United States House Committee on Energy and Commerce

Testimony of Jack Dorsey
Chief Executive Officer
Twitter, Inc.

September 5, 2018

Chairman Walden, Ranking Member Pallone, and Members of the Committee:

Thank you for the opportunity to appear before the Committee today so I may speak to you and the American people.

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

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

Today, I hope my testimony before the Committee will demonstrate the challenges that we are tackling as a global platform. Twitter is approaching these challenges with a simple question: How do we earn more trust from the people using our service? We know the way earn more trust around how we make decisions on our platform is to be as transparent as possible. We want to communicate how our platform works in a clear and straightforward way.

There are other guiding objectives we consider to be core to our company. We must ensure that all voices can be heard. We must continue to make improvements to our service so that everyone feels safe participating in the public conversation – whether they are speaking or simply listening. And we must ensure that people can trust in the credibility of the conversation and its participants.

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

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

My testimony today will provide important information about our service: (1) an explanation of our commitment to improve the health on Twitter; (2) the algorithms that shape the experience of individuals who use Twitter; (3) an update on Twitter's work on Russian interference in the 2016 elections; and (4) information on recent malicious activity Twitter saw on the platform.

I. TWITTER'S COMMITMENT TO HEALTH

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

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

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

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

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

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

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

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

II. ALGORITHMS SHAPING THE TWITTER EXPERIENCE

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

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

A. Timeline Ranking and Filtering

For nearly a decade, the Twitter home timeline displayed Tweets from accounts an individual follows in reverse chronological order. As the volume of content on Twitter continually increased, individuals using the platform told us they were not always seeing useful or relevant information, or were missing important Tweets, and that their home timeline sometimes felt noisy. Based on this feedback, in 2016 we introduced a new ranking feature to the home timeline. This feature creates a better experience for people using Twitter by showing people the Tweets they might find most interesting first. Individuals on Twitter can disable this feature in their settings and return to a reverse chronological timeline at any time. When the feature is disabled, our content suggestions are relatively minimal.

Depending on the number of accounts an individual follows, not all content from all followed accounts may appear in the home timeline. Many people using Twitter follow hundreds or even thousands of Twitter accounts. While Twitter strives to create a positive experience with the ranked timeline, people opening Twitter may still feel as if they have missed important Tweets. If that happens, people can always opt to return to a reverse chronological timeline or view content from people they follow by visiting their profiles directly. We also continue to invest in improving our machine learning systems to predict which Tweets are the most relevant for people on our platform.

In addition to the home timeline, Twitter has a notification timeline that enables people to see who has liked, Retweeted and replied to their Tweets, as well as who mentioned or followed them. We give individuals on Twitter additional controls over the content that appears in the notifications timeline, since notifications may contain content an individual on Twitter has not chosen to receive, such as mentions or replies from someone the individual does not follow. By default, we filter notifications for quality, and exclude notifications about duplicate or potentially spammy Tweets. We also give individuals on the platform granular controls over specific types of accounts they might not want to receive notifications from, including new accounts, accounts the individual does not follow, and accounts without a confirmed phone or email address.

B. Conversations

Conversations are happening all the time on Twitter. The replies to any given Tweet are referred to as a "conversation." Twitter strives to show content to people that we think they will be most interested in and that contributes meaningfully to the conversation. For this reason, the replies, grouped by sub-conversations, may not be in chronological order. For example, when ranking a reply higher, we consider factors such as if the original Tweet author has replied, or if a reply is from someone the individual follows.

C. Safe Search

Twitter's search tools allow individuals on Twitter to search every public Tweet on Twitter, going back to my very first Tweet in 2006. There are many ways to use search on Twitter. An individual can find Tweets from friends, local businesses, and everyone from well-known entertainers to global political leaders. By searching for topic keywords or hashtags, an individual can follow ongoing conversations about breaking news or personal interests. To help people understand and organize search results and find the most relevant information quickly, we offer several different versions of search.

By default, searches on Twitter return results in "Top mode." Top Tweets are the most relevant Tweets for a search. We determine relevance based on the popularity of a Tweet (*e.g.*, when a lot of people are interacting with or sharing via Retweets and replies), the keywords it contains, and many other factors. In addition, "Latest mode" returns real-time, reverse-chronological results for a search query.

We give people control over what they see in search results through a "Safe Search" option. This option excludes potentially sensitive content from search results, such as spam, adult content, and the accounts an individual has muted or blocked. Individual accounts may mark their own posts as sensitive as well. Twitter's safe search mode excludes potentially sensitive content, along with accounts an individual may have muted or blocked, from search results in both Top and Latest. Safe Search is enabled by default, and people have the option to turn safe search off, or back on, at any time.

D. Behavioral Signals and Safeguards

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

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

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

Through early testing in markets around the world, Twitter has already seen a recent update to this approach have a positive impact, resulting in a 4 percent drop in abuse reports from search and 8 percent fewer abuse reports from conversations. That metric provided us with strong evidence that fewer people are seeing Tweets that disrupt their experience on Twitter.

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

E. Additional Context to High-Profile Incidents

Conservative voices have a strong presence on Twitter. For example, in 2017, there were 59.5 million Tweets about Make America Great Again or MAGA. According to the Pew Research Center, people on Twitter used #MAGA an average of 205,238 times per day from Election Day 2016 through May 1, 2018. It was the third most Tweeted hashtag in 2017. Another top hashtag on Twitter is #tcot, or Top Conservatives on Twitter, with 8.4 million Tweets in 2017. During the annual Conservation Political Action Committee (CPAC) conference in February 2018, #CPAC and #CPAC2018 were tweeted 1.2 million times in a four day period. And Twitter's political sales team works with hundreds of active conservative advertisers.

Twitter enabled the White House and media broadcasters to have a dynamic experience on Twitter, publishing and promoting live video event pages to millions of people on Twitter during President Trump's State of the Union address in 2017. In total, more than 39 media broadcasters including ABC, Bloomberg, CBS, FoxNews, PBS NewsHour, Reuters, Univision,

and USA Today participated. Additionally, the White House and Senate GOP both published the entire live video on Twitter reaching over 3.4 million viewers.

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

An analysis of accounts for Members of Congress that were affected by this search issue demonstrate there was no negative effect on the growth of their follower counts. To the contrary, follower counts of those Members of Congress spiked. Twitter can make the results of this internal analysis available to the Committee upon request.

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

Some critics have raised concerns regarding the impact that mass block lists can have on our algorithms. Our behavioral signals take into account only blocks and mutes that are the result of direct interactions among people on Twitter. That means that, while blocks that result from interactions with others on Twitter are factored into the discoverability of content, blocks that derive from mass block lists have minimal effect on the platform beyond those who have blocked particular other individuals on the platform.

In preparation for this hearing and to better inform the members of the Committee, our data scientists analyzed Tweets sent by all members of the House and Senate that have Twitter accounts for a 30 day period spanning July 23, 2018 until August 13, 2018. We learned that, during that period, Democratic members sent 10,272 Tweets and Republican members sent 7,981. Democrats on average have more followers per account and have more active followers. As a result, Democratic members in the aggregate receive more impressions or views than Republicans.

Despite this greater number of impressions, after controlling for various factors such as the number of Tweets and the number of followers, and normalizing the followers’ activity, we observed that there is no statistically significant difference between the number of times a Tweet

by a Democrat is viewed versus a Tweet by a Republican. In the aggregate, controlling for the same number of followers, a single Tweet by a Republican will be viewed as many times as a single Tweet by a Democrat, even after all filtering and algorithms have been applied by Twitter. Our quality filtering and ranking algorithm does not result in Tweets by Democrats or Tweets by Republicans being viewed any differently. Their performance is the same because the Twitter platform itself does not take sides.

III. TWITTER'S WORK ON RUSSIAN INTERFERENCE IN THE 2016 ELECTION

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 conversations 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 the combined information is more powerful in combating these threats.

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, 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. Additional information on the accounts associated with the Internet Research Agency is included below.

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.

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.

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.

B. Combating Malicious Automation

In the last year, Twitter 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. Many of these product changes are designed to combat spam and malicious automation.

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 then requiring confirmation that a human is controlling the account. 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 percent more accounts year-over-year for violating our 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, a result that demonstrates 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.

Finally, this summer, we made an important step to increase confidence in follower counts by removing locked accounts from follower counts across profiles globally, 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. As a result, the number of followers displayed on many profiles went down. We were transparent about these changes which impacted many people who use Twitter across the political spectrum and are a key part of our information quality efforts.

IV. RECENT ACTIVITY ON THE PLATFORM

Twitter continues to see bad faith actors attempt 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. 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; we then identified additional problematic accounts by matching other behavioral signals. Some of these accounts appeared to pretend to be people in the U.S. and discuss U.S. social commentary. In most cases, the accounts that appeared to suggest a U.S. affiliation or target U.S. audiences 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 engage via virtual private network.

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 and our peer companies on this issue.

Twitter has been in close contact with our industry peers about the malicious accounts located within Iran—we have received detailed information from them that 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.

* * *

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

Thank you, and I look forward to your questions.