

WRITTEN TESTIMONY OF TWITTER CEO JACK DORSEY (@JACK)
U.S. HOUSE COMMITTEE ON ENERGY & COMMERCE
MARCH 25, 2021

Twitter’s purpose is to serve the public conversation. While much has changed in the world since we started fifteen years ago, we believe our mission is more important than ever.

Every day Twitter grapples with complex considerations on how to address extremism and misinformation. How do we prevent harm, while also safeguarding free expression and the right of diverse individuals to express a range of views? How do we develop policies that can be built at scale and adapt rapidly, especially given diverse regulatory models around the world? What role should our company play in determining these pivotal questions? What information should we rely on when making decisions? How do we earn the trust of those who use our service?

These are even harder questions in an increasingly polarized world, which has consequently heightened concerns about information sources. Quite simply, a trust deficit has been building over the last several years, and it has created uncertainty — here in the United States and globally. That deficit does not just impact the companies sitting at the table today but exists across the information ecosystem and, indeed, across many of our institutions.

This Committee has expressed interest in what we are doing to combat “falsehoods about the COVID-19 vaccine” and “debunked claims of election fraud.” We have [COVID-19](#) and [vaccine misinformation](#) policies, as well as a [COVID information hub](#). Our [civic integrity](#) and [platform manipulation](#) policies are available on our [Help Center](#), along with information on our bans on [state-controlled media](#) advertising and [political advertising](#). As a follow-up to our [preliminary post-election update](#), we are conducting a review of the 2020 U.S. election, the findings of which we intend to share.

Our efforts to combat misinformation, however, must be linked to earning trust. Without trust, we know the public will continue to question our enforcement actions. I believe we can earn trust by focusing on the following: enhancing transparency, ensuring procedural fairness, enabling algorithmic choice, and strengthening privacy.

Building & Earning Trust

Every day, millions of people around the world Tweet hundreds of millions of Tweets, with one set of rules that applies to everyone and every Tweet. We strive to implement policies impartially and at scale. We built our policies primarily around the promotion and protection of three fundamental human rights — freedom of expression, safety, and privacy.

At times, these rights can conflict with one another. As we develop, implement, and enforce our policies, we must balance these rights. Additionally, our policies must be adaptable to changes in behavior and evolving circumstances. This is why we must be transparent, embrace procedural fairness and choice, and protect privacy.

Transparency

While Twitter has made significant progress with respect to transparency, we know that we can do more to strengthen our efforts. People who use our service should understand our processes — how potential violations of our rules are reported and reviewed, how content-related decisions are made, and what tools are used to enforce these decisions. Publishing answers to questions

like these will continue to make our internal processes both more robust and more accountable to the people we serve.

Twitter's open nature means our enforcement actions are plainly visible to the public, even when we cannot reveal the private details behind individual accounts that have violated our rules. We use a combination of machine learning and human review to assess potential violations of the Twitter Rules. We take a behavior-first approach, meaning we look at how accounts behave before we review the content they are posting. If an account owner breaks our Rules and may be required to delete a Tweet, we have worked to build better in-app notices to communicate with both the account that reports a Tweet and the account that posted it with additional information about our actions. In January, we published our biannual update to the [Twitter Transparency Center](#), with additional data about actions we have taken to disrupt state-backed information operations, to enforce our COVID-19 policy, and take action on Tweets that violate our Rules.

In addition to ensuring transparency around our decisions, we are seeking ways to enhance transparency around how we develop our content moderation policies. In recent months, for example, there have been increased questions about how we should address policy violations from world leaders. As a result, we are currently re-examining our approach to world leaders and are soliciting [feedback](#) from the public. Our feedback period is currently open and our survey will be available in more than a dozen languages to ensure a global perspective is reflected.

Procedural Fairness (Accountability & Reliability)

Twitter is focused on advancing procedural fairness in our decision-making. 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 an automated system — are inevitable and why we strive to make appeals easier. We believe that all companies should be required to provide those who use their service with straightforward processes to appeal decisions that impact them.

Algorithmic Choice

We believe that people should have transparency or meaningful control over the algorithms that affect them. We recognize that we can do more to provide algorithmic transparency, fair machine learning, and controls that empower people. The machine learning teams at Twitter are studying techniques and developing a roadmap to ensure our present and future algorithmic models uphold a high standard when it comes to transparency and fairness.

We also provide people control over algorithms that affect their core experience on Twitter. We have invested heavily in building systems that organize content to show individuals relevant information that improves their experience. With 192 million people last quarter using Twitter daily in dozens of languages and countless cultural contexts, we rely upon machine learning algorithms to help us organize content by relevance to provide a better experience for the people who use our service.

Privacy

We have always believed that privacy is a fundamental human right. We believe that individuals should understand the personal data that is shared with companies and have the tools to help

them control their information. To help people better understand their options, we have created the [Twitter Privacy Center](#), which acts as a hub for information about our privacy and data protection work.

We are constantly working to improve the controls people have to manage their personal data and experience on Twitter. In addition, we continue to support efforts to pass strong federal privacy legislation to safeguard important privacy rights.

Innovations to Address Misinformation

We also recognize that addressing harms associated with misinformation requires innovative solutions. Content moderation in isolation is not scalable, and simply removing content fails to meet the challenges of the modern Internet. This is why we are investing in two experiments — [Birdwatch](#) and [Bluesky](#). Both are aimed at improving our efforts to counter harmful misinformation.

Birdwatch

In January, we launched the “Birdwatch” pilot, a community-based approach to misinformation. Birdwatch is expected to broaden the range of voices involved in tackling misinformation, and streamline the real-time feedback people already add to Tweets. We hope that engaging diverse communities here will help address current deficits in trust for all. More information on Birdwatch can be found [here](#). We expect data related to Birdwatch will be publicly available at [Birdwatch Guide](#), including the algorithm codes that power it.

Bluesky

Twitter is also funding Bluesky, an independent team of open source architects, engineers, and designers, to develop open and decentralized standards for social media. This team has already created an initial review of the ecosystem around protocols for social media to aid this effort. Bluesky will eventually allow Twitter and other companies to contribute to and access open recommendation algorithms that promote healthy conversation and ultimately provide individuals greater choice. These standards will support innovation, making it easier for startups to address issues like abuse and hate speech at a lower cost. Since these standards will be open and transparent, our hope is that they will contribute to greater trust on the part of the individuals who use our service. This effort is emergent, complex, and unprecedented, and therefore it will take time. However, we are excited by its potential and will continue to provide the necessary exploratory resources to push this project forward.

Conclusion

As we look to the future, I agree with this Committee that technology companies have work to do to earn trust from those who use our services. For Twitter, that means tackling transparency, procedural fairness, algorithmic choice, and privacy. I think that this approach will be a growing trend across all companies and organizations, both big and small. I look forward to your questions.