

Twitter out to crack down on abusive tweets

April 16 2015, by Glenn Chapman



Twitter said on April 16, 2015 that it is cracking down on mean, hateful or menacing tweets that cross the red line from free speech into abuse

Twitter said Thursday that it is cracking down on mean, hateful or menacing tweets that cross the red line from free speech into abuse.

Twitter is overhauling its safety policy and beefing up the team responsible for enforcing it, along with investing "heavily" in ways to detect and limit the reach of abusive content, general counsel Vijaya Gadde said in an column published by the Washington Post.

"We need to do a better job combating abuse without chilling or silencing speech," Gadde said.

Twitter last month modified its rules to ban 'revenge porn'—the tweeting of intimate or revealing pictures or video of people without their permission.

The San Francisco-based micro-blogging site is also taking steps to curtail the use of anonymously created Twitter accounts to intimidate or silence targeted people.

"We are changing our approach to this problem, in some ways that won't be readily apparent and in others that will be," Gadde said.

Twitter has tripled the size of the team responsible for protecting users of the service, resulting in a five-fold increase in the speed of response to complaints, according to the general counsel.

"We are also overhauling our safety policies to give our teams a better framework from which to protect vulnerable users," Gadde said.

Changes included expanding the definition of banned "abuse" to include indirect threats of violence.

"As some of our users have unfortunately experienced firsthand, certain types of abuse on our platform have gone unchecked because our policies and product have not appropriately recognized the scope and extent of harm inflicted by abusive behavior," Gadde said.

"Even when we have recognized that harassment is taking place, our response times have been inexcusably slow and the substance of our responses too meager. This is, to put it mildly, not good enough."

Instagram targets terror

Facebook last month updated its "community standards" guidelines, giving users more clarity on acceptable posts relating to nudity, violence, hate speech and other contentious topics.

Facebook-owned smartphone photo and video sharing service Instagram followed suit on Thursday with a similar overhaul of its rules about what is deemed unacceptable.

"It was time for a refresh; to streamline it and provide a better explanation," Instagram director of public policy Nicky Jackson Colaco told AFP, citing the services global growth since being acquired by Facebook three years ago in a deal valued at a billion dollars.

"We are setting expectations for what kinds of content we think are acceptable to share on Instagram and what could happen if you violated the policy."

Instagram boasts more than 300 million active users worldwide, while Facebook lays claim to about 1.38 billion active monthly users.

Instagram guidelines ban nudity, along with threats, and hate speech.

The new community guidelines state that "sharing graphic images for sadistic pleasure or to glorify violence is never allowed."

Facebook's updated community doctrine states that the world's biggest social network will not allow a presence from groups advocating "terrorist activity, organized criminal activity or promoting hate."

The moves come with Twitter, Facebook, Instagram and other social media struggling with defining acceptable content and freedom of

expression, and with these services increasingly linked to radical extremism and violence.

"What we come back to is what we want our platform to be used for and what we don't want it to be used for," Jackson Colaco said, noting that Instagram was created in a 'post 9/11 world.'

"Instagram is not a place to support or praise terrorism, organized crime, or hate groups," the new community guidelines stated.

"We remove content that contains credible threats or hate speech, content that targets private individuals to degrade or shame them, personal information meant to blackmail or harass someone, and repeated unwanted messages."

© 2015 AFP

Citation: Twitter out to crack down on abusive tweets (2015, April 16) retrieved 19 April 2024 from <https://phys.org/news/2015-04-twitter-abusive-tweets.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.