United States Senate Judiciary Committee

Testimony of Jack Dorsey
Chief Executive Officer
Twitter, Inc.

November 17, 2020

Chairman Graham, Ranking Member Feinstein, and Members of the Committee:

Thank you for the opportunity today to speak to the Committee and to the American people.

Twitter's purpose is to serve the public conversation. People from around the world come together on Twitter in an open and free exchange of ideas. We want to make sure conversations on Twitter are healthy and that people feel safe in expressing their points of view. We do our work with the recognition that free speech and safety are interconnected.

Today's hearing, *Breaking the News: Censorship, Suppression, and the 2020 Election*, was called, in part, as a response to enforcement decisions relating to Tweets by @NYPost on October 14, 2020, and concerns of this Committee regarding "Big Tech," censorship, and competition. The @NYPost example demonstrates the complexity of content moderation and policy enforcement decisions. The world has changed since Section 230 of the Communications Decency Act of 1996 became law, but the fundamentals of online speech that led to its passage largely remain.

Knowing that overly burdensome government regulatory schemes are not always nimble nor quick and can have unintended consequences, I encourage Congress to work with industry and civil society to build upon Section 230's foundation, whether it be through additions to Section 230, industry-wide self-regulation best practices, or a new legislative framework. By doing so, we can build an adaptable future Internet that people trust by empowering technology companies to continually make necessary changes to policies, services, and products, as well as experiment and learn, to improve their platforms and services.

Working together we can also avoid potential pitfalls. For example, completely eliminating Section 230 or prescribing reactionary government speech mandates will neither address concerns nor align with the First Amendment. Indeed, such actions could have the opposite effect, likely resulting in increased removal of speech, the proliferation of frivolous lawsuits, and severe limitations on our collective ability to address harmful content and protect people online.

Likewise, amending the law solely through carve-outs will inevitably favor large incumbents with vast resources who may willingly embrace such changes as it would leave only a small number of giant and well-funded technology companies. For innovation to thrive, we must not entrench the largest companies further.

The challenges that technology companies face on the Internet continue to change, requiring us to be agile in updating our policies and also make unprecedented investments to safeguard the public conversation. I would like to share what Twitter is doing to address your concerns and earn trust from those who use our services, which may help create a blueprint of solutions for the broader technology community. We also have taken steps to respond to an increasing demand from our consumers to provide context around misinformation – including our efforts around civic integrity and combating efforts to undermine the US 2020 election, which I will discuss today.

Building & Earning Trust

Three weeks ago, I told the Senate Committee on Commerce, Science and Transportation that I believe the best way to address our mutually-held concerns is to require the publication of moderation processes and practices, a straightforward process to appeal decisions, and best efforts around algorithmic choice, while protecting the privacy of the people who use our service. These are achievable in short order.

<u>Transparency</u>

We believe increased transparency is the foundation to promote healthy public conversation on Twitter and to earn trust. It is critical that people understand our processes and that we are transparent about what happens as a result. Content moderation rules and their potential effects, as well as the process used to enforce those rules, should be simply explained and understandable by everyone. We believe that companies like Twitter should publish their moderation process. We should be transparent about how cases are reported and reviewed, how decisions are made, and what tools are used to enforce. Publishing answers to questions like these will make our process more robust and accountable to the people we serve.

At Twitter, we use a combination of machine learning and humans to review reports and determine whether they violate the Twitter Rules. We take a behavior-first approach, meaning we look at how accounts behave before we review the content they are posting. Twitter's open nature means our enforcement actions are plainly visible to the public, even when we cannot reveal the private details of individual accounts that have violated our Rules. We have worked to build better in-app notices where we have removed Tweets for breaking our Rules. We also communicate with both the account that reports a Tweet and the account that posted it with additional detail on our actions. That said, we know we can continue to improve to further earn the trust of the people using Twitter.

We also know that an important part of transparency is acknowledging when our policies require updating because of new or unanticipated circumstances and acting quickly to make the necessary changes. The @NYPost situation is a prime example of this. In 2018, we created a policy to prevent Twitter from being used to spread hacked materials. This policy was informed by conversations with the US government about foreign state misinformation and disinformation and the use of hacked materials or materials of dubious origin being used to manipulate the electorate and influence the outcome of an election. These warnings from government partners were also repeated in advance of the 2018 US Midterm Election and 2020 US Election.

It was against this backdrop that we enforced our Hacked Materials Policy against very specific content shared by the @NYPost. Under this policy, people on Twitter were blocked from sharing certain links from the @NYPost, publicly or privately, as those specific articles contained the source materials themselves. References to the contents of the materials or discussion about the materials were not restricted under the policy. After hearing from journalists and others, we quickly updated our policy to limit its scope to only cover the removal of materials shared by hackers directly. This action, however, did not allow us to reinstate the @NYPost Tweets as we do not retroactively review enforcement actions when we update our policies. In order to address the unique facts in the @NYPost case, we determined that we should change our practices to allow for circumstances when actions on a specific account have led to a policy change. Accordingly, we updated the relevant policy, informed @NYPost, and the newspaper's account was restored. While we may have taken longer than some would have wanted to take these actions, we believe that this process and associated review have helped us create strong and more transparent policies.

Advancing Procedural Fairness

As a company, Twitter is focused on advancing the principle of procedural fairness in our decision-making across the board. We strive to give people an easy, clear way to appeal decisions we make that they think are not right. Mistakes in enforcement — made either by a human or algorithm — are inevitable, and why we strive to make appeals easier. We believe that all companies should be required to provide a straightforward process to appeal decisions. This makes certain people can let us know when we do not get it right, so that we can fix any mistakes and make our processes better in the future.

Procedural fairness at Twitter also means we ensure that all decisions are made without using political viewpoints, party affiliation, or political ideology, whether related to automatically ranking content on our service or how we develop or enforce the Twitter Rules. Our Twitter Rules are not based on ideology or a particular set of beliefs. We believe strongly in being impartial, and we strive to enforce our Twitter Rules fairly.

Algorithmic Choice

We believe that people should have choices about the key algorithms that affect their experience online. At Twitter, we want 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 to show individuals the most relevant information for that individual first. With 187 million people last quarter 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.

In December 2018, Twitter introduced an icon located at the top of everyone's timelines that allows individuals using Twitter to easily switch to a reverse chronological order ranking of the Tweets from accounts or topics they follow. This improvement gives people more control over the content they see, and it also provides greater transparency into how our algorithms affect

what they see. It is a good start. We believe this points to an exciting, market-driven approach where people can choose what algorithms filter their content so they can have the experience they want. We are inspired by the approach suggested by Dr. Stephen Wolfram, Founder and Chief Executive Officer of Wolfram Research, in his testimony before the Senate Committee on Commerce, Science and Transportation Subcommittee on Communications, Technology, Innovation, and the Internet in June 2019. Enabling people to choose algorithms created by third parties to rank and filter their content is an incredibly energizing idea that is within reach.

We also recognize that we can do even more to improve to provide greater algorithmic transparency and fair machine learning. The machine learning teams at Twitter are studying these techniques and developing a roadmap to ensure our present and algorithmic models uphold a high standard when it comes to transparency and fairness. We believe this is an important step in ensuring fairness in how we operate and we also know that it is critical that we be more transparent about our efforts in this space.

Protecting Privacy

In addition to the principles I have outlined to address content moderation issues in order to better serve consumers, it is also critical to protect the privacy of the people who use online services. We believe privacy is a fundamental human right, not a privilege. We offer a range of ways for people to control their privacy experience on Twitter, from offering pseudonymous accounts to letting people control who sees their Tweets to providing a wide array of granular privacy controls. Our privacy efforts have enabled people around the world using Twitter to protect their own data.

That same philosophy guides how we work to protect the data people share with Twitter. We empower the people who use our service to make informed decisions about the data they share with us. We believe individuals should know, and have meaningful control over, what data is being collected about them, how it is used, and when it is shared.

We believe that individuals should control the personal data that is shared with companies and provide them with the tools to help them control their data. Through the account settings on Twitter, we give people the ability to make a variety of choices about their data privacy, including limiting the data we collect, determining whether they see interest-based advertising, and controlling how we personalize their experience. In addition, we provide them with the ability to access information about advertisers that have included them in tailored audiences to serve them ads, demographic and interest data about their account from ad partners, and information Twitter has inferred about them.

Twitter's Civic Integrity Work Around the 2020 Elections

Throughout the 2020 election, we've seen record-levels of election-related conversations on Twitter. Our teams have and will continue to actively work to protect the integrity of this public conversation. We have taken a three-pronged approach to our work around the election, focusing our efforts on protecting our services through our policies, products, and partnerships. We will

produce a longer-form retrospective of all of our work around the 2020 US Election in early 2021, but here is an initial post-election assessment.

Policy Updates

In the lead up to the 2020 elections, we made significant enhancements to our policies to protect the integrity of the election. Most notably, this year, we updated our civic integrity policy to more comprehensively enforce labeling or removing of false and misleading information. The updated policy, which we not only announced publicly but also briefed the Presidential campaigns, civil society, and other stakeholders on, covers the following activities:

- False or misleading information about how to participate in an election or civic process;
- Content intended to intimidate or dissuade people from participating;
- Misrepresentation about affiliation (for ex., a candidate or political party);
- Content that causes confusion about laws and regulations of a civic process, or officials and institutions executing those civic processes;
- Disputes of claims that could undermine public confidence in the election (e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results); and
- Content that misleads about outcomes (e.g., claiming victory before results are in, inciting unlawful conduct to prevent the procedural or practical implementation of election results).

The civic integrity policy augmented and enhanced other important rules aimed at preventing interference with the election. Twitter banned all political advertising in 2019, only allowing some cause-based advertising for non-partisan civic engagement, in line with our belief that the reach of political speech should be earned, not bought. Additionally, we adopted rules prohibiting deceptively shared synthetic or manipulated media, sometimes referred to as "deep fakes," that may lead to serious offline harm; and labeling deceptive or synthetic media to provide additional context. Moreover, we have rules prohibiting platform manipulation, impersonation, hateful conduct, ban evasion, and attributed activity, among other harmful activities. We have also labeled specific government and state-media accounts from UN P-5 nation states, and plan to expand this effort in the near future.

Providing Context to Limit the Risk of Harmful Misinformation

As we noted in a blog published last week, we applied labels to add context and limit the risk of harmful election misinformation spreading without important context because the public told us they wanted us to take these steps. An initial assessment of our efforts from October 27th to November 11th has found the following:

- **Approximately 300,000 Tweets** have been labeled under our Civic Integrity Policy for content that was disputed and potentially misleading. These represent **0.2% of all US election-related Tweets** sent during this time period.
- **456** of those Tweets were also covered by a warning message and had engagement features limited (Tweets could be Quote Tweeted but not Retweeted, replied to, or liked).

- **Approximately 74%** of the people who viewed those Tweets saw them after we applied a label or warning message.
- We saw an estimated **29% decrease in Quote Tweets** of these labeled Tweets due in part to a prompt that warned people prior to sharing.

We also got ahead of potentially misleading information by showing everyone on Twitter in the US a series of pre-bunk prompts. These prompts, which were seen 389 million times, appeared in people's home timelines and in Search, and reminded people that the announcement of election results were likely to be delayed, and that voting by mail is safe and legitimate. Our efforts to safeguard the conversation on Twitter about the 2020 US elections continue unabated.

Product Changes

In the weeks leading up to and during election week, we implemented significant product changes intended to increase context and encourage more thoughtful consideration before Tweets are amplified. We are continuing to assess the impact of these product changes to fully understand the effect on the public conversation, which will help guide our work going forward, but I wanted to mention some of our findings today.

We encouraged people to add their own commentary when amplifying content by prompting Quote Tweets instead of Retweets. This change introduced some friction, and gave people an extra moment to consider why and what they were adding to the conversation. The change slowed the spread of misleading information by virtue of an overall reduction in the amount of sharing on the service. We observed a 23% decrease in Retweets and a 26% increase in Quote Tweets, but on a net basis the overall number of Retweets and Quote Tweets combined decreased by 20%.

In addition, we stopped providing "liked by" and "followed by" Tweet recommendations from accounts you don't follow in the Home Timeline and through notifications. While we had initially hoped that this would reduce the potential for misleading information to spread on our service, we did not observe a statistically significant difference in misinformation prevalence as a result of this change (nor any meaningful reduction in abuse reports). Instead, we found that pausing these recommendations prevented many people from discovering new conversations and accounts to follow, and we have since reverted the change.

Partnerships

A core part of our civic integrity efforts included partnerships that allowed us to share information, gather input from experts, and better gain context on how misinformation was being spread and impacting the public conversation. These partnerships included leaders in civic tech, our peers, federal, state, and local governments organizations (e.g., National Association of Secretaries of State, National Association of State Election Directors, Department of Homeland Security, Federal Bureau of Investigation, Department of Justice, Office of the Director of National Intelligence, and elections officials across the country), news organizations, and civil society, among others.

****

We want to be very clear that we do not see our job in this space as done. Our work here continues and our teams are learning and improving how we address these challenges and earn the trust of the people who use Twitter. I look forward to continuing to work with you on solutions and building the guideposts for the future Internet. Thank you for the opportunity to appear today.