United States Senate Select Committee on Intelligence

Testimony of Jack Dorsey
Chief Executive Officer
Twitter, Inc.

September 5, 2018

Chairman Burr, Vice Chairman Warner, and Members of the Committee:

I am grateful for the opportunity to appear here today.

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.

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. We have implemented significant improvements since we last appeared before the Committee in November, and we will continue to undertake important steps in the coming months and years.

I look forward to sharing our work with the members of this Committee and listening to your recommendations on how best to increase the health of our platform and its role in our democracy from manipulation by hostile foreign actors.

From Twitter's perspective, this threat is not limited solely to elections or politics. Instead, we view it as a challenge to the fundamental health of our platform, and by extension, to the global public conversation that Twitter serves. We commit to continuing to confront that challenge together.

I. RUSSIAN INTERFERENCE IN THE 2016 ELECTION AND LESSONS LEARNED

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. 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.

A. Retrospective Review

Last fall, we conducted a comprehensive retrospective review of platform activity related to the 2016 election. To better understand the nature of the threat and ways to address future attempts at manipulation, we examined activity on the platform during a 10-week period preceding and immediately following the 2016 election (September 1, 2016 to November 15, 2016). We focused on identifying accounts that were automated, linked to Russia, trying to get unearned attention, and Tweeting election-related content, and we compared activity by those accounts to the overall activity on the platform. We reported the results of that analysis in November 2017, and we updated the Committee in January 2018 about the findings from our ongoing review.

As we reported in January 2018, we identified 50,258 automated accounts that were Russian-linked and Tweeting election-related content, representing less than two one-hundredths of a percent (0.016%) of the total accounts on Twitter at the time. Of all election-related Tweets that occurred on Twitter during that period, these malicious accounts constituted approximately one percent (1.00%), totaling 2.12 million Tweets. Additionally, in the aggregate, automated, Russian-linked, election-related Tweets from these malicious accounts generated significantly fewer impressions (*i.e.*, views by others on Twitter) relative to their volume on the platform. Additional information on the accounts associated with the Internet Research Agency is included below.

Twitter is committed to ensuring that promoted accounts and paid advertisements are free from hostile foreign influence. In connection with the work we did in the fall, 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 into election and civic engagement.

B. Insights from Our Review

Although the volume of malicious election-related activity that we could link to Russia was relatively small, we strongly believe that any such activity on Twitter is unacceptable. 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. Tied to this is our commitment to providing tools for journalism to flourish by creating and maintaining a platform that helps to provides people with high-quality, authentic information in a healthy and safe environment.

We also recognize that, as a private company, there are threats that we cannot understand and address alone. We must continue to work together with our elected officials, government partners, industry peers, outside experts, and other stakeholders so that the American people and the global community can understand the full context in which these threats arise.

II. IMPROVEMENTS TO TWITTER

We have made the health of Twitter our top priority, and our efforts will be measured by how we help encourage more healthy debate, conversations, and critical thinking on the platform. Conversely, abuse, automation, and manipulation will detract from the health of our platform. 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 the 2018 election, described below.

A. Combating Malicious Automation and Protecting Conversation Health

Using the insights from our retrospective review, 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 users 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.

B. Corporate Reorganization and Formation of a Dedicated Cross-Functional Analytical Team

Our improvements include important structural changes. I 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.

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.

C. Political Conversations Dashboard

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.

Through real-time review and detection of anomalous and potentially malicious automated or human-coordinated activity, the team will work to identify and address any attempts by bad faith actors to interfere with the electoral process, and will be better informed about where and how to deploy resources to proactively review potential malicious activity. Accounts will be escalated for review in real-time if exhibiting anomalous patterns of behavior. These efforts will significantly improve our ability to detect malicious automated and human-coordinated activity surrounding political content as well as the speed with which we address those issues.

D. Candidate Verification

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).

E. Election Labels

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.

F. Advertising and Promoted Content

As we learned from our 2016 retrospective review and the important work of your Committee, bad faith actors have attempted to influence the electoral process by propagating paid content on the platform, including political advertisements and promoted Tweets. As we reported in the fall, 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, 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.

G. Engagement with Key Stakeholders

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.

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, 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.

Consistent with our longstanding commitment to serving the public conversation, we partnered with experts at the University of Oxford and Leiden University to better evaluate our work on conversation health, 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.

Last October, Twitter barred advertising from Russia Today and Sputnik, both of which the U.S. Intelligence Community determined to have interfered with the election on behalf of the Russian government. We also devoted the $1.9 million these accounts spent on the platform to research. The first recipients of those funds include the Kofi Annan Foundation's Global Commission on Elections, Democracy, and Security, the Atlantic Council, the EU Disinfolab and the Reporters Committee for Press Freedom.

We also collaborate with a number of non-governmental organizations that are focused on voter registration, civic engagement, and media literacy, including RockTheVote, Democracy Works, TurboVote Challenge, HeadCount, DoSomething, and Ballotpedia.

H. Additional Safety Measures for Accessing Public Tweet Data

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").

We recognize that access to that data could be manipulated, so we have taken steps to prevent the use of our 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.

III. RECENT ACTIVITY ON THE PLATFORM

Twitter continues to see bad faith actors continue their attempts to manipulate and divide people on Twitter. Two such examples include recent activity related to new malicious activity by the Russian Internet Research Agency and malicious accounts located in Iran.

A. Malicious Accounts Affiliated with the Russian Internet Research Agency

Twitter has seen recent activity on the platform affiliated with the Russian Internet Research Agency. As we reported to the Committee in January 2018, we continue to identify accounts that we believe may be linked to the Internet Research Agency ("IRA"). As of today, we have suspended a total of 3,843 accounts we believe are linked to the IRA. And we continue to build on our contextual understanding of these accounts to improve our ability to find and suspend this activity as quickly as possible in the future, particularly as groups such as the IRA evolve their practices in response to suspension efforts across the industry.

As an example of Twitter's ongoing efforts, Twitter identified 18 accounts in March 2018 we believe to be linked to the Internet Research Agency uncovered by our ongoing additional reviews. These accounts were created and registered after the 2016 election. These accounts used false identifies purporting to be Americans, and created personas focused on divisive social and political issues. The accounts represented both sides of the political spectrum. We continue to work with our law enforcement partners on this investigation.

B. Malicious Accounts Located in Iran

In August 2018, we were notified by an industry peer about possible malicious activity on their platform. After receiving information from them, we began an investigation on our platform to build out our understanding of these networks. We immediately notified law enforcement on this matter as soon as we discovered malicious activity.

We initially identified accounts based on indicators such as phone numbers and email addresses. Some of these accounts appeared to pretend to be U.S. person and discuss U.S. social commentary. In most cases, the accounts that appeared to suggest a U.S. affiliation or target U.S. person were created after the 2016 election. These accounts were in violation of our platform manipulation policies, and were engaged in coordinated activity intended to propagate messages artificially across accounts.

These accounts appear to be located in Iran. This is indicated by, for example, accounts related by an Iranian mobile carrier or phone number or Iranian email address on the account. Although Twitter is blocked in Iran, we may see people active on our service via a virtual private network, or VPN.

We suspended 770 accounts for violating Twitter policies. 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 a year old. One advertiser ran $30 in ads in 2017. Those ads did not target the U.S. and the billing address was located outside of Iran. We will remain engaged with law enforcement on this issue.

Twitter has been in close contact with our industry peers on this matter and received detailed information from them about the malicious accounts located with Iran, which has assisted us in our investigation, and we have shared our own details and work with other companies. We expect this process will continue and that the industry can continue to build on this effort and assist with this ongoing investigation.

*   *   *

Our core mission is to serve the public conversation. It is why we exist. We must promote and maintain the health of that conversation. The people who use our service must 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 taking the steps I have outlined above, we continue our efforts to address those threats posed by hostile foreign governments and foster an environment conducive to healthy, meaningful conversations on our platform. This work is essential, and today's hearing better equips us to confront this new threat to our platform and democracies across the globe.

I look forward to answering your questions.