Twitter—written evidence (FEO0010)

Communications and Digital Committee
Inquiry into Freedom of expression online

Our rules

Twitter is reflective of real conversations happening in the world and that sometimes includes perspectives that may be offensive or controversial to others. While we welcome everyone to express themselves on our service, we will not tolerate behaviour that harasses, threatens, or uses fear to silence the voices of others. Our Twitter Rules are in place to help ensure everyone feels safe expressing their beliefs and we strive to enforce them with uniform consistency.

We empower people to understand different sides of an issue and encourage dissenting opinions and viewpoints to be discussed openly. This approach allows many forms of speech to exist on our service and, in particular, promotes counter-speech: speech that presents facts to correct misstatements or misperceptions, points out hypocrisy or contradictions, warns of offline or online consequences, denounces hateful or dangerous speech, or helps change minds and disarm.

Context matters. When determining whether to take enforcement action, we may consider a number of factors, including (but not limited to) whether:

- the behaviour is directed at an individual, group, or protected category of people;
- the report has been filed by the target of the abuse or a bystander;
- the user has a history of violating our policies;
- the severity of the violation;
- the content may be a topic of legitimate public interest.

Violators can appeal permanent suspensions if they believe we made an error. They can do this through the platform interface or by filing a report. Upon appeal, if we find that a suspension is valid, we respond to the appeal with information on the policy that the account has violated.

Transparency

Transparency is foundational to the kind of Internet that we all want to see - empowering consumers, building trust and strengthening democracies. Twitter is the only major service to make public conversation data available via an API for the purposes of study. Making this type of data available to researchers has resulted in a number of important benefits and, in November, we worked with Demos[1] to highlight why greater data transparency should be at the centre of our societal response to online harms, such as information operations. In October, we similarly highlighted how data transparency had enabled the government's Anti Muslim Hatred Working Group[2] to conduct critical research. Our work to increase transparency efforts across the company is tireless and constant, and we welcome opportunities to support publicly available data being used to advance research objectives

---

[1]   https://blog.twitter.com/en_us/topics/company/2020/nation-states-exerting-power-online-sharing-data-can-guard-again.html
[2]   https://blog.twitter.com/en_gb/topics/company/2020/twitteruk-amhwguk-working-partnership.html

on a wide range of topics and in a safe, compliant way with the public's basic expectation of privacy.

Part of our transparency efforts also include our biannual Twitter Transparency Report, which we've produced since July 2012 to share global trends across a number of areas of our enforcement on Twitter, including the Twitter Rules and legal requests we receive.

The report is ever-evolving. In August 2020, for example, we launched the Twitter Transparency Centre.[3] Our goal with this evolution is make our transparency reporting more easily understood and accessible to the general public. This includes:

- A new website that includes all our disclosed data in one place;
- Data visualizations making it easier to compare trends over time;
- Country comparison module;
- Tooltips to help explain key terms and provide more insights on the terms we use;
- History of transparency milestones and updates;
- New metrics and methodology on the enforcement of the Twitter Rules (from July 2018 through December 2019);
- New policy categories to better align with the Twitter Rules.

We have also started to include state-backed information operations datasets, which were first released to the public in 2018 to empower research and awareness of these campaigns. We now host over 35 different datasets that we believe are connected to state-backed information operations, and hosted our first research workshop on the data with the Carnegie Partnership for Countering Influence Operations in July 2020.

We remain deeply committed to transparency at Twitter, and look forward to continuing to work with government, civil society and the wider community on these important issues.

Privacy

Twitter was founded upon a set of core values that continues to define how the company develops. Among those values are defending and respecting the user's voice - and a commitment to both freedom of expression and user privacy. We give all users transparency and meaningful controls over the data we collect, how it is used, and when it is shared.

It is important to note we are a service that, in pursuit of protecting user privacy, engages in data minimisation. Our approach has been developed in consultation with leading NGOs working on the frontlines of these issues worldwide. Some of Twitter's most vulnerable users do not use their own names out of fear, or a need to maintain some degree of pseudonymity. This includes cases such as journalists in dangerous environments, whistleblowers who need anonymity to do their vital work freely and safely, and those practicing religions in countries that do not tolerate their faith. In democratic societies, pseudonymity may be used to highlight issues faced by vulnerable communities, victims of crimes, or where there is a risk of personal retribution for speaking out. Indeed, many of the first voices to speak out in what has become a wider scandal have done so behind some degree of pseudonymity.

Balancing both a diverse range of perspectives, while protecting users and the integrity of our site, requires vigilance and poses challenges. Indeed, while Twitter does not require people to use their real name, a number of processes are in place to ensure this is not

---

[3]   https://transparency.twitter.com/

abused. For example, our fake accounts policy forbids people from using misleading account information in order to engage in spamming, abusive, or disruptive behavior, including attempts to manipulate the conversations on Twitter. Similarly, our Impersonation Policy provides that a user cannot impersonate individuals, groups, or organizations in a manner that is intended to or does mislead, confuse, or deceive others.

Platform design

Much of our work now is focused on how we can ensure the design of Twitter encourages healthy conversation, and how we can use technology to reduce the burden on victims of abuse.

In 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. Tweets with the latter two settings will be labelled and the reply icon will be grayed out for people who cannot 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. In August 2020, we announced updates to the experiment following feedback from users, with prompts now including more info on why you received it.

More than 50% of abusive content that is enforced is now surfaced proactively to our teams for review, instead of relying on reports from people on Twitter.

Regulation

Protecting the Open Internet is a key objective for Twitter. There are two key areas to consider:

1. Avoid entrenching the dominance of the biggest players by protecting competition;
2. Focus on how content is discovered and amplified, and less on removal alone.

By delivering unprecedented economic and technological progress, the Open Internet has led to wider access to information and opportunities to speak that are core to participatory, democratic societies.

With repressive regimes around the world using shutdowns, throttling, and draconian legal frameworks to stifle online freedoms, the norms that inform new legislation have never been more important.

If legislation serves to cement the position of larger companies, it will harm the Open Internet, innovation, and consumer choice irreparably. In sum, robust competition and guaranteeing a fair playing field are essential.

Legislative frameworks should consider impacts to competition and innovation. GDPR, for example, has had positive effects on consumer privacy but a Tilburg University study found it had adverse effects on competition[4] by strengthening large companies. If regulation

unconsciously creates a framework where big companies continue to consolidate and expand market power, there's risk of undermining the foundational principles of the open, participatory, and democratic Internet.

The debate is too often framed through the prism of content removal alone. Without question, some content – including terrorist content and child sexual abuse material – must be removed expeditiously. But removing content alone cannot be the sole paradigm of Internet policy.

Government, industry, and NGOs must consider our shared responsibility in offering the public more context, de-amplifying certain types of content, and ensuring the reach of political speech is earned and not bought or manipulated. In order to protect open, civic-minded societies, we need a dynamic, sophisticated and nuanced approach to content moderation.

We believe regulation can be rights-based. And we believe it should hold corporate power to account, while rebuking authoritarian models of Internet governance, and declaring to the world that fundamental online freedoms and competition will be protected.

We welcome the government's stated commitment to Freedom of Expression in the Full Response to the Online Harms White Paper. We look forward to continuing to engage with government and civil society to ensure this is fulfilled as the regulatory proposals are developed further.

Access to the free and open Internet

As we continue to face the global coronavirus pandemic, one of the elemental aspects powering our shared resolve is the connected nature of our world, connectivity that is driven by the Internet. However, a recent report from Access Now[5] shows an increase in Internet shutdowns globally, with 213 documented shutdowns in 2019 and a concerning trend toward more targeted and prolonged shutdowns.

Sometimes known as (Internet) 'blackouts' or 'kill switches', shutdowns can be highly targeted, focused on specific groups of individuals for the purposes of repressing dissent, to larger shutdowns which can deprive entire communities of essential information and isolate people from their support networks. The deliberate throttling of Internet speeds - increasingly a tool of repression - creates friction and barriers to entry for those who need it most.

To help raise awareness of this important issue, we have supported the #KeepItOn campaign. #KeepItOn is a coalition of more than 200 organisations - ranging from research centers to rights and local advocacy groups, detection networks, and media organizations - located within 75 countries around the world, fighting to end Internet shutdowns globally.

Having access to the free and open Internet is a right that many have come to see as essential. It is a right that is increasingly inseparable from free expression, self-determination, and self-actualization.

Our right to access news, connection, and medical information on the open Internet is more vital than ever during this pandemic. For some, it may be existential. It acts as a

---

[4]   https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3598130
[5]   https://www.accessnow.org/cms/assets/uploads/2020/02/KeepItOn-2019-report-1.pdf

distribution channel for the most up-to-date information on how to prevent the spread of the virus; a directory that points us to the services we need to stay healthy; and a place where we can gather with friends and family. And with our economies reliant on digital technologies and connectivity, an open Internet is a foundational pillar on which recovery must be built.

Shutdowns can only impede this progress, unnecessarily walling off targeted communities and entire populations. We have also seen a worrying trend towards regulatory action that could see instances of shutdowns increase further. The long-term result could be an Internet that is less open, less free, and less empowering for all.

8 January 2021