Written Representation 153

Name: Twitter Inc

Received: 7 Mar 2018

Chairman Charles Chong, and Distinguished Members of the Select Committee:

Twitter appreciates the importance of the Select Committee's work on "Deliberate Online Falsehoods - Causes, Consequences, and Countermeasures", and we thank the Committee for the opportunity to provide a written submission. We welcome this inclusive engagement, and we look forward to continuing our deep involvement.

Our written submission stands together with a separate industry submission from the Asia Internet Coalition (AIC), co-authored by Twitter and our peers. We have previously expressed our views on information quality issues in similar legislative contexts elsewhere internationally and on our corporate blogs[1], which are in the public domain. To complement these as well as the AIC submission to the Select Committee, we have structured this submission as follows:

- First, we provide an overview of Twitter, our core values, and the value we strive to bring to the world;
- Second, we summarise progress made to safeguard Twitter from malicious manipulation, enforce our Rules, and to improve the quality of information on Twitter; and
- Lastly, we share some of the ways stakeholders can work together to address these issues.

Overview of Twitter and our Values

Twitter currently has 330 million monthly active users around the world, more than a million of whom are located in Singapore. The volume of activity on Twitter globally is enormous: our users generate hundreds of millions of Tweets per day.

---

[1] https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html; https://blog.twitter.com/official/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html; and
https://blog.twitter.com/official/en_us/topics/company/2018/2016-election-update.html

In Singapore, people look to Twitter every day to see what's happening. Our users engage with entertainment and sports influencers, such as actor Aaron Aziz (@AaronDwiAziz), and pop stars JJ Lin (@JJ_Lin) and Stefanie Sun (@Stefsunyanzi). They engage with brands as Singapore Airlines (@singaporeair) and business leaders such as Tan Min-Liang (@minliangtan). Politics and news are important verticals for Singaporean users to obtain accurate and timely information, be it directly from PM Lee Hsien Loong (@leehsienloong) and the Singapore government (@govsingapore), or from Channel NewsAsia (@ChannelNewsAsia) and the Straits Times (@STcom). Twitter's value in breaking critical information and news, especially during times of crises and unforeseen disruptions to infrastructure, has also been leveraged by SMRT Corporation (@SMRT_Singapore) and the National Environment Agency (@NEAsg), both of which have a significant reach in Singapore.

Fundamentally, Twitter is committed to providing a service that keeps people informed in real time about what's happening in the world. As such, we care deeply about deliberate online falsehoods and their potentially harmful effect on civic and political discourse that is core to our mission. We are committing to help increase the health, openness and civility of public conversations on Twitter and to hold ourselves publicly accountable towards progress [2].

The term "fake news" has become a catchphrase used to describe everything from manufactured news stories to incorrect information, from state-supported propaganda to news or points of view people don't like. The scholar Claire Wardle of First Draft, a non-profit focused on media literacy and improving information quality online, has proposed a typology[3] for misinformation (inadvertent sharing of false information) and disinformation (deliberate creation and sharing of false information). Quality, peer-reviewed research like this illustrates the need for society to better understand the new information quality challenges we are all confronting, and the different responses required.

In the face of these complexities, Twitter's open and real-time nature is a powerful antidote to the spreading of all types of false information. This is important because we cannot distinguish whether every single Tweet from every person is truthful or not. No single company, governmental or non-governmental actor, should be the arbiter of truth. Instead, we see journalists, experts and engaged citizens Tweeting side-by-side to affirm, correct, and challenge public discourse in seconds; Twitter is also a place for journalists to find eyewitness accounts of events and verify them - even when they are not on the ground. These vital interactions happen on Twitter every day.

<u>Twitter in times of emergency and exigent circumstances</u>

During emergencies and times of crises, Twitter's open and real-time features become critical tools for our users, governments, and nonprofits to share and exchange information. This is especially since our users turn to Twitter to find information that is accurate, credible, and most importantly, timely. We want to ensure factual information is elevated across our

---

[2] https://twitter.com/jack/status/969234275420655616
[3] https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79

platform in times of crisis. We thus work with key stakeholders throughout the year, conducting outreach to government agencies and nonprofits, verifying authentic accounts, as well as training for capacity-building and to improve implementation of best practices.

Twitter has also put in place systems to work with law enforcement to deal with exigent circumstances, while keeping our commitment to our users around the world. In line with our privacy policy, we may disclose account information to law enforcement in response to a valid emergency disclosure request. We evaluate emergency disclosure requests on a case-by-case basis in compliance with relevant law. If we receive information that provides us with a good faith belief that there is an exigent situation involving the danger of death or serious physical injury to a person, we may provide the information necessary to prevent that harm, if we have it.

As part of our commitment to transparency, we have since 2012 also published a biannual Transparency Report that provides trends and insights into the legal requests we receive in the aggregate, on a country and thematic basis, including both emergency and non-emergency requests.

Automation and Twitter's Information Quality Initiative

One area that merits further understanding is the **dissemination techniques** being used to spread misleading content online, such as malicious automation. This is where companies, like Twitter, are taking active steps to stop malicious accounts and spammy content from spreading.

Twitter is continuing its effort to detect and prevent malicious automation by leveraging our technological capabilities and investing in initiatives aimed at understanding and addressing behavioral patterns associated with such accounts. For example, in early 2017, we launched the Information Quality initiative, an effort aimed at enhancing the strategies we use to detect and stop malicious automation, improve machine learning to spot spam, and increase the precision of our tools designed to prevent such content from contaminating our platform.

Automation – i.e., bots – is not prohibited on Twitter; in fact, it often serves a useful and important purpose. Automation is essential for certain informational content, particularly when time is of the essence, including for law enforcement or public safety notifications, e.g. earthquake and typhoon warnings. Automation is also used to provide customer service for a range of companies online, including on Twitter.

Automation can, however, also be used for malicious purposes, most notably in generating spam - unwanted content consisting of multiple postings either from the same account or from multiple coordinated accounts. Twitter's Rules take an expansive view of spam because it negatively impacts the experience of providing useful, timely, and appropriate information.

Our efforts under the Information Quality initiative have produced clear results in terms of our ability to detect and block such content. We now detect and block approximately 523,000

suspicious logins daily for being generated through malicious automation. In December 2017, our systems identified and challenged more than 6.4 million suspicious accounts globally per week - a 60% increase in our detection rate from October 2017.

We developed new techniques for identifying malicious automation (such as near-instantaneous replies to Tweets, non-random Tweet timing, and coordinated engagement). We improved our phone verification process and introduced new challenges, including reCAPTCHAs to help validate that a human is in control of an account. We enhanced our capabilities to link accounts that were formed by the same person or that are working in concert. And we are improving how we detect when accounts may have been hacked or compromised.

This year, we plan to build upon our 2017 improvements, investing further in machine-learning capabilities that help us detect and mitigate the effect on users of fake, coordinated, and automated account activity.

This emphasis on machine-learning capabilities is not new to Twitter, and we have been at the forefront in addressing the evolving challenge of terrorist exploitation online. In early 2016, the technological tools we had at our disposal detected approximately one third of terrorism-related accounts that were removed. In 2017, this rose to 95% of terrorism accounts being suspended as a result of our enhanced detection tools, with 75% of those suspended prior to sending their first Tweet.

Supporting the Information Ecosystem

While enforcing the Twitter Rules and addressing malicious automation and attempts to game the Twitter service to propagate disinformation is a response, it isn't the long-term solution. The long-term solution must also include the active involvement of governments, civil society, and NGOs in addressing media literacy as well. The information ecosystem of the future will also depend upon informed and media-savvy consumers. And we must prepare the next generation to thrive in such an environment.

Globally, our partners Common Sense Media, the National Association for Media Literacy, the Family Online Safety Institute and Connect Safely, amongst others, have helped us to craft materials and conduct workshops to help our users learn how to process online information and discern between sources of news. We focus on elements like verification of sources, critical thinking, active citizenship online, and the breaking down of digital divides.

In Singapore, we have, with the help of our partners, built an excellent track record in supporting digital literacy across society. This February on Safer Internet Day, Twitter co-launched the Media Literacy Council's 2018 'Better Internet Campaign'. We will continue to partner throughout the year, hosting and conducting workshops for parents together with others like the National Library Board and MENDAKI. This outreach to parents complements our ongoing work to build up discernment in Singaporean youth, through the National Youth Council, the National Council of Social Services, as well as educational institutions such as the Institute of Technical Education. From Singapore and across the region, we have also

supported non-profits such as the Digital Quotient Institute, in amplifying and supporting the publication of their digital literacy educational materials.

Twitter also partners with journalistic NGOs for training and outreach initiatives, including Reporters without Borders, the Committee to Protect Journalists, and the Reporters Committee for Freedom of the Press. We will continue to work with reporters and media organizations to ensure that Twitter's real-time capacity for dispelling untruths is built into the approach of newsrooms and established media outlets.

We also recognise that government organisations can be important sources of credible and authoritative information, especially in times of crises and emergency. We frequently conduct training and outreach initiatives in Singapore, to help ensure that government organisations can utilise Twitter effectively, so that they can add value to the discourse online and promote continuous progress in civic engagement and democratic accountability.

Conclusion

We hope this written submission, together with the industry submission by the Asia Internet Coalition, will be useful inputs to the Select Committee's work. Twitter is committed to working with you, the Singapore government, our industry partners, and other stakeholders to ensure that we have a better understanding of the issues at stake, and can find the best way to approach this together. Working with the broader community, we will continue to test, to learn, to share, and to improve, so that our platform remains effective and safe.

*   *   *