Written evidence submitted by Twitter (OSB0072)

We believe deeply in, and advocate for, freedom of expression and open dialogue - but that means little as an underlying philosophy if voices are silenced because people are afraid to speak up. With this in mind - we welcome the government's focus on online safety, and this Committee's work to consider the draft Online Safety Bill.

As debate around the world focuses on how to solve public policy challenges related to the technology industry, our approach to regulation and public policy issues is centered on protecting the Open Internet. We define the Open Internet as a global and singular internet that is open to all and promotes diversity, competition, and innovation.

We believe that the Open Internet has driven unprecedented economic, social and technological progress, and while not without significant challenges, it has also led to greater access to information and greater opportunities to speak that are now core to an open society.

We support smart regulation that is forward thinking, understanding that a one-size-fits all approach fails to consider the diversity of the online environment, and poses a threat to innovation. We have welcomed the opportunity to participate in the Online Safety Bill consultation process over recent years.

Objectives

Our view is that regulatory frameworks that look at system-wide processes, as opposed to individual pieces of content, will be able to better reflect the diversity of our online environment and the challenges of scale that modern communications services involve - and we are therefore supportive of the government's original stated commitment to this approach.

Similarly, we welcome Ofcom's designation as the regulator for Online Harms. As we stated in our submission to the White Paper back in 2019, we think that Ofcom is the most appropriate and qualified body to be designated as the independent regulatory authority.

We do, however, think that the Bill in its present form fails to achieve a key objective: providing clarity to UK internet users - and providers - on what speech is and is not allowed online. What's more, as Ellen Judson, senior researcher at Demos has stated: "*This Bill is a jigsaw: not only internally complex, but so much of what it means for the world relies on things that don't exist yet - secondary legislation, Codes of Practice - that can't even exist until the Bill is finalised, which makes scrutiny difficult.*"

Fundamentally, the consequence of this approach is confusion for internet users on what speech will and will not be permitted, a significant lack of clarity for service providers, the potential for freedom of expression to be curtailed - and delays on implementation as we await Ofcom or secondary legislation filling in these critical details.

More broadly, we would welcome further work on reconciling the objectives of the Online Safety Bill with the government's aim of promoting fair competition in the technology sector. We support the government's work to better coordinate digital policy, including through the Plan for Digital Regulation. Competition, however, is absolutely critical for our industry to thrive; we believe that the Open Internet is at risk of being less open as it becomes less competitive and people have less choice. Globally, we are urging regulators to factor into their decisions a test of whether proposed measures, such as those in the Online Safety Bill, further enhance the dominance of existing players; or set insurmountable compliance barriers and costs for smaller companies and new market entrants.

What's more, we believe that greater engagement with the public, especially young people and vulnerable groups, can help address a number of present challenges and tensions with the Bill - including through social media itself. Public feedback processes we have run, for instance, have provided vital input on our policies from a broad range of stakeholders - a standalone user's perspective expressed in a Tweet is just as valid as a formal submission. This year, nearly 49,000 people from around the globe took time to share their feedback on how content from world leaders, for instance, should be handled on our service.

A key learning from other pieces of legislation around the world is the importance of exchange and learning, rather than focusing predominantly on enforcement action. All of us - industry, government and civil society - want the online world to be a safer place. We have seen elsewhere that an enforcement-led approach may lead to companies paying a fine - but is not grounded in supporting industry-wide improvement, nor consideration of alternative approaches to meeting overall objectives. Developments in the online world are volatile and ever-evolving - and so meaningful engagement between companies and the regulator outside of enforcement action will be key. Ofcom does appear to envisage this kind of positive relationship - but we would welcome more detail in the Bill on what such collaboration and engagement looks like in practice. Clear guardrails must be put in place, and full assessments of potential unintended consequences should be undertaken before regulatory action is pursued.

Finally, regulation must give companies the ability to react to upcoming issues, and adjust and apply the response to them as long as they serve the overall objective. Clear definitions - but with the flexibility to innovate on solutions, and adapt systems and processes over time - will be critical to future-proofing the regulation.

Content and Services in Scope

When it comes to content moderation, we already remove both illegal and harmful legal content - millions of Tweets and accounts every single year. The challenge with the Bill at present, however, is a lack of definitions on what exactly is expected. In other areas of law (such as the Malicious Communications Act), the challenges posed by overly vague definitions are well-documented (see, for example, the 2018 Law Commission report). Clear definitions are critical to avoid ambiguity, help those within scope fully understand what is required to comply with the law and, crucially, ensure that UK citizens know exactly what is and is not permissible - while trying to protect freedom of expression and ensure that the reaction from service providers is not just to remove large amounts of content for fear of being in breach.

As a group of LGBTQ+ campaigners led by Stephen Fry said at the beginning of September: *"The new law introduces the "duty of care" principle and would give internet companies extensive powers to delete posts that may cause 'harm.' But because the law does not define what it means by 'harm' it could result in perfectly legal speech being removed from the web."*

Fundamentally, if Ofcom is to have legitimacy and credibility for decisions taken on legal speech in particular, it will be important to have acquired Parliamentary approval to decide the specific content at odds with a Duty of Care on legal harms.

The proposed exemption for 'content of democratic importance' in particular requires far greater clarity; and the proposed carveout for journalists, though well-intentioned, may have unintended consequences. As the Bill is currently drafted, it introduces uncertainty through the inclusion of these categories of 'protected' speech without defining them in any detail. Leaving it to secondary legislation and Ofcom to resolve these issues means critical decisions about what speech is permitted and protected are not made by Parliament through primary legislation - undermining democratic oversight and accountability on key issues of free expression. For instance, would this create a loophole that people suspended from Twitter would be able to challenge their suspension if they ran for election or established a political party?

Journalism is the lifeblood of Twitter - we believe in it, we advocate for it, and we seek to protect it. What's more, we recognise that sometimes it may be in the public interest to allow people to view Tweets that would otherwise be taken down, and have developed policies and processes accordingly. The challenge with translating this to regulation is the absence of a clear definition of what constitutes 'journalistic content.' Every day we see Tweets with screenshots of newspaper front pages, links to blogs, updates from journalists and firsthand accounts of developing events. Crucially, there are accounts we have suspended for breaking our rules who

have described themselves as 'journalists.' Similarly, we have previously seen examples where journalistic content has included visible links to terrorist material, such as that produced by ISIS. Indeed, after the Christchurch mosque shootings, a number of news organisations broadcast the attacker's videos in full. Is the expectation that services should not remove this content? The lack of detail around these provisions risks significant confusion and potentially undermines the overall objectives of the Bill.

If Parliament wishes to establish a category of content based on democratic importance or journalistic content, it seems only right that Parliament should define what that is. Without doing so, it risks confusion not just for news publishers and for services like ours, but for the people using them. These are vital questions - and ones that cannot be avoided or passed to a regulator, or private services, to resolve - not least because of potential ramifications beyond this legislation, and broader issues impacting the freedom of the press. Greater clarity on both protected, as well as prohibited, content would also assist in helping companies better balance the duties on freedom of expression and on 'legal but harmful' content.

These issues are further complicated by the discretion given to the Secretary of State in the Bill to not just modify codes of practice, but to also designate (at any stage) what constitutes 'legal but harmful' content - even that which goes beyond the already ambiguous definition of harm set out (content for which there is a "material risk" of having "significant adverse physical or psychological impact on an adult of ordinary sensibilities").

As the Carnegie UK Trust have set out: "*To meet the UK's international commitments on free speech, there should be a separation of powers between the Executive and a communications regulator. The draft Bill takes too many powers for the Secretary of State. These should be reduced, removing in particular the Secretary of State's power to direct Ofcom to modify its codes of practice to bring them in line with government policy.*"

Far more clarity on both these powers, and also enforcement penalties such as 'business disruption measures,' is essential. People around the world have been blocked from accessing Twitter and other services by multiple governments under the false guise of 'online safety,' impeding peoples' rights to access information online. The government should be mindful of setting a precedent - if the UK wants to lead the online debate globally, it must also set the highest standards of transparency and due process in its own legislation.

In terms of services in scope, greater consideration should be given to the impact on smaller companies of the resources required to comply with regulations. This is especially true across the digital ecosystem, but also even within Category 1 where there is already a range of size and type of service - both some of the largest companies that have ever existed in the world, as

well as their challengers. It will be important to avoid developing requirements that only a very small handful of companies can fulfil. It is also not difficult to envisage a situation in which compliance becomes a competitive advantage rather than a feasible reality for more than that very small handful of companies. Again, the issues around freedom of expression are highlighted in that small providers may err on the side of caution and remove content, rather than being able to have systems in place to assess the nuance that so much of this content will undoubtedly have.

Finally, it is imperative that Ofcom ensure less established or well-known companies, who may be hosting profoundly harmful content - but may not receive public complaints or attention, or indeed make data available for research - are captured. This is something we raised in our original White Paper submission in 2019. We believe greater detail on how 'risk' will be measured in Category 1 - and the scope to include such platforms - may help address this.

Algorithms and user agency

Safe design does not need to only involve how content is removed, and we welcome the Committee's specific focus on algorithms and user agency. Indeed, when it comes to user agency, we believe far more consideration could be given to the opportunities of algorithmic control and choice.

For us, Responsible Machine Learning (ML) consists of the following pillars:

- Taking responsibility for our algorithmic decisions
- Equity and fairness of outcomes
- Transparency about our decisions and how we arrived at them
- Enabling agency and algorithmic choice

Technical solutions alone do not resolve the potential harmful effects of algorithmic decisions. Our Responsible ML working group is interdisciplinary and is made up of people from across the company, including technical, research, trust and safety, and product teams.

Public feedback is particularly important as we assess the fairness and equity of the automated systems we use. Better, more informed decisions are made when the people who use Twitter are part of the process. One way that we have put this into practice is with Birdwatch, a new community-driven approach to help address misleading information on Twitter. Birdwatch allows people to identify information in Tweets they believe is misleading, and write notes that provide informative context. As we develop algorithms that power Birdwatch - such as

reputation and consensus systems - we aim to publish that code publicly in the Birdwatch Guide. The initial ranking system for Birdwatch is already available here.

Over the coming months, we plan to build on this thinking further. We're conducting in-depth analysis and studies to assess the existence of potential harms in the algorithms we use. Here are some analyses you will have access to in the upcoming months:

- A gender and racial bias analysis of our image cropping (saliency) algorithm (published May 2021 here)
- A fairness assessment of our Home timeline recommendations across racial subgroups
- An analysis of content recommendations for different political ideologies across seven countries

This must be balanced, however, with the challenges and limitations of algorithmic transparency. As Nick Pickles, our Public Policy Director for Strategy, Development & Partnerships, said in January at the Home Affairs Select Committee: *"The challenge is that if you have the algorithm and you do not have the data, then you have code but you might not be able to reproduce what is happening on the platform. That question of just giving someone code might satisfy a notion of transparency but, if it does not inform, empower and enable something, providing code alone will not give you the public policy outcome you want."*

What's more, a fundamental risk of algorithmic transparency is that of providing insight to bad actors into how a system can be gamed. If we were, for instance, to publish how our algorithms identify abusive Tweets, this would immediately risk highlighting to people how they would be able to get around our defensive measures.

In the long term, therefore, rather than focusing exclusively on the potential applicability of algorithmic transparency, we believe the practical policy objective should be to give people far more control over the algorithms that affect their everyday lives. For example, Twitter in 2018 introduced the ability to turn off our home timeline ranking algorithm, returning people to a reverse-chronological order of Tweets. Giving people more control over the content they see is also an important way to strike a balance between content moderation and personal choice, particularly given that there are many areas of legal speech that some people find offensive or objectionable, but also have a critical role in public debate.

The role of Ofcom

We are supportive of Ofcom's designation of the regulator, something we stated in our original White Paper submission in 2019. We believe it is the most appropriate and qualified body to be designated as an independent regulatory authority.

We have, however, raised above our concerns with the powers given to the Secretary of State of the day (see above). We were also disappointed that transparency as set out in the Bill appears limited to formal Transparency Reports. At Twitter, transparency is embodied in our open APIs, our information operations archive, and our disclosures in the Twitter Transparency Center. Tens of thousands of researchers access Twitter data we have made available over the past decade via our APIs. Most recently, we have offered a dedicated Covid-19 endpoint to empower public health research, and a new academic platform to encourage cutting edge research using Twitter data. Our archive of state-linked information operations is a unique resource and offers experts, researchers and the public insight into these activities. This bill is an opportunity to set out a clearer framework for such disclosures.

This transparency is one of the reasons you hear more about reports featuring Twitter as core to the research methodology - we empower it. In the long term, we believe a greater openness across the industry would be invaluable in delivering the transparency and accountability we all want to see. What's more, transparency requirements appear limited to individual companies in the Bill. Further openness from public bodies and government - such as CTIRU or the Cross-Whitehall Counter Disinformation Unit - about the requests they are sending to technology companies would help build our collective understanding of, and trust in, the overall ecosystem.

Recommendations

As the Bill proceeds with pre-legislative scrutiny, we are calling for:

- Clarity in the Bill of exactly what 'legal but harmful' content is expected to be removed or protected. This should include clear, robust definitions for what constitutes 'content of democratic importance' and 'journalistic content' that provides the public, service providers and the regulator clear guidance on Parliament's intent;
- Robust guardrails on the powers of the Secretary of State;
- A clear framework to ensure that the competition of the online ecosystem is not damaged and barriers to entry for new services are not insurmountable;
- Consideration of a wider range of platform design interventions to deliver online safety, such as greater control over and choice between algorithms and the importance of open standards;

- A sustained role for the public to engage in the development of this regulation, especially young people and vulnerable groups - including through social media itself. Public feedback processes we have run, for instance, have provided vital input on our policies from a broad range of stakeholders - a standalone user's perspective expressed in a Tweet is just as valid as a formal submission. This year, nearly 49,000 people from around the globe took time to share their feedback on how content from world leaders, for instance, should be handled on our service.

We have already shared with government concerns about both the expertise required for, and the technical feasibility of, some of these proposals. This will be even more important as the Bill is finalised and the regulation comes into force.

We believe that it is critical during the pre-legislative process that independent technology specialists can advise on the technical feasibility of proposed solutions. Reconciling well-intended objectives with both practical challenges and policy tradeoffs is especially key within the realms of technology law and policy, as we have seen with previous legislation. Leveraging strong technical expertise within the process can help more rapidly resolve some of these tensions, while avoiding implementation problems later down the line - and protecting against vendors over-selling the potential and feasibility of their products and services, while embedding proprietary standards and tools.

September 2021