We believe these types of actors engaged with the representatives' accounts-- the impact of this coordinated behavior, in combination with our implementation of search auto-suggestions, caused the representatives' accounts to not show up in auto-suggestions."* (emphasis added).**

For additional context, the bullet no. 3 referenced above pointed out: "How other accounts interact with you (e.g. who mutes you, who follows you, who retweets you, who blocks you, etc)".

Yet, in response to a question from Rep. Walberg about what "specific signals or actions of other accounts interacting with the representatives' account would you suggest contributed to the auto-suggest issue," you asserted that the "behaviors we were seeing were actual violations of our terms of service."

Please explain in detail whether the auto-suggest search issue was a consequence of "interaction" issue described in Twitter's blog post or violations of Terms of Service as you

described in your testimony, or both? If it was a consequence of Terms of Service violations, how and why would such violations render affected accounts invisible to Twitter users?

In July 2018, we acknowledged that some accounts (including those of Republicans and Democrats) were not being auto-suggested even when people were searching for their specific name. Our usage of the behavioral signals within search was causing this to happen. Specifically, if an account had a large number of followers who violated our terms of service, it impacted the visibility of the account. To be clear, this only impacted our search auto-suggestions. The accounts, their Tweets, and surrounding conversation about those accounts were still showing up in search results. Once identified, this issue was promptly resolved within 24 hours. This impacted 600,000 accounts across the globe and across the political spectrum. And most accounts affected had nothing to do with politics at all. In addition to fixing the search auto-suggestion function, Twitter is continuing to improve our systems so they can better detect these issues and correct for them.

Twitter had made a change to how one of our behavior based algorithms works in search results. When people used search, our algorithms were filtering out those that had a higher likelihood of being abusive from the "Latest" tab by default. Those search results were visible in "Latest" if someone turned off the quality filter in search, and they were also in Top search and elsewhere throughout the product. Twitter decided that a higher level of precision is needed when filtering to ensure these accounts are included in "Latest" by default. Twitter therefore turned off the algorithm. As always, we will continue to refine our approach and will be transparent about why we make the decisions that we do.

8. Mr. Dorsey, in response to a question about the Meghan McCain incident and the inadequacies of Twitter's abuse prioritization mechanism, you indicated "[i]n this particular case, the reason why was because [the violent and physical harm element] *was captured within an image rather than the tweet text itself*" (emphasis added). Is currently Twitter without the technological tools to police harmful and abusive content embedded in either images, .gifs, links, videos, and audio clips? If yes to any, how do human reviewers police harmful and abusive content embedded in either images, .gifs, links, videos, and audio clips?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report

Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

Twitter employs extensive content detection technology to identify and police harmful and abusive content embedded in various forms of media on the platform. We use PhotoDNA and hash matching technology, particularly in the context of child sexual exploitation material and terrorism. We use these technologies to identify known violative content in order to surface it for agent review, however, if it is the first time that an image has been seen, it would not necessarily be subject to our technology. It is important to note that we continually expand our databases of known violative content.

9. Mr. Dorsey, what policies and procedures does Twitter have in place to notify accounts and users when their messages or content have been removed, suspended, banned, or otherwise rendered invisible? What steps going forward does Twitter intend to take to better notify accounts and users?

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Failure to do so may result in Twitter taking one or more of the following enforcement actions: (1) requiring an individual to delete prohibited content before he or she can again create new posts and interact with other Twitter users; (2) temporarily limiting an individual's ability to create posts or interact with other Twitter users; (3) requesting an individual verify account ownership with a phone number or email address; or (4) permanently suspending an account or related accounts.

We have improved the transparency of our enforcement and believe that increasing understanding of our rules can help educate people who use Twitter. For example one tool we use to enforce our rules is putting people in a read-only state until they delete a Tweet that violates our Terms of Service. Accounts that we put into period of limited functionality subsequently generate 25% fewer abuse reports, and 65% of these accounts are in this read-only state on only one instance.

Accounts under investigation or which have been detected as sharing content in violation with the Twitter Rules may have their account or visibility limited in various parts of Twitter, including search.

10. Mr. Dorsey, the Twitter Rules preclude users from posting graphic violence and adult content on the Twitter platform.

a. Please provide Twitter's definition of "graphic violence" and outline the specific factors Twitter uses to determine whether content meets this definition.

b. Please provide Twitter's definition of "adult content" and outline the specific factors Twitter uses to determine whether content meets this definition.

Twitter allows some forms of graphic violence or adult content in Tweets marked as containing sensitive media. We consider graphic violence to be any form of gory media related to death, serious injury, violence, or surgical procedures. We consider adult content to be any media that is pornographic and may be intended to cause sexual arousal.

Individuals on our platform may not use such content in live video, the person's profile, or header images. Additionally, Twitter may sometimes require an individual to remove excessively graphic violence.

While we want people to feel empowered to share media that reflects their creativity or individuality, or to show what's happening in the world, we have heard feedback from people that they don't want to be exposed to sensitive media inadvertently. Additionally, research has shown that repeated exposure to violent content online may negatively impact an individual's wellbeing.

For this reason, we place media containing adult content, graphic violence, or hateful imagery behind an interstitial, or warning message. This enables us to advise viewers that they may view sensitive media if they click through, helping those who do not wish to see it make an informed decision. However, an individual can not include this type of content in live video, or in profile or header images. All other instances of this sensitive content should be marked as sensitive media.

11. Mr. Dorsey, how does Twitter review the platform to identify content that meets Twitter's definition of graphic violence or adult content?

a. Is content on the platform proactively scanned or reviewed by Twitter before it is flagged by a Twitter user ?

i. If yes, is the screening done by algorithms or humans?

1. If the screening is done by algorithms, please outline the factors and signals the algorithms use to determine what is considered "graphic violence" and "adult content" as defined by Twitter.

2. If the screening is done by humans, please provide the number of employees responsible for reviewing content on the Twitter platform to identify "graphic violence" or "adult content."

ii. If no, why does Twitter solely rely on users to flag potentially violative content?

We encourage individuals who Tweet media containing graphic violence and adult content to mark their account as sharing potentially sensitive media via the safety page within account settings. When this content is reported to our team for review, we will manually apply an interstitial, or warning message. If an individual repeatedly uploads sensitive media that is mislabeled, Twitter may permanently adjust that individual's account settings to label media they Tweet as containing material that may be sensitive.

Including sensitive media in live video, header or profile images is a violation of our media policy. Those in violation will be required to remove this media from their account and may be asked to verify their contact information or serve a period of time-out before they can use their account again. Subsequent violations may result in permanent account suspension.

If an individual encounters media in Tweets that he or she believes should be treated as sensitive under Twitter's media policy, we ask the individual to report it by reporting a Tweet and selecting that the Tweets displays a sensitive image. We subsequently provide recommendations to the individual who reported the Tweet for additional actions he or she can take to improve the Twitter experience. It is possible that if an individual sees something he or she doesn't like, and Twitter has not placed a warning label before it, it's possible that it doesn't meet our threshold for a warning on the media.

Twitter reviews reports of media flagged by individuals using our platform to determine if that media requires a warning message in order to comply with Twitter's media policies. We caution individuals that by flagging media they are bringing it to the attention of the Twitter team. The flagged content will not automatically receive a warning message or be removed from the site.

Twitter may also use automated techniques to detect and label potentially sensitive media, and also utilize automated techniques to detect and label accounts that frequently tweet potentially sensitive media.

12. Mr. Dorsey, the Twitter Rules indicate that some content that may be violative of the Twitter Rules will be allowed but marked as "sensitive media."

What factors does Twitter consider when deciding whether content should be prohibited and removed from the Twitter platform or should be allowed and remain on the platform, but marked as "sensitive media?"

Is this decision made by algorithms or humans?

i. If it is made by algorithms, what factors or signals do the algorithms consider when making this decision?

ii. If it is made by humans, please provide the number of employees responsible for making this decision and identify what factors those employees consider.

Please see answers 10 and 11.

13. Mr. Dorsey, the Twitter Rules state that "at times, [Twitter] may prevent certain content from trending" and that it may be content that either "violates Twitter Rules" or "may attempt to manipulate trends."

a. When content does not violate the Twitter Rules or attempt to manipulate trends, does Twitter ever prevent such content from "trending?"

i. If so, please provide an example of content that Twitter has not allowed to "trend" despite it being in compliance with the Twitter Rules.

b. Please explain what factors Twitter considers when determining whether content "attempts to manipulate trends."

We want trends to promote healthy discussions on Twitter. This means that, at times, we may prevent certain content from trending. These include trends that contain profanity or adult or graphic references; incite hate on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease; or violate the Twitter Rules.

In certain cases, we may also consider the newsworthiness of the content, or if it is in the public interest, when evaluating potential violations. In these cases, the content may continue to trend on our platform. Note that even if we prevent the hashtag or the content from appearing on the list of trending topics, an individual may still be able to access conversations around that content on Twitter.

Keeping Twitter safe and free from malicious automation is a top priority for us. One of the most common violations we see is the use of multiple accounts and the Twitter developer platform to attempt to artificially amplify or inflate the prominence of certain Tweets. To be clear: Twitter prohibits any attempt to use automation for the purposes of posting or disseminating spam, and such behavior may result in enforcement actions.

In January 2018, we announced that as part of our Information Quality efforts we would be making changes to TweetDeck and the Twitter Application Program Interface, or API, to limit the ability of individuals to perform coordinated actions across multiple accounts. These changes

are an important step in ensuring we stay ahead of malicious activity targeting the crucial conversations taking place on Twitter — including elections in the United States and around the world.

Twitter does not allow developers and the individuals with whom they work to simultaneously post identical or substantially similar content to multiple accounts. For example, a service is not permitted to allow their users to select several accounts they control from which to publish a given Tweet. This applies regardless of whether the Tweets are published to Twitter at the same time, or are scheduled/queued for future publication. Bulk, aggressive, or very high-volume automated Retweeting is not permitted under the Automation Rules, and may be subject to enforcement actions. And simultaneously performing actions such as Likes, Retweets, or follows from multiple accounts is also prohibited.

The use of any form of automation (including scheduling) to post identical or substantially similar content, or to perform actions such as Likes or Retweets, across many accounts is not permitted. For example, applications that coordinate activity across multiple accounts to simultaneously post Tweets with a specific hashtag (e.g. in an attempt to cause that topic to trend) are prohibited.

Posting duplicative or substantially similar content, replies, or mentions over multiple accounts under a developer's control, or creating duplicate or substantially similar accounts, with or without the use of automation, is never allowed. Posting multiple updates (on a single account or across multiple accounts a developer controls) to a trending or popular topic (for instance, through the use of a specific hashtag) with an intent to subvert or manipulate the topic, or to artificially inflate the prominence of a hashtag or topic, is never allowed.

14. Mr. Dorsey, the Twitter Rules provide that users may not engage in hateful conduct. Please provide Twitter's definition of "hateful conduct" and outline the specific factors Twitter uses to determine whether content meets this definition.

a. When content is flagged as "hateful conduct", what factors does Twitter consider when deciding whether to remove the content?

i. Does Twitter consider data points or information about a user unrelated to the specific content to make this determination?

Twitter's mission is to give everyone the power to create and share ideas and information, and to express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.

We recognize that if people experience abuse on Twitter, it can jeopardize their ability to express themselves. Research has shown that some groups of people are disproportionately targeted with abuse online. This includes; women, people of color, lesbian, gay, bisexual, transgender, queer, intersex, asexual individuals, marginalized and historically underrepresented

communities. For those who identity with multiple underrepresented groups, abuse may be more common, more severe in nature and have a higher impact on those targeted.

We are committed to combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalized. For this reason, we prohibit behavior that targets individuals with abuse based on protected category.

An individual on the platform is not permitted to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

We do not allow individuals to use hateful images or symbols in their profile image or profile header. Individuals on the platform as not allowed to use the username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate towards a person, group, or protected category.

Under this policy, we take action against behavior that targets individuals or an entire protected category with hateful conduct. Targeting can happen in a number of ways, for example, mentions, including a photo of an individual, referring to someone by their full name, etc.

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent account suspension. If an account is engaging primarily in abusive behavior, or is deemed to have shared a violent threat, we will permanently suspend the account upon initial review.

Please see the response to question five for additional information regarding our violent extremist group policy.

15. Mr. Dorsey, when a Twitter user flags content as potentially violative of the Twitter Rules, what specific steps does Twitter undertake to determine whether that content should be removed from the platform?

In order to maintain a safe environment for individuals on Twitter, we may suspend accounts that violate the Twitter Rules. Most of the accounts we suspend are suspended because they are automated or fake accounts, and they introduce security risks for Twitter and all of the individuals who use our service. These types of accounts are contrary to our Twitter Rules. Unfortunately, sometimes a real person's account gets suspended by mistake, and in those cases we work with the person to make sure the account is unsuspended. If we suspect an account has

been hacked or compromised, we may suspend it until it can be secured and restored to the account owner in order to reduce potentially malicious activity caused by the compromise.

As described above, we may suspend an account if it has been reported to us as violating our Rules surrounding abuse. When an account engages in abusive behavior, like sending threats to others, we may suspend it temporarily or, in some cases, permanently. An individual may be able to unsuspend his or her own account by providing a phone number or confirming an email address.

An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her account for a specific period of time or may be asked to verify certain information before proceeding.

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 engage with the account holder via email to resolve the suspension.

16. Mr. Dorsey, what specific steps does Twitter undertake to determine whether a Twitter user's account should be suspended?

Please see the answer to question 15.

17. Mr. Dorsey, what specific steps does Twitter undertake to determine whether a Twitter user' s account should be banned?

Please see the answer to question 15.

18. Mr. Dorsey, does Twitter only use content on the platform to make decisions about suspending or removing Twitter accounts or does Twitter use data from outside entities to make those decisions?

a. If yes, please identify the outside entities Twitter relies on for information about Twitter users to make these decisions.

Please see the response to question five for additional information concerning our violent extremist group policy and the circumstances under which we consider off-platform activity.

We take pride in Twitter being a platform where a diverse range of opinions can be held and discussed, but we will not tolerate groups or individuals associated with them who engage in and promote violence against civilians both on and off the platform. Accounts affiliated with groups in which violence is a component of advancing their cause risk having a chilling effect on opponents and bystanders. The violence that such groups promote could also have dangerous

consequences offline, jeopardizing their targets' physical safety.

We prohibit the use of Twitter's services by violent extremist groups – i.e., identified groups subscribing to the use of violence as a means to advance their cause, whether political, religious, or social. Groups that are prohibited on our services are those that identify as such or engage in activity — both on and off the platform — that promotes violence. An individual on Twitter may not affiliate with organizations that – whether by their own statements or activity both on and off the platform – use or promote violence against civilians to further their causes.

We consider violent extremist groups to be those which identify through their stated purpose, publications, or actions, as an extremist group; have engaged in, or currently engage in, violence (and/or the promotion of violence) as a means to further their cause; and target civilians in their acts (and/or promotion) of violence.

19. Mr. Dorsey, in late August, you were interviewed by CNN and you indicated "[Twitter has] changed a lot. But [Twitter hasn't] changed the underlying fundamentals." Please identify the underlying fundamentals of Twitter.

The mission we serve as Twitter, Inc. is to give everyone the power to create and share ideas and information instantly without barriers. Our business will always follow that mission in ways that improve – and do not detract from – a free and global conversation.

Twitter is the best place to see what's happening and what people are talking about. Every day, instances of breaking news, entertainment, sports, politics, big events and everyday interests happen first on Twitter. Twitter is where the full story unfolds with live commentary and where live events come to life unlike anywhere else. Our primary service can be accessed on a variety of mobile devices, at twitter.com and via SMS.

20. Mr. Dorsey, in late August, you were interviewed by CNN and you indicated "We've seen abuse, we've seen trolling, we've seen harassment, we've seen misinformation." Did you not anticipate this potential, either when you launched the platform or at any other point since launching?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report

Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

21. Mr. Dorsey, as the company grew, were there ever discussions about implementing greater security and enhanced transparency measures?

a. If yes, who was involved in those conversations and, if no, why not?

Discussions regarding security and transparency have occurred since the founding of Twitter and are ongoing. We have dedicated Information Security and Platform teams who focused on these issues as part of their core missions. Twitter recommends to the individuals on its platform certain best security practices in order to help keep their accounts secure. These include the use of a strong password that is not reused on other websites, the use login verification, and requiring email and phone number to request a reset password link or code. Twitter also suggests that individuals on the platform be cautious of suspicious links and always make sure an individualis on twitter.com before he or she enters login information. We caution people to never give their username and password out to third parties, especially those promising to grow followers, make money, or verify an account.

Our biannual Twitter Transparency Report highlights trends in legal requests, intellectual property-related requests, and email privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from worldwide government and non-government legal requests we have received for account information. Removal requests are also included in the Transparency Report and include worldwide legal demands from governments and other authorized reporters, as well as reports based on local laws from trusted reporters and non-governmental organizations, to remove or withhold content.

The purpose of Twitter is to serve the public conversation, and we do not make value judgments on personal beliefs. We are focused on making our platform—and the technology it relies upon—better and smarter over time and sharing our work and progress with this Committee and the American people. We think increased transparency is critical to promoting healthy public conversation on Twitter and earning trust.

22. Mr. Dorsey, Twitter agreed to a "code of conduct" on Hate Speech under threat from

the European Union. Has it changed any of its policies as it affects the U.S. based on this agreement, and has Twitter ever removed any U.S. based accounts on behalf of any foreign government?

Twitter's mission is to give everyone the power to create and share ideas and information, and to express their opinions and beliefs without barriers. Free expression is a human right – we believe that everyone has a voice, and the right to use it. Our role is to serve the public conversation, which requires representation of a diverse range of perspectives.

We recognize that if people experience abuse on Twitter, it can jeopardize their ability to express themselves. Research has shown that some groups of people are disproportionately targeted with abuse online. This includes; women, people of color, lesbian, gay, bisexual, transgender, queer, intersex, asexual individuals, marginalized and historically underrepresented communities. For those who identity with multiple underrepresented groups, abuse may be more common, more severe in nature and have a higher impact on those targeted.

We are committed to combating abuse motivated by hatred, prejudice or intolerance, particularly abuse that seeks to silence the voices of those who have been historically marginalized. For this reason, we prohibit behavior that targets individuals with abuse based on protected category.

An individual on the platform is not permitted to promote violence against or directly attack or threaten other people on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or serious disease. We also do not allow accounts whose primary purpose is inciting harm towards others on the basis of these categories.

We do not allow individuals to use hateful images or symbols in your profile image or profile header. Individuals on the platform as not allowed to use the username, display name, or profile bio to engage in abusive behavior, such as targeted harassment or expressing hate towards a person, group, or protected category.

Under this policy, we take action against behavior that targets individuals or an entire protected category with hateful conduct. Targeting can happen in a number of ways, for example, mentions, including a photo of an individual, referring to someone by their full name, etc.

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to the severity of the violation and an individual's previous record of rule violations. For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to longer read-only periods and may eventually result in permanent account suspension. If an account is engaging primarily in abusive behavior, or is deemed to have shared a violent threat, we will permanently suspend the account upon initial review.

In regard to the removal of accounts, our biannual Twitter Transparency Report highlights trends in legal requests, intellectual property-related requests, and email privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from governments worldwide and non-government legal requests we have received for account information. Removal requests are also included in the Transparency Report and include worldwide legal demands from governments and other authorized reporters, as well as reports based on local laws from trusted reporters and non-governmental organizations, to remove or withhold content.

23. Mr. Dorsey, with the exclusion of the "Network of Enlightened Women," which Twitter appears to have added recently and have no record opposing online censorship, none of the 48 groups and individuals on the Twitter Trust and Safety Council lean even mildly right-of-center. Does Twitter have plans to include more pro-free speech or conservative voices in their product and policy decisions?

On Twitter, every voice has the power to shape the world. We see this power every day, from activists who use Twitter to mobilize citizens to content creators who use Twitter to shape opinion.

To ensure people can continue to express themselves freely and safely on Twitter, we must provide more tools and policies. With hundreds of millions of Tweets sent per day, the volume of content on Twitter is massive, which makes it extraordinarily complex to strike the right balance between fighting abuse and speaking truth to power. It requires a multi-layered approach where each individual on our platform has a part to play, as do the community of experts working for safety and free expression.

The Twitter Trust and Safety Council provides input on our safety products, policies, and programs. Twitter works with safety advocates, academics, and researchers; grassroots advocacy organizations that rely on Twitter to build movements; and community groups working to prevent abuse.

As we develop products, policies, and programs, our Trust & Safety Council help us tap into the expertise and input of organizations at the intersection of these issues more efficiently and quickly. Our diverse list of partners include safety advocates, academics, and researchers focused on minors, media literacy, digital citizenship, and efforts around greater compassion and empathy on the Internet; grassroots advocacy organizations that rely on Twitter to build movements and momentum; and community groups with an acute need to prevent abuse, harassment, and bullying, as well as mental health and suicide prevention.

These organizations are chosen based on their areas of focus and subject matter expertise, rather than due to any particular political ideology. If the Chairman has additional suggestions for groups that may be able to provide additional expertise on the Trust and Safety Council, we welcome that input.

24. Mr. Dorsey, what percentage of Twitter content moderation reviews are conducted by

actual human beings rather than via automated processes and artificial intelligence?

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

25. Mr. Dorsey, how many employees have been hired and are currently assigned to content moderation and reviews? Currently, what training and instruction do reviewers receive to ensure Twitter users were in compliance with the company's Terms of Service? What salary range does Twitter pay employees assigned to content moderation and reviews?

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Failure to do so may result in Twitter taking one or more of the following enforcement actions: (1) requiring an individual to delete prohibited content before he or she can again create new posts and interact with other Twitter users; (2) temporarily limiting an individual's ability to create posts or interact with other Twitter users; (3) requesting an individual verify account ownership with a phone number or email address; or (4) permanently suspending an account or related accounts.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. In order to maintain a safe environment for individuals on Twitter, we may suspend accounts that violate the Twitter Rules. Most of the accounts we suspend are suspended because they are automated or fake accounts, and they introduce security risks for Twitter and all of the individuals who use our service. These types of accounts are contrary to our Twitter Rules. Unfortunately, sometimes a real person's account gets suspended by mistake, and in those cases we work with the person to make sure the account is unsuspended. If we suspect an account has been hacked or compromised, we may suspend it until it can be secured and restored to the account owner in order to reduce potentially malicious activity caused by the compromise.

As described above, we may suspend an account if it has been reported to us as violating our Rules surrounding abuse. When an account engages in abusive behavior, like sending threats to others, we may suspend it temporarily or, in some cases, permanently. An individual may be able to unsuspend his or her own account by providing a phone number or confirming an email address.

An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her

account for a specific period of time or may be asked to verify certain information before proceeding.

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 engage with the account holder via email to resolve the suspension.

26. Mr. Dorsey, does Twitter have a list of specific users it surveils more closely than other Twitter users because of a history of use of fake accounts?

We have updated the Twitter Rules to provide clearer guidance around several key issues, including fake account, attributed activity, and distribution of hacked materials. We have heard feedback that people think our rules about spam and fake accounts only cover common spam tactics like selling fake goods. As platform manipulation tactics continue to evolve, we are updating and expanding our rules to better reflect how we identify fake accounts, and what types of inauthentic activity violate our guidelines. We now may remove fake accounts engaged in a variety of emergent, malicious behaviors. Some of the factors that we will take into account when determining whether an account is fake include the use of stock or stolen avatar photos, use of stolen or copied profile bios, and use of intentionally misleading profile information, including profile location

Additionally, as per the Twitter Rules, if we are able to reliably attribute an account on Twitter to an entity known to violate the Twitter Rules, we will take action on additional accounts associated with that entity. We are expanding our enforcement approach to include accounts that deliberately mimic or are intended to replace accounts we have previously suspended for violating our rules. Further, our rules prohibit the distribution of hacked material that contains private information or trade secrets, or could put people in harm's way. We are also expanding the criteria for when we will take action on accounts which claim responsibility for a hack, which includes threats and public incentives to hack specific people and accounts. Commentary about a hack or hacked materials, such as news articles discussing a hack, are generally not considered a violation of this policy.

Twitter has increased our ability to identify and remove individuals who have previously been banned from our service.

27. Mr. Dorsey, how do you define a "fake account" and what steps does Twitter under take to identify fake accounts?

a. Once a fake account is identified, what specific steps does Twitter undertake to verify that an account should be removed?

Twitter continues to develop the detection tools and systems needed to combat malicious automation on our platform. Twitter has refined its detection systems. Twitter prioritizes

identifying suspicious account activity, such as exceptionally high-volume Tweeting with the same hashtag or mentioning the same @handle without a reply from the account being addressed, and requires an individual using the platform to confirm control. Twitter has also increased its use of challenges intended to catch automated accounts, such as reCAPTCHAs, that require individuals to identify portions of an image or type in words displayed on screen, and password reset requests that protect potentially compromised accounts. Twitter is also in the process of implementing mandatory email or cell phone verification for all new accounts.

Our efforts have been effective. Due to technology and process improvements, we are now removing 214% more accounts year-over-year for violating our platform manipulation policies. For example, over the course of the last several months, our systems identified and challenged between 8.5 million and 10 million accounts each week suspected of misusing automation or producing spam. Spam can be generally described as unsolicited, repeated actions that negatively impact other people. This includes many forms of automated account interactions and behaviors as well as attempts to mislead or deceive people. This constitutes more than three times the 3.2 million we were catching in September 2017. We thwart 530,000 suspicious logins a day, approximately double the amount of logins that we detected a year ago.

These technological improvements have brought about a corresponding reduction in the number of spam reports from people on Twitter, evidence to us that our systems' ability to automatically detect more malicious accounts and potential bad faith actors than they did in the past. We received approximately 25,000 such reports per day in March of this year; that number decreased to 17,000 in August.

We also removed locked accounts from people's follower counts, to ensure these figures are more reliable. Accounts are locked when our systems detect unusual activity and force a password change or other challenge. If the challenge has not been met or the password has not been changed within a month, the account is locked, barring it from sending Tweets, Retweets or liking posts from others.

28. Mr. Dorsey, in January of this year, the New York Times published an article entitled "The Follower Factory" that revealed some Twitter users purchase followers and retweets from a company who specializes in selling automated bot accounts. Is it a violation of the Twitter User Agreement to purchase followers and retweets ?

a. Since the publication of the New York Times' article "The Follower Factory" what steps has Twitter taken to address the issue of users purchasing fake accounts and retweets?

b. What are the harms and concerns about artificially inflating the number of followers on social media by paying for it?

Twitter Rules specifically prohibit users from creating, purchasing, and selling fake accounts. Twitter considers such activity as prohibited spam, which includes creating fake accounts, account interactions, or impressions; selling, purchasing, or attempting to artificially

inflate account interactions (such as followers, Retweets, likes, etc.); and using or promoting third-party services or applications that claim to provide user accounts with more followers, Retweets, or likes (such as follower trains, sites promising "more followers fast," or any other site that offers to automatically add followers or engagements to your account or Tweets).

Twitter proactively enforces these rules, including by requiring users to confirm their phone numbers or pass a reCAPTCHA challenge to confirm that the account is being operated by a real person. On a weekly basis, Twitter challenges or suspends millions of suspicious or potentially fake accounts, including upon registration and log in.

After the New York Times article was published, Twitter used techniques, including pattern identification, to identify potential fake accounts, including those described in the article. These techniques must constantly evolve because the characteristics of fake accounts evolve constantly to evade detection. Twitter also ran searches to identify copycat accounts, or accounts that post content copied from actual users. This process can also lead to identification of fake accounts.

29. Mr. Dorsey, the Twitter Rules provide that "if voices are silenced because people are afraid to speak up" then the underlying philosophy of free speech and expression means little.

a. Do algorithms make the determination that certain content silences voices and, if so, what factors are considered by the algorithm ?

b. Do humans make the determination that certain content silences voices and, if so, what factors are considered?

Twitter does not use political ideology to make any decisions -- whether the decision is made by algorithms or humans. Political ideology does not factor into decisions regarding ranking content on our service or how we enforce our rules. We believe strongly in being impartial, and we strive to enforce our rules impartially. We do not shadowban anyone based on political ideology. In fact, from a simple business perspective and to serve the public conversation, Twitter is incentivized to keep all voices on the platform.

30. Mr. Dorsey, the Twitter Rules specifically state that users may not use the platform for "any unlawful purpose or in furtherance of illegal activities." If Twitter determines content is unlawful or in furtherance of illegal activities, does Twitter notify law enforcement?

We have well-established relationships with law enforcement agencies, particularly in the arena of hostile foreign action, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of Homeland Security's Election Security Task Force. We look forward to continued cooperation with them on these issues, as only they have access to information critical to our joint efforts to stop bad faith actors.

Information sharing and collaboration are critical to Twitter's success in preventing hostile foreign actors from disrupting meaningful political conversations on the platform. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We each possess information the other does not have, and our combined information is more powerful in combating these threats together.

31. Mr. Dorsey, has Twitter, or any third parties it engages, audited its input data and ranking system to determine exactly how much bias each contributes to output bias in resulting Twitter search results? With what frequency does Twitter or others does conduct such audits or studies?

We want Twitter to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets per day on Twitter, we ha have invested heavily in building systems that organize content on Twitter to show individuals using the platform the most the relevant information for that individual first. We want to do the work for our customers to make it a positive and informative experience. With 335 million people using Twitter every month in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance.

To preserve the integrity of our platform and to safeguard our democracy, Twitter has also employed technology to be more aggressive in detecting and minimizing the visibility of certain types of abusive and manipulative behaviors on our platform. The algorithms we use to do this work are tuned to prevent the circulation of Tweets that violate our Terms of Service, including the malicious behavior we saw in the 2016 election, whether by nation states seeking to manipulate the election or by other groups who seek to artificially amplify their Tweets.

Despite the success we are seeing with our use of algorithms to combat abuse, manipulation, and bad faith actors, we recognize that even a model created without deliberate bias may nevertheless result in biased outcomes. Bias can happen inadvertently due to many factors, such as the quality of the data used to train our models. In addition to ensuring that we are not deliberately biasing the algorithms, it is our responsibility to understand, measure, and reduce these accidental biases. This is an extremely complex challenge in our industry, and algorithmic fairness and fair machine learning are active and substantial research topics in the machine learning community. The machine learning teams at Twitter are learning about these techniques and developing a roadmap to ensure our present and future machine learning models uphold a high standard when it comes to algorithmic fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it's critical that we be more transparent about our efforts in this space.

32. Mr. Dorsey, what suggestions do you have for raising the awareness of Twitter users by signaling bias in Twitter search and timeline results?

Twitter does not use political ideology to make any decisions -- whether related to ranking content on our service or how we enforce our rules. We believe strongly in being

impartial, and we strive to enforce our rules impartially. We do not shadowban anyone based on political ideology. In fact, from a simple business perspective and to serve the public conversation, Twitter is incentivized to keep all voices on the platform.

33. Mr. Dorsey, Twitter announced August 30th a new issue ads policy and certification process in the U.S. When will the company begin to enforce its new issue ads policy? Please describe in detail what was the previous issue ads policy and certification process, if there was one?

Twitter first implemented an updated Political Campaigning Policy to provide clearer guidance about how we define political content and who can promote-political content on our platform. Under the revised policy, advertisers who wish to target the United States with federal political campaigning advertisements are required to self-identify as such and certify that they are located within the United States. Foreign nationals will not be permitted to serve political ads to individuals who identify as being located in the United States.

Twitter accounts that wish to target the U.S. with federal political campaigning advertisements must also comply with a strict set of requirements. Among other things, the account's profile photo, header photo, and website must be identical to the individual's or organization's online presence. In addition, the advertiser must take steps to verify that the address used to serve advertisements with content related to a federal political campaign is genuine.

To further increase transparency and better educate those who access promoted content, accounts serving ads with content related to a federal political campaign will now be visually identified and contain a disclaimer. This feature will allow people to more easily identify federal political campaign advertisements, quickly identify the identity of the account funding the advertisement, and immediately tell whether it was authorized by the candidate.

In June, we launched the Ads Transparency Center, which is open to everyone on Twitter and the general public, and currently focuses on electioneering communications. Twitter requires extensive information disclosures of any account involved in federal electioneering communications and provides specific information to the public via the Ads Transparency Center, including:

- Purchases made by a specific account;
- All past and current ads served on the platform for a specific account;
- Targeting criteria and results for each advertisement;
- Number of views each advertisement received; and
- Certain billing information associated with the account.

These are meaningful steps that will enhance the Twitter experience and protect the health of political conversations on the platform.

In addition,Twitter's efforts to provide transparency continue with the launch of a U.S.-specific Issue Ads Policy and certification process, which occurred prior to the 2018 U.S. midterm elections. The new policy impacts advertisements that refer to an election or a clearly identified candidate or advertisements that advocate for legislative issues of national importance. To provide people with additional information about individuals or organizations promoting issue ads, Twitter has established a process that verifies an advertiser's identity and location within the United States. These advertisements will also be included in the Ads Transparency Center. We are also examining how to adopt political campaigning and issue ads policies globally. We remain committed to continuing to improve and invest resources in this space.

34. Mr. Dorsey, how many advertisers -- either individuals, organizations, or campaigns -- have applied and received certification under Twitter's new process?

As of November 19, 2018, 234 electioneering and 321 issue advertisers were certified in the United States. Twitter serves the public conversation by promoting health and earning the trust of the people who use our service. We cannot succeed unless the American people have confidence in the integrity of the information found on the platform, especially with respect to information relevant to elections and the democratic process. In addition to the advertising policies described in the answer to question 33, and to promote transparency and assist our stakeholders in identifying messages from elected officials and those who are running for office, we have made a concerted effort to verify all major party candidates for both federal and key state positions. Through verification – a blue badge that appears next to a person's Twitter handle throughout the platform – we let people know that accounts of public interest are the authentic accounts (as opposed to impersonation or parody accounts).

In addition, we have developed a new U.S. election label to identify political candidates. The label includes information about the office the candidate is running for, the state the office is located in, and the district number, if applicable. Accounts of candidates who have qualified for the general election and who are running for governor or for the U.S. Senate or House of Representatives will display an icon of a government building. These new features are designed to instill confidence that the content people are viewing is reliable and accurately reflects candidates' and elected officials' positions and opinions.

35. Mr. Dorsey, please explain in detail specifically how will campaign and issue ads be labeled or highlighted in a user's timeline? How many campaign and issue ads are currently labeled under the new policy and process? From their timeline can a user now click on the promoted ads to immediately see information about the advertiser's identity and location, like a "Learn More" button?

Please see the response to question 33.

36. Mr. Dorsey, under Twitter's August 30th issues ad policy, will both campaign and issue

ads be viewable in Twitter's Ads Transparency Center? Please explain what specific detail will be viewable to a user?

Yes, both campaign and issue ads are viewable in the Twitter's Ads Transparency Center. In June, we launched the Ads Transparency Center, which is open to everyone on Twitter and the general public, and currently focuses on electioneering communications. Twitter requires extensive information disclosures of any account involved in federal electioneering communications and provides specific information to the public via the Ads Transparency Center, including:

- Purchases made by a specific account;
- All past and current ads served on the platform for a specific account;
- Targeting criteria and results for each advertisement;
- Number of views each advertisement received; and
- Certain billing information associated with the account.

37. Until 1987, the Fairness Doctrine required broadcasters to provide a right of reply to ensure the presentation of balanced views on issues of public importance, during a time when broadcast was a dominant news source for most Americans. While the FCC repealed the Fairness Doctrine, the principle of providing balanced perspectives is still important in journalism. When amplifying certain topics on the platform, does Twitter seek to ensure that it promotes a balanced variety of viewpoints on a particular issue? If so, what steps are taken to ensure representation of a variety of viewpoints, including those that are less popular with users?

Twitter's purpose is to serve the public conversation. We are an American company that serves our global audience by focusing on the people who use our service, and we put them first in every step we take. Twitter is used as a global town square, where people from around the world come together in an open and free exchange of ideas. We must be a trusted and healthy place that supports free and open discussion.

Twitter has publicly committed to improving the collective health, openness, and civility of public conversation on our platform. Twitter's health is measured by how we help encourage more healthy debate, conversations, and critical thinking, including exposure to diverse perspectives. Conversely, abuse, malicious automation, and manipulation detracts from the health of our platform. We are committed to hold ourselves publicly accountable towards progress of our health initiative.

There are other guiding objectives we consider to be core to our company. We must ensure that all voices can be heard. We must continue to make improvements to our service so that everyone feels safe participating in the public conversation – whether they are speaking or

simply listening. And we must ensure that people can trust in the credibility of the conversation and its participants.

Let me be clear about one important and foundational fact: Twitter does not use political ideology to make any decisions, whether related to ranking content on our service or how we enforce our rules. We believe strongly in being impartial, and we strive to enforce our rules impartially. We do not shadowban anyone based on political ideology. In fact, from a simple business perspective and to serve the public conversation, Twitter is incentivized to keep all voices on the platform.

38. In 1996, 10 years before the original Twitter platform was released, Section 230 of the Communications Decency Act was enacted, distinguishing interactive online platforms from traditional publishers by setting up a safe harbor protecting them from the lawsuits publishers may face over third party generated content. This was based on the premise that interactive services are essentially neutral platforms, not exercising the full editorial judgment wielded by a publisher. Although like the internet message boards that were prevalent when Section 230 was enacted, Twitter gets much of its content from the users, given the curating power of the Twitter algorithms, coupled with other efforts on the company's part to proactively promote or de-prioritize particular content or users, are you exercising editorial judgment?

It is important to note that like any other private citizen, Twitter has a First Amendment right to free speech. Twitter speaks when it enforces our rules about what content should and should not be allowed on our platform. For example, by prohibiting members of violent extremist groups from using the platform, Twitter is exercising its First Amendment right to declare its disapproval of content from, or the message of, such groups.

39. In *Six4three, LLC v. Facebook, Inc.,* a case concerning alleged anticompetitive effects of its content management practices, Facebook recently argued that it is a publisher and that its editorial decisions are therefore protected by the First Amendment.

a. Would you characterize Twitter in the same way?

b. Should content moderation efforts that make value judgments about quality, veracity, or tone, similar to decisions made by publishers, affect the applicability of Section 230, which specifies that Internet platforms cannot be treated as publishers?

Private companies do have First Amendment rights, including rights to free speech, and these rights should be considered.

40. Earlier this year, the Allow States and Victims to Fight Online Sex Trafficking Act of 2017 (FOSTA) was signed into law. This legislation represented the first successful attempt to amend Section 230 since it was enacted in 1996.

a. Has Twitter changed any of its procedures with regard to detecting and removing sex trafficking content as a result of FOSTA?

b. Has the legislation impacted Twitter's bottom line?

c. Does the Twitter user experience differ in other markets where internet platforms are not protected with safe harbor legislation similar to Section 230?

Sex trafficking is an illegal act. An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content.

Additionally, Twitter prohibits the promotion of adult sexual content globally. This policy applies to Twitter's paid advertising products. Examples of adult sexual content include pornography, escort services and prostitution, full and partial nudity, dating sites which focus on facilitating sexual encounters or infidelity, dating sites in which money, goods or services are exchanged in return for a date; penis enlargement products & services and breast enhancement services, modeled clothing that is sexual in nature, mail order bride services, sex toys, and host and hostess clubs. All advertisers must comply with this advertising policy and with all applicable laws and regulations.

41. In your testimony, you indicated that Wall Street did not approve of Twitter's recent actions to de-activate suspicious accounts, ultimately lowering the number of followers of certain accounts. While Twitter can afford to do this in the short term, Twitter, as a publicly held company, has a fiduciary duty to its shareholders. How do you expect the shifted focus on the health of the conversation taking place on the platform, rather than maximizing user engagement, to impact the company?

The purpose of Twitter is to serve the public conversation. We serve our global audience by focusing on the needs of the people who use our service, and we put them first in every step we take. We want to be a global town square, where people from around the world come together in an open and free exchange of ideas. We must be a trusted and healthy place that supports free and open democratic debate.

Twitter is committed to improving the collective health, openness, and civility of public conversation on our platform. Twitter's is built and measured by how we help encourage more healthy debate, conversations, and critical thinking. Conversely, abuse, malicious automation, and manipulation detracts from it. We are committing Twitter to hold ourselves publicly accountable towards progress.

We do see the health initiative and objective as a long-term growth vector for the company. Although it does have short-term implications on daily active users, we ultimately believe that we are driving this towards a growth vector. We believe it is important to solve issues around unhealthy behavior on the platform and better the experience more broadly on

Twitter. This is an extremely important initiative to us, not only for the experience of Twitter, but we believe the long-term growth of the platform, and we are proud of our progress so far.

42. For some events or topics, Twitter features custom emojis that automatically appear when a user types a certain hashtag, which is a form of promoting or prioritizing certain content. How is it decided to allow a custom emoji for a hashtag, and who designs the emojis? Can users apply for this service? Are the designs done in consultation with the event coordinators or users actively promoting these hashtags?

On Twitter, advertisers can design custom branded emojis that are triggered when a specific hashtag is used. We know people quickly move through their timeline and a Tweet needs to stand out to get attention. We call this "stopping power" and branded emojis help marketers achieve it. The amount of attention ads receive increases by almost 10% when branded emojis are included in the ad.

When branded emojis are paired with a promoted video, the emotional connection and interest in the ad increases six-fold, as people are more focused on the ad. Further, campaigns with branded emojis extend a brand's presence across Twitter in a way that is personal and authentic to the brand. In fact, the median number of earned media generated is 5.3 million Tweet impressions, representing a 420% increase compared to the earned media baseline.

Twitter also supports emojis that can catalyze conversations around important national or international events.

43. The "trending topics" list is one important method Twitter uses to promote or prioritize certain content. Is a topic chosen to be featured in this list strictly based on how many posts are made on a specific hashtag within a period of time, or do any human decisions intervene? Anecdotally, some users have observed that hashtags with fewer tweets sometimes appear on trending lists while those with more tweets do not.

a. A "tailored trends" list is generated based on a user's location and who they follow on Twitter. How does Twitter use location in determining which trends or content to display to an individual user? What is the source of the location data being used?

When individuals on Twitter log into their accounts, they have immediate access to a range of tools and account settings to access, correct, limit, delete or modify the personal data provided to Twitter and associated with the account, including public or private settings, marketing preferences, and applications that can access their accounts. These data settings can be used to better personalize the individual's use of Twitter and allow him or her the opportunity to make informed choices about whether Twitter collects certain data, how it is used, and how it is shared.

For example, individuals can change the personalization and data settings for their Twitter account, including:

- Whether interest-based advertisements are shown to an individual on and off the Twitter platform;
- How Twitter personalizes an individual's experience across devices;
- Whether Twitter collects and uses an individual's precise location;
- Whether Twitter personalizes their experience based on places they have been; and
- Whether Twitter keeps track of the websites where an individual sees Twitter content.

An individual on Twitter can disable all personalization and data setting features with a single master setting prominently located at the top of the screen.

44. Rep. Walberg stated in his question to you that ISPs have certain transparency requirements that require the ISPs to disclose if they are altering traffic, and asked you if a similar idea would be helpful in the tech industry. You replied, "That [transparency] is a good idea and would help earn people's trust." What should those transparency requirements look like with regard to content moderation practices?

Twitter is committed to the open exchange of information. First published on July 2, 2012, our biannual Twitter Transparency Report highlights trends in legal requests, intellectual property-related requests, and email privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from worldwide government and non-government legal requests we have received for account information. Removal requests are also included in the Transparency Report and include worldwide legal demands from governments and other authorized reporters, as well as reports based on local laws from trusted reporters and non-governmental organizations, to remove or withhold content.

The Transparency Report also includes information on government requests to remove content that may violate Twitter's Terms of Service (TOS) under the following Twitter Rules categories: abusive behavior, copyright, promotion of terrorism, and trademark. It does not include legal demands, regardless of whether they resulted in a TOS violation, which will continue to be published in our removal request section report. As we take an objective approach to processing global TOS reports, the fact that the reporters in these cases happened to be government officials had no bearing on whether any action was taken under our Rules. We continue to look for ways to improve The Transparency Report.

45. Mr. Dorsey, understanding that in your testimony you stated that algorithmic bias is a relatively "new problem" and that Twitter is "early" in its learning process in terms of managing it, what research, development, and other tool deployment is Twitter currently undertaking to understand, manage, and address algorithmic bias on its platform?

We want Twitter to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets per day on Twitter, we ha have invested heavily in building systems that organize content on Twitter to show individuals using the platform the most the relevant information for that individual first. We want to do the work for our customers to make it a positive and informative experience. With 335 million people using Twitter every month in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance.

To preserve the integrity of our platform and to safeguard our democracy, Twitter has also employed technology to be more aggressive in detecting and minimizing the visibility of certain types of abusive and manipulative behaviors on our platform. The algorithms we use to do this work are tuned to prevent the circulation of Tweets that violate our Terms of Service, including the malicious behavior we saw in the 2016 election, whether by nation states seeking to manipulate the election or by other groups who seek to artificially amplify their Tweets.

Despite the success we are seeing with our use of algorithms to combat abuse, manipulation, and bad faith actors, we recognize that even a model created without deliberate bias may nevertheless result in biased outcomes. Bias can happen inadvertently due to many factors, such as the quality of the data used to train our models. In addition to ensuring that we are not deliberately biasing the algorithms, it is our responsibility to understand, measure, and reduce these accidental biases. This is an extremely complex challenge in our industry, and algorithmic fairness and fair machine learning are active and substantial research topics in the machine learning community. The machine learning teams at Twitter are learning about these techniques and developing a roadmap to ensure our present and future machine learning models uphold a high standard when it comes to algorithmic fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it's critical that we be more transparent about our efforts in this space.

46. Mr. Dorsey, in your testimony, you stated that Twitter reviewers look for indicators of "fairness" and "impartiality" when reviewing algorithmic output to try and identify bias.

a. Please provide Twitter' s definition of "fairness" as it applies to Twitter's oversight of its algorithms.

b. Please provide Twitter's definition of "impartiality" as it applies to Twitter's oversight of its algorithms.

Twitter uses a range of behavioral signals to determine how Tweets are organized and presented in the home timeline, conversations, and search based on relevance. Twitter relies on behavioral signals—such as how accounts behave and react to one another—to identify content that detracts from a healthy public conversation, such as spam and abuse. Unless we have determined that a Tweet violates Twitter policies, it will remain on the platform, and is available in our product. Where we have identified a Tweet as potentially detracting from healthy conversation (*e.g.*, as potentially abusive), it will only be available to view if you click on "Show more replies" or choose to see everything in your search setting.

Some examples of behavioral signals we use, in combination with each other and a range of other signals, to help identify this type of content include: an account with no confirmed email address, simultaneous registration for multiple accounts, accounts that repeatedly Tweet and mention accounts that do not follow them, or behavior that might indicate a coordinated attack. Twitter is also examining how accounts are connected to those that violate our rules and how they interact with each other. The accuracy of the algorithms developed from these behavioral signals will continue to improve over time.

These behavioral signals are an important factor in how Twitter organizes and presents content in communal areas like conversation and search. Our primary goal is to ensure that relevant content and Tweets contributing to healthy conversation will appear first in conversations and search. Because our service operates in dozens of languages and hundreds of cultural contexts around the globe, we have found that behavior is a strong signal that helps us identify bad faith actors on our platform.

The behavioral ranking that Twitter utilizes does not consider in any way political views or ideology. It focuses solely on the behavior of all accounts. Twitter is always working to improve our behavior-based ranking models such that their breadth and accuracy will improve over time. We use thousands of behavioral signals in our behavior-based ranking models—this ensures that no one signal drives the ranking outcomes and protects against malicious attempts to manipulate our ranking systems.

The Honorable Michael C. Burgess

1. Mr. Dorsey, you stated during the hearing that Twitter is building technologies so that it doesn't have to wait on reports from victims to act on violent or threatening content.

a. When do you anticipate implementation of these new technologies?

b. Currently, how many reports of inappropriate or threatening content does Twitter receive per day?

i. How many of these reports result in action?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

In the cases of violent threats, Twitter recommends that in addition to reporting the abusive content to the platform, the individual considers contacting local law enforcement and we provide a tool that allows the individual to generate an email with all of the relevant necessary information to submit a law enforcement report. Local law enforcement agencies can accurately assess the validity of the threat, investigate the source of the threat, and respond to concerns about physical safety. If Twitter is contacted by law enforcement directly, we can work with them and provide the necessary information for their investigation of the threat. We continuously deploy new technological tools to identify Tweets and accounts that violate our Terms of Service.

2. Mistakes do happen, and we understand that, but we would like to understand what the process is at Twitter if a mistake happens. That process is still very unclear.

a. If a user does not believe they have violated the Twitter Rules, what recourse does that user have? What is the internal process at Twitter to review an appeal? Are employees deciding whether they agree with a users ' posts?

b. There are high profile Twitter users who seem to have a different level of recourse from, for example, a constituent of mine who uses Twitter for news and conversations with friends. Are there different channels for review in these cases?

In order to maintain a safe environment for individuals on Twitter, we may suspend accounts that violate the Twitter Rules. Most of the accounts we suspend are suspended because they are automated or fake accounts, and they introduce security risks for Twitter and all of the individuals who use our service. These types of accounts are contrary to our Twitter Rules. Unfortunately, sometimes a real person's account gets suspended by mistake, and in those cases we work with the person to make sure the account is unsuspended. If we suspect an account has been hacked or compromised, we may suspend it until it can be secured and restored to the account owner in order to reduce potentially malicious activity caused by the compromise.

We may suspend an account if it has been reported to us as violating our Rules surrounding abuse. When an account engages in abusive behavior, like sending threats to others, we may suspend it temporarily or, in some cases, permanently. An individual may be able to unsuspend his or her own account by providing a phone number or confirming an email address.

An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her account for a specific period of time or may be asked to verify certain information before proceeding.

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 engage with the account holder via email to resolve the suspension.

3. Mr. Dorsey, I understand that Twitter is conducting internal investigations that will result in a transparency report.

a. When do you anticipate being able to share this information with the Energy and Commerce Committee?

Twitter is committed to the open exchange of information. First published on July 2,

2012, our biannual Twitter Transparency Report highlights trends in legal requests, intellectual property-related requests, and email privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from worldwide government and non-government legal requests we have received for account information. Removal requests are also included in the Transparency Report and include worldwide legal demands from governments and other authorized reporters, as well as reports based on local laws from trusted reporters and non-governmental organizations, to remove or withhold content.

The Transparency Report also discloses information on third-party requests that compel Twitter to remove content for legal reasons ("legal requests") under our Country Withheld Content ("CWC") policy. Governments (including law enforcement agencies), organizations chartered to combat discrimination, and lawyers representing individuals are among the many complainants that submit legal requests included below. For example, we may receive a court order requiring the removal of defamatory statements in a particular jurisdiction, or law enforcement may ask us to remove prohibited content such as Nazi symbols in Germany.

In December 2017, Twitter updated its in-product messaging about withheld content to better explain why content has been withheld. Subsequently, we began to differentiate between legal demands (e.g., court orders) and reports based on local law(s) (e.g., reports alleging the illegality of particular content in a certain country). To further increase transparency, this change is also reflected in the report below.

The Transparency Report also includes information on government requests to remove content that may violate Twitter's Terms of Service (TOS) under the following Twitter Rules categories: abusive behavior, copyright, promotion of terrorism, and trademark. It does not include legal demands, regardless of whether they resulted in a TOS violation, which will continue to be published in our removal request section report. As we take an objective approach to processing global TOS reports, the fact that the reporters in these cases happened to be government officials had no bearing on whether any action was taken under our Rules.

The Transparency Report also includes the total number of Digital Millennium Copyright Act (DMCA) takedown notices and counter notices received for Twitter and Periscope content, along with data about the top five copyright reporters across both platforms. The Vine app was transitioned in January of 2017. Trademark notices include reports of alleged Trademark Policy violations received for Twitter and Periscope.

4. Mr. Dorsey, I've long been concerned about the role the Internet and online platforms, such as Google, Facebook, Pinterest and Twitter, have played in enabling access to deadly and illegal controlled substances. Given the current opioid crisis, would you please describe in detail what specific policies and procedures Twitter has in place to crack down on illegal online sales and marketing of opioids and pharmaceutical drugs.

a. Does Twitter have policies and procedures in place to disable the ability to use the Twitter search function for the sales and marketing of controlled

substances? If yes, please outline in detail those policies and procedures. If no, does Twitter plan on implementing such measures and when?

b. Does Twitter have policies and procedures in place to report to Federal, State or international law enforcement information Twitter receives indicating that an individual or organization is engaged in the sale or marketing of controlled substances? If yes, please outline in detail those policies and procedures. If no, does Twitter plan on implementing such measures and when?

c. Has Twitter established a 24/7 point of contact with whom Federal, State or international law enforcement can communicate directly if law enforcement has information indicating that an individual or organization is engaged in the sale or marketing of controlled substances? If you have, please outline in detail what person or department is Twitter is responsible, and provide their contact information. If you have not, are you planning on implementing such measures and when?

d. Does Twitter have specific information in its "Help Center" for users to report the sale or marketing of controlled substances on the platform, similar to the help modules on "online abuse" and "self-harm and suicide"? If yes, please outline in detail what they are. If no, does Twitter plan on implementing such measures and when?

Twitter agrees that addiction to opioids and the overdoses that result are incredibly serious. Twitter appreciates the Committee's strong leadership on this public health crisis. Our terms of service are strong against illegal opioid sales: An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content.

We have well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We each possess information the other does not have, and our combined information is more powerful in combating these threats together. We have continuous coverage to address reports from law enforcement around the world and have a portal to swiftly handle law enforcement requests rendered by appropriate legal process. Additional information about our policies and procedures for law enforcement requests can be found here: https://help.twitter.com/en/rules-and-policies/twitter-law-enforcement-support.

Twitter has participated in events convened by White House to address this public health crisis, including the Opioid Summit and Be Best launch. We have participated in summit

organized by the Food and Drug Administration, with researchers, industry peers, and various governmental stakeholders, including Drug Enforcement Agency, the Department of Justice, and Health and Human Services. And we have hosted a series of events on opioids, including an event on "Combating the Opioid Epidemic" at our Washington, D.C. Twitter offices with the Surgeon General and Republican members of Congress.

Additionally, we provided a custom emoji to support the Drug Enforcements Efforts around #TakebackDay. We tweeted from official Twitter handles to drive engagement and awareness around the event designed to encourage people to safely dispose of unwanted prescription drugs. Twitter also provides ongoing Twitter trainings to nongovernmental organizations focused on recovery efforts, including Lilly's Place and Young People in Recovery, to maximize their reach and effectiveness. We provide these stakeholders information on their use the platform to help people in need, provide education around abuse and recovery, and increase fundraising on Twitter.

The Honorable Robert E. Latta

1. Mr. Dorsey, in your written statement you indicated Twitter conducted an internal analysis of the Members of Congress affected by the auto-suggest search issue, and that you would make that information available to the Committee, if requested. Please provide the internal analysis done by Twitter with respect to the auto-suggest search issue.

In July 2018, we publicly acknowledged that some accounts were not being auto-suggested even when people were searching for their specific name. Our technology was using a decision-making criteria that considered the behavior of people following these accounts. Specifically, the followers of these accounts had a higher proportion of abusive behavior on the platform.

This issue impacted 600,000 accounts across the globe. The vast majority of impacted accounts were not political in nature. The issue impacted 53 accounts of politicians in the U.S., representing 0.00883 percent of total affected accounts. This subset of affected accounts includes 10 accounts of Republican Members of Congress. The remainder of impacted political accounts relate to campaign activity and affected candidates across the political spectrum.

To be clear, this issue is limited solely to the search auto-suggestion function. The accounts, their Tweets, and all surrounding conversation about those accounts remained displayed in search results. Once identified, this issue was promptly resolved within 24 hours.

An analysis of accounts for Members of Congress that were affected by this search issue demonstrates that there was no negative effect on the growth of their follower counts. To the contrary, follower counts of impacted Members of Congress increased.

Twitter will not reveal the names of the impacted accounts due to privacy and security concerns, but we can disclose information concerning the accounts of those who have publicly discussed they were impacted.

We have enclosed information concerning follower count data displayed in graph format for the official accounts of Congressmen Matt Gaetz, Jim Jordan, Mark Meadows, and Devin Nunes. The graphs contain publicly available data on the follower counts for the three month period surrounding the incident, which occurred in late July. The other impacted members of Congress also saw their followers increase during this time period.

2. Mr. Dorsey, in your testimony you indicated that Twitter has made more than 30 policy and product changes since the beginning of last year to improve health on your platform. Please provide a complete accounting of all 30-plus policy and product changes made.

Twitter recently developed and launched more than 30 policy and product changes designed to foster information integrity and protect the people who use our service from abuse and malicious automation. Twitter has made a number of improvements specifically in

preparation for the 2018 U.S. midterm election. Described below are highlights of our election integrity work, and we are happy to follow up with you in greater detail.

Twitter continues to develop the detection tools and systems needed to combat malicious automation on our platform. Twitter has refined its detection systems. Twitter prioritizes identifying suspicious account activity, such as exceptionally high-volume Tweeting with the same hashtag or mentioning the same @handle without a reply from the account being addressed, and requires an individual using the platform to confirm control. Twitter has also increased its use of challenges intended to catch automated accounts, such as reCAPTCHAs, that require individuals to identify portions of an image or type in words displayed on screen, and password reset requests that protect potentially compromised accounts. Twitter is also in the process of implementing mandatory email or cell phone verification for all new accounts.

Our improvements include important structural changes. We recently reorganized the structure of the company to allow our valued employees greater durability, agility, invention, and entrepreneurial drive. The reorganization simplified the way we work, and enabled all of us to focus on health of our platform.

In particular, we have created an internal cross-functional analytical team whose mission is to monitor site and platform integrity. Drawing on expertise across the company, the analytical team can respond immediately to escalations of inauthentic, malicious automated or human-coordinated activity on the platform. The team's work enables us to better understand the nature of the malicious activity and mitigate it more quickly.

Our cross-functional team has developed a political conversations dashboard to evaluate the integrity of political conversations on the platform in the aggregate, focusing primarily (but not exclusively) on elections in the United States in the near term. For example, this dashboard surfaces information about sudden shifts in sentiment around a specific conversation, suggesting a potential coordinated campaign of activity, as well as information about groups of potentially linked accounts that are posting about the same topic.

Twitter serves the public conversation by promoting health and earning the trust of the people who use our service. We cannot succeed unless the American people have confidence in the integrity of the information found on the platform, especially with respect to information relevant to elections and the democratic process. To promote transparency and assist our stakeholders in identifying messages from elected officials and those who are running for office, we have made a concerted effort to verify all major party candidates for both federal and key state positions. Through verification – a blue badge that appears next to a person's Twitter handle throughout the platform – we let people know that accounts of public interest are the authentic accounts (as opposed to impersonation or parody accounts).

In addition, we have developed a new U.S. election label to identify political candidates. The label includes information about the office the candidate is running for, the state the office is located in, and the district number, if applicable. Accounts of candidates who have qualified for the general election and who are running for governor or for the U.S. Senate or House of Representatives will display an icon of a government building. These new features are designed

to instill confidence that the content people are viewing is reliable and accurately reflects candidates' and elected officials' positions and opinions.

We have devoted considerable resources to increasing transparency and promoting accountability in the ads served to Twitter customers. Twitter first implemented an updated Political Campaigning Policy to provide clearer guidance about how we define political content and who can promote political content on our platform. Under the revised policy, advertisers who wish to target the United States with federal political campaigning advertisements are required to self-identify as such and certify that they are located within the United States. Foreign nationals will not be permitted to serve political ads to individuals who identify as being located in the United States.

Twitter accounts that wish to target the U.S. with federal political campaigning advertisements must also comply with a strict set of requirements. Among other things, the account's profile photo, header photo, and website must be identical to the individual's or organization's online presence. In addition, the advertiser must take steps to verify that the address used to serve advertisements with content related to a federal political campaign is genuine.

To further increase transparency and better educate those who access promoted content, accounts serving ads with content related to a federal political campaign will now be visually identified and contain a disclaimer. This feature will allow people to more easily identify federal political campaign advertisements, quickly identify the identity of the account funding the advertisement, and immediately tell whether it was authorized by the candidate.

In June, we launched the Ads Transparency Center, which is open to everyone on Twitter and the general public, and currently focuses on electioneering communications. Twitter requires extensive information disclosures of any account involved in federal electioneering communications and provides specific information to the public via the Ads Transparency Center

In addition,we recently announced the next phase of our efforts to provide transparency with the launch of a U.S.-specific Issue Ads Policy and certification process. The new policy impacts advertisements that refer to an election or a clearly identified candidate or advertisements that advocate for legislative issues of national importance. To provide people with additional information about individuals or organizations promoting issue ads, Twitter has established a process that verifies an advertiser's identity and location within the United States. These advertisements will also be included in the Ads Transparency Center. We are also examining how to adopt political campaigning and issue ads policies globally. We remain committed to continuing to improve and invest resources in this space.

To further address malicious automation and abuse on the platform, we have also recently updated our developer policies, which govern the access and use of public Tweet data made available to developers and other third parties through our application programming interfaces ("APIs").

The Honorable Cathy McMorris Rodgers

1. Your company is an immense supporter of net neutrality regulations being applied to Internet Service Providers (ISP's) based on Title II of the Telecommunications Act written in 1934. We can all agree to the basic principles of Net Neutrality and that there shouldn't be blocking, throttling or discrimination by those companies but I also believe that Title II is the wrong approach.

a. Do you believe ISPs and Edge Providers have a different relationship to their customers? If so, do you believe that ISPs and Edge Providers have different responsibilities regarding the throttling and blocking of content?

b. Can you identify who your customers are?

c. Do you believe that standard consensus rules of the road regarding ISP throttling and blocking should apply only ISPs?

d. Would you agree to have your company adhere to the same regulations you are advocating the ISP's live by? And if not, why?

We disagree with the decision by the Federal Communications Commission to weaken net neutrality provisions. American democracy is built on freedom of expression. As Benjamin Franklin said, "Freedom of speech is a principal pillar of a free government: When this support is taken away, the constitution of a free society is dissolved." Increasingly, Americans are going online to enter the public sphere. According to a March 2017 poll, 57 percent of Americans shared their political opinions or feelings via social media channels. Nearly 80 percent believe social media channels impact public policy outcomes, and more than half said social media had some impact on their voting decision.

Twitter allows every user to have a voice. As a global platform for free expression and live conversation, Twitter has become a significant medium to give voice to the voiceless and to speak truth to power. Our open platform facilitates conversations and connections that would have been unlikely or impossible before the first Tweet was sent in 2006. From popular uprisings that challenge oppressive regimes during the #ArabSpring to local tragedies that highlight social injustice in #Ferguson, Missouri, these moments could not have emerged from local connections to reach an international audience without the free and open conversation that Twitter facilitates.

Furthermore, Twitter has advanced how elected officials and candidates communicate with constituents and the electorate, allowing them to connect in real-time on a global scale without permission from large corporate gatekeepers. At least 856 Twitter accounts belong to heads of state and government and foreign ministers in 178 countries, representing 92 percent of all United Nations member states, with a combined audience of 356 million followers. Because of platforms like Twitter, virtually anyone can share his or her views on issues of local or national importance with government leaders. As a result, leaders have a stronger understanding of both the people whom they represent and issues of civic interest as well as increased

accountability.

Net Neutrality is one of the most important free expression issues of our time. Without Net Neutrality, Internet Service Providers could charge content providers gratuitous new fees to reach Internet users and discriminate in the prioritized delivery of such content, frustrating the free flow of information. Without Net Neutrality in force, ISPs would be able to block content they don't like, reject apps and content that compete with their own offerings, and arbitrarily discriminate against content providers by prioritizing the Internet traffic of some over others.

ISPs and edge providers have a different relationships with their users. Because broadband Internet access is increasingly provided by fewer and fewer companies in a concentrated ISP marketplace, smaller and noncommercial voices may be impacted adversely because they are unable to pay a broadband toll for "fast lane" service. Allowing ISPs to determine what content is relegated to the backwaters of the Internet in second or third-tier status could reduce the visibility and impact of important voices in the local, national, or global media mix. In this context, Net Neutrality is a modern-day safeguard to protect freedom of expression. Without strong rules supporting Net Neutrality, Twitter's core value of promoting and supporting the freedom of expression of our users is at risk.

2. On Privacy, there is clearly a significant public debate about private companies' access to, storage of, and monetization of people's data.

a. What do you believe is the role of consumers regarding control of their information?

b. Do you believe that we need greater transparency of private companies' privacy and data usage practices?

c. Do you believe consumers should have greater control over their information? What do you make of proposed opt-in policies for Internet Service Providers?

d. Do you believe ISPs and Edge Providers like you should be subject to the same or different standards as it relates to consumer privacy? If so, can you elaborate?

To ensure such trust, the privacy of the people who use our service is of paramount importance to Twitter. We believe privacy is a fundamental right, not a privilege. Privacy is part of Twitter's DNA. Since Twitter's creation over a decade ago, we have offered a range of ways for people to control their experience on Twitter, from creating pseudonymous accounts to letting people control who sees their Tweets, in addition to a wide array of granular privacy controls. This deliberate design has allowed people around the world using Twitter to protect their privacy.

That same philosophy guides how we work to protect the data people share with Twitter. Twitter empowers the people who use our services to make informed decisions about the data they share with us.

Twitter believes individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared. Twitter is always working to improve transparency into what data is collected and how it is used. Twitter designs its services so that individuals can control the personal data that is shared through our services. People that use our services have tools to help them control their data. For example, if an individual has registered an account, through their account settings they can access, correct, delete or modify the personal data associated with their account.

The Honorable David McKinley

1. A recent study in the American Journal of Public Health analyzed a number of tweets over 5 months and identified nearly 2000 sites linked to illicit online drug sales on Twitter. Your website states that this is prohibited. Can you tell me how many of those 2000 sites are still up?

Twitter agrees that addiction to opioids and the overdoses that result are incredibly serious. Twitter appreciates the Committee's strong leadership on this public health crisis. Our rules of service are strong on illegal opioid sales: An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content. We remove any accounts that are linked to illicit online drug sales as soon as they are brought to our attention. We continuously improve our ability to locate and remove accounts claiming to sell drugs on our platform.

We have well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We each possess information the other does not have, and our combined information is more powerful in combating these threats together.

Twitter has participated in events convened by White House to address this public health crisis, including the Opioid Summit and Be Best launch. We have participated in summit organized by the Food and Drug Administration, with researchers, industry peers, and various governmental stakeholders, including Drug Enforcement Agency, the Department of Justice, and Health and Human Services. And we have hosted a series of events on opioids, including an event on "Combating the Opioid Epidemic" at our Washington, D.C. Twitter offices with the Surgeon General and Republican members of Congress.

Additionally, we provided a custom emoji to support the Drug Enforcements Efforts around #TakebackDay. We tweeted from official Twitter handles to drive engagement and awareness around the event designed to encourage people to safely dispose of unwanted prescription drugs. Twitter also provides ongoing Twitter trainings to nongovernmental organizations focused on recovery efforts, including Lilly's Place and Young People in Recovery, to maximize their reach and effectiveness. We provide these stakeholders information on their use the platform to help people in need, provide education around abuse and recovery, and increase fundraising on Twitter.

2. Can you explain Twitter's current process for locating and removing criminal and prohibited activity on the platform? What actions are you taking to make this process more efficient and effective?

Anyone can report illegal behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is harmful. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. An individual can also contact local law enforcement and we provide a tool that allows the individual to generate an email with all of the relevant necessary information to submit a law enforcement report. Local law enforcement agencies can accurately assess the validity of the threat, investigate the source of the threat, and respond to concerns raised by the Tweet. If Twitter is contacted by law enforcement directly, we can work with them and provide the necessary information for their investigation of the threat.

We also recognize that collaboration with our industry peers and civil society is critically important to addressing common threats and that it has been successful in meeting shared challenges. In June 2017, for example, we launched the Global Internet Forum to Counter Terrorism (the "GIFCT"), a partnership among Twitter, YouTube, Facebook, and Microsoft.

The GIFCT facilitates, among other things, information sharing; technical cooperation; and research collaboration, including with academic institutions. In September 2017, the members of the GIFTC announced a multimillion dollar commitment to support research on terrorist abuse of the Internet and how governments, tech companies, and civil society can respond effectively. We are looking to establish a network of experts that can develop these platform-agnostic research questions and analysis that consider a range of geopolitical contexts.

The GIFCT has created a shared industry database of "hashes"—unique digital "fingerprints"—for violent terrorist imagery or terrorist recruitment videos or images that have been removed from our individual services. 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 even block extremist content before it is posted in the first place.

The database now contains more than 88,000 hashes. Instagram, Justpaste.it, LinkedIn, Oath, and Snap have also joined this initiative, and we are working to add several additional companies in 2018. Twitter also participates in the Technology Coalition, which shares images to counter child abuse. The database works to surface content for human review against each platform's respective terms of service. This is essential to take into account the context, for example academic or news media use.

Because each platform is unique, there are many elements of our coordinated work that do not translate easily across platforms. Although we share with other companies our approach to addressing shared threats, including certain signals that we use to identify malicious content,

solutions applicable to the Twitter platform are not always applicable to other companies. We describe our tools as “in-house and proprietary” to distinguish them from tools that are developed by and licensed from third-party vendors.

The Honorable Gus Bilirakis

1. I have heard from my local school districts that they must consistently respond to threats of school violence. In your response you mentioned that the company conducts outreach to local entities and law enforcement officials when you see anything impacting potential physical security threats. What are the specific implementations of that process?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

In the cases of violent threats, Twitter recommends that in addition to reporting the abusive content to the platform, the individual considers contacting local law enforcement and we provide a tool that allows the user to generate an email with all of the relevant necessary information to submit a law enforcement report. Local law enforcement agencies can accurately assess the validity of the threat, investigate the source of the threat, and respond to concerns about physical safety. If Twitter is contacted by law enforcement directly, we can work with them and provide the necessary information for their investigation of the threat. Twitter may also contact law enforcement proactively if we become aware of an exigent situation that is likely to result in imminent harm or death.

2. How does Twitter determine the efficacy of the company's process for responding to potential threats?

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Failure to do so may result in Twitter taking one or more of the following enforcement actions: (1) requiring an individual to delete prohibited content before he or she can again create new posts and interact with other Twitter users; (2) temporarily limiting an individual's ability to create posts or interact with other Twitter users; (3) requesting an individual verify account ownership with a phone number or email address; or (4) permanently suspending an account or related accounts.

Accounts under investigation or which have been detected as sharing content in violation with the Twitter Rules may have their account or visibility limited in various parts of Twitter, including search. Our policies and enforcement options evolve continuously to address emerging behaviors online.

Our enforcement options have expanded significantly over the years. We originally had only one enforcement option: account suspension. Since then, we've added a range of enforcement actions and now have the ability to take action at the Tweet, Direct Message, and account levels. Additionally, we take measures to educate individuals that have violated our rules about the specific tweet(s) in violation and which policy has been violated. We also continue to improve the technology we use to prioritize reports that are most likely to violate our rules and last year we introduced smarter, more aggressive witness reporting to augment our approach.

We have embarked on an effort to better measure the health of the conversation on our platform and will be transparent with our results.

3. You said you are always looking for ways to improve the process for quickly alerting the company to potential threats and responding with outreach to the proper authorities, and would be open to an implementation that can be scaled. In what ways would you consider working directly with local school districts and institutions, which play a key role in protecting its student body?

Twitter is continuing to explore and invest in what more we can do with our technology, enforcement options, and policies – not just in the U.S., but to everyone we serve around the world. We are committed to continuing to improve and to holding ourselves accountable as we work to make Twitter better for everyone.

We are open to ideas, from local school districts and other institutions, to better use of our technology to catch people working to evade a Twitter suspension and identifying malicious, automated accounts, and more quickly activating our team to ensure a human review element continues to be present within our all of automated processes.

Local school districts and other institutions should always raise any concerns to local law enforcement. In line with our Privacy Policy, we may disclose account information to law enforcement in response to a valid emergency disclosure request. Twitter evaluates emergency disclosure requests on a case-by-case basis in compliance with relevant law (e.g., 18 U.S.C. § 2702(b)(8) and Section 8 Irish Data Protection 1988 and 2003). If we receive information that

provides us with a good faith belief that there is an exigent emergency involving the danger of death or serious physical injury to a person, we may provide information necessary to prevent that harm, if we have it. If there is an exigent emergency that involves the danger of death or serious physical injury to a person that Twitter may have information necessary to prevent, law enforcement officers can submit an emergency disclosure request through our Legal Request Submissions site (the quickest and most efficient method).

4. You mention that Twitter takes a data-driven approach to arranging Moments, based primarily on the amount of conversation happening on a particular topic or event, and then afterwards going into "impartiality" to provide many differing perspectives. Are there editorial standards the company has set for what makes a Moment, and if so, what are they? If no standards, why not, and would you consider using an official standard?

Moments are curated stories showcasing the very best of what's happening on Twitter. The Moments guide on Twitter is customized to show a person on the platform current topics that are popular or relevant, so an individual can discover what is unfolding on Twitter in an instant.

Twitter has standards and policies for our curators that are intended to reduce bias, and increase accuracy and quality. Additional information concerning these policies is located here: https://help.twitter.com/en/rules-and-policies/twitter-moments-guidelines-and-principles.

An individual can navigate to the Moments tab to see customized Moments from Today. An individual can select Moments categorized by News, Sports, Entertainment, Fun, and more. When a person the platform sees a Moment he or she would like to explore, the individual can click it to view the entire story. A person the platform can like a moment, or reply, Retweet, and like the Tweet captured by the Moment.

People on the platform create their own individual Twitter Moments. Twitter Moments are curated stories about what's happening around the world—powered by Tweets. It's easy to individuals to create their own story with Twitter Moments. There are three ways an individual can begin creating his or her own Moment via twitter.com. The person can access Moments through the Moments tab, on the profile page, or through a Tweet detail. To get started, an individual only needs is a title, description, Tweets, and a selected cover image.

5. Regarding algorithms used to flag malicious behavior, what is the "error" rate, meaning how often are tweets overturned upon review by an employee content reviewer?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

In the cases of violent threats, Twitter recommends that in addition to reporting the abusive content to the platform, the individual considers contacting local law enforcement and we provide a tool that allows the individual to generate an email with all of the relevant necessary information to submit a law enforcement report. Local law enforcement agencies can accurately assess the validity of the threat, investigate the source of the threat, and respond to concerns about physical safety. If Twitter is contacted by law enforcement directly, we can work with them and provide the necessary information for their investigation of the threat.

6. How many employees do you have to review content and make decisions on policing behavior? How long does it take on average to get an "error" overturned? Is there notification to user during the steps of this process?

In order to maintain a safe environment for individuals on Twitter, we may suspend accounts that violate the Twitter Rules. Most of the accounts we suspend are suspended because they are automated or fake accounts, and they introduce security risks for Twitter and all of the individuals who use our service. These types of accounts are contrary to our Twitter Rules. Unfortunately, sometimes a real person's account gets suspended by mistake, and in those cases we work with the person to make sure the account is unsuspended. If we suspect an account has been hacked or compromised, we may suspend it until it can be secured and restored to the account owner in order to reduce potentially malicious activity caused by the compromise.

We may suspend an account if it has been reported to us as violating our Rules surrounding abuse. When an account engages in abusive behavior, like sending threats to others, we may suspend it temporarily or, in some cases, permanently. An individual may be able to unsuspend his or her own account by providing a phone number or confirming an email address.

An account may also be temporarily disabled in response to reports of automated or abusive behavior. For example, an individual may be prevented from Tweeting from his or her account for a specific period of time or may be asked to verify certain information before proceeding.

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 engage with the account holder via email to resolve the suspension.

7. Is the content reviewing algorithm technology unique to Twitter or is this something that is developing industry wide? If the former and it shows promise, is it something they are willing to share with competitors to provide a uniform industry standard?

Although we cannot speak to the innovative technologies being employed at other companies, Twitter has strengthened its information sharing with its industry peers, particularly to address shared challenges such as violent extremism, child sexual exploitation, and state-sponsored disinformation operations. Information sharing and collaboration are critical to Twitter's success in preventing hostile foreign actors from disrupting meaningful political conversations on the platform. We recognize the value of inputs we receive from our industry peers about hostile foreign actors. We have shared and remain committed to sharing information across platforms to better understand and address the threat of hostile foreign interference with the electoral process.

The Honorable Susan Brooks

1. During the hearing, I talked about a bill of mine that passed into law in 2015, the OHS Social Media Improvement Act. The bill created a working group of relevant stakeholders to put their heads together to look at the impact social media has on preparedness, response, and recovery. The working group, housed in OHS' Science and Technology Directorate, also looks at how to counteract misinformation spread via Twitter during disasters and how first responders and communities can effectively use social media during times of crisis. This working group has to-date made 3 reports, which highlight countering false information in disasters and emergencies, best practices for incorporating social media into exercises, and how to operationalize social media for public safety. You said you were not aware of this working group, but would be willing to consider these reports. Below I have provided links to the reports. If you would like more information about the group, please reach out to my Legislative Assistant, Mimi Strobel (202.225.2276 or mimi.strobel@mail.house.gov).

- **April 2016 - "From Concept to Reality: Operationalizing Social Media for Preparedness, Response and Recovery;" (https://www.dhs.gov/sites/default/files/publications/SMWG From-Concept-to Reality-Operationalizing-Social-Media-508.pdf);**
- **March 2017 - "Best Practices for Incorporating Social Media into Exercises;" (https://www.dhs.gov/publication/best-practices-incorporating-social-mediaexercises); and,**
- **March 2018 - "Countering False Information on Social Media in Disasters and Emergencies" (https://www.dhs.gov/sites/default/files/publications/SMWG Countering-False_Info-Social-Media-Disasters-EmergenciesMar2018-508.pdf).**

Thank you for this information. Twitter agrees that our speed and reach are critical in disaster response. We work cooperatively with governments around the world to better use the platform to keep their constituencies safe. Additional information regarding our Crisis Response and Natural Emergency work can be found here: https://about.twitter.com/content/dam/about-twitter/values/twitter-for-good/Twitter-Crisis-Response-One-Pager.pdf

The Honorable Earl "Buddy" Carter

1. Mr. Dorsey, you mentioned during my questioning that you wanted to reboot the verification process. Can you please describe the steps that Twitter is taking to overhaul that process? What will the new process look like? What is the timeline?

Verification was meant to authenticate identity and voice but it is interpreted as an endorsement or an indicator of importance. Twitter recognizes that we have created this confusion and need to resolve it. Beginning in November 2017, We have paused all general verifications while we work to resolve this issue.

Verification has long been perceived as an endorsement. We gave verified accounts visual prominence on the service which deepened this perception. We should have addressed this earlier but did not prioritize the work as we should have. This perception became worse when we opened up verification for public submissions and verified people who we in no way endorse.

Twitter is currently working on a new authentication and verification program. In the meantime, we are not accepting any public submissions for verification and have introduced new guidelines for the program.

2. Mr. Dorsey, you mentioned that there is a new behavioral trend regarding algorithms and the illegal opioid sales present on your platform. You also mentioned that you need to look at how your algorithms are determining when they see this activity and can take action. Please describe how you will be reviewing your algorithms for their behavior and what changes you will be implementing.

Twitter agrees that addiction to opioids and the overdoses that result are incredibly serious. Twitter appreciates the Committee's strong leadership on this public health crisis. Our rules of service are strong on illegal opioid sales: An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content.

We have well-established relationships with law enforcement agencies, and we look forward to continued cooperation with them on these issues, as often they have access to information critical to our joint efforts to stop bad faith actors. The threat we face requires extensive partnership and collaboration with our government partners and industry peers. We each possess information the other does not have, and our combined information is more powerful in combating these threats together.

Twitter has participated in events convened by White House to address this public health crisis, including the Opioid Summit and Be Best launch. We have participated in summit organized by the Food and Drug Administration, with researchers, industry peers, and various governmental stakeholders, including Drug Enforcement Agency, the Department of Justice, and Health and Human Services. And we have hosted a series of events on opioids, including an

event on "Combating the Opioid Epidemic" at our Washington, D.C. Twitter offices with the Surgeon General and Republican members of Congress.

Additionally, we provided a custom emoji to support the Drug Enforcements Efforts around #TakebackDay. We tweeted from official Twitter handles to drive engagement and awareness around the event designed to encourage people to safely dispose of unwanted prescription drugs. Twitter also provides ongoing Twitter trainings to nongovernmental organizations focused on recovery efforts, including Lilly's Place and Young People in Recovery, to maximize their reach and effectiveness. We provide these stakeholders information on their use the platform to help people in need, provide education around abuse and recovery, and increase fundraising on Twitter.

3. Mr. Dorsey, intellectual property violations not only hurt those who directly develop this content, but also those industries that have developed around them. Across Georgia, thousands of well-paying jobs have been developed as a result of this growth. How is Twitter taking action to address intellectual property theft or rebroadcasting to keep up with changing practices? A quick search found numerous accounts that have been up for years with thousands of followers that continuously tweet out these links. How has Twitter not addressed these accounts and what steps will you take moving forward?

Twitter responds to copyright complaints submitted under the Digital Millennium Copyright Act ("DMCA"). 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. Note that not all unauthorized uses of copyrighted materials are infringements.

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 user 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.

The Honorable Frank Pallone, Jr.

1. At the hearing, I asked you about Twitter' s training and resources for content moderation. You could not provide specific responses at the time. Please provide specific responses to the following questions:

a. How many human content moderators does Twitter employ in the U.S. and how much do they get paid?

b. How many hours of training is given to them to ensure consistency in their decisions?

Twitter strives to provide an environment where people can feel free to express themselves. If abusive behavior happens, Twitter wants to ensure that it is easy for people to report it to us. In order to ensure that people feel safe expressing diverse opinions and beliefs, Twitter prohibits behavior that crosses the line into abuse, including behavior that harasses, intimidates, or uses fear to silence another's voice.

Twitter has approximately 4,000 employees worldwide and at various times in their work, all will focus on improving the health of the conversation on Twitter. Our team is continuously trained, not only as we are onboarded, but particularly following changes to the Twitter Terms of Service.

Anyone can report abusive behavior directly from a Tweet, profile, or Direct Message. An individual navigates to the offending Tweet, account, or message and selects an icon that reports that it is abusive or harmful. Other options are available, for example posting private information or a violent threat. Multiple Tweets can be included in the same report, helping us gain better context while investigating the issues to resolve them faster. For some types of report Twitter also prompts the individual to provide more information concerning the issue that is being reported.

Twitter uses a combination of machine learning and human review to adjudicate abuse reports and whether they violate our rules. Context matters when evaluating abusive behavior and determining appropriate enforcement actions. Factors we may take into consideration include, but are not limited to whether: the behavior is targeted at an individual or group of people; the report has been filed by the target of the abuse or a bystander; and the behavior is newsworthy and in the legitimate public interest. Twitter subsequently provides follow-up notifications to the individual that reports the abuse. We also provide recommendations for additional actions that the individual can take to improve his or her Twitter experience, for example using the block or mute feature.

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Failure to do so may result in Twitter taking one or more of the following enforcement actions: (1) requiring an individual to delete prohibited content before he or she can again create new posts and interact with other Twitter users; (2) temporarily limiting an

individual's ability to create posts or interact with other Twitter users; (3) requesting an individual verify account ownership with a phone number or email address; or (4) permanently suspending an account or related accounts.

Accounts under investigation or which have been detected as sharing content in violation with the Twitter Rules may have their account or visibility limited in various parts of Twitter, including search. Our policies and enforcement options evolve continuously to address emerging behaviors online.

2. What steps is Twitter taking to improve the consistency of its enforcement and the metrics that demonstrate improvement?

Our enforcement options have expanded significantly over the years. We originally had only one enforcement option: account suspension. Since then, we've added a range of enforcement actions and now have the ability to take action at the Tweet, Direct Message, and account levels. Additionally, we take measures to educate individuals that have violated our rules about the specific tweet(s) in violation and which policy has been violated. We also continue to improve the technology we use to prioritize reports that are most likely to violate our rules and last year we introduced smarter, more aggressive witness reporting to augment our approach.

We believe an important component of improving the health on Twitter is to measure the health of conversation that occurs on the platform. This is because in order to improve something, one must be able to measure it. By measuring our contribution to the overall health of the public conversation, we believe we can more holistically approach our impact on the world for years to come.

Earlier this year, Twitter began collaborating with the non-profit research center Cortico and the Massachusetts Institute of Technology Media Lab on exploring how to measure aspects of the health of the public sphere. As a starting point, Cortico proposed an initial set of health indicators for the United States (with the potential to expand to other nations), which are aligned with four principles of a healthy public sphere. Those include:

- Shared Attention: Is there overlap in what we are talking about?
- Shared Reality: Are we using the same facts?
- Variety: Are we exposed to different opinions grounded in shared reality?
- Receptivity: Are we open, civil, and listening to different opinions?

Twitter strongly agrees that there must be a commitment to a rigorous and independently vetted set of metrics to measure the health of public conversation on Twitter. And in order to develop those health metrics for Twitter, we issued a request for proposal to outside experts for their submissions on proposed health metrics, and methods for capturing, measuring, evaluating and reporting on such metrics. Our expectation is that successful projects will produce

peer-reviewed, publicly available, open-access research articles and open source software whenever possible.

As a result of our request for proposal, we are partnering with experts at the University of Oxford and Leiden University and other academic institutions to better measure the health of Twitter, focusing on informational echo chambers and unhealthy discourse on Twitter. This collaboration will also enable us to study how exposure to a variety of perspectives and opinions serves to reduce overall prejudice and discrimination. While looking at political discussions, these projects do not focus on any particular ideological group and the outcomes will be published in full in due course for further discussion.

The Honorable Debbie Dingell

1. What is the minimum number of individuals you allow to be targeted for advertisements?

Tailored audiences is a product feature that advertisers use to target existing people on the platform and customers. For example, Advertisers can reach existing customers by uploading a list of email addresses. Advertisers can also target those that have recently visited the advertisers' websites or reach those that have taken specific action in an application, such as installation or registration.

Twitter informs individuals on the platform about Tailored Audiences in several ways. For example, Twitter describes this activity in its Privacy Policy, an "Ads info" footer on twitter.com, and the "Why am I seeing this ad?" section of the drop down menu on Twitter ads themselves. Each of these locations describe interest-based advertising on Twitter and explain how to use the associated privacy controls. In addition, the "Your Twitter Data" tool allows individuals on Twitter to download a list of advertisers that have included them in a Tailored Audience.

If people on Twitter do not want Twitter to show them Tailored Audience ads on and off of Twitter, there are several ways they can turn off this feature: using their Twitter settings, they can visit the Personalization and data settings and adjust the Personalize ads setting; if they are on the web, they can visit the Digital Advertising Alliance's consumer choice tool at optout.aboutads.info to opt out of seeing interest-based advertising from Twitter in their current browser; if they do not want Twitter to show them interest-based ads in Twitter for iOS on their current mobile device, they can enable the "Limit Ad Tracking" setting in their iOS phone's settings; and if they do not want Twitter to show them interest-based ads in Twitter for Android on their current mobile device, they can enable "Opt out of Ads Personalization" in an Android phone's settings.

In addition to explaining Tailored Audiences to people on the platform, offering them several ways to disable the feature, and enabling them to view the advertisers who have included them in Tailored Audiences, as described above, the Tailored Audience legal terms require that advertisers have secured all necessary rights, consents, waivers, and licenses for use of data.

Advertisers are also required to provide all people from whom the data is collected with legally-sufficient notice that fully discloses the collection, use, and sharing of the data that is provided to Twitter for purposes of serving ads targeted to people's interest, and legally sufficient instructions on how they can opt out of interest-based advertising on Twitter.

2. Do you have any existing partnerships with device makers?

Twitter does have agreements in place with a wide range of device manufacturers around the world, which grant rights to pre-install applications to use Twitter on mobile devices. These agreements are meant to eliminate the need for end-users to manually go through the download

and installation process upon device purchase and activation; they do not otherwise grant access to Twitter user information. They provide a user experience similar to that in the United States, where consumers see various applications pre-populated on their phones when initializing service.

Twitter also has agreements in place to enable device manufacturers to provide a "Find Your Friends"-like feature on mobile phones, which allows users to find the Twitter handles of their existing contacts. This feature takes a user's contact list (email addresses and phone numbers) and returns associated Twitter handles via an API. This integration does not return or provide email addresses or phone numbers. Like within the Twitter application, this feature respects user choice: any Twitter user can turn off their discoverability setting, and they will not be discoverable through the application program interface, or API.

It is important to note that all Twitter APIs honor the long-standing principle of defending the privacy of our users, reflecting our core values as a company. As stated in our Developer Agreement and Policy, we prohibit the use of Twitter content for targeting, segmenting, or profiling individuals based on health, negative financial status or condition, political affiliation or beliefs, race, ethnicity, religious or philosophical affiliation or beliefs, sex life or sexual orientation, trade union membership, data relating to any alleged or actual commission of a crime, or any other sensitive categories of personal information prohibited by law.

We also prohibit any entity accessing Twitter content from conducting any analyses or research that isolates a group of individuals or any single individual for any unlawful or discriminatory purpose or in a manner that would be inconsistent with our users' reasonable expectations of privacy.

While Twitter does not sell private user data, we note that users of our commercial data platform seeking access to public data must complete a rigorous review and approval process before we grant them access to Twitter data via our enterprise and premium APIs, and are subject to regular reviews and policy checks once they have access. API users who are found to be in violation of our policies are subject to enforcement actions, including immediate termination of API access.

3. When you look at the performance of your AI, what is an acceptable error rate?

We want Twitter to provide a useful, relevant experience to all people using our service. With hundreds of millions of Tweets per day on Twitter, we ha have invested heavily in building systems that organize content on Twitter to show individuals using the platform the most the relevant information for that individual first. We want to do the work for our customers to make it a positive and informative experience. With 335 million people using Twitter every month in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance.

Twitter is continuing to improve our systems so they can better detect any issues that

arise and correct for them. We will continue to refine our approach and will be transparent about why we make the decisions that we do.

American technology companies have thrived, in part, because we have had the freedom to experiment, learn and iterate at great speed. We are continuously improving our machine learning tools, but an expectation that we will not, or cannot make mistakes, is unrealistic. We will continue to tinker and learn from errors and will strive to be as transparent as possible along the way.

4. Will you be deploying any of these Al programs to combat election propaganda?

Yes. To preserve the integrity of our platform and to safeguard our democracy, Twitter has also employed algorithmic technology to be more aggressive in detecting and minimizing the visibility of certain types of abusive and manipulative behaviors on our platform. The algorithms we use to do this work are tuned to prevent the circulation of Tweets that violate our Terms of Service, including the malicious behavior we saw in the 2016 election, whether by nation states seeking to manipulate the election or by other groups who seek to artificially amplify their Tweets. Additional information regarding our election integrity efforts can be found here: https://about.twitter.com/en_us/values/elections-integrity.html

5. Because influence operations are always changing tactics, how often do you retrain your models to adapt to these changes?

We continuously improve our approaches to combat malicious automation and curtail efforts to manipulate the conversation on Twitter. Twitter uses a range of behavioral signals to determine how Tweets are organized and presented in the home timeline, conversations, and search based on relevance. Twitter relies on behavioral signals—such as how accounts behave and react to one another—to identify content that detracts from a healthy public conversation, such as spam and abuse. Unless we have determined that a Tweet violates Twitter policies, it will remain on the platform, and is available in our product. Where we have identified a Tweet as potentially detracting from healthy conversation (*e.g.*, as potentially abusive), it will only be available to view if you click on "Show more replies" or choose to see everything in your search setting.

Some examples of behavioral signals we use, in combination with each other and a range of other signals, to help identify this type of content include: an account with no confirmed email address, simultaneous registration for multiple accounts, accounts that repeatedly Tweet and mention accounts that do not follow them, or behavior that might indicate a coordinated attack. Twitter is also examining how accounts are connected to those that violate our rules and how they interact with each other. The accuracy of the algorithms developed from these behavioral signals will continue to improve over time.

6. Does Twitter have the capacity to monitor IP addresses that access multiple accounts to look for networks of malicious activity?

Twitter continues to develop the detection tools and systems needed to combat malicious automation on our platform. Twitter has refined its detection systems. Twitter prioritizes identifying suspicious account activity, such as exceptionally high-volume Tweeting with the same hashtag or mentioning the same @handle without a reply from the account being addressed, and requires an individual using the platform to confirm control. Twitter has also increased its use of challenges intended to catch automated accounts, such as reCAPTCHAs, that require individuals to identify portions of an image or type in words displayed on screen, and password reset requests that protect potentially compromised accounts. Twitter is also in the process of implementing mandatory email or cell phone verification for all new accounts.

IP addresses can be an important signal for our automated technology, but there are also many other signal that can be useful, such as phone numbers and email addresses.

7. Like you, I believe understanding how AI comes to certain decisions is vital to building public trust. Although the study of this is in its infancy, what work is Twitter doing to move towards explainable AI?

This is an extremely complex challenge in our industry, and algorithmic fairness and fair machine learning are active and substantial research topics in the machine learning community. The machine learning teams at Twitter are developing a roadmap to ensure our present and future machine learning models uphold a high standard when it comes to algorithmic fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it's critical that we be more transparent about our efforts in this space.

The Honorable Jerry McNerney

1. As I noted at the hearing, some social media platforms have recently been accused of facilitating discriminatory advertising-such as housing and employment ads. I want to know what Twitter is doing to identify and remedy any potential similar practices. During our exchange at the hearing, you stated "[w]e do regular audits of how our ads are targeted and how they're delivered and we work to make sure that we have fairness within them."

a. Are these audits performed by Twitter employees or are they conducted by third parties?

b. What role do algorithms have in carrying out these audits?

c. Does each audit examine an individual ad? If not, how many ads are made part of each audit?

d. How often are these audits conducted?

e. When did Twitter begin conducting these audits?

f. When you said "fairness within them" what did you mean by this?

g. Do these audits specifically examine whether an ad is being targeted in a discriminatory way?

h. Do these audits specifically examine whether any discriminatory effects may have resulted from Twitter's own ad optimization process?

i. Are there any other steps that Twitter takes to determine whether ads are being targeted by advertisers in a discriminatory way? Please explain in detail what these steps entail.

j. Are there any other steps that Twitter takes to determine whether discriminatory effects have occurred as a result of its own ad optimization process? Please explain in detail what these steps entail.

k. What happens once an instance is identified in which an ad has been targeted in a discriminatory way or ultimately delivered in a way that resulted in discriminatory effects?

Twitter does not allow discriminatory advertising, such as in housing or employment, on our platform. An individual who uses Twitter may not use our service for any unlawful purposes or in furtherance of illegal activities. By using Twitter, an individual agrees to comply with all applicable laws governing his or her online conduct and content. Because discrimination in

housing and employment is against the law, an individual using Twitter may not engage in this illegal activity on our platform.

In addition to this general prohibition against illegal activities by all individuals using Twitter, there are specific Twitter Ads Policies that apply to paid advertising products on the platform and prohibit discrimination. Our Twitter Ads Policies, as well as our legal agreements that govern the use of Twitter for advertising, expressly state that advertisers on Twitter are responsible for their Twitter Ads. This means all advertisers must follow all applicable laws and regulations, create honest ads, and advertise safely and respectfully. Our policies and agreements with advertisers require all advertisers to follow the law. The Twitter Privacy Policy emphasizes the specific prohibition on using our advertising service in discriminatory ways, stating: "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."

Twitter takes violations of our Twitter Ads policies, the Twitter Rules, and Terms of Service seriously. We will examine reported violations and take appropriate action, which may include removal of offending advertisements and advertisers from the Twitter Ads platform. Additional information about our ads policies can be found here: https://business.twitter.com/en/help/ads-policies/other-policy-requirements/policies-for-keyword-targeting.html

2. Does Twitter have any protocols in place for what actions should be taken when it is discovered that ads are being targeted in a discriminatory way or being delivered in a way that ultimately results in discriminatory effects? If so, please provide a copy of the documents that describe these protocols.

The Twitter Terms of Service, Twitter Master Service Agreement (for advertising), Twitter Ads Policies, and Twitter Privacy Policy each prohibit discriminatory advertising behavior, as described above. Twitter has additional rules for advertisers using Twitter's tailored audiences products. Tailored audiences is a product feature that advertisers use to target existing people on the platform and customers. For example, advertisers can reach existing customers by uploading a list of email addresses. Advertisers can also target those that have recently visited the advertisers' websites or reach those that have taken specific action in an application, such as installation or registration.

In addition to explaining Tailored Audiences to people on the platform, offering them several ways to disable the feature, and enabling them to view the advertisers who have included them in Tailored Audiences, the Tailored Audience legal terms require that advertisers have secured all necessary rights, consents, waivers, and licenses for use of data.

Advertisers are also required to provide all people from whom the data is collected with legally-sufficient notice that fully discloses the collection, use, and sharing of the data that is provided to Twitter for purposes of serving ads targeted to people's interest, and legally sufficient instructions on how they can opt out of interest-based advertising on Twitter.

Twitter prohibits the creation of tailored audiences based on any sensitive information, which includes: alleged or actual commission of a crime; health; genetic and/or biometric data; negative financial status or condition; racial or ethnic origin; religious or philosophical affiliation or beliefs; or sex life.

3. During the hearing, you acknowledged that it is possible for discrimination to result from how advertisers are able to target ads on your platform (specifically, that they are able to establish criteria that includes and excludes categories of users). Yet, as the company's CEO, you were unable to answer my question if Twitter has ever taken down an ad because of potential discriminatory effects and instead told me that you would have to follow up to get the information. It was incredibly troubling that you did not know the answer to this question after acknowledging that discrimination is possible.

a. Has Twitter ever taken down any ads because they were being targeted in a discriminatory way? If so, how many?

b. Has Twitter ever taken down any ads because of discriminatory effects that resulted from Twitter' s own ad optimization process? If so, how many?

c. Has Twitter ever taken down any ads because the content of the ad was discriminatory? If so, how many?

The Twitter Terms of Service, Ads Policies, and Privacy Policy each prohibit discriminatory advertising behavior, as described in the responses to questions one and two. There are additional rules that govern the uses of keyword targeting in the Twitter timeline. Advertisers using keyword targeting in timeline may not select keywords that target sensitive categories.

Twitter prohibits keyword targeting based on any sensitive information, which includes: alleged or actual commission of a crime; health; genetic and/or biometric data; negative financial status or condition; racial or ethnic origin; religious or philosophical affiliation or beliefs; or sex life.

Further, advertisers may not create advertisements which assert or imply knowledge of personally identifiable or sensitive information, even when the ad has been created and targeted

without using such information. Advertisers using keyword targeting must adhere to this policy and all other Twitter Ads policies in order to advertise.

Targeting people who use Twitter based on sensitive categories is a violation of our Twitter Ads policies. Twitter takes violations of its Twitter Ads Policies, the Twitter Rules and Terms of Service seriously. We will examine reported violations and take appropriate action, which may include removal of offending advertisements and advertisers from the Twitter Ads platform.

4. One reason that it is difficult for us to know if ads on Twitter' s platform are having discriminatory effects is because there is no real way for watchdog groups to help identify potential bias. When I asked you whether there is a way for watchdog groups to examine how non-political ads are being targeted, you stated "[y]es, our Ads Transparency Center is comprehensive of all ads." However, upon reviewing the information made available in Twitter's Ads Transparency Center, it appears that no information is available about how non-political ads are being targeted.

a. On what date do you plan to make available information about how non-political ads are being targeted and who is seeing the ads?

b. Why does the Ads Transparency Center only include information about ads during the last seven days? Do you plan to extend this time period?

Twitter first implemented an updated Political Campaigning Policy to provide clearer guidance about how we define political content and who can promote-political content on our platform. Under the revised policy, advertisers who wish to target the United States with federal political campaigning advertisements are required to self-identify as such and certify that they are located within the United States. Foreign nationals will not be permitted to serve political ads to individuals who identify as being located in the United States.

Twitter accounts that wish to target the U.S. with federal political campaigning advertisements must also comply with a strict set of requirements. Among other things, the account’s profile photo, header photo, and website must be identical to the individual’s or organization’s online presence. In addition, the advertiser must take steps to verify that the address used to serve advertisements with content related to a federal political campaign is genuine.

To further increase transparency and better educate those who access promoted content, accounts serving ads with content related to a federal political campaign will now be visually identified and contain a disclaimer. This feature will allow people to more easily identify federal political campaign advertisements, quickly identify the identity of the account funding the advertisement, and immediately tell whether it was authorized by the candidate.

In June, we launched the Ads Transparency Center, which is open to everyone on Twitter and the general public, and currently focuses on electioneering communications. Twitter requires extensive information disclosures of any account involved in federal electioneering communications and provides specific information to the public via the Ads Transparency Center, including:

- Purchases made by a specific account;
- All past and current ads served on the platform for a specific account;
- Targeting criteria and results for each advertisement;
- Number of views each advertisement received; and
- Certain billing information associated with the account.

These are meaningful steps that will enhance the Twitter experience and protect the health of political conversations on the platform.

In addition,we recently announced the next phase of our efforts to provide transparency with the launch of a U.S.-specific Issue Ads Policy and certification process. The new policy impacts advertisements that refer to an election or a clearly identified candidate or advertisements that advocate for legislative issues of national importance. To provide people with additional information about individuals or organizations promoting issue ads, Twitter has established a process that verifies an advertiser's identity and location within the United States. These advertisements will also be included in the Ads Transparency Center. We are also examining how to adopt political campaigning and issue ads policies globally. We remain committed to continuing to improve and invest resources in this space.

For non-political ads, when individuals on Twitter log into their accounts, they have immediate access to a range of tools and account settings to access, correct, limit, delete or modify the personal data provided to Twitter and associated with the account, including public or private settings, marketing preferences, and applications that can access their accounts. Twitter also informs individuals on the platform about advertising in several ways. For example, Twitter describes advertising activity in its Privacy Policy, an "Ads info" footer on twitter.com, and the "Why am I seeing this ad?" section of the drop down menu on Twitter ads themselves.

5. When I asked you if Twitter is running any educational campaigns to inform users about how their data is being used, you stated "[n]ot at the moment, but we should be looking at that …"

a. What specific steps has Twitter taken since September 5, 2018 towards running educational campaigns that let users know what information it collects about them and how that information is used?

b. Has Twitter taken any other steps since September 5, 2018 to improve how it

informs users about what information is being collected and how that information is used?

c. Will you commit to running educational campaigns so that users are able to better understand what information is being collected about them and how it is used?

d. If so, on what date can we expect Twitter to launch these educational campaigns?

Twitter's purpose is to serve the public conversation. We serve our global audience by focusing on the people who use our service, and we put them first in every step we take. People around the world use Twitter as a "town square" to publicly, openly, and freely exchange ideas. We must be a trusted and healthy place in order for this exchange of ideas and information to continue.

To ensure such trust, the privacy of the people who use our service is of paramount importance to Twitter. We believe privacy is a fundamental right, not a privilege. Privacy is part of Twitter's DNA. Since Twitter's creation over a decade ago, we have offered a range of ways for people to control their experience on Twitter, from creating pseudonymous accounts to letting people control who sees their Tweets, in addition to a wide array of granular privacy controls. This deliberate design has allowed people around the world using Twitter to protect their privacy.

That same philosophy guides how we work to protect the data people share with Twitter. Twitter empowers the people who use our services to make informed decisions about the data they share with us. We believe individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared.

Twitter recently updated our Privacy Policy to include callouts, graphics, and animations designed to enable people to better understand the data we receive, how it is used, and when it is shared.

6. Many consumers do not realize that Twitter makes inferences about them. Buried at the bottom of one of its webpages titled "Your Twitter Data" is a link to "inferred interests from Twitter."

a. Aside from how ads are targeted, what are other ways in which "inferred interests from Twitter" are used to personalize a user's experience?

b. Is each user able to see all of the "inferred interests from Twitter" that have been made about them?

c. If users are unable to see all of the "inferred interests from Twitter," how does Twitter determine which interests users are able to see?

d. When making these inferences, does Twitter take into account the user's online activity on Twitter's platform as well online activity off of its platform?

e. Will you commit to making available to users all of the "inferred interests from Twitter" that have been made about them?

Although the information people share on Twitter is generally public, Twitter also receives non-public personal information. For example, a person creating a Twitter account must provide the platform with his or her email address or phone number. Twitter will also receive standard log information, such as the device being used and the Internet Protocol (IP) address. People who use the service may also choose to share additional information with Twitter including, for example, their address book contacts in order to connect with people they know, help others find and connect with them, and better recommend content to them and others.

In addition, and consistent with nearly all other online platforms, Twitter uses cookies and other similar technologies, such as pixels or local storage, to operate its services and help provide individuals on the platform with a better, faster, and safer experience. Cookies are small files that websites place on a computer as an individual browses the web. Like many websites, Twitter uses cookies to discover how people are using the services and to make them work better. Twitter also uses cookies to help serve people more relevant content based on where they have seen Twitter content on the web, and to serve targeted advertising. Twitter provides individuals with additional control over whether their data is used for these purposes.

In order to show people the most interesting and relevant content, Twitter may infer information about individuals based on their activity on the platform and other information. This includes inferences such as what topics people may be interested in, how old a person is, what languages a person speaks, and whether the content of one account may be of interest to others on the platform. For example, Twitter may infer that an individual is a basketball fan based on accounts the individual follows and suggest content related to the National Basketball Association. Inferences assist Twitter in offering better services and personalizing the content Twitter shows, including advertisements.

Twitter uses the data it receives to deliver, measure, and improve services in a variety of ways, including: protecting the services; authentication and security; remembering preferences; improving analytics and research, including Twitter Ads and Twitter buttons and widgets; customizing Twitter services with more relevant content like tailored trends, stories, advertisements, and suggestions for people to follow; and assisting in delivering advertisements, measuring their performance, and making them more relevant.

Twitter informs individuals on the platform about Tailored Audiences in several ways. For example, Twitter describes this activity in its Privacy Policy, an "Ads info" footer on twitter.com, and the "Why am I seeing this ad?" section of the drop down menu on Twitter ads themselves. Each of these locations describe interest-based advertising on Twitter and explain

how to use the associated privacy controls. In addition, the "Your Twitter Data" tool allows individuals on Twitter to download a list of advertisers that have included them in a Tailored Audience.

If people on Twitter do not want Twitter to show them Tailored Audience ads on and off of Twitter, there are several ways they can turn off this feature: using their Twitter settings, they can visit the Personalization and data settings and adjust the Personalize ads setting; if they are on the web, they can visit the Digital Advertising Alliance's consumer choice tool at optout.aboutads.info to opt out of seeing interest-based advertising from Twitter in their current browser; if they do not want Twitter to show them interest-based ads in Twitter for iOS on their current mobile device, they can enable the "Limit Ad Tracking" setting in their iOS phone's settings; and if they do not want Twitter to show them interest-based ads in Twitter for Android on their current mobile device, they can enable "Opt out of Ads Personalization" in an Android phone's settings.

In addition to explaining Tailored Audiences to people on the platform, offering them several ways to disable the feature, and enabling them to view the advertisers who have included them in Tailored Audiences, as described above, the Tailored Audience legal terms require that advertisers have secured all necessary rights, consents, waivers, and licenses for use of data.

Advertisers are also required to provide all people from whom the data is collected with legally-sufficient notice that fully discloses the collection, use, and sharing of the data that is provided to Twitter for purposes of serving ads targeted to people's interest, and legally sufficient instructions on how they can opt out of interest-based advertising on Twitter.

7. Many consumers do not realize that Twitter collects "inferred interests from partners" about them. This information is similarly buried at the very bottom of the webpage titled "Your Twitter Data."

a. Is each user able to view all of the "inferred interests from partners" that are used to reach them on Twitter?

b. If users are unable to view all of the "inferred interests from partners," how does Twitter determine which ones users are able to see?

c. Will you commit to making available to users all of the "inferred interests from partners"?

Twitter recently updated our Privacy Policy to include callouts, graphics, and animations designed to enable people to better understand the data we receive, how it is used, and when it is shared.

Twitter also provides a toolset called Your Twitter Data. Your Twitter Data tools provide individuals accessible insights into the type of data stored by Twitter, such as username, email address, and phone numbers associated with the account and account creation details. The

birthdays and locations of individuals are also shown in the tool if they have previously been provided to Twitter.

Individuals using the Your Twitter Data tool can also see and modify certain information that Twitter has inferred about the account and device such as gender, age range, languages, and interests. People on Twitter can review inference information, advertisers who have included them in tailored audiences, and demographic and interest data from external advertising partners. The Your Twitter Data tool also allows people with a Twitter account to download a copy of their relevant data from Twitter. We recently updated the download feature of the Your Twitter Data tool to include additional information. Since that update on May 25, 2018, we have seen approximately 586,000 people around the world use the tool to download 560 terabytes of data.

There is a version of this tool available to individuals who do not have a Twitter account, or for those logged out of the account.

8. Many consumers do not realize that Twitter can track their activity across the web.

a. Is there a place on Twitter where users can go and see in plain text information about the websites that the user has visited?

b. Will you commit to making this information available to users in plain text?

When individuals on Twitter log into their accounts, they have immediate access to a range of tools and account settings to access, correct, limit, delete or modify the personal data provided to Twitter and associated with the account, including public or private settings, marketing preferences, and applications that can access their accounts. These data settings can be used to better personalize the individual's use of Twitter and allow him or her the opportunity to make informed choices about whether Twitter collects certain data, how it is used, and how it is shared.

For example, individuals can change the personalization and data settings for their Twitter account, including:

- Whether interest-based advertisements are shown to an individual on and off the Twitter platform;
- How Twitter personalizes an individual's experience across devices;
- Whether Twitter collects and uses an individual's precise location;
- Whether Twitter personalizes their experience based on places they have been; and
- Whether Twitter keeps track of the websites where an individual sees Twitter content.

An individual on Twitter can disable all personalization and data setting features with a single master setting prominently located at the top of the screen.

People on the platform can also deactivate their accounts. Deactivated Twitter accounts, including the display name, username, Tweets, and public profile information, are no longer viewable on Twitter.com, Twitter for iOS, and Twitter for Android.

9. During the hearing, I asked whether Twitter stores previously collected data about users after the user chooses to no longer have his or her activity tracked. You said "I believe it's erased, but we'll have to follow up with the details."

a. If a user disables "Track where you see Twitter content across the web," does Twitter still store any previously collected information?

I also asked you if you would commit to erasing data when users opt out of the data being collected. You said "yes, but let just make sure I understand the constraints and ramifications of that." I want to make sure that you still stand by your commitment.

b. Will you commit to giving users the option to have this information permanently deleted?

Twitter believes individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared. Twitter is always working to improve transparency into what data is collected and how it is used. Twitter designs its services so that individuals can control the personal data that is shared through our services. People that use our services have tools to help them control their data. For example, if an individual has registered an account, through their account settings they can access, correct, delete or modify the personal data associated with their account.

10. It is my understanding that Twitter also collects information both on and off its platform about non-users.

a. How are non-users able to see all of the information that Twitter collects about them?

b. How are non-users able to delete all of the information that Twitter collects about them?

Twitter believes individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared. Twitter is always working to improve transparency into what data is collected and how it is used. Twitter designs its services so that individuals can control the personal data that is shared through our services. People that use our services have tools to help them control their data. For example, if an individual has registered an account, through their account settings they can access, correct, delete or modify the personal data associated with their account.

Twitter recently updated our Privacy Policy to include callouts, graphics, and animations designed to enable people to better understand the data we receive, how it is used, and when it is shared.

Twitter also provides a toolset called Your Twitter Data. Your Twitter Data tools provide individuals accessible insights into the type of data stored by Twitter, such as username, email address, and phone numbers associated with the account and account creation details. The birthdays and locations of individuals are also shown in the tool if they have previously been provided to Twitter.

Individuals using the Your Twitter Data tool can also see and modify certain information that Twitter has inferred about the account and device such as gender, age range, languages, and interests. People on Twitter can review inference information, advertisers who have included them in tailored audiences, and demographic and interest data from external advertising partners. The Your Twitter Data tool also allows people with a Twitter account to download a copy of their relevant data from Twitter. We recently updated the download feature of the Your Twitter Data tool to include additional information. Since that update on May 25, 2018, we have seen approximately 586,000 people around the world use the tool to download 560 terabytes of data.

There is a version of this tool available to individuals who do not have a Twitter account, or for those logged out of the account. Instructions on how users with or without a Twitter account can access this tool can be found in our help center here: https://help.twitter.com/en/managing-your-account/accessing-your-twitter-data

The Honorable John Sarbanes

Media reports have indicated Twitter has become increasingly sensitive to allegations it suffers from anti-conservative basis -- an allegation that continues to be advanced by many conservatives. Some, including myself , have raised concern that Republican elected officials and conservative leaders are intentionally drumming up accusations of political bias at social media companies like Twitter in an effort to "work the refs," securing favorable treatment as recompense for non-existent bias.

As our Committee works to understand better the policies in place at your company to prevent political bias, I am committed to better understanding how exactly those policies are being developed. To that end, I greatly appreciate Twitter's thorough and candid responses to the following questions.

1. It was reported in June that, in an effort to respond to complaints of bias against conservatives on Twitter, you met with conservative leaders and Republican officials, including Senator Ted Cruz and Trump Administration official Mercedes Schlapp.

a. Did these reported meetings take place? If so, what was the impetus for scheduling these meetings? Were they requested by you or by the persons with whom you met?

b. What topics were discussed at these meetings? What specific requests were made of you at these meetings? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

c. To what extent was potential future regulation of social media platforms generally or Twitter specifically discussed with current government officials? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

d. To what extent have internal deliberations at Twitter regarding these accusations of bias focused on potential future regulation of social media platforms generally or Twitter specifically? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

e. Have any similar meetings with progressive or Democratic leaders taken place, been offered, or been requested? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

f. Have any similar meetings with victims' rights, anti-harassment, civil rights, or anti-hate group advocates, or any other individuals or groups concerned with rampant racist and misogynist harassment on Twitter taken place, been offered, or been requested? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

g. Have any similar meetings with election security , campaign finance, or good government advocates, or any other individuals or groups concerned with Twitter's role in the functioning of our democratic institutions taken place, been offered, or been requested? Please provide any documents, meeting notes, or memoranda relating to the deliberation.

Twitter officials have in the past and will continue in the future to meet with government and political representatives across the political spectrum. We have good partnerships with civil society groups around the globe, including with partners focused on election integrity. In every interaction, we strive to uphold the values of the company.

Twitter does not use political ideology to make any decisions, whether related to ranking content on our service or how we enforce our rules. We believe strongly in being impartial, and we strive to enforce our rules impartially. We do not shadowban anyone based on political ideology. In fact, from a simple business perspective and to serve the public conversation, Twitter is incentivized to keep all voices on the platform.

Twitter plays an important role in our democracy and governments around the world. In the United States, all 100 Senators, 50 governors, and nearly every member of the House of Representatives currently reach their constituents through Twitter accounts. Our service has enabled millions of people around the globe to engage in local, national, and global conversations on a wide range of issues of civic importance. We also partner with news organizations on a regular basis to live-stream congressional hearings and political events, providing the public access to important developments in our democracy. The notion that we would silence any political perspective is antithetical to our commitment to free expression.

2. After reported "anti-conservative" bias, the accusations of which are questionable, at Facebook, specifically regarding its "Trending" section, Facebook removed human editors from the process and used algorithms to moderate the Trending section. Almost immediately, the Trending section was overrun with false information and hoaxes masquerading as news stories.

a. To what extent has Facebook's experience with charges of anti-conservative bias, its overcorrection, and its resulting issues with "fake news" informed Twitter' s decision-making regarding similar issues?

b. What value do you, Twitter as a company, and Twitter's algorithms place on veracity? How does Twitter balance competing concerns over equal treatment of political viewpoints and legitimate concerns about misinformation and harassment?

c. Is the deliberate spreading of misinformation a cause for suspending or banning a Twitter account? Please provide any documents, meeting notes, or memoranda relating to Twitter's formal policies on the subject.

d. Does Twitter have rules regarding the veracity of promoted content? If so, what guidelines and processes are in place to evaluate the veracity of promoted content? If not, have there been any internal discussions or deliberations about moderating promoted content to prevent the spread of misinformation?

Twitter is committed to help increase the collective health, openness, and civility of public conversation, and to hold ourselves publicly accountable towards progress. At Twitter, health refers to our overall efforts to reduce malicious activity on the service, including malicious automation, spam, and fake accounts. Twitter has focused on measuring health by evaluating how to encourage more healthy debate, and critical thinking.

The platform provides instant, public, global messaging and conversation, however, we understand the real-world negative consequences that arise in certain circumstances. Twitter is determined to find holistic and fair solutions. We acknowledge that abuse, harassment, troll armies, manipulation through bots and human-coordination, misinformation campaigns, and increasingly divisive echo chambers occur.

We have learned from situations where people have taken advantage of our service and our past inability to address it fast enough. Historically, Twitter focused most of our efforts on removing content against our rules. Today, we have a more comprehensive framework that will help encourage more healthy debate, conversations, and critical thinking.

We believe an important component of improving the health on Twitter is to measure the health of conversation that occurs on the platform. This is because in order to improve something, one must be able to measure it. By measuring our contribution to the overall health of the public conversation, we believe we can more holistically approach our impact on the world for years to come.

Earlier this year, Twitter began collaborating with the non-profit research center Cortico and the Massachusetts Institute of Technology Media Lab on exploring how to measure aspects of the health of the public sphere. As a starting point, Cortico proposed an initial set of health indicators for the United States (with the potential to expand to other nations), which are aligned with four principles of a healthy public sphere. Those include:

- Shared Attention: Is there overlap in what we are talking about?
- Shared Reality: Are we using the same facts?
- Variety: Are we exposed to different opinions grounded in shared reality?
- Receptivity: Are we open, civil, and listening to different opinions?

Twitter strongly agrees that there must be a commitment to a rigorous and independently vetted set of metrics to measure the health of public conversation on Twitter. And in order to develop those health metrics for Twitter, we issued a request for proposal to outside experts for their submissions on proposed health metrics, and methods for capturing, measuring, evaluating and reporting on such metrics. Our expectation is that successful projects will produce peer-reviewed, publicly available, open-access research articles and open source software whenever possible.

As a result of our request for proposal, we are partnering with experts at the University of Oxford and Leiden University and other academic institutions to better measure the health of Twitter, focusing on informational echo chambers and unhealthy discourse on Twitter. This collaboration will also enable us to study how exposure to a variety of perspectives and opinions serves to reduce overall prejudice and discrimination. While looking at political discussions, these projects do not focus on any particular ideological group and the outcomes will be published in full in due course for further discussion.

In regards to advertising, Twitter has stricter rules and higher standards for promoted content. Additional information on prohibited content policies can be found here: https://business.twitter.com/en/help/ads-policies/prohibited-content-policies/inappropriate-content.html

3. Relatedly, after Twitter took steps to remove "trolls" and fake accounts, conservatives accused Twitter of executing a purge of conservative accounts. In reality, Twitter was responding to concerns about widespread use of bots and fake accounts during the 2016 election.

a. What do you and Twitter as a company consider Twitter's responsibilities and obligations to our democratic institutions in light of your platform's growing importance to campaigning and governing?

b. To what extent are you and Twitter as a company concerned that allegations of anti-conservative bias are an attempt to prevent Twitter from taking necessary steps to prevent the abuse of the platform that occurred during the 2016 election?

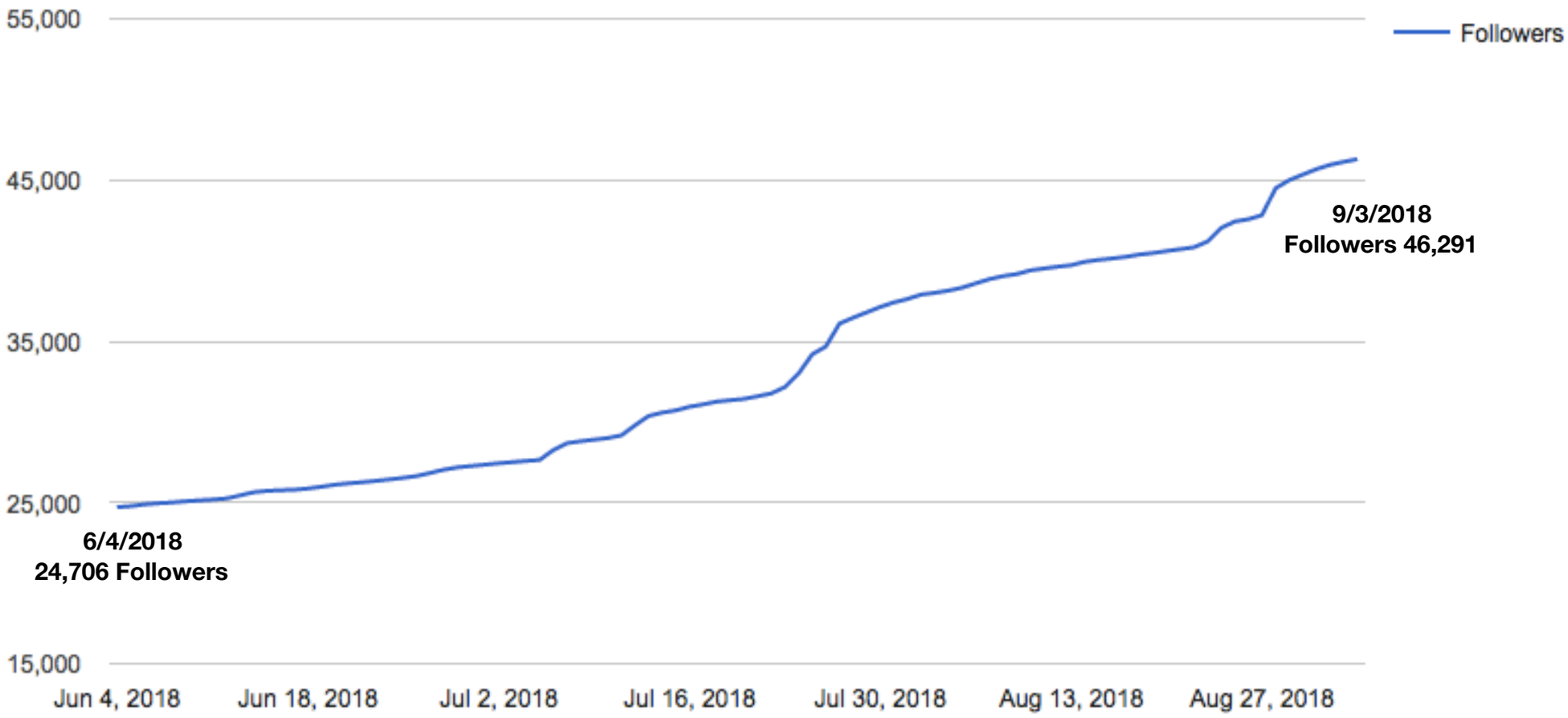
Please provide responses to these questions and any relevant documents, meeting notes, or memoranda relating to these matters.

Twitter is committed to improving the collective health, openness, and civility of public conversation on our platform. Twitter's is built and measured by how we help encourage more healthy debate, conversations, and critical thinking. Conversely, abuse, malicious automation, and manipulation detracts from it. We are committing Twitter to hold ourselves publicly accountable towards progress.

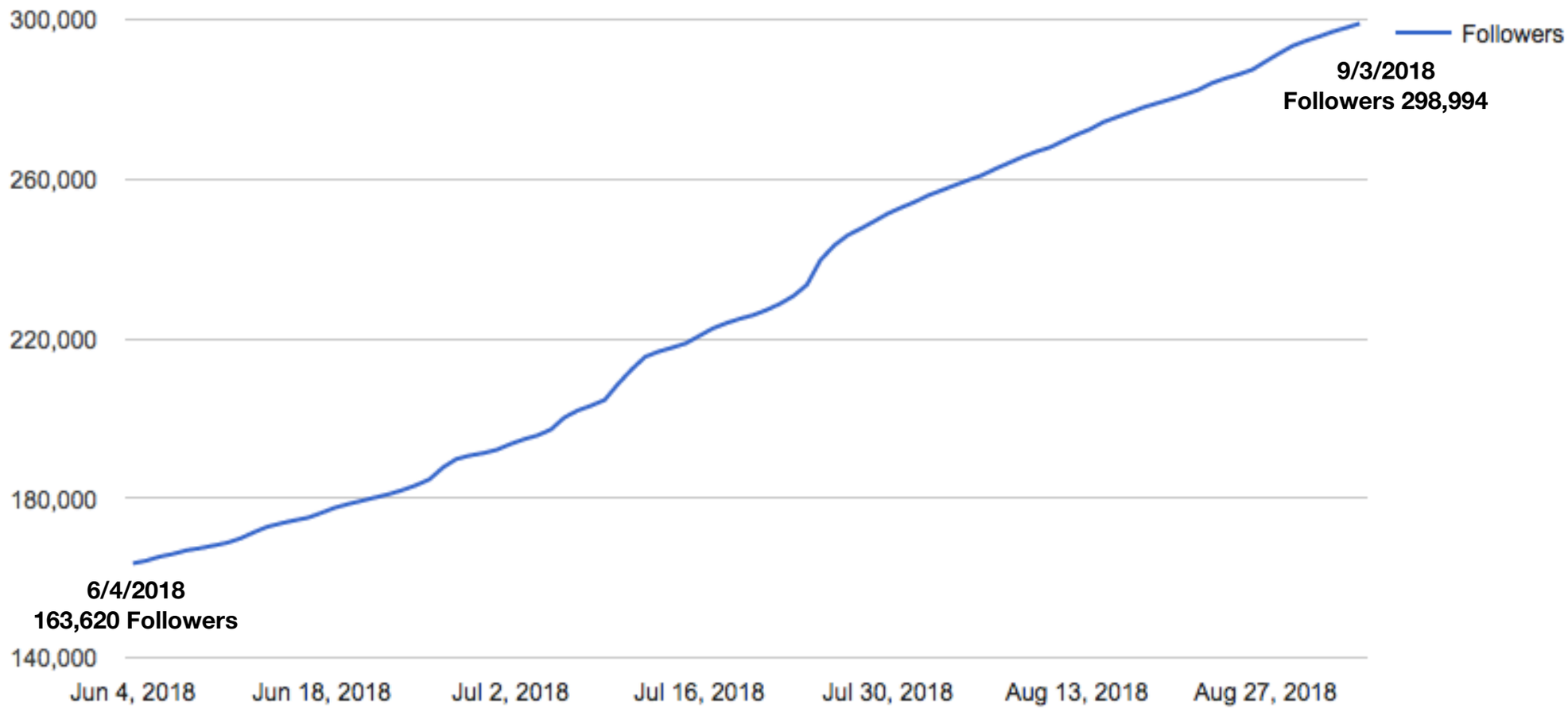
The public conversation occurring on Twitter is never more important than during elections, the cornerstone of our democracy. Our service shows the world what is happening,

democratizes access to information and—at its best—provides people insights into a diversity of perspectives on critical issues; all in real-time. We work with commitment and passion to do right by the people who use Twitter and the broader public. Any attempts to undermine the integrity of our service is antithetical to our fundamental rights and undermines the core tenets of freedom of expression, the value upon which our company is based. This issue affects all of us and is one that we care deeply about as individuals, both inside and outside the company. Additional information regarding our elections integrity efforts can be found here: https://about.twitter.com/en_us/values/elections-integrity.html

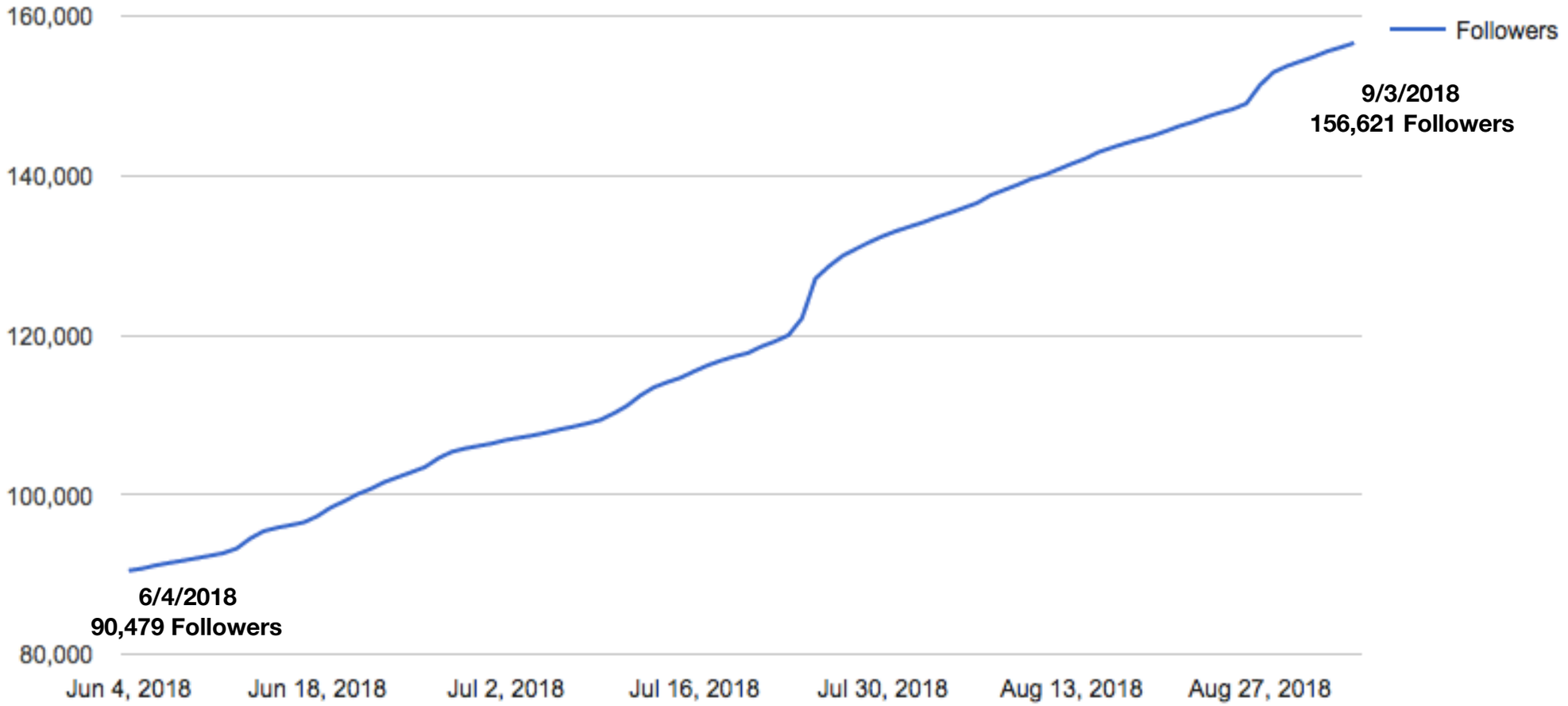
@MattGaetz


55,000
45,000
35,000
25,000
15,000
Followers
9/3/2018
Followers 46,291
6/4/2018
24,706 Followers
Jun 4, 2018
Jun 18, 2018
Jul 2, 2018
Jul 16, 2018
Jul 30, 2018
Aug 13, 2018
Aug 27, 2018

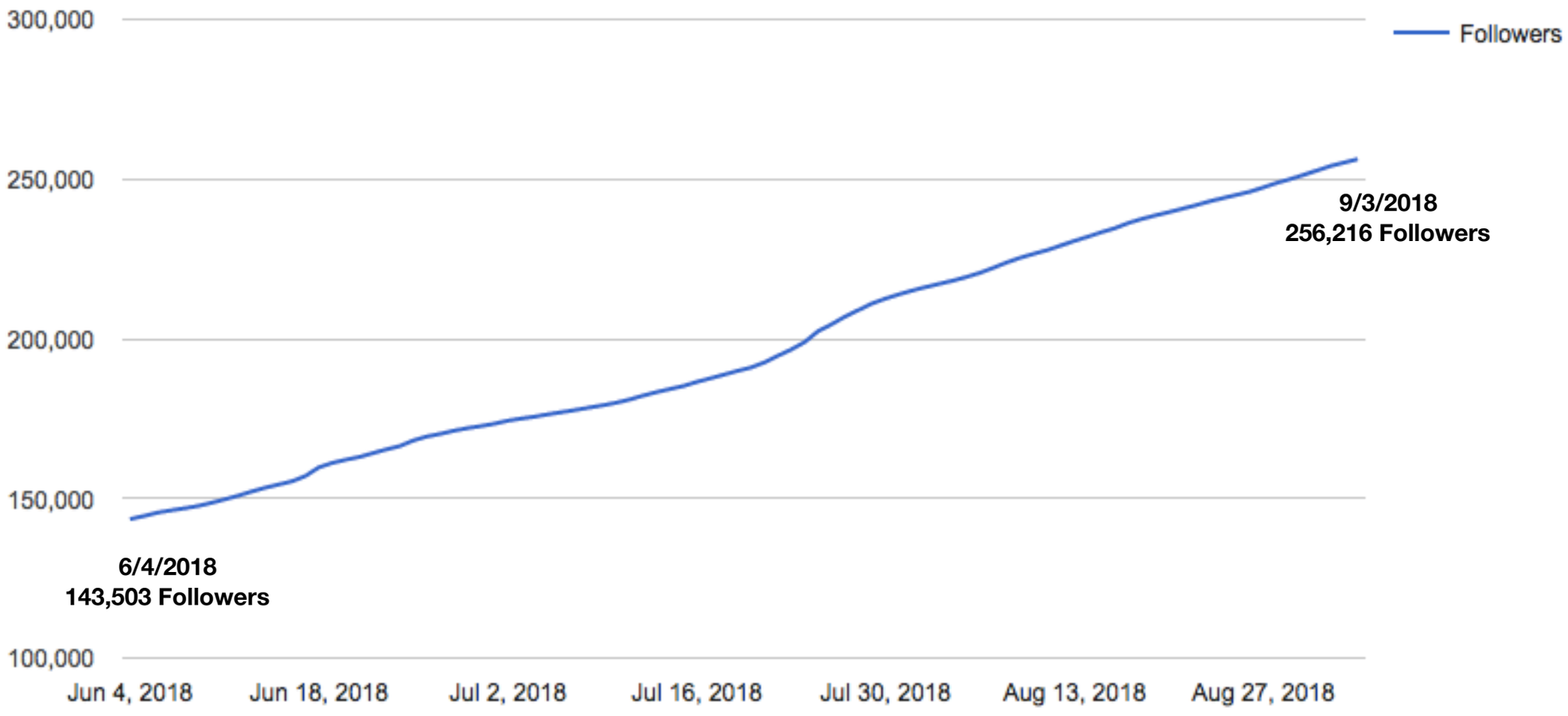
@JimJordan


300,000
260,000
220,000
180,000
140,000
Followers
9/3/2018
Followers 298,994
6/4/2018
163,620 Followers
Jun 4, 2018
Jun 18, 2018
Jul 2, 2018
Jul 16, 2018
Jul 30, 2018
Aug 13, 2018
Aug 27, 2018

@MarkMeadows



@DevinNunes


300,000
250,000
200,000
150,000
100,000
Followers
9/3/2018
256,216 Followers
6/4/2018
143,503 Followers
Jun 4, 2018
Jun 18, 2018
Jul 2, 2018
Jul 16, 2018
Jul 30, 2018
Aug 13, 2018
Aug 27, 2018