Questions for the Record
Senate Select Committee on Intelligence
Hearing on Foreign Influence Operations Using Social Media September 17, 2018
Questions for the Record for Mr. Jack Dorsey, Chief Executive Officer, Twitter.

[From Vice Chairman Warner]

1. According to reports about social media usage during the Catalan Independence Referendum in Spain, "[Sputnik and RT], both financed by the Kremlin, managed to see their links shared more than those from Spanish public networks EFE and RTVE, or those of private international publications such as The Guardian and CNN." This information operation utilized Russian bots on Twitter and came almost a year after the Russian interference in the U.S. election which used similar tactics.

- **Given that similar tactics were used and this event happened almost a year after the 2016 US election, why was Twitter not able to detect and stop this information operation by Russian-linked operatives?**

We remain vigilant about identifying and eliminating abuse on the platform perpetrated by hostile foreign actors, and we will continue to invest in resources and leverage our technological capabilities to do so. Twitter's main focus is promoting healthy public discourse through protection of the democratic process. Twitter continues to engage in intensive efforts to identify and combat state-sponsored hostile attempts to abuse social media for manipulative and divisive purposes. We now possess a deeper understanding of both the scope and tactics used by malicious actors to manipulate our platform and sow division across Twitter more broadly. Our efforts enable Twitter to fight this threat while maintaining the integrity of peoples' experience on the service and supporting the health of conversation on our platform. Our work on this issue is not done, nor will it ever be.

Any amount of election interference in any election is unacceptable to us. 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.

We conducted a comprehensive analysis of accounts that promoted election-related Tweets on the platform throughout 2016 in the form of paid ads. We reviewed nearly 6,500 accounts and our findings showed that approximately one-tenth of one-percent—only nine of the total number of accounts—were Tweeting election-related content and linked to Russia. The two most active accounts out of those nine were affiliated with Russia Today ("RT"), which Twitter subsequently barred from advertising on Twitter. And Twitter is donating the $1.9 million that RT spent globally on advertising to academic research and partnerships focused on election and civic engagement.

In any election, we take action on accounts that break our rules. This included accounts Tweeting about the Catalan referendum. Some accounts outside of Spain did engage in the conversation, including sharing international media links. As part of our standard review patterns and related enforcement actions, we took action on a number of spammy accounts, including some posting at a high volume. We cannot make definitive attributions of these accounts. As is always the case, we are committed to protecting the integrity of the public conversation and that is never more important than during elections. We will always err on the side of transparency. We recently made the full database of potentially state-backed interference on Twitter available with the goal of empowering researchers to conduct independent, investigatory analysis, including the Catalan referendum.

2. This is a screenshot of an advertisement run on Twitter during the 2016 election season, suggesting it was possible to vote via text message:

- **Have you conducted analysis into who created these ads, and which country or countries those individuals are located in, and if so, can you share your findings?**
- **Were these voter discouragement ads targeted at people of any particular race or ethnic group?**
- **Since 2016, have you changed your policies and operations in any way to disallow similar ads that create confusion as to where or how to vote?**

Twitter unequivocally condemns the use of our platform for any election interference activity. Tweets aimed at suppressing voter turnout generally are surfaced through reports from people using our service. This content is reviewed, then promptly removed as illegal interference with voting rights: the content is either restricted as inaccessible pending deletion by the individual (i.e., other individuals on the platform are unable to see the content) or the responsible accounts are permanently suspended. In addition, in order to proactively surface additional Tweets with a given text-to-vote meme, Twitter utilizes technology for identifying instances where the same image appears across multiple Tweets. Content identified through this process is then subject to manual review.

Depending on the number of violations for any given account disseminating voter suppression Tweets, Twitter will either restrict access to the Tweet or suspend the account. During the period leading up to the 2016 election, for example, Twitter labeled and restricted access to the vote-to-text Tweets pursuant to the Twitter User Agreement, which contains the Twitter Terms of Service, Twitter Privacy Policy, and Twitter Rules. According to the unlawful use provision of the Twitter Rules, individuals are prohibited from using Twitter's "service for any unlawful purpose or in furtherance of illegal activities" and "[i]nternational users agree to comply with all local laws regarding online conduct and acceptable content."

Because the Tweet in question appeared to mislead people into believing that they could vote online or vote by text, Twitter viewed the Tweets as an unlawful interference with the voting process. Twitter labeled as "restricted pending deletion" a total of 918 such Tweets from 529 Twitter accounts, which rendered the Tweets inaccessible and disabled the accounts' ability to use the platform until those Tweets were deleted. In connection with this activity, Twitter also suspended 106 of those accounts, a majority of which were found to be in violation of the Twitter Rules prohibiting spam, including posting duplicate content over multiple accounts or multiple duplicate updates on one account. In a few instances, however, Twitter suspended accounts of people who shared the voting-related content and had previous, but otherwise unrelated, violations of the Twitter Rules against abusive behavior.

The specific Tweet identified in the question was an organic Tweet, not a Promoted Tweet, and they could not be targeted to any particular individual on the platform. In this specific instance, Twitter identified and suspended this Tweet on November 6, 2016. We permanently suspended the account -- which we believe was located within the United States -- on November 7, 2016 as it violated our rules against repeatedly posting content with the intent to deceive.

Twitter identified, but did not take action against, an additional 286 Tweets of the relevant content from 239 Twitter accounts. With respect to those Tweets, Twitter determined that they propagated the content in order to refute the message and alert other people that the information is false and misleading. And partly as a result of our enforcement decisions, those refuting Tweets generated significantly greater engagement across the platform compared to the Tweets spreading the misinformation—eight times as many impressions, engagement by ten times as many people on the platform, and twice as many replies.

Since 2016, we have taken a number of important steps to safeguard the integrity of elections. As part of these efforts, we have developed well-established relationships with law enforcement agencies active in this arena, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of 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.

Additionally, to further promote information sharing and to tap into the experience and expertise of active stakeholders, we recently updated a Partner Support Portal. Our goal is to expedite our response to reports from people active in the election arena. This includes election support organizations, U.S.-based research organizations, and universities and academics who study the spread of misinformation in the media. Reports from accounts within this select group are expedited and can be actioned promptly.

3. Twitter has a tool called Tailored Audiences, which is similar to Facebook's Custom Audiences and allows advertisers to upload lists of specific users to target them with ads. Your policies state that advertisers who use Tailored Audiences must obtain consent from users about the data they have acquired.

- **How do you ensure that all of your advertisers actually follow those policies when there are bad actors like the Internet Research Agency who are willing to break the law and illicit data freely available to them?**

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.

4. Mr. Dorsey, you have indicated your company’s strong support for the Honest Ads Act. Thank you for your support and your efforts to largely abide by the terms of that legislation.

- **Do you support passage of the Honest Ads act into law?**
- **Have you seen evidence – in either the Russian context or any recent disruptions – that your new policies on ad transparency have helped stop foreign purchases of political ads on your platform?**

Twitter supports the goals of the Honest Ads Act. Through our own initiative, we have announced voluntary, industry-leading steps to improve transparency and accountability in our ads platform that strongly aligns with the goals and standards in the Act. In fact, in some cases, our new transparency requirements go further than the draft legislation—for example, by requiring transparency for all advertisers regardless of topic, and by committing to the inclusion of advertisements for candidates on state and local levels.

We do have suggestions for potential improvements of the bill. First, we want to be sure that the proposed requirements, including in-ad disclosure language, are sufficiently flexible to account for character-constrained platforms like Twitter. Second, we hope that legislation on this topic would clarify that, while the duty to collect and display disclosure information lies with the platforms, the duty to provide accurate information lies with the advertisers.

5. What is Twitter’s current policy on the posting or promotion of hacked emails on your platform?

Twitter rules prohibit the distribution of hacked material that contains private information or trade secrets, or could put people in harm’s way. According to our Rules, Twitter does not permit the use of our services to directly distribute content obtained through hacking that contains personally identifiable information, may put people in imminent harm or danger, or contains trade secrets. Direct distribution of hacked materials includes posting hacked content on Twitter (for instance, in the text of a Tweet, or in an image), or directly linking to hacked content hosted on other websites.

We also expanded 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. We also may suspend accounts in which Twitter is able to reliably attribute a hack to the account distributing that content. Commentary about a hack or hacked materials, such as news articles discussing a hack, are generally not considered a violation of this policy.

6. Europe has established new rules for data protection and privacy for European citizens (General Data Protection Regulation, or GDPR). These new rules include required data portability, the right to be forgotten online, a 72-hour data breach disclosure requirement, and first-party consent requirements.

- **How is Twitter complying with GDPR?**
- **Are there protections that will flow to U.S. users as a result?**
- **What lessons should we be learning from the European experiment with data protection?**
- **Should we consider policy solutions like first-party consent?**
- **Why shouldn't companies be required to obtain explicit and informed consent before collecting or processing user data like in Europe?**

Twitter has undertaken a variety of internal and public facing updates to comply with the obligations imposed by the coming into force of the General Data Protection Regulation.This includes, for example, appointment of a Global Data Protection Officer, providing mechanisms to allow people to download their data from Twitter, mechanisms to allow people to contact Twitter's Office of Data Protection, and ensuring internal systems and processes exist to support Twitter's compliance with the GDPR.

The GDPR was developed over many years and thus, the underlying goals are commendable. Notably the GDPR's objectives of protecting consumers by providing for data protection that includes core tenants from the Federal Information Processing Standards developed in the 1970s in the U.S. Most of the internal and external facing updates Twitter has undertaken for compliance with the GDPR apply to all people who use Twitter's services irrespective of where they reside.

There are areas that should examined carefully before considering adoption of the same regulations in the U.S. For example, the language around automated decision making may prove restrictive for business and innovation. Similarly, the GDPR's language around the commonly described right to be forgotten does not comport with the First Amendment.

Twitter believes that informed consent should be obtained for data processing. Twitter believes people should know and have control over the types of data that are received about them by data processors, how it is used, and when it is shared. However, Twitter does not believe that a blanket opt-in consent requirement should be imposed. This can lead to operational and technical difficulties. For example, to provide a person with a landing page to a service in their language, their IP address is processed to determine their approximate location to infer language. Required opt-in consent for such processing would make such useful features difficult to provide and result in friction for consumers. Thus, the type of consent mechanism used should be informed by the type of service, the type of data at issue, when in the use of the service the consent is being solicited, and the information and controls available to the consumer.

7. Do you think Twitter might benefit from more independent insight into anonymized activity?

- **Isn't there a public interest in better understanding how your platform works and how users interact on social media?**

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.

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. On August 24, 2018, Twitter hosted our industry peers to discuss data sharing about hostile foreign actors regarding 2018 election security. We continue to meet in regular cadence with our industry peers about election integrity efforts.

We also have well-established relationships with law enforcement agencies active in this arena, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of 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.

Additionally, we committed to the United States Congress and the public to provide regular updates and information regarding our investigation into foreign interference in political conversations on Twitter. Since that time, we have shared examples of these types of content posted on Twitter by the Internet Research Agency (IRA) and provided the public with a direct notice if they interacted with these accounts. In August, we also disclosed details of another attempted influence campaign we identified as potentially located within Iran.

In line with our strong principles of transparency and with the goal of improving understanding of foreign influence and information campaigns, on October 17, 2018, Twitter released the full, comprehensive archives of the Tweets and media that are connected with these two previously disclosed and potentially state-backed operations on the service. We are making this data available with the goal of encouraging open research and investigation of these behaviors from researchers and academics around the world.

These large datasets d 3,841 accounts affiliated with the IRA, originating in Russia, and 770 other accounts, potentially originating in Iran. They include more than 10 million Tweets and more than 2 million images, GIFs, videos, and Periscope broadcasts, including the earliest on-Twitter activity from accounts connected with these campaigns, dating back to 2009.

8. The fact that Twitter failed to anticipate misuse is extremely troubling.

- **Why should we have confidence that you are any more prepared to handle issues of misuse now?**
- **How are you better protecting the users of your products?**
- **You have indicated that Twitter is now more fully addressing potential threats to new products before launching them.**
- **Why was this not a part of Twitter’s process previously?**

Twitter is committed to protecting the integrity of elections. We have made recent improvement to three critical areas of our election integrity efforts: (1) Updates to the Twitter Rules (2) Detection and Enforcement; and (3) Product Improvements.

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.

We have seen positive results from our investments in conversational health and information integrity. We continue to enforce our rules against intentionally misleading election-related content. In August, we removed approximately 50 accounts misrepresenting themselves as members of various state Republican parties. We have also taken action on Tweets sharing media regarding elections and political issues with misleading or incorrect party affiliation information. We continue to partner closely with the RNC, DNC, and state election institutions to improve how we handle these issues. In August, we removed 770 accounts engaging in coordinated behavior which appeared to originate in Iran. Our investigation into this activity continues, and we will share further updates on our findings with law enforcement, our industry peers, and the public.

Our automated detections continue to identify and challenge millions of potentially spammy and automated accounts per week. In the first half of September, we challenged an average of 9.4 million accounts each week. As a result of our proactive detections and enforcements, we have continued to see a decline in the average number of spam-related reports we receive from individuals each day — from an average of approximately 17,000 per day in May, to approximately 16,000 per day in September. We are continuing to roll out improvements to our proactive enforcements against common policy violations, including building new proprietary systems to identify and remove ban evaders at speed and scale.

Finally, we continue to make improvements to the Twitter product to help people stay informed and to see the best content first. We heard feedback that people want an easy way to see the most recent Tweets in their home timeline. We recently updated the timeline personalization setting to allow people to select a strictly reverse-chronological experience, without recommended content and recaps. This ensures you have more control of how you experience what's happening on our service. We are continuing to roll out new features to show people context about accounts on Twitter. In May, we launched an election labels beta for candidates in the 2018 U.S. midterm elections. We are also going to send candidates a message prompt to ensure they have two-factor authentication enabled on their account so it is safe and secure.

We are also offering electoral institutions increased support via an elections-specific support portal, which is designed to ensure we receive and review critical feedback about emerging issues as quickly as possible. We will continue to expand this program ahead of the elections and will provide information about the feedback we receive in the near future. As part of our civic engagement efforts, we are building conversation around the hashtag #BeAVoter with a custom emoji, sending U.S.-based individuals a prompt in their home timeline with information on how to register to vote, and drawing attention to these conversations and resources through the top US trend. This trend is being promoted by @TwitterGov, which will create even more access to voter registration information, including election reminders and an absentee ballot FAQ.

9. At our most recent public hearing with experts on social media, all of our witnesses opined that Russian influence operations are ongoing and currently using several social media platforms, including Twitter.

- **Do you believe that the Russian-linked operatives continue to utilize Twitter for information operations to undermine our democracy?**
- **Have you seen non-IRA, Russian-linked activity on your platform conducting similar types of information operations?**
- **What percentage of Russian-linked activity do you think the IRA represents?**
- **Have you seen evidence of additional Russian-linked troll farms?**
- **Have you identified any troll farms backed by countries other than Russia?**
- **Do you anticipate additional account take-downs in the weeks ahead?**
- **Will you commit to notifying the public should you identify other foreign influence operations?**
- **Will you alert users when they've been exposed to these types of operations?**

It is clear that information operations and coordinated inauthentic behavior will not cease. These types of tactics have been around for far longer than Twitter has existed — they will adapt and change as the geopolitical terrain evolves worldwide and as new technologies emerge. For our part, we are committed to understanding how bad-faith actors use our services. We will continue to proactively combat nefarious attempts to undermine the integrity of Twitter, while partnering with civil society, government, our industry peers, and researchers to improve our collective understanding of coordinated attempts to interfere in the public conversation.

Our dedicated site integrity team, in partnership with a diverse range of committed organizations and personnel across the company, continue to invest heavily in this area. We are constantly seeking to improve our own ability to detect, understand, and neutralize these campaigns as quickly and robustly as technically possible. Twitter has learned from 2016 and more recently from other nation's elections how best to protect the integrity of our elections. Better tools, stronger policy, and new partnerships are already in place. We intend to understand the efficacy of these measures to continue to get better.

[From Senator Feinstein]
10. Twitter has taken action against hundreds of foreign accounts conducting influence operations. However, it is concerning that in the context of the most recent examples from August 21st, action required input from the cybersecurity company FireEye – rather than Twitter finding the subject accounts exclusively through its own internal processes.

- **In the recent case of the Iranian-associated influence campaign, did an external company have to alert you to the activity; and if so, why?**
- **What specific steps are you taking to enhance your ability to find and mitigate influence operations?**

On August 21, 2018, working with our industry peers, Twitter suspended 770 accounts from Twitter for engaging in coordinated manipulation. Based on our analysis, it appears that many of these accounts originated from Iran. As with all investigations, we are committed to engaging with other companies and relevant law enforcement entities. Our goal is to assist investigations into these activities and where possible, we will provide the public with transparency and context on our efforts.

Fewer than 100 of the 770 suspended accounts claimed to be located in the U.S. and many of these were sharing divisive social commentary. On average, these 100 accounts Tweeted 867 times, were followed by 1,268 accounts, and were less than one year old.
In line with our strong principles of transparency and with the goal of improving understanding of foreign influence and information campaigns, on October 21, 2018, we released the full, comprehensive archives of the Tweets and media that are connected with these two previously disclosed and potentially state-backed operations on our service. We are making this data available with the goal of encouraging open research and investigation of these behaviors from researchers and academics around the world.

Independent analysis of this activity by researchers is a key step toward promoting shared understanding of these threats. To support this effort, we have provided early access to a small group of researchers with specific expertise in these issues. Working with law enforcement and the authorities will always be our first priority, but we strongly believe that this level of transparency can enhance the health of the public conversation on the internet. This is our singular mission.

11. As has been illustrated with the actions Twitter took in August, stopping Russian and Iranian-associated influence accounts requires close coordination between the government, social media companies, other private sector entities, and even the public. This construct has been useful in the past; in 2016, Twitter and other social media companies created a shared database of videos and images to counter online terrorist propaganda.

- **Do you believe there is a need for better information sharing between the social media companies?**
- **What is prohibiting your company from sharing more with your peers, government actors, and the public with respect to foreign information 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. On August 24, 2018, Twitter hosted our industry peers to discuss data sharing about hostile foreign actors regarding 2018 election security. These conversations continue to occur with regular cadence in the lead-up to the 2018 midterm election.

12. One of the major criticisms against the referenced countering extremist content database is that there is little information about how it operates and how effective it is in preventing prohibited content from being uploaded again.

- **Have your companies agreed on a common standard for what constitutes prohibited extremist or terrorist content? If not, why not?**
- **Would a shared standard and the deployment of similar software used to detect spam and copyrighted material, facilitate the automated blocking of such content across all four platforms?**
- **In the interest of transparency, would you make this database open to the public or researchers to know which images are prohibited?**

We agree 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.

13. Where Twitter has identified content as advancing a foreign influence campaign, will you commit to providing public access to a library of all ads that target users based on demographics (what content, purchased by whom, targeting whom)? If not, why not?

Twitter is committed to providing greater transparency to our account holders and the public, particularly as it relates to election integrity. In the future, we commit to releasing all the accounts and related content associated with potential information operations as appropriate. Following our investigation into the propaganda effort by the Internet Research Agency (IRA), we notified approximately 1.4 million individuals on our platform who interacted with this malicious content.

Twitter sent notices to people on the platform with an active email address who our records indicate are based in the U.S. and fall into at least one of the following categories:

- People who directly engaged during the election period with the 3,814 IRA-linked accounts we identified, either by Retweeting, quoting, replying to, mentioning, or liking those accounts or content created by those accounts;
- People who were actively following one of the identified IRA-linked accounts at the time those accounts were suspended; and

- People who opt out of receiving most email updates from Twitter and would not have received our initial notice based on their email settings.

[From Senator Wyden]
14. Since the 2016 election, has any foreign government, or anyone that Twitter believes to be acting on the behalf of a foreign government, used Twitter to promote or amplify misleading or "hoax" content to users in the United States (for example, claims that a national tragedy did not occur or was perpetrated by our own government)?

- **If yes, please provide a detailed accounting of each case, including the suspected foreign entity, and the number of users that saw or interacted with the content (e.g. clicked or shared).**
- **What steps has Twitter taken to inform its users, the public, and the United States Government of each case that you have listed in response to the previous question?**

In early 2018, we committed to the United States Congress and the public to provide regular updates and information regarding our investigation into foreign interference in political conversations on Twitter. Since that time, we have shared examples of these types of content posted on Twitter by the Internet Research Agency (IRA) and we notified approximately 1.4 million individuals on our platform who interacted with this malicious content. In August, we also disclosed details of another attempted influence campaign we identified as potentially located within Iran.

In line with our strong principles of transparency and with the goal of improving understanding of foreign influence and information campaigns, on October 21, 2018, we released the full, comprehensive archives of the Tweets and media that are connected with these two previously disclosed and potentially state-backed operations on our service. We made this data available with the goal of encouraging open research and investigation of these behaviors from researchers and academics around the world.

These large datasets comprise 3,841 accounts affiliated with the IRA, originating in Russia, and 770 other accounts, potentially originating in Iran. They include more than 10 million Tweets and more than 2 million images, GIFs, videos, and Periscope broadcasts, including the earliest on-Twitter activity from accounts connected with these campaigns, dating back to 2009.

Additionally, we have well-established relationships with law enforcement agencies active in this arena, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of 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.

15. In September 2017, Twitter confirmed that it had found approximately 200 accounts linked to the same Russian groups that had purchased ads on Facebook. In August 2018, Twitter confirmed that it had suspended 770 accounts for "coordinated manipulation."

- **In addition to the cases listed above, has any foreign government, their agent, or an entity acting on the behalf of a foreign government, created Twitter accounts or written tweets, that masquerade as American for the purpose of influencing political debate or policymaking within the United States, not limited to elections?**
- **Has any other foreign entity, even if it is not known to be acting on behalf of a foreign government, created Twitter accounts, or written tweets, that masquerade as American for the purpose of influencing political debate or policymaking within the United States, not limited to elections?**
- **If the answer to either of the previous two questions is yes, please provide a detailed accounting of each case, including the foreign government (if applicable), the issue, and the number of users that saw or interacted with the content (e.g. clicked or shared).**
- **What steps has Twitter taken to inform users, the public, and the United States Government of any case listed in response to the previous question?**

Twitter is unaware of any instances, beyond the Russian-linked and Iran-affiliated accounts we have already disclosed publicly, of foreign government or entity acting on the behalf of a foreign government that have created Twitter accounts or written tweets, that masquerade as American for the purpose of influencing political debate or policymaking within the United States.

16. Twitter, like several other major technology companies, warns users when it believes their accounts may have been targeted by foreign governments.

- **In each of the past five years, how many times has Twitter notified users located in the United States that their accounts were targeted by a foreign government?**
- **Prior to being notified by Twitter, how many of these accounts had some form of two-factor authentication enabled on their accounts?**
- **Prior to being notified by Twitter, how many of these accounts were secured with a two-factor authentication security key?**

Twitter is committed to providing greater transparency to our account holders and the public, particularly as it relates to election integrity. Following our investigation into the propaganda effort by the Internet Research Agency (IRA), we notified approximately 1.4 million individuals on our platform who interacted with this malicious content.

Twitter sent notices to people on the platform with an active email address who our records indicate are based in the U.S. and fall into at least one of the following categories:

- People who directly engaged during the election period with the 3,814 IRA-linked accounts we identified, either by Retweeting, quoting, replying to, mentioning, or liking those accounts or content created by those accounts;
- People who were actively following one of the identified IRA-linked accounts at the time those accounts were suspended; and
- People who opt out of receiving most email updates from Twitter and would not have received our initial notice based on their email settings.

Twitter does not have data on the the number of accounts with a two-factor authentication key that interacted with with the IRA, although in this instance the security of the accounts that interacted with the IRA were not compromised.

17. In each of the past five years, how many times has Twitter notified users believed by Twitter to be elected officials or their staff in the United States that their accounts were targeted by a foreign government?

- **Prior to being notified by Twitter, how many of these accounts had some form of two-factor authentication enabled on their accounts?**
- **Prior to being notified by Twitter, how many of these accounts were secured with a two-factor authentication security key?**

We will provide additional information to the Committee concerning the targeting of elected officials or their staff in the United States in a more secure setting.

18. In each of the past five years, how many user accounts, if any, have been compromised, such that someone other than the user gained access to the user's non-public account data?

- **How many of these accounts had some form of two-factor authentication enabled on their accounts.**
- **How many of these accounts were secured with a two-factor authentication security key?**

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. A relatively small number of people using Twitter within the United States have two-factor authentication enabled. Since May 2018, Twitter estimates that approximately 3 million accounts may have potentially been impacted by data breaches, although there is no indication these have been associated with foreign government activity.

19. In each of the past five years, how many user accounts were compromised, such that someone other than the user gained access to the user's non- public account data, by adversaries that Twitter believes may be a foreign government or are working with a foreign government?

- **How many of these accounts had some form of two-factor authentication enabled on their accounts?**
- **How many of these accounts were secured with a two-factor authentication security key?**

Please see the response to question 18.

20. Twitter has a tool called Tailored Audiences, which is similar to Facebook's Custom Audiences and allows advertisers to upload lists of specific users to target them with ads. Twitter's policies state that advertisers who use Tailored Audiences must obtain consent from users about the data they have acquired.

- **How does Twitter ensure that all of its advertisers actually follow those policies?**
- **Is Twitter aware of any advertisements targeted with Tailored Audiences that appear to be designed to discourage any United States citizen from voting?**
    - **If yes, please provide a full accounting of each case, including the advertisement, what Twitter knows about the party that purchased the advertising, and the number of users that saw or interacted with the content (e.g. clicked).**
    - **If the answer to the question above is yes, were these voter discouragement ads targeted at people of any particular race or ethnic group?**
    - **Were these voter discouragement ads predominantly targeted at people expected to vote for one party or the other?**

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 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 people on the platform to download a list of advertisers that have included them in a Tailored Audience.

If people 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 individuals 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 is not aware of any advertisements targeted with Tailored Audiences that appear to be designed to discourage any United States citizen from voting.

21. Has any foreign government, their agent, or other foreign entity ever used Tailored Audiences to target individuals in the United States?

- **If yes, please provide a full accounting of each case, including the party that purchased the advertising, the foreign government sponsor (if applicable), and the number of users that saw or interacted with the content (e.g. clicked or shared).**

In 2017, the U.S. intelligence community named Russia Today (RT) and Sputnik as implementing state-sponsored Russian efforts to interfere with and disrupt the 2016 U.S.

Presidential election. We made the policy decision to off-board advertising from all accounts owned by RT and Sputnik based on the retrospective work we had conducted around the 2016 U.S. election and the U.S. intelligence conclusion that both RT and Sputnik attempted to interfere with the election on behalf of the Russian government.

In 2014, @RT_com used Tailored Audiences to deliver advertisements totaling $8,487 and @RTUKnews used Tailored Audiences to deliver advertisements totaling $165 in 2015.

Based on our internal investigation of their behavior as well as their inclusion in the January 2017 intelligence community report, Twitter decided to take the $1.9 million we were projected to have earned from RT global advertising since they became an advertiser in 2011, which includes the $274,100 in 2016 U.S.-based advertising, and donated those funds to support external research into the use of malicious automation and misinformation, with an initial focus on elections and automation.

22. Does Twitter have a policy of shutting down accounts that seek to suppress voting?

- **If yes, to what kind of content does Twitter apply that policy (e.g., content discouraging people from voting, content providing inaccurate information on how or when to vote, etc.)?**

Twitter unequivocally condemns the use of our platform for any election interference activity. Tweets aimed at suppressing voter turnout generally are surfaced through reports from people using our service. This content is reviewed, then promptly removed as illegal interference with voting rights: the content is either restricted as inaccessible pending deletion by the individuals (i.e., other individuals on the platform are unable to see the content) or the responsible accounts are permanently suspended. In addition, in order to proactively surface additional Tweets with a given text-to-vote meme, Twitter utilizes technology for identifying instances where the same image appears across multiple Tweets. Content identified through this process is then subject to manual review.

Depending on the number of violations for any given account disseminating voter suppression Tweets, Twitter will either restrict access to the Tweet or suspend the account. During the period leading up to the 2016 election, for example, Twitter labeled and restricted access to the vote-to-text Tweets pursuant to the Twitter User Agreement, which contains the Twitter Terms of Service, Twitter Privacy Policy, and Twitter Rules. According to the unlawful use provision of the Twitter Rules, people are prohibited from using Twitter's "service for any unlawful purpose or in furtherance of illegal activities" and "[i]nternational users agree to comply with all local laws regarding online conduct and acceptable content."

Additionally, 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

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.

23. Has the Internet Research Agency ever used Tailored Audiences to target users, in the United States or elsewhere, with advertisements?

Twitter is not aware of any efforts by the Internet Research Agency to use Tailored Audiences to target individuals, in the United States or elsewhere, with advertisements.

24. Does Twitter believe that any of the content created by the Russian Internet Research Agency was designed to discourage anyone from voting?

Twitter did not see that content created by the Russian Internet Research Agency that constituted voter suppression.

25. Can users opt out of being targeted with Tailored Audiences?

- **If no, why not?**

Yes. If individuals 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 your 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 your Android phone's settings.

26. For several years, Twitter has allowed its customers to protect their accounts from hacking through the use of two-factor authentication. Since June, Twitter has also supported the use of a physical security token as an enhanced form of two-factor authentication. However, two-factor authentication remains an opt-in feature for Twitter users.

- **Does Twitter require that its employees use two-factor authentication for their work accounts?**
- **If yes, does Twitter require, like Google, that employees use a security key?**

We will provide additional information to the Committee concerning our security protocols for employees in a more secure setting.

27. Do you have two-factor authentication enabled for your personal Twitter and personal email accounts?

- **If yes, are you using a security key?**

Yes, two-factor authentication is enabled and we will provide additional information to the Committee concerning our security protocols in a more secure setting.

28. What percentage of Twitter's U.S. customers have enabled any form of two-factor authentication?

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 individual is 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.

We will provide additional information to the Committee concerning the use of two-factor authentication in the U.S. in a more secure setting.

29. What percentage of Twitter's U.S. customers have enabled enhanced two-factor authentication using a security key?

We will provide additional information to the Committee concerning the use of two-factor authentication using a security key in the U.S. in a more secure setting.

30. Since May, Twitter now specially identifies the accounts of individuals running for public office.

- **What percentage of these Twitter accounts currently have any form of two-factor authentication enabled?**
- **What percentage are using a security key?**
- **What specific outreach, if any, has Twitter engaged in to encourage elected officials and individuals running for public office to enable two-factor authentication on their official and personal Twitter accounts?**

Twitter recommends to the individuals on its platform certain best security practices in order to help keep their accounts secure, including those individuals who are running for public office. Twitter has 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.

In our correspondence with candidates participating in the election label program, Twitter encouraged them to review all of our security best practices. We stated: "As a friendly reminder, we highly recommend that you review our account safety and security best practices" and included a link to the relevant content. We have also distributed detailed information that includes security best practices to he political parties and other election stakeholders to encourage them learn about all of our integrity efforts.

We will provide additional information to the Committee concerning the use of two-factor authentication of badged candidates in the U.S. and their use of security keys in a more secure setting.

31. Twitter will place a blue verification badge on accounts “of public interest” which have been verified as authentic by Twitter.

- **Does Twitter currently require that verified accounts enable two-factor authentication?**

  No, two-factor authentication is not required.

[From Senator Heinrich]
32. In July 2018, Twitter acknowledged it has a problem with fake and automated accounts, or bots, and announced that in the final three months of 2017, the company had suspended 58 million accounts, another 70 million in May and June, and continuing at a rate of a million per day.

- **How has Twitter improved detection of automated accounts? Has this been a technical challenge?**

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 people 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. We will continue to undertake important steps to improve detection of automated accounts in the coming months and years.

33. According to the New York Times, Twitter's "purge" of fake and automated accounts resulted in over 300,000 followers lost for President Trump's Twitter account, about .58% of his total.

- **Following the purge, does Twitter estimate that any fake or automated accounts still follow President Trump on Twitter?**
- **How accurately can Twitter or other Twitter audit sites estimate the number of real and fake Twitter followers for any particular account?**

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 peopleto 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, including President Trump, 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.

34. In reply to a question about whether Twitter users should be notified whether they are communicating with a human or a machine, you testified that "we can identify these automations, we can label them, and I think that is useful context and it's an idea that we have been considering over the past few months. It's really a question of the implementation, but we are interested in it and we are going to do something along those lines." You also noted that Twitter is looking at "expanding that transparency report around suspensions of any account."

- **Do you have any more information on what a transparency report on numbers of and suspension of automated account activity might look like or when Twitter might begin issuing such reports?**
- **What are the challenges in distilling such information?**

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.

The forthcoming Transparency Report will also include information on automated manipulation.

35. Twitter estimates that fewer than 8.5 percent of its users use automation tools, yet it has recently announced the suspension of millions of accounts, which calls that estimate into question.

- **What is Twitter's latest estimate of numbers of its accounts that are automated?**
- **What is Twitter's estimate of numbers of automated accounts that are used maliciously, as opposed to for positive purposes?**
- **Why is it difficult to provide these kinds of estimates?**

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 peopleto 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.

36. Twitter has disputed estimates by outside researchers that up to 15 percent of its accounts are bots rather than real people.

- **Is Twitter collaborating with academics and the research community in order to better quantify the extent of its bot problem?**

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.

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. On August 24, 2018, Twitter hosted our industry peers to discuss data sharing about hostile foreign actors regarding 2018 election security. We continue to meet in regular cadence with our industry peers about election integrity efforts.

We also have well-established relationships with law enforcement agencies active in this arena, including the Federal Bureau of Investigation Foreign Influence Task Force and the Department of 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.

Additionally, we committed to the United States Congress and the public to provide regular updates and information regarding our investigation into foreign interference in political conversations on Twitter. Since that time, we have shared examples of these types of content posted on Twitter by the Internet Research Agency (IRA) and provided the public with a direct notice if they interacted with these accounts. In August, we also disclosed details of another attempted influence campaign we identified as potentially located within Iran.

In line with our strong principles of transparency and with the goal of improving understanding of foreign influence and information campaigns, on October 17, 2018, Twitter released the full, comprehensive archives of the Tweets and media that are connected with these two previously disclosed and potentially state-backed operations on the service. We are making this data available with the goal of encouraging open research and investigation of these behaviors from researchers and academics around the world.

These large datasets consist of 3,841 accounts affiliated with the IRA, originating in Russia, and 770 other accounts, potentially originating in Iran. They include more than 10 million Tweets and more than 2 million images, GIFs, videos, and Periscope broadcasts, including the earliest on-Twitter activity from accounts connected with these campaigns, dating back to 2009.

[From Senator Lankford]
37. How do you verify that third-parties with access to Twitter's data do not violate the company's terms of use?

We recognize that access to that data could be manipulated, so we have taken steps to prevent the use of our application programming interfaces ("APIs") for products and services that are abusive or that disrupt the health of conversations. Those to whom we grant access to our APIs are prohibited from using the data to manipulate conversations or otherwise abuse the data. Between April and June 2018 alone we removed more than 143,000 applications that we determined to be in violation of our developer policies. Most violated our policies against producing spam via APIs. And we continue to invest in and improve our detection tools to stop misuse of public Twitter data.

In July 2018, we introduced a new measure designed to increase developers' accountability for applications that create and engage with Twitter content and accounts. Twitter now reviews and conducts compliance checks of all developers' stated use of the data that they wish to access. We have also added new protections aimed to prevent the registration of low quality and spam-generating applications. We believe that these additional steps will help protect the integrity of our platform.

38. What are Twitter's platform threat detection capabilities?

- **What are the limitations of Twitter's ability to detect threats to the platform?**
- **How much of Twitter's platform threat detection is outsourced?**

Twitter has 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.

To supplement its own analyses, Twitter's analytical team also receives and responds to reports from across the company and from external third parties. The results from all of the team's analyses are shared with key stakeholders at Twitter and provide the basis for policy changes and product initiatives and removal of accounts.

The primary focus of the cross-functional analytical team is election readiness. Leading up to and during the 2018 election period, the team will examine, respond to, and escalate instances of suspected inauthentic, election-related coordinated activity in political conversation and conduct in-depth analyses of relevant Twitter data.

[From Senator Harris]
39. In his November 2017 testimony to the Committee, Twitter's general counsel stated that "false or spam accounts represent less than 5% of our [Monthly Active Users] (MAU). On July 7, 2018, the Washington Post reported that Twitter suspended over 70 million accounts deemed fake or suspicious in May and June. Additionally, on July 17th, the Associated Press reported Twitter suspended 58 million accounts in the final three months of 2017.

- **How many accounts has Twitter suspended, in total, since November of 2016?**
- **What percentage of Twitter's total registered users does that represent?**
- **What percentage of your active users does that represent?**
- **How many of the suspended accounts claimed a location in the United States but had technical access that suggested a foreign location?**
- **How many of the suspended accounts connected to Twitter from an IP address in a foreign country?**
- **How many of the suspended accounts used a Virtual Private Network (VPN)?**
- **How many of the suspended accounts were automated?**
- **How many of the suspended automated accounts used Twitter's application program interface (API)?**
- **How many malicious automated accounts used "headless" browsers, i.e., browsers without a visual user interface, or other methods of device impersonation?**
    - **What steps have you taken to detect such 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 people 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.

40. Do you stand by previous estimates of false or spam accounts, including for previous quarters in your SEC filings?

Yes.

41. Do you intend to revise or update testimony previously provided to the Committee concerning Twitter's estimates of the proportion of MAU comprising false or spam accounts?

No.

42. There are machine learning techniques that can create entirely fake videos, called "deepfakes." These deepfakes often depict people saying things they never said or portray events that never occurred.

- **Are deepfakes a violation of Twitter's terms of use?**
- **What is Twitter doing to identify deepfakes on its platform and to alert users when they may be seeing deepfakes?**
- **How many deepfakes has Twitter identified on its platform to date?**

Twitter is aware of deepfakes in the context of intimate media on the platform. Deepfakes in the context of intimate media are clear violations of our terms of services and our intimate media policy. Twitter suspends any account we identify as the original poster of intimate media that has been produced or distributed without the subject's consent. We also suspend any account dedicated to posting this type of content.

43. Can Twitter commit to:

- **Assessing how foreign disinformation campaigns can use deepfakes;**
- **Developing a strategy to combat it; and,**
- **Reporting its findings and efforts to the Committee by the end of the year?**

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.

We appreciate the continued partnership with the Committee, and we share your concern about malicious foreign efforts to manipulate and divide people in the United States and throughout the world, including through the use of foreign disinformation campaigns that rely upon the use of deepfakes. We will continue to share our ongoing work to safeguard elections with the members of this Committee.