

Twitter allows some users to flag 'misleading' content

18 August 2021



Some Twitter users from the United States, South Korea and Australia will be able to report "misleading" posts.

Twitter on Tuesday announced a new feature to allow users to flag content that could contain misinformation, a scourge that has only grown during the pandemic.

"We're testing a feature for you to report Tweets that seem misleading - as you see them," the social network said from its safety and security account.

Starting Tuesday, a button would be visible to some users from the United States, South Korea and Australia to choose "it's misleading" after clicking "report tweet."

Users can then be more specific, flagging the misleading tweet as potentially containing misinformation about "health," "politics" and "other."

"We're assessing if this is an effective approach so we're starting small," the San Francisco-based company said.

"We may not take action on and cannot respond to

each report in the experiment, but your input will help us identify trends so that we can improve the speed and scale of our broader misinformation work."

Twitter, like Facebook and YouTube, regularly comes under fire from critics who say it does not do enough to fight the [spread of misinformation](#).

But the platform does not have the resources of its Silicon Valley neighbors, and so often relies on experimental techniques that are less expensive than recruiting armies of moderators.

Such efforts have ramped up as Twitter toughened its misinformation rules during the Covid-19 pandemic and during the US [presidential election](#) between Donald Trump and Joe Biden.

For example, Twitter began blocking users in March who have been warned five times about spreading false information about vaccines.

And the network began flagging tweets from Trump with a banner warning of their misleading content during his 2020 re-election campaign, before the then-president was finally banned from the website for posting incitements to violence and messages discrediting the election results.

Moderators are ultimately responsible for determining which content actually violates Twitter's terms of use, but the network has said it hopes to eventually use a system that relies on both human and automated analysis to detect suspicious posts.

Concern around Covid-19 vaccine misinformation has become so rampant that in July Biden said Facebook and other platforms were responsible for "killing" people in allowing false info around the shots to spread.

He walked back the remarks the clarify that the false information itself is what could harm or even

kill those who believe it.

© 2021 AFP

APA citation: Twitter allows some users to flag 'misleading' content (2021, August 18) retrieved 29 November 2021 from <https://techxplore.com/news/2021-08-twitter-users-flag-content.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.