Twitter, Inc.

United States Senate Committee on the Judiciary, Subcommittee on Crime and Terrorism

Update on Results of Retrospective Review of Russian-Related Election Activity

January 19, 2019

As we explained last fall, Twitter undertook a retrospective review of activity on our platform to assess Russian efforts to influence the 2016 U.S. election through malicious automation, coordinated human activity, and advertising. We committed at that time to pursue the review beyond the Committee's immediate inquiries, and we noted that we would update and supplement our findings with any additional information to report. Twitter took those commitments seriously.

Since our initial analysis, we have incorporated additional data, signals and other inputs; refined aspects of our methodology; and searched for additional meaningful connections in the data. The results of that work, which supplement our testimony, are presented here.

We have identified more accounts that appear to be associated with the Internet Research Agency ("IRA") and additional instances of relevant automated activity on our platform that indicate a connection to Russia. The results also confirm the broad conclusions of our earlier work: our analysis indicates that automated election-related content associated with Russian signals represented a very small fraction of the overall activity on Twitter in the ten-week period preceding the 2016 election.

As described in greater detail below, we have now identified an additional 1,062 accounts that appear to be IRA-linked. In total, during the relevant time period, the 3,814 identified IRA-linked accounts posted 175,993 Tweets, approximately 8.4% of which were election related.

Also as detailed below, through our supplemental analysis, we have identified an additional 13,512 accounts, for a total of 50,258 automated accounts that we identified as Russian-linked and Tweeting election-related content. This represents approximately two one-hundredths of a percent (0.016%) of the total accounts on Twitter at the time. The 2.12 million election-related Tweets that we identified through our retrospective review as generated by Russian-linked, automated accounts constituted approximately one percent (1.00%) of the overall election-related Tweets on Twitter at the time. Those 2.12 million Tweets received only one-half of a percent (0.49%) of impressions on election-related Tweets, based on impressions generated within the first seven days of posting by users logged into the system. In the aggregate, automated, Russian-linked, election-related Tweets generated significantly fewer impressions relative to their volume on the platform.

Last fall, Twitter emphasized its view that congressional hearings were an important step toward gaining a greater appreciation for how social media platforms, working hand-in-hand with the public and private sectors, can prevent the propagation of extremist content and disinformation—both generally and, of critical importance, in the context of the electoral

process. Our presentation of the latest results of our retrospective review reflects Twitter's continued commitment to those goals.

In furtherance of our commitment toward working with the public to provide transparency into this important issue for our democracy, we have also undertaken to provide notice to users in the United States who followed, liked or Retweeted IRA content on Twitter during the election time period. We believe it is consistent with Twitter's commitment to transparency to provide these insights, and helps increase awareness of the challenge these matters present now and in the future.

I. Retrospective Reviews of Malicious Automated and Human-Coordinated Activity in the 2016 Election

A. Human-Coordinated Russian-Linked Accounts

In our written testimony, we noted that we analyzed the accounts that we had identified through information obtained from third-party sources as linked to the IRA at that time. We have continued to refine that work, examining additional information in order to uncover other accounts that we believe have a sufficiently strong connection to attribute them to the IRA.

In the fall, we reported that we had found 2,752 IRA-linked accounts. We noted that this was an active area of inquiry and that we planned to update the Committee as we continued the analysis. Through that continued review, we have identified 1,062 more IRA-linked accounts during the relevant period, for a total of 3,814 such accounts. Those accounts posted 175,993 Tweets, approximately 8.4% of which were election-related. Many of their Tweets—just over 53%—were automated.

While automation may have increased the volume of content created by these accounts, IRA-linked accounts exhibited non-automated patterns of activity that attempted more overt forms of broadcasting their message. Some of those accounts represented themselves as news outlets, members of activist organizations, or politically engaged Americans. A few spent very small amounts (approximately $20 total) to promote content unrelated to the election. We have seen evidence of the accounts actively reaching out to journalists and prominent individuals (without the use of automation) through mentions. Some of the accounts appear to have attempted to organize rallies and demonstrations, and several engaged in abusive behavior and harassment. All 3,814 IRA-linked accounts were suspended for Terms of Service violations, and all but a few compromised accounts that have subsequently been restored to their legitimate account owners remain suspended.

B. Malicious Automated Activity

For our review of Twitter's core product, we analyzed election-related activity from the period preceding and including the election (September 1, 2016 to November 15, 2016) in order to identify content that appears to have originated from automated accounts associated with Russia. We then assessed the results to discern trends, evaluate our existing detection systems, and identify areas for improvement and enhancement of our detection tools.

1. Methodology

We took a broad approach for purposes of our review of what constitutes an election-related Tweet, relying on annotations derived from a variety of information sources, including Twitter handles, hashtags, and Tweets about significant events. For example, Tweets mentioning @HillaryClinton and @realDonaldTrump received an election-related annotation, as did Tweets that included #primaryday and #feelthebern. In total, our review has now encompassed nearly 212 million Tweets annotated in this way out of the total corpus of nearly 18.2 billion Tweets posted during this time period (excluding Retweets).

To ensure that we captured as many relevant automated accounts as possible in our review, Twitter analyzed the data not only using the detection tools that existed at the time the activity occurred, but also using newer and more robust detection tools that have been implemented since then or were developed for purposes of this review. We compared the results to determine whether these detection tools are able to capture automated activity that our 2016 techniques could not. These analyses relied on objective, measurable signals, such as the timing of Tweets and engagements, to classify a given action as automated.

We took a similarly expansive approach to defining what qualifies as a Russian-linked account. Because there is no single characteristic that reliably determines geographic origin or affiliation, we relied on a number of criteria, such as whether the account was created in Russia, whether the user had a Russian phone carrier or a Russian email address associated with the account, whether the user's display name contains a significant number of Cyrillic characters, and whether the user has logged in from any Russian IP address, even a single time. We considered an account to be Russian-linked if it had even one of the relevant criteria. As clarification, while we initially considered using frequency of Tweeting in Russian as a potential signal, ultimately this signal was not included in either phase of our analysis. For purposes of both the original and supplemental analysis, we focused on account sign-up language as the language signal, as it represents the language displayed to the user in their interface with Twitter.

Despite the breadth of our approach—and even in light of the additional signals we were able to rely on for the purpose of our supplemental analysis—there are technological limits to what we can determine based on the information we can detect regarding a user's origin. In the course of our analysis—and based in part on work conducted by our Information Quality team—we observed that a high concentration of automated engagement and content originated from data centers and users accessing Twitter via Virtual Private Networks ("VPNs") and proxy servers. In fact, based on our analysis at the time of the hearing, nearly 12% of Tweets created during the election originated with accounts that had a masked/indeterminate location on the day they were posted. Use of such facilities obscures the actual origin of traffic. Although our conclusions are thus necessarily contingent on the limitations we face, and although we recognize that there may be other methods for analyzing the data, we continue to believe our approach is the most effective way to capture as accurate an understanding as possible of activity on our system.

2. Analysis and Key Findings

Applying the methodology described above, and using detection tools we have since implemented or developed to further this review, we identified 50,258 accounts that generated automated, election-related content and had at least one of the characteristics we used to associate an account with Russia.

During the relevant period, those accounts generated approximately 2.12 million automated, election-related Tweets, which collectively received approximately 454.7 million impressions generated within the first seven days of posting.[1]

Because of the scale on which Twitter operates, it is important to place those numbers in context:

- The 50,258 automated accounts that we identified as Russian-linked and Tweeting election-related content represent approximately two one-hundredths of a percent (0.016%) of the total accounts on Twitter at the time.

- The 2.12 million election-related Tweets that we identified through our retrospective review as generated by Russian-linked, automated accounts constituted approximately one percent (1.00%) of the overall election-related Tweets on Twitter at the time.

- Those 2.12 million Tweets received only one-half of a percent (0.49%) of impressions on election-related Tweets (based on impression generated within the first seven days of posting). In the aggregate, automated, Russian-linked, election-related Tweets generated significantly fewer impressions relative to their volume on the platform.

In 2016, we detected and labeled some, but not all, of those Tweets using our then-existing anti-automation tools. Specifically, in real time, we detected and labeled as automated over half of the Tweets (approximately 1.34 million) from approximately half of the accounts (23,601), representing 0.63% of overall election-related Tweets and 0.20% of election-related Tweet impressions generated within the first seven days of posting.

Thus, our continued analysis has reinforced our earlier determination that the number of accounts we could link to Russia and that were Tweeting election-related content was small in comparison to the total number of accounts on our platform during the relevant time period. Similarly, the volume of automated, election-related Tweets that originated from those accounts remained small in comparison to the overall volume of election-related activity on our platform. And those Tweets generated significantly fewer impressions in the aggregate and on average compared to other election-related Tweets.

[1] Because Tweet impressions are generated by user views—and because users' Twitter timelines are constantly updated with new (and generally the most recent) Tweets—the vast majority of impressions on a Tweet are generated within the first seven days of posting.

3. Level of Engagement

In an effort to better understand the impact of Russian-linked accounts on broader conversations on Twitter, we continued to examine those accounts' volume of engagements with election-related content using additional signals.

We first reviewed the accounts' engagement with Tweets from @HillaryClinton and @realDonaldTrump. Our data showed that, during the relevant time period, @HillaryClinton Tweets were Retweeted approximately 8.6 million times. Of those Retweets, 47,846—or 0.55%—were from Russian-linked automated accounts. Tweets from @HillaryClinton received approximately 19.2 million likes during this period; 119,730—or 0.62%—were from Russian-linked automated accounts. The volume of engagements with @realDonaldTrump Tweets from Russian-linked automated accounts was higher, but still relatively small. The Tweets from the @realDonaldTrump account during this period were Retweeted more than 11 million times; 469,537—or 4.25%—of those Retweets were from Russian-linked, automated accounts. Those Tweets received approximately 28.8 million likes across our platform; 517,408—or 1.8%—of those likes came from Russian-linked automated accounts.

We also reviewed engagement between automated or Russia-linked accounts and the @Wikileaks, @DCLeaks_, and @GUCCIFER_2 accounts. The amount of automated engagement with these accounts ranged from 47.5% to 72.7% of Retweets and 37% to 64% of likes during this time—substantially higher than the average level of automated engagement, including with other high-profile accounts. The volume of automated engagements from Russian-linked accounts was lower overall. Our data show that, during the relevant time period, @Wikileaks Tweets were Retweeted approximately 5.65 million times. Of these Retweets, 196,836—or 3.48%—were from Russian-linked automated accounts. The Tweets from @DCLeaks_ during this time period were Retweeted 6,774 times, of which 2.47% were from Russian-linked automated accounts. The Tweets from @GUCCIFER_2 during this time period were Retweeted approximately 24,000 times, of which 2.32% were from Russian-linked automated accounts.

We also analyzed data concerning Tweets promoting the #PodestaEmails hashtag, which originated with Wikileaks' publication of thousands of emails from the Clinton campaign chairman John Podesta's Gmail account. We found that slightly under 5% of Tweets containing #PodestaEmails came from accounts with potential links to Russia, and that those Tweets accounted for less than 20% of impressions generated within the first seven days of posting. The core of the hashtag was propagated by Wikileaks, whose account sent out a series of 118 original Tweets containing variants on the hashtag #PodestaEmails referencing the daily installments of the emails released on the Wikileaks website. In the two months preceding the election, around 64,000 users posted approximately 484,000 unique Tweets containing variations of the #PodestaEmails hashtag. Our automated spam detection systems identified in real time approximately 25% of those Tweets, hiding them from searches. Based on information we had available at the time we submitted our written testimony, we know that approximately 75% of impressions on the trending topic within the first seven days were views by U.S.-based users. A significant portion of these impressions, however, are attributable to a handful of high-profile accounts, primarily @Wikileaks. At least one heavily-Retweeted Tweet came from another potentially Russia-linked account that showed signs of automation.

With respect to #DNCLeak, which concerned the disclosure of leaked emails from the Democratic National Committee, approximately 26,500 users posted around 154,800 unique Tweets with that hashtag in the relevant period. Of those Tweets, roughly 3% were from potentially Russian-linked accounts. Our automated systems at the time detected, labeled, and hid just under half (47%) of all the original Tweets with #DNCLeak. Of the total Tweets with the hashtag, 0.95% were hidden and also originated from accounts that met at least one of the criteria for a Russian-linked account. Those Tweets received 0.35% of overall Tweet impressions within the first seven days after posting. We learned that a small number of Tweets from several large accounts were principally responsible for the propagation of this trend. In fact, several of the most-viewed Tweets with #DNCLeak were posted by @Wikileaks, an account with millions of followers.

* * *

We hope that this presentation of the latest results of our retrospective review demonstrates our continued commitment to working with you, our industry partners, and other stakeholders to further public understanding of the role of social media in the 2016 election.