United States Senate Committee on the Judiciary
Subcommittee on the Constitution

Testimony of Carlos Monje
Director, Public Policy and Philanthropy
Twitter, Inc.

April 10, 2019

Chairman Cruz, Ranking Member Hirono, and Members of the Subcommittee:

Thank you for the opportunity to appear before you today.

Twitter is an American company, and Twitter's purpose is to serve the public conversation. Twitter is an open communications platform. We welcome perspectives and insights from diverse sources and embrace being a platform where the open and free exchange of ideas can occur. Every day on Twitter we see this play out on topics as diverse as sporting events, award shows, natural disasters, political movements, and the latest music.

We put the people who use our service first in every step we take. To support the many voices on Twitter, we have rules in place that are designed to ensure the safety and security of the people who come to our service. Safety and free expression go hand in hand, both online and in the real world. If people do not feel safe to speak, they very often will not.

These two guideposts, free expression for all perspectives and rules of the road to promote safety, are not only in our users' interests, but also paramount to sustaining our business. People come to Twitter to discover and talk about what is happening, and they want to hear from multiple perspectives. Conversely, people will not use our service if it is not a healthy space.

Today, I hope my testimony before the Committee will demonstrate our commitment to the free flow of information and the sharing of diverse perspectives and viewpoints. We want to communicate how our platform works in a clear and straightforward way.

Let me be clear about some important and foundational facts: Twitter does not use political viewpoints, perspectives, or party affiliation to make any decisions, whether related to automatically ranking content on our service or how we develop or enforce our rules. Our rules are not based on ideology or a particular set of beliefs. Instead, the Twitter Rules are based on behavior.

We believe strongly in being impartial, and we strive to enforce our rules dispassionately. We work extremely hard to make sure our algorithms are fair and endeavor to be transparent and fix issues when we make mistakes. The open nature of Twitter means that our enforcement actions are plainly visible to the public, even when we cannot always reveal the private details of individual accounts who have broken our rules. We do this to protect the privacy of the individuals who use our platform. And we strive to become more transparent when we remove a Tweet by providing explanations to individuals regarding which specific rules were broken.

My testimony today will provide important information about our service: (1) protecting diverse perspectives on Twitter; (2) additional context on some high-profile incidents; (3) the algorithms that shape the experience of individuals who use Twitter; and (4) Twitter's application of rules and policies.

I. PROTECTING DIVERSE PERSPECTIVES ON TWITTER

Every day, we see elected representatives around the world using Twitter to communicate with their constituents, fellow elected representatives, and with international leaders. In the United States, every senator, governor, House member, and mayors of the 25 largest cities have Twitter accounts. Millions of people around the globe have taken to Twitter to engage in local, national, and global conversations on a wide range of issues of civic importance. We also partner with news organizations to live-stream prominent 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.

Twitter continues to be one of the most popular platforms for conservative voices and movements in the United States. For example, in 2018, there were 32.6 million Tweets about Make America Great Again or MAGA. It was the fifth most Tweeted hashtag in the U.S. in 2018. Globally, the top 10 most mentioned accounts in 2018 included @realdonaldtrump and @POTUS, accounts for President Donald Trump. And Twitter's political sales team works with hundreds of active conservative advertisers.

Our Government and Elections Team also provides Twitter support and regular best-practices trainings for members of Congress — on both sides of the aisle. Providing this support to all elected officials, regardless of political party, is consistent with our commitment to serving the public conversation around political speech from various viewpoints.

Twitter also supports the White House and media broadcasters to have a dynamic experience on Twitter, publishing live video event pages to millions of people on Twitter during

President Trump's State of the Union address in 2019. In total, more than 22 media broadcasters including ABC, CBS, NBC News, PBS NewsHour, Reuters, Univision, and USA Today participated, reaching approximately 2.7 million live viewers. Additionally, the White House and Senate GOP both published the entire live video on Twitter reaching more than 4.6 million viewers. There were 5 million Tweets regarding the 2019 State of the Union. As a subset of that total, Twitter developed an emoji hashtag #SOTU that was Tweeted nearly 1.7 million times. The purpose of an emoji hashtag is to make it easier for people to discover and participate in the conversation about this topic. Although emoji hashtags are typically created as a paid advertisement, Twitter provided it without charge to encourage open discourse.

In preparation for this hearing and to better inform the members of the Subcommittee, our data scientists analyzed Tweets sent by all members of the House and Senate that have Twitter accounts for a five-week period spanning February 7, 2019, until March 17, 2019. We learned that, during that period, Democratic members sent 8,665 Tweets and Republican members sent 4,757. 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 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 algorithms do 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.

II. ADDITIONAL CONTEXT TO HIGH-PROFILE INCIDENTS

A. Auto-Suggest Issue

In July 2018, we acknowledged that some accounts (including those of Republicans and Democrats) were not being auto-suggested when people were searching for their specific name. This happened because Twitter had made a change to how one of our behavior-based algorithms worked in search results. A more detailed explanation of our behavior-based algorithms is included in Section III. When people used search, our algorithms were filtering out of auto-complete those accounts that had a higher likelihood of being abusive. Those search results remained visible if someone turned off the quality filter in search, and they were also visible elsewhere throughout the product.

Our change in the 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 available in search results. Once identified, this issue was resolved within 24 hours. In addition to fixing the search auto-suggestion function, we continue to carefully evaluate potential product changes for unintended consequences such as this.

This issue impacted 600,000 accounts across the globe. The vast majority of impacted accounts were not political in nature. The issue impacted 53 accounts of politicians in the U.S., representing 0.00883 percent of total affected accounts. This subset of affected accounts includes 10 accounts of Republican Members of Congress. The remainder of impacted political accounts relate to campaign activity and affected candidates across the political spectrum.

An analysis of accounts for Members of Congress that were affected by this search issue demonstrates there was no negative effect on the growth of their follower counts. To the contrary, follower counts of those Members of Congress spiked. Twitter has made this internal analysis available to the House Committee on Energy and Commerce, and we have submitted copies to this Subcommittee.

This functionality was not what we intended, and we removed this signal from our search suggestions as soon as we became aware of this issue.

It is important to note that these behavior-based algorithms are designed to reduce the visibility of abusive content, and the initial results of these behavioral filters showed a reduction in abuse reports of 8 percent from conversations and 4 percent drop in abuse reports from search results. But this technology is constantly evolving, and we know that we will continually learn and adapt to achieve the best outcome for our users. As always, we will continue to refine our approach, evaluate unintended consequences, and will be transparent about the reasons underpinning our decisions.

B. Rules Violations

Twitter takes violations of the Twitter Rules and Terms of Service seriously. We want to ensure that we police our platform in meaningful ways using automated systems, and those efforts are not always visible to the public. Additionally, we do not always share publicly the reason we take action on a particular account to protect the privacy of our users.

In the recent instance regarding the account @UnplannedMovie, the account was caught in our automated systems used to detect ban evasion. Ban evasion occurs when an individual registers for a new account despite having been suspended previously for breaking our rules. We reinstated the @UnplannedMovie account as soon as it was brought to our attention that the new account was not intended for similar violative activity. Followers of a specific account are replenished over time following reinstatement, and we are not hiding follower counts or disallowing certain people from following this account. If users searched for and followed the account during this time, it appeared as if the account was unfollowed. Individuals who followed the account during that time period were automatically restored as a follower to that account once it stabilized. Ultimately, the hashtag #unplannedmovie became a trending topic on Twitter.

In other instances, Twitter employs extensive content detection technology to identify and police harmful and abusive content embedded in various forms of media on the platform. We use PhotoDNA and hash matching technology, particularly in the context of child sexual exploitation material and terrorism. From January to June 2018, we removed 487,363 unique accounts due to violations of our rules prohibiting child sexual exploitation material, 97 percent of which were identified through our internal tools. Additionally, during the same period, we suspended 205,156 accounts as violations of our prohibitions regarding promotion of terrorism, 91 percent of which were identified internally. We do not share publicly the reasons an individual's account has been removed in most of these cases for privacy reasons and to ensure we do not interfere with a potential investigation by law enforcement.

C. Sensitive Content Controls

Some commentators have raised concerns about the limiting of specific Tweets that fall under our "sensitive" content controls. The Twitter Rules and Twitter Media Policy limit the types of content that may be shared on Twitter and describe requirements for users who choose to share potentially sensitive content on Twitter. For example, when adult content, graphic violence, or hateful imagery appears in Tweets, we may place this content behind an interstitial advising viewers to be aware that they will see sensitive media if they click through. This allows us to identify potentially sensitive content that some people may not wish to see.

Every user has the ability to mark their account as "sensitive" based on the content they share, and every user has the choice of whether they will see a warning for sensitive content or not. When an individual on Twitter has this setting enabled, people who visit a specific profile may see a message that the account may include potentially sensitive content and inquiring if the individual wants to view it. This setting enables individuals on Twitter to control their own experience and protects them from seeing sensitive content without first having made a choice to click through the warning, or to never see warnings.

D. Political Advertisements

We develop policies governing advertisements that run on Twitter that strive to balance allowing our advertisers to communicate their message with protecting individuals on the platform from potentially distressing content. Striking the right balance is particularly challenging in the realm of political advertising. We see a range of groups across the political spectrum utilize our advertising products, and all are bound by the same Ads Policies and Twitter Rules.

Some critics have raised concerns regarding Twitter's error in initially delaying a political advertisement promoted by Senator Marsha Blackburn. The advertisement ultimately ran on Twitter, the initial video was never blocked, and her original Tweet was never censored.

In the advertisement, Senator Blackburn referenced ending the "sale of baby body parts" by Planned Parenthood, and Twitter reviewers responded to user reports that the ad was inflammatory. Our team made the wrong call. Following an appeal from Senator Blackburn's media firm, we reviewed the initial decision. We relied upon additional context that there were no graphic images portrayed and that the concerning language was a very small portion of the overall advertisement. We then reversed the decision and apologized.

III. 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 every day on Twitter, we have invested heavily in building systems that organize content on Twitter to show individuals using the platform the most relevant information for that individual first. With 126 million people using Twitter each day in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance. If an individual wants to see Twitter without any algorithms applied, they have a single accessible control in order to view their timeline in reverse chronological order.

To preserve the integrity of our platform and to protect conversations on the platform from manipulation, Twitter also employs tools and technology to detect and minimize the visibility of certain types of abusive and manipulative behaviors on our platform.

A. Timeline Ranking and Filtering

For nearly a decade, the Twitter home timeline displayed Tweets from accounts an individual follows in reverse chronological order. When individuals originally opened Twitter, they saw the most recently posted Tweet first. As the volume of content on Twitter increased, individuals using the platform told us they were not always seeing useful or relevant information or were missing important Tweets. 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.

In December 2018, Twitter introduced a sparkle icon located at the top of individuals' timelines to more easily switch on and off reverse chronological timeline. As described above, the algorithms we employ are designed to help people see the most relevant Tweets. The icon now allows individuals using Twitter to easily switch to chronological order ranking of the Tweets from only those accounts they follow. This improvement allows individuals on Twitter to see how algorithms affect what they see, and enables greater transparency into the technology we use to rank Tweets.

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. 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. 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 an individual clicks on "Show more replies" or choose to see everything in his or her 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.

Some critics have described the sum of all of this work as a banning of conservative voices. Once again, we restate that this is unfounded and false. In fact, our approach of focusing on behavior is a robust defense against bias, as it requires us to define and act upon bad conduct, not any specific language or type of speech. Our purpose is to serve the conversation, not to make value judgments on personal beliefs.

IV. TWITTER'S APPLICATION OF RULES AND POLICIES

Content moderation on a global scale is a new challenge not only for our company, but also across our industry. While we continue to improve in efficiency and effectiveness in our content moderation practices, mistakes do occur. When we become aware of mistakes, we act promptly to correct them. We now offer people who use Twitter the ability to more easily file an appeal from within the Twitter app when we tell them which Tweet has broken our rules. This makes the appeal process quicker and easier for users. We also allow individuals to file a report through a web form that can be accessed at http://help.twitter.com/appeals. We also continue to improve our transparency around the actions we take, including better in-app notices where we have removed Tweets for breaking our rules. We also communicate with both the account who reports a Tweet and the account which posted it with additional detail on our actions. These steps are all a part of our continued commitment to transparency, and we will continue to better inform individuals who use Twitter on our work in these areas.

As policymakers and experts examine policies around content moderation, it is critical to note the importance of preserving Section 230 of the Communications Decency Act (CDA § 230). When it enacted CDA § 230 more than 20 years ago as part of the Telecommunications Act of 1996, Congress made the judgment that companies like Twitter that host content provided by others should have the latitude to make editorial decisions without becoming legally responsible for that content. CDA § 230 is a foundational law that has enabled American leadership in the tech sector worldwide.

It is the protection that allows us to proactively moderate content around activities such as child sexual exploitation, terrorism, voter suppression, and illicit drug sales. Without these tools, platforms would either cease to moderate content, including content that could relate to offline harm, or over-moderate content, resulting in less speech. Some lawmakers have suggested carving out CDA § 230 for political speech, and I want to be clear that any such measures could have unwanted consequences for people across the political spectrum who use Internet platforms to share their views. Eroding CDA § 230 creates risks of liability for companies that make good-faith efforts to moderate bad faith actors and could result in greater restrictions around free expression.

* * *

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 everyone. Simply put: Twitter would not be Twitter if everyone had the same viewpoints and ideology. We strive to balance the safety of Twitter users and freedom of expression every day. We are working to be more clear about our rules and transparent about our enforcement.

Thank you, and I look forward to your questions.