

Twitter moves to stem violent threats, abuse

April 21 2015



Twitter began implementing a new policy aimed at curbing use of the network to incite violence, and to crack down on abuse and harassment

Twitter on Tuesday began implementing a new policy aimed at curbing use of the social network to incite violence, and to crack down on abuse and harassment on the service.

The new rules are the latest implemented by social networks aiming to stem violence and harassment while attempting to safeguard freedom of online speech.

"We need to ensure that voices are not silenced because people are afraid to speak up," said Twitter's head of product management Shreyas Doshi in a blog post.

"To that end, we are today announcing our latest product and policy updates that will help us in continuing to develop a platform on which users can safely engage with the world at large."

Some of the changes were outlined last week in a Washington Post column by Twitter general counsel Vijaya Gadde.

On Tuesday, Doshi noted that Twitter had updated its policy on violent threats "so that the prohibition is not limited to 'direct, specific threats of violence against others' but now extends to 'threats of violence against others or promot(ing) violence against others.'"

He said the previous policy "was unduly narrow and limited our ability to act on certain kinds of threatening behavior."

Twitter's new policy also allows the social network to "lock" abusive accounts for specific periods of time, which could help crack down on so-called cyberbullying and spam.

"This option gives us leverage in a variety of contexts, particularly where multiple users begin harassing a particular person or group of people," Doshi said.

Doshi also noted that Twitter would be testing "a product feature to help us identify suspected abusive Tweets and limit their reach."

The new tool, which was not described in detail, could help Twitter identify abusive activity by looking at "a wide range of signals and context that frequently correlates with abuse including the age of the

account itself, and the similarity of a tweet to other content that our safety team has in the past independently determined to be abusive," said Doshi.

This tool "will not affect your ability to see content that you've explicitly sought out, such as tweets from accounts you follow, but instead is designed to help us limit the potential harm of abusive content," he noted.

Twitter's actions follow initiatives by Facebook as well as Facebook-owned Instagram that aim to crack down on abusive conduct and use of the platforms to promote violence.

Social networks have been struggling with defining acceptable content and freedom of expression, and radical extremism and [violence](#) increasingly linked to these services.

Facebook said last month it will not allow the social network to be used to promote terrorism or hate speech as it unveiled a wide-ranging update of its "community standards."

© 2015 AFP

Citation: Twitter moves to stem violent threats, abuse (2015, April 21) retrieved 19 April 2024 from <https://phys.org/news/2015-04-twitter-stem-violent-threats-abuse.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.