United States Senate Committee on Commerce, Science, and Transportation
Hearing Titled "Mass Violence, Extremism, and Digital Responsibility"

September 18, 2019

Written Testimony of Nick Pickles
Director, Public Policy Strategy
Twitter, Inc.

Chairman Wicker, Ranking Member Cantwell, and Members of the Committee:

At Twitter, our mission is to serve the public conversation. Twitter is a place where people from around the world come together in an open and free exchange of ideas. We have made the health of our service the top priority. Conversely, abuse, malicious automation, hateful conduct, violent extremist and terrorist content terrorism, and manipulation will detract from the health of our platform.

Tackling terrorism, violent extremism, and preventing violent attacks require a whole of society response, including from social media. It has long been a priority of Twitter to remove this content from the service. Let me be clear: Twitter has no incentive to keep terrorist and violent extremist content available on our platform. Such content does not serve our business interests, breaks our rules, and is fundamentally contrary to our values.

Communities in America and around the world have been impacted by incidents of mass violence, terrorism, and violent extremism with tragic frequency in recent years. These events demand a robust public policy response from every quarter. We acknowledge that the technology companies play a critical role, however, it is important to recognize content removal online cannot alone solve these issues.

We welcome the opportunity to continue to work with you on the Committee, our industry peers, government, academics, and civil society to find the right solutions. Partnership is essential.

My statement today will provide information and deeper context on: (I) Twitter's work to protect the health of the public conversation, including combating terrorism, violent extremist groups, and hateful conduct; (II) our policies relating to weapons and weapon accessories; and (III) our partnerships and societal engagement.

I.   TWITTER'S POLICIES ON TERRORIST CONTENT, VIOLENT EXTREMIST GROUPS, AND HATEFUL CONDUCT

All individuals accessing or using Twitter's services must adhere to the policies set forth in the Twitter Rules. Accounts under investigation or that have been detected as sharing content in violation with the Twitter Rules may be required to remove content, or in serious cases, will see their account permanently suspended. Our policies and enforcement options evolve continuously to address emerging behaviors online.

A.   *Policy on Terrorism*

Individuals on Twitter are prohibited from making specific threats of violence or wish for the serious physical harm, death, or disease of an individual or group of people. This includes, but is not limited to, threatening or promoting terrorism.

We suspended more than 1.5 million accounts for violations related to the promotion of terrorism between August 1, 2015, and December 31, 2018. In 2018, a total of 371,669 accounts were suspended for violations related to promotion of terrorism. More than 90 percent of these accounts are suspended through our proactive measures.

We have a zero-tolerance policy and take swift action on ban evaders and other forms of behavior used by terrorist entities and their affiliates. In the majority of cases, we take action at the account creation stage — before the account even Tweets.

Government and law enforcement reports constituted less than 0.1 percent of all suspensions in the last reporting period. Continuing the trend we have seen for some time, the number of reports we received from governments of terrorist content from the second half of last year decreased by 77 percent compared to the previous reporting period covering January through June 2018.

We are reassured by the progress we have made, including recognition by independent experts. For example, Dublin City University Professor Maura Conway found in a detailed study that "ISIS's previously strong and vibrant Twitter community is now...virtually non-existent."

B.   *Policy on Violent Extremist Groups*

In December 2017, we broadened our rules to encompass accounts affiliated with violent extremist groups. Our prohibition on the use of Twitter's services by violent extremist groups —

i.e., identified groups subscribing to the use of violence as a means to advance their cause — applies irrespective of the cause of the group.

Our policy states:

Violent extremist groups are those that meet all of the below criteria:
- *identify through their stated purpose, publications, or actions as an extremist group;*
- *have engaged in, or currently engage in, violence and/or the promotion of violence as a means to further their cause; and*
- *target civilians in their acts and/or promotion of violence.*

An individual on Twitter may not affiliate with such an organization — whether by their own statements or activity both on and off the service — and we will permanently suspend those who do so.

We know that the challenges we face are not static, nor are bad actors homogenous from one country to the next in how they behave. Our approach combines flexibility with a clear, consistent policy philosophy, enabling us to move quickly while establishing clear norms of unacceptable behavior.

Since the introduction of our policy on violent extremist groups, we have taken action on 186 groups under this policy and permanently suspended 2,217 unique accounts. Ninety-three of these groups advocate violence against civilians alongside some form of extremist white supremacist ideology.

C. *Policy on Hateful Conduct*

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

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

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

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

D.   *Investing in Tech: Behavior vs. Content*

Twitter's philosophy is to take a behavior-led approach, utilizing a combination of machine learning and human review to prioritize reports and improve the health of the public conversation. That is to say, we increasingly look at how accounts behave before we look at the content they are posting. This is how we seek to scale our efforts globally and leverage technology even where the language used is highly context specific. Twitter employs extensive content detection technology to identify potentially abusive content on the service, along with allowing users to report content to us either as an individual or a bystander.

For abuse, this strategy has allowed us to take three times the amount of enforcement of action on abuse within 24 hours than this time last year. We now proactively surface over 50 percent of abusive content we remove using our technology compared to 20 percent a year ago. This reduces the burden on individuals to report content to us. Since we started using machine learning three years ago to reduce the visibility on abusive content:

- 80 percent of all replies that are removed were already less visible;
- Abuse reports have been reduced by 7.6 percent;
- The most visible replies receive 45 percent less abuse reports;
- 100,000 accounts were suspended for creating new accounts after a suspension during January through March 2019 — a 45 percent increase from the same time last year;
- 60 percent faster response to appeals requests with our new in-app appeal process;
- 3 times more abusive accounts suspended within 24 hours after a report compared to the same time last year; and
- 2.5 times more private information removed with a new, easier reporting process.

II. TWITTER POLICIES REGARDING WEAPONS AND WEAPON ACCESSORIES

Although Twitter's service does not have an e-commerce function, our Rules prohibit the selling, buying, or facilitating transactions in weapons, including firearms, ammunition, and explosives, and instructions on making weapons, such as bombs or 3D printed weapons. We will take appropriate action on any account found to be engaged in this activity, including permanent suspension of accounts where appropriate.

As stated publicly in our advertising policies, Twitter does not allow the use of our promoted products for the purpose of promoting weapons and weapon accessories globally. We explicitly ban advertising of guns, including airsoft guns, air guns, blow guns, paintball guns, antique guns, replica guns, and imitation guns. Twitter also prohibits the use of our promoted products for gun parts and accessories, including gun mounts, grips, magazines, and ammunition. We also do not allow the advertising of the rental of guns (other than from shooting ranges), stun guns, taser guns, mace, pepper spray or other similar self defense weapons. Additionally, we do not permit the advertising of a variety of weapons including swords, machetes, and other edged/bladed weapons; explosives, bombs and bomb making supplies and/or equipment; fireworks, flamethrowers and other pyrotechnic devices; and knives, including butterfly knives, fighting knives, switchblades, disguised knives, and throwing stars.

We do allow advertising related to the discussion of public policy issues pertaining to firearms. Twitter requires extensive information disclosures of any account involved in political issue advertising and provides specific information to the public via our Ads Transparency Center. Such advertisements are distinctly labeled as political issue promoted tweets. Organizations on both sides of the debate have utilized Twitter's promoted products and continue to do so, within the boundaries of our advertising policies.

III. PARTNERSHIPS AND SOCIETAL ENGAGEMENT

We work closely with the Federal Bureau of Investigation, along with law enforcement and numerous public safety authorities around the world. As our partnerships deepen, we are able to better respond to the changing threats we all face, sharing valuable information and promptly responding to valid legal requests for information.

A. *Cooperation with Law Enforcement*

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

Twitter informs individuals using the platform that we may preserve, use, or disclose an individual's personal data if we believe that it is reasonably necessary to comply with a law, regulation, legal process, or governmental request; to protect the safety of any person; to protect the safety or integrity of our platform, including to help prevent spam, abuse, or malicious actors on our services, or to explain why we have removed content or accounts from our services; to address fraud, security, or technical issues; or to protect our rights or property or the rights or property of those who use our services.

Twitter retains different types of information for different time periods, and in accordance with our Terms of Service and Privacy Policy. Given Twitter's real-time nature, some information (e.g., Internet Protocol logs) may only be stored for a very brief period of time.

Some information we store is automatically collected, while other information is provided at the user's discretion. Though we do store this information, we cannot guarantee its accuracy. For example, the user may have created a fake or anonymous profile. Twitter doesn't require real name use, email verification, or identity authentication.

Once an account has been deactivated, there is a very brief period in which we may be able to access account information, including Tweets. Content removed by account holders (e.g., Tweets) is generally not available.

Twitter accepts requests from law enforcement to preserve records, which constitute potentially relevant evidence in legal proceedings. We will preserve, but not disclose, a temporary snapshot of the relevant account records for 90 days pending service of valid legal process.

Twitter may honor requests for extensions of preservation requests, but encourage law enforcement agencies to seek records through the appropriate channels in a timely manner, as we cannot always guarantee that requested information will be available.

Our biannual Twitter Transparency Report highlights trends in enforcement of our Rules, legal requests, intellectual property-related requests, and email privacy best practices. The report also provides insight into whether or not we take action on these requests. The Transparency Report includes information requests from governments worldwide and non-government legal requests we have received for account information. In 2018, we received 4,323 requests from United States authorities, relating to 13,086 accounts,

B. *Industry Collaboration*

Collaboration with our industry peers and civil society is also critically important to addressing common threats from terrorism globally. In June 2017, 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 GIFCT announced a significant financial commitment to support research on terrorist abuse of the Internet and how governments, tech companies, and civil society can respond effectively. Our goal is to establish a network of experts that can develop platform-agnostic research questions and analysis that consider a range of geopolitical contexts.

Technological collaboration is a key part of GIFCT's work. In the first two years of GIFCT, two projects have provided technical resources to support the work of members and smaller companies to remove terrorist content.

First, the shared industry database of "hashes" — unique digital "fingerprints" — for violent terrorist propaganda now spans more than 100,000 hashes. 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 block extremist content before it is posted.

Second, a year ago, Twitter began working with a small group of companies to test a new collaborative system. Because Twitter does not allow files other than photos or short videos to be uploaded, one of the behaviors we saw from those seeking to promote terrorism was to post links to other services where people could access files, longer videos, PDFs, and other materials. Our pilot system allows us to alert other companies when we removed an account or Tweet that linked to material that promoted terrorism hosted on their service. This information sharing ensures the hosting companies can monitor and track similar behavior, taking enforcement action

pursuant with their individual policies. This is not a high-tech approach, but it is simple and effective, recognizing the resource constraints of smaller companies.

Based on positive feedback, the partnership has now expanded to 12 companies and we have shared more than 14,000 unique URLs with these services. Every time a piece of content is removed at source, it means any link to that source — wherever it is posted — will no longer be operational.

We are eager to partner with additional companies to expand this project, and we look forward to building on our existing partnerships in the future.

Finally, GIFCT has established a real-time crisis response process that allows us to respond to a violent act quickly to ensure that we share valuable information to limit the spread of terrorist and violent extremist content.

C.   *The Christchurch Call to Action*

In the months since a terrorist attack in Christchurch, New Zealand, New Zealand Prime Minister Jacinda Ardern has led the international policy debate, and that work has culminated in the Christchurch Call. Twitter's Chief Executive Officer Jack Dorsey attended the launch of the Christchurch Call in Paris, meeting with the Prime Minister to express our support and partnership with the New Zealand Government.

Because terrorism cannot be solved by the tech industry alone, the Christchurch Call is a landmark moment and an opportunity to convene governments, industry, and civil society to unite behind our mutual commitment to a safe, secure open, global Internet. It is also a moment to recognize that however or wherever evil manifests itself, it affects us all.

In fulfilling our commitments in the Call, we will take a wide range of actions. We continue to invest in technology to prioritize signals, including user reports, to ensure we can respond as quickly as possible to a potential incident, building on the work we have done to harness proprietary technology to detect and disrupt bad actors proactively.

As part of our commitment to educate users about our rules and to further prohibit the promotion of terrorism or violent extremist groups, we have updated our rules and associated materials to be clearer on where these policies apply. This is accompanied by further data being provided in our transparency report, allowing public consideration of the actions we are taking under our rules, as well as how much content is detected by our proactive efforts.

Twitter will take concrete steps to reduce the risk of livestreaming being abused by terrorists, while recognizing that during a crisis these tools are also used by news organizations, citizens and governments. We are investing in technology and tools to ensure we can act even faster to remove video content and stop it spreading.

Finally we are committed to continuing our partnership with industry peers, expanding on our URL sharing efforts along with wider mentoring efforts, strengthening our new crisis protocol arrangements, and supporting the expansion of GIFCT membership.

D.   *Partnerships with Civil Society*

In tandem with removing content, our wider efforts on countering violent extremism going back to 2015 have focused on bolstering the voices of non-governmental organizations and credible outside groups. These organizations and groups can use our uniquely open service to spread positive and affirmative campaigns that seek to offer an alternative to narratives of hate. Ideologies can only be successfully countered by those who have the credibility to take on the core messages being propagated, and if these core messages go unchallenged the removal of content will always be an incomplete response. These groups do critical work and policy makers should continue to find ways to broaden support for these efforts.

We have partnered with organizations delivering counter and alternative narrative initiatives across the globe and we encourage the Committee to consider the role of government in supporting the work of credible messengers in this space at home and abroad. Twitter has also delivered capacity building workshops to a range of organizations who seek to provide positive, alternative messages and work with communities and individuals at risk.

E.   *A Whole of Society Response*

The challenges we face as a society are complex, varied, and constantly evolving. These challenges are reflected and often magnified by technology. The push and pull factors influencing individuals vary widely, there is no common catalyst to action and there is no one solution to prevent an individual turning to violence. This is a long-term problem requiring a long-term response, not just the removal of content.

We are committed to playing our part. We will continue to seek to proactively remove terrorist and violent extremist content, work with industry peers to respond quickly in a crisis and to support smaller companies in tackling these challenges.

While we strictly enforce our policies, removing all discussion of particular viewpoints, no matter how uncomfortable society may find them, does not eliminate the ideology underpinning them. There is a risk such an approach moves these views into darker corners of the Internet where they cannot be challenged and held to account. As our peer companies improve in their efforts, this content continues to migrate to less-governed platforms and services often not at the forefront of public discussions. We are committed to learning and improving, but every part of the online ecosystem has a part to play.

Furthermore, not every issue will be one where the underlying factors can be addressed by public policy interventions led by technology companies.

* * *

We stand ready to assist the Committee in its important work regarding the issue of the tools that Internet companies can employ to stop the spread of mass violence on our services.