Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

29 September 2020

Dear Lord Evans,

Thank you for your letter, and for giving us the opportunity to update the Committee since our last letter, provided in December 2019. Whilst there will always be more to do, we continue to make considerable progress and strive to facilitate healthy dialogue on the service, empower individuals to express diverse opinions and beliefs, and prohibit behaviour that intimidates, harasses, or is otherwise intended to shame or degrade others.

Transparency

Building on our previous Transparency Reports, last month we launched the Twitter Transparency Center. This covers a broader array of our transparency efforts, including sections covering information requests, removal requests, copyright notices, trademark notices, email security, Twitter Rules enforcement, platform manipulation, and state-backed information operations. Highlights from the latest reporting period (July - December 2019) show:

- There was a 95% increase in the number of accounts actioned for violations of our abuse policy during this reporting period - the largest increase in the number of accounts actioned under these policies in a single reporting period.
- Our Hateful Conduct policy expanded to include a new dehumanization policy on July 9, 2019. Since this update, this reporting period saw a 54% increase in the number of accounts actioned for violations.

Critically, transparency is at our core - Twitter is the only major service that makes public conversations available for study, and researchers have used our public API (technical interface

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

to allow access to our public data) over the past decade to improve our collective understanding of a wide range of topics - including on online harms.

As one example, we have shared the HateLab case study on our website, where researchers from Cardiff University created a Dashboard to provide a real-time look at hate speech online. Professor Matthew Williams, HateLab's Principal investigator, noted: *"One positive trend we see is that there's significantly more healthy conversations going on than toxic ones, but the toxic conversations get more press. We want to make the positive conversations more effective."*

User options

More recently, we have also introduced a number of product changes and experiments. Critically, on 11 August 2020, we made conversation controls available to all users following a trial in the spring. Before you Tweet, you can now choose who can reply with three options: 1) everyone (standard Twitter, and the default setting), 2) only people you follow, or 3) only people you mention (and, therefore, an option of turning replies off in full). Tweets with the latter two settings will be labeled and the reply icon will be grayed out for people who can't reply. Our trial identified that people who face abuse find these settings helpful - those who have submitted abuse reports are three times more likely to use these settings.

Concurrently, we have also been running an experiment (presently available to a limited number of users) with a prompt that gives you the option to revise your reply before it's published if it uses language that could be harmful. On 10 August 2020, we announced updates to the experiment following feedback from users, with prompts now including more information on why you received it.

Supporting people in public life

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

We are acutely aware that many high profile users can, at times, be particularly vulnerable to abuse and harassment. In February 2020, we created a series of videos with high-profile influencers in the UK to talk about the experiences they have had and safety tools they have used. Garnering 1.8 million impressions - and with an average view through rate of 57% - we saw that this was a way to use engaging and authentic content to reach a wider range of users on the service.

In February, we were also pleased to host London Women Stand with Glitch, Change.org, and The Parliament Project, a full-day workshop in central London to support women with an interest in entering politics. Women of all ages, backgrounds and political affiliations came together for a day full of inspiration, information and motivation, to explore their political purpose, community activism and develop confidence in using your voice online and offline. We provided workshops on how to become an elected politician, online safety and digital self care, and how to campaign effectively. Information about London Women Stand is available here.

Partnerships

In my previous letter, I outlined our approach to supporting politicians in the UK**.** With UK Parliament specifically, we have a partnership with the Members' Security Support Service - where they access a specific portal (our Partner Support Portal) to make a wide range of reports that go directly to a dedicated team within Twitter. That process is supplemented by a weekly phone meeting between Twitter and the Parliamentary Security Department. It is a valuable opportunity to ask any questions, unblock any issues and flag upcoming events, where I can, for instance, subsequently let our Safety Team know we may see higher levels of reports.

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

We have been pleased to continue our partnership in 2020, and identify opportunities for further collaboration. Prior to the lockdown, for instance, we participated with peer companies in a safety drop-in session in Parliament for Members and their staff; similarly over the summer, we worked together to promote the new Conversations Control tool among Parliamentarians.

Similarly, we have also maintained our collaboration with the Web and Social Media Team in the Scottish Parliament. Having delivered a series of Twitter safety training sessions with MSPs (with my counterpart at Facebook) in 2019, we are currently planning to deliver a follow-up safety session virtually - and to provide safety training as part of the formal induction process for new MSPs following the election next year.

2019 General Election

Ahead of the 2019 UK General Election, we established a cross-functional team to proactively protect the integrity of the election-related conversation on our platform and maximise Twitter as the go-to place for people to see what's happening, participate in the conversation, and track the campaign trail. As part of our election integrity efforts, we effectivley worked to support partner escalations and to identify potential threats from malicious actors.

At the beginning of the election campaign, we worked with our counterparts at Facebook and Google to create a 'one-stop shop' advice centre for candidates on the Internet Association website. This was in addition to sharing our bespoke General Election and safety resources with all MPs, the main political parties, DCMS, NCSC, the Electoral Commission and the Home Office, who we also worked with to share information to all 43 police forces

More broadly, we now have roughly 40 UK partners on our Partner Support Portal, able to escalate a wide range of issues to a dedicated team within Twitter - ranging from reports of

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

abuse to support queries. Over the election campaign, we received 497 reports from partners on the Partner Support Portal, with an action rate of 93.3%.

Context for users

An important part of our work to serve the public conversation is providing people with context so they can make informed decisions about what they see and how they engage on Twitter. Twitter provides an unmatched way to connect with, and directly speak to public officials and representatives. This direct line of communication with leaders and officials has helped to democratise political discourse and increase transparency and accountability.

In 2019, we took steps to protect that discourse because we believe political reach should be earned not bought, by banning all state-backed media advertising and political advertising. In addition, this summer we expanded the types of political accounts we label. We now add new labels to the following categories of Twitter accounts:

1. Accounts of key government officials, including foreign ministers, institutional entities, ambassadors, official spokespeople, and key diplomatic leaders. At this time, our focus is on senior officials and entities who are the official voice of the state abroad;
2. Accounts belonging to state-affiliated media entities, their editors-in-chief, and/or their senior staff.

Twitter provides an important space for people to come to hear directly from elected officials and candidates for office, to find breaking news, and increasingly, for information on when and how to vote in elections. We will continue to focus on building tools that better enable people to find quality news and have informative conversations on Twitter.

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

Civic integrity

As more people seek ways to vote and express their fundamental civil rights safely during the COVID-19 pandemic, the need for this type of information has only grown. Our existing Civic Integrity Policy targets the most directly harmful types of content, namely those related to:

1. Information or false claims on how to participate in civic processes
2. Content that could intimidate or suppress participation
3. False affiliation

However, in recognition of the changing circumstances of how people will vote in and beyond 2020, and in line with our commitment to protecting the integrity of the election conversation, we have now expanded this framework. The goal is to further protect against content that could suppress voting and help stop the spread of harmful misinformation that could compromise the integrity of an election or other civic process. People who use our service have told us that non-specific, disputed information that could cause confusion about an election should be presented with more context. Therefore, false or misleading information intended to undermine public confidence in an election or other civic process will be removed or labelled. This includes, but is not limited to:

1. False or misleading information that causes confusion about the laws and regulations of a civic process, or officials and institutions executing those civic processes.
2. Disputed claims that could undermine faith in the process itself, e.g. unverified information about election rigging, ballot tampering, vote tallying, or certification of election results.
3. Misleading claims about the results or outcome of a civic process which calls for or could lead to interference with the implementation of the results of the process, e.g. claiming

Twitter UK
20 Air Street
London
W1B 5AN

twitter.com

victory before election results have been certified, inciting unlawful conduct to prevent a peaceful transfer of power or orderly succession.

In line with our existing enforcement approach, Tweets that are labeled under this expanded policy will have reduced visibility across the service, meaning that we will not amplify the Tweets on a number of surfaces across Twitter.

Going forwards

We continue to welcome opportunities to work with government and civil society. We would be happy to answer any further questions if you would like to arrange a meeting.

Yours sincerely,

Katy Minshall
Head of UK Government, Public Policy and Philanthropy