Written supplementary follow up evidence submitted by Twitter (OSB0225)

11 November 2021

Dear Chair,

Thank you for inviting us to give oral evidence at your Committee last month. We said that we would follow up on a number of points.

Trends

As discussed in the hearing, we have policies in place that specifically apply to trends. An overview of how trends work is available here. These policies apply to hashtags that:

- Contain profanity or adult/graphic references.
- Incite fear on the basis of race, ethnicity, national origin, sexual orientation, gender, gender identity, religious affiliation, age, disability, or disease.
- Violate any other Twitter Rules.

We do not share externally the terminology prohibited from trending. This is to prevent attempts to game the system through, for instance, developing new alternative spellings of words. However, we can confirm that the list is not event specific and covers a global, wide range of terms, including those referenced in the committee discussion. This has been in place for several years. We also take steps to ensure that deliberate variations of words (swapping letters for numbers, for example) are detected by our systems. During the hearing, it was noted that screenshots had been circulated and, as agreed, we are happy to receive examples to cross-reference against our system.

Safety by Design

We support a safety by design approach, as discussed in the hearing. In our recent Open Internet Position Paper, which we published last month (available here) we put forward five principles to inform the (global) policy debate that apply to product design in several areas.

- The Open Internet is global, should be available to all, and should be built on open standards and the protection of human rights.
- Trust is essential and can be built with transparency, procedural fairness, and privacy protections.
- Recommendation and ranking algorithms should be subject to human choice and control.
- Competition, choice, and innovation are foundations of the Open Internet and should be protected and expanded, ensuring incumbents are not entrenched by laws and regulations.
- Content moderation is more than just leave up or take down. Regulation should allow for a range of interventions (e.g. platform changes, adding labels, hiding content), while setting clear definitions for categories of content.

With that in mind, when it comes to what a regulator might focus on and on how the draft Bill might better facilitate these, key recommendations would therefore include:

- A principle based approach would allow the greatest impact while being flexible to a wide range of services and innovations. The more granular the regulation, the less likely it is to stand the test of time, as technology improves, behaviour changes and new services emerge.
- Risk mitigations should be proportionate - it is never possible to fully eliminate all risks for all people.
- Allowing companies to develop solutions that work best for their service and the people who use them. Recognising that content moderation and content organisation are two different spheres of work, particularly when content is recommended without a positive signal to seek it out, policymakers should prioritise empowering people to have control over algorithms they interact with and ultimately drive an ability to make our own choices between algorithms. Choice can also help foster greater understanding and awareness of how algorithms impact people's online experiences, leading to greater digital literacy.
- This is also an important part of how to tackle a range of risks, where subjective personal judgements may differ between people. Particularly for content that is legal, the more choice and control people have over their own experience, the better we can balance between reducing risk and protecting free expression.
- A major question for the regulator to consider is the technological barriers to deploying safety tools and systems. The technologies that underpin the ability to address and remove the most harmful content and respond to further harms remain in proprietary silos, becoming exponentially more effective as businesses scale, further enhancing dominance and undermining competition. Content moderation technology is one of the most significant barriers to entry, particularly as regulators set ever stricter requirements on the time taken to remove harmful content. Policymakers should encourage and facilitate a fundamental change in the availability of proactive technologies and the data that underpin them to enable new services and tools to be

made more accessible to a greater range of services, including providing a robust legal framework for information sharing.

Flashing imagery designed to trigger seizures

We welcomed you raising this issue during the hearing. We have been working with the Epilepsy Society on the abhorrent attempts to send content intended to trigger seizures since early last year. This behaviour is against our rules and we will take action on anyone engaged in this issue. As highlighted in evidence to the Committee that this activity is not currently criminalised, we feel this is an important step to take action against those engaging in this abhorrent behavior.

With regard to Twitter's product, we provide people on Twitter with the option of preventing media from autoplaying in their Timelines, which is an important control for people at risk of epilepsy and we are exploring additional steps to help protect the people on Twitter from this type of media.

Critically, this is an issue that also depends on the actions of third party GIF providers, who host the gifs used across a wide range of services. We continue to urge these providers to be proactive in removing this content from their libraries and better understanding the risks the content poses. it may be of benefit for the committee to hear directly from them on this topic.

We prevent any GIFs from appearing when someone searches for 'seizure' in GIF search. In July last year, we took the decision to further ban three search terms, 'epileptic', 'photosensitive' and 'photosensitivity' from our GIF search function.

We have already connected with the Epilepsy Society to arrange a further meeting.

Thank you again for inviting us to participate in your inquiry.

17 November 2021