

Annex IV: Twitter Email Response

From: Nick Pickles

Sent: 04 August 2020 03:01

To: Deborah Brown [REDACTED]; Namratha Somayajula [REDACTED]

Belkis Wille [REDACTED]

Cc: Cynthia Wong [REDACTED]

Subject: Twitter response to HRW letter

Dear Deborah, Belkis and Namratha,

Thank you for your letter regarding Twitter content used in your research and our approach to terrorist or violent extremist content (TVEC). We appreciate the role organizations like HRW play in documenting potential abuses, promoting accountability, and ensuring that counterterrorism measures respect human rights.

Pursuant to the U.S. Stored Communications Act (18 U.S.C. 2701 *et seq.*), Twitter is prohibited from disclosing users' content absent an applicable exception to the general bar on disclosure. This law allows U.S. law enforcement to compel disclosure of content with a valid and properly scoped search warrant, but there is no such mechanism for disclosure to entities who are unable to obtain a warrant (whether governmental or non-governmental).

Unfortunately, this means we cannot provide copies of the content you have identified for archival purposes. However, Twitter is supportive of efforts through the Global Internet Forum to Counter Terrorism (GIFCT)'s working group on legal frameworks to consider potential avenues to allow greater access to content for appropriate uses.

Identification and removal of TVEC Content

Twitter's philosophy is to take a behavior-led approach, utilizing a combination of machine learning and human review to prioritize reports and improve the health of the public conversation. That is to say, we increasingly look at how accounts behave before we look at the content they are posting. Twitter also employs content detection technology to identify potentially abusive content on the service, along with allowing users to report content. This is how we seek to scale our efforts globally and leverage technology even where the language used is highly context specific.

In certain situations, behaviour identification may allow us to take automated action - for example, accounts clearly tied to those that have been previously suspended, often identified through technical data. However, we recognise the risks of false positives in this work and humans are in the loop for decisions made using content and where signals are not strong enough to automate. Signals based on content analysis are part of our toolkit, but not used in isolation to remove accounts and we agree with concerns raised by civil society and academics that current technology is not accurate enough to fully automate these processes. We would not use these systems to block content at upload, but do use them to prioritise human review.

User Privacy and Responding to Legal Process

As stated above, Twitter generally cannot provide content data without appropriate legal process. In accordance with legal requirements, we also comply with preservation requests by preserving responsive user data pending service of valid legal process compelling disclosure.

Twitter retains different types of information for different lengths of time, and in accordance with our Terms of Service and [Privacy Policy](#). Our priority is ensuring that people understand the information we collect, how long we keep it for, what we use it for, and when we may share it.

You can find more details about our approach to the legal process in our [law enforcement guidelines](#), and we disclose information about legal requests we receive in our biannual [Transparency Report](#).

We recognize that while the current legal framework is intended to protect users' privacy, this means that NGOs and journalists may face significant barriers in accessing removed content. We would welcome further discussion of how to address these competing interests, including within multi stakeholder forums like the GIFCT.

Best wishes,

Nick

--

 Nick Pickles || Director, Public Policy Strategy and Development