

Written evidence submitted by Twitter [MISS0070]

Dear Chair,

Thank you for inviting us to provide a written submission in response to your inquiry on body image. We welcome the opportunity to provide an update and respond to the questions asked.

Overview of Twitter

Twitter is a public, real-time global micro-blogging service, where people can see every side of a topic, discover news, share their perspectives, and engage in conversation. Twitter is available in more than 40 languages around the world, and can be accessed via [Twitter.com](https://twitter.com), and an array of mobile devices via Twitter owned and operated mobile applications (e.g. Twitter for iPhone and Twitter for Android). From breaking news and entertainment, to sports, politics, and everyday interests, Twitter users provide insight into every angle of a story. On Twitter, users can join the open conversation, including with photos, video clips, and live-streaming. As of Q3 2020, we have 187 million global mDAU (Daily Active Users).

Our Rules

The [Twitter Rules](#) prohibit users from engaging in the targeted harassment of someone, or inciting other people to do so. We consider abusive behavior an attempt to harass, intimidate, or silence someone else's voice. In order to facilitate healthy dialogue on the platform, and empower individuals to express diverse opinions and beliefs, we prohibit behavior that harasses or intimidates, or is otherwise intended to shame or degrade others. In addition to posing risks to people's safety, abusive behavior may also lead to physical and emotional hardship for those affected.

When determining the penalty for violating this policy, we consider a number of factors including, but not limited to, the severity of the violation and an individual's previous record of rule violations. For example, we may ask someone to remove the violating content and serve a period of time in read-only mode before they can Tweet again. Subsequent violations will lead to

Written evidence submitted by Twitter [MISS0070]

longer read-only periods and may eventually result in permanent suspension. If an account is engaging primarily in abusive behavior, we may permanently suspend the account upon initial review.

We also have rules regarding altered media on Twitter. Our [policy](#) states that you may not deceptively promote synthetic or manipulated media that are likely to cause harm. In addition, we may label Tweets containing synthetic and manipulated media to help people understand their authenticity and to provide additional context.

Twitter is constantly reviewing its rules and policies given the always-evolving nature of the content on the service. The [Twitter Trust and Safety Council](#) brings together independent expert organisations from around the world that advocate for safety and advise us as we develop our products, programs, and the Twitter Rules. At the end of 2019, we expanded the Council to include experts from even more countries and diverse perspectives.

The Council is composed of several advisory groups, each focused on issues that are critical to the health of the public conversation. We also set up temporary groups on specific topics. The groups are: Online Safety and Harassment, Human and Digital Rights, Child Sexual Exploitation, Suicide Prevention and Mental Health, and Dehumanization.

We look forward to reviewing the Committee's findings on its inquiry into body image.

Advertising

Advertisers on Twitter are responsible for their [Twitter Ads](#). This means following all applicable laws and regulations, creating honest ads, and advertising safely and respectfully.

Our [Unacceptable Business Practices Policy](#) prohibits the promotion of unacceptable business practices globally, including misleading, false, or unsubstantiated claims. In addition, we prohibit a range of products and services being promoted to minors, including dietary supplement products, under our [Prohibited Content for Minors Policy](#).

Twitter requires people using the service to be 13 years of age or older. Furthermore, Twitter provides advertisers demographic targeting options which allow brands to only target advertising at specific age groups, e.g. over 18s. Unlike broadcasting platforms and on street advertising, for example, online advertising can use audience insights to enable our partner brands to target their advertising to specific demographics, including age groups, on Twitter. To make this possible, we collect age data based on the birthdates people have entered on their Twitter profile. Where we do not have a user's birthday, we use our age inference tools, based on a number of attributes such as the accounts they follow and their interests. This makes it easier for advertisers and others with content not suitable for minors to advertise on Twitter - for example, brands can only advertise alcohol to users aged over 18.

Written evidence submitted by Twitter [MISS0070]

Similarly, [age screening](#) is a way for brands and others to determine online whether a follower meets a minimum age requirement, in a way that is consistent with relevant industry or legal guidelines. It requires new Twitter followers to enter their birth date before being permitted to follow a brand's account. If a user has already entered their date of birth on their profile page, it will be detected and they will be able to immediately follow the account. If they have not already entered their date of birth, they will be asked to enter their date of birth on their profile page settings.

Partnerships and support for users

In 2020, Twitter partnered with the UK's eating disorder charity, Beat, to launch a new prompt to help those impacted by eating disorders. The feature is an expansion of Twitter's #ThereIsHelp initiative, which provides resources to vulnerable people and encourages them to reach out and get support when they need it. When people in the UK search for content related to eating disorders and associated information on Twitter, they will be served a prompt, which directs them to support services provided by our partner, Beat.

Beat's National Helpline exists to encourage and empower people to get help quickly, which is vital considering the sooner someone starts treatment, the greater their chance of recovery. This prompt is available to help direct people to reach out to find help and support they need by signposting those impacted by eating disorders to contact Beat's helpline staff online, by DM on Twitter, or by phone.

More broadly, Twitter provides a platform for a range of campaigns promoting and supporting positive body image and self-esteem, with individuals and organisations using the platform to ignite meaningful conversations on these topics. Dove UK ran a Twitter campaign in 2020 for example, using the hashtag [#SelfEsteemAtHome](#) which focused on sharing tips for using social media in inspiring and positive ways.

Charities like Changing Faces have used Twitter to drive positive and accurate representation of people with visible difference. In 2019, their #PledgeToBeSeen campaign called on companies and brands to represent more people with a visible difference - [it reached 2 billion people through social media](#). We have been pleased to engage with Changing Faces in recent years, including delivering [joint presentations](#) and [participating](#) in their campaigners programme.

We continue to sponsor and contribute to media literacy initiatives around the world. Globally, our partners [UNESCO](#), Common Sense Media, the National Association for Media Literacy, the Family Online Safety Institute, and Connect Safely (amongst others) have helped us develop materials and conduct workshops to help people learn how to better process online information and discern between sources of news.

Algorithms

Written evidence submitted by Twitter [MISS0070]

Transparency, explainability, and consumer choice are critical principles that we are prioritising. Please see below information about how our algorithms work and the data we collect, as well as our ongoing work in these areas.

- On the Twitter [home timeline](#), you may have the ranking algorithm turned on. Historically, you would just see, in reverse chronological order, tweets from accounts you followed. In 2016, we launched a ranking algorithm, following user feedback - research has shown that people find Twitter more useful when they are shown the most relevant Tweets first. The key aspect, however, is that we give all users the option to turn the ranking algorithm off. If you are on mobile, in the top-right corner there is a sparkle icon where you can turn the algorithm off - and once again see tweets in reverse chronological order from accounts that you follow.
- In [conversations](#), we strive to show content to people that we think they will be most interested in and that contributes meaningfully to the conversation. For this reason, the replies, grouped by sub-conversations, may not be in chronological order. For example, when ranking a reply higher, we consider factors such as if the original Tweet author has replied, or if a reply is from someone the individual follows. We also take a range of behavioural signals into account, including into considering replies we might automatically hide. Some examples of behavioral signals we use to help identify this type of content include: an account with no confirmed email address, simultaneous registration for multiple accounts, accounts that repeatedly Tweet and mention accounts that do not follow them, or behavior that might indicate a coordinated attack. As a result we have seen 8% fewer abuse reports from conversations.
- More broadly, we give all users transparency and controls over the data we collect, how it is used, and when it is shared. In the 'Your Twitter Data' section of users' settings, for instance, users can see the types of data stored by Twitter - such as username, email, phone number and account creation details. If a user has shared a birthday or location with us that will also be shown here. Through this tool, users can modify certain information we may have inferred such as gender, age range, language, and location. Users can also download a copy of their data from Twitter through this tool, and review other inference information - such as which advertisers have included you in their tailored audiences, and demographic and interest data from external ads partners. An individual can disable all personalisation and data settings with a single master toggle within their settings.

Bias in Machine Learning (ML) systems is an industry-wide issue, and one we are committed to improving on Twitter. We are aware of our responsibility, and want to work towards making it easier for everyone to understand how our systems work. While no system can be completely free of bias, we will continue to minimise bias through deliberate and thorough analysis, and share updates as we progress in this space. To further understand the impact of algorithms, we are partnering with researchers at the University of California Berkeley to establish a new research initiative focused on studying and improving the performance of ML in social systems,

Written evidence submitted by Twitter [MISS0070]

such as Twitter. Studying the societal impact of algorithms is a growing area of research in which Twitter will continue to support and participate.

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, for example, 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 labeled 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 information on why you received it.

Research and 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](#) to highlight why greater data transparency should be at the centre of our societal response to online harms.

At Twitter, we value and encourage collaborations with academic researchers that keep us accountable, while helping us tackle new challenges through discoveries and innovations. We believe in the value of an open exchange of information and know that this openness and collaboration with third party researchers will improve our service and increase learning around the world. We currently work with countless researchers and institutions globally and continue to learn from these trusted partners. Examples include:

- Twitter hosts the industry's largest data archive of [state backed information operations](#).
- Twitter has worked with independent researchers who get the opportunity to analyse our state backed data pre-publication
- Twitter released a new endpoint into our [developer lab](#) that allows approved researchers to indepthly study the public COVID-19 conversation

Written evidence submitted by Twitter [MISS0070]

- Twitter publishes a [biannual transparency report](#) on the enforcement we have taken against our rules and policies.

The Twitter Application Programming Interface (API) enables programmatic access to Twitter in unique and advanced ways. This can be used to analyse, learn from, and interact with Tweets, Direct Messages, users, and other key Twitter resources. Twitter data can advance research objectives on topics as diverse as the global conversations happening on Twitter. We will soon be introducing a new Academic Research product track which has been designed to serve the unique needs and challenges of academic researchers. In addition to new endpoints and features tailored to the goals of researchers who use our API in their studies, this product track will include additional levels of access to our v2 API, with the flexibility to grow and change access levels based on current needs.

Regulation and Online Harms

We welcome opportunities to continue to collaborate and engage with the different regulatory authorities in the UK. We endeavour to be as proactive as possible to make ourselves available to respond to any queries or discuss any concerns.

With the ASA, we have designated points of contact who make themselves available on an ongoing basis to respond to any questions. We welcome opportunities to partner on shared objectives where possible; following a request from the ASA, for example, we updated our [advertising policies page](#) in 2019 to include links to the CAP code and ASA Guidance. The aim is to raise awareness of the UK Advertising Code, particularly among SMEs that either do not know the Code exists, or do not believe it applies to them. We would be pleased to continue to work with the ASA and to discuss any concerns they have.

On regulation more broadly, 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

Written evidence submitted by Twitter [MISS0070]

had adverse effects on competition by strengthening large companies. If regulation unconsciously creates a framework where big companies continue to consolidate and expand market power, there is a 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 in the Full Response to the Online Harms White Paper to unlock innovation across digital markets, while also ensuring we keep people safe online and promote a thriving democracy. We look forward to continuing to engage with government and civil society to ensure this is fulfilled as the regulatory proposals are developed further.

Thank you again for inviting Twitter to participate.

Yours sincerely,

Katy Minshall
Head of UK Public Policy

January 2021