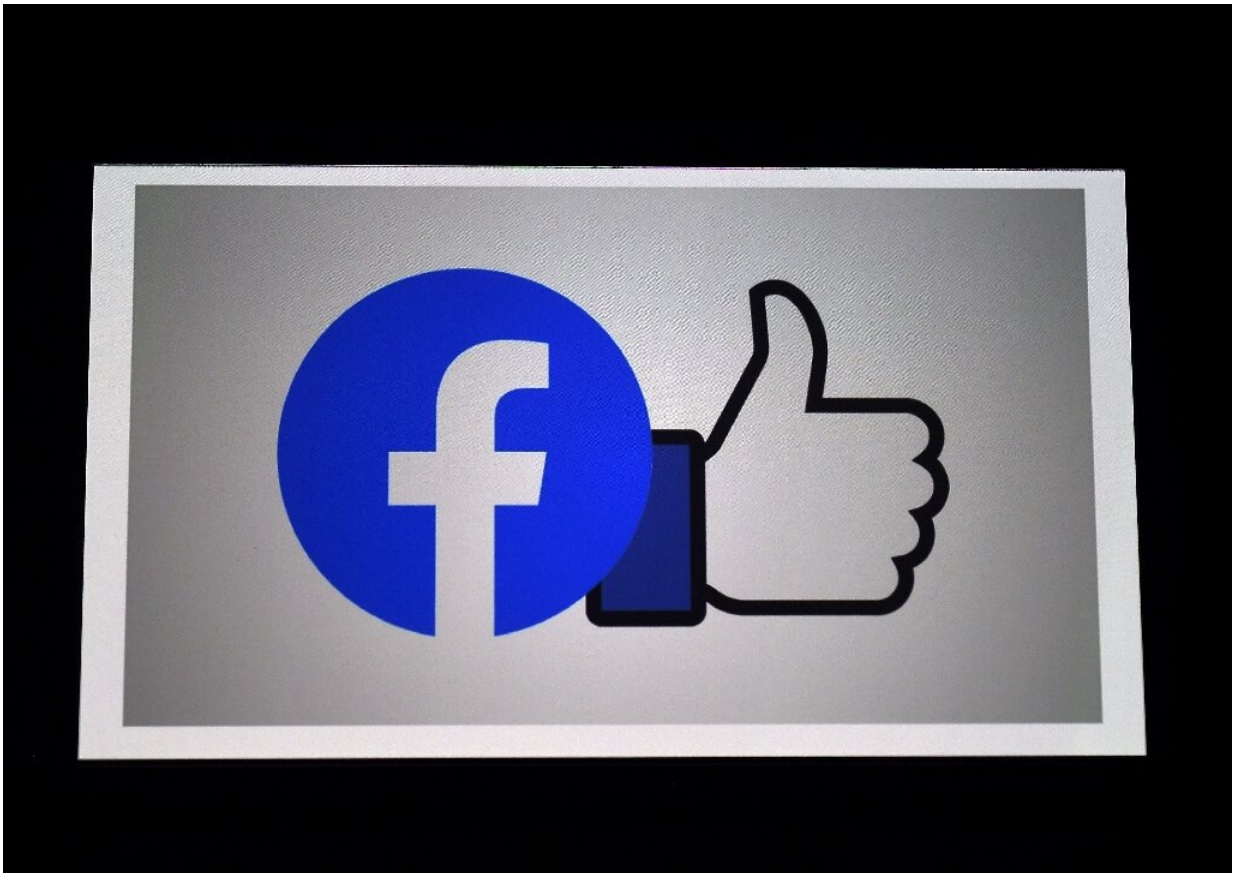


Facebook reports progress catching hate speech

August 11 2020



Facebook says its automated systems improved in detecting and removing hateful and abusive content

Facebook on Tuesday reported progress in catching abusive content on

the platform as it relied more on automated systems during the pandemic.

The leading social network released its latest enforcement report as it announced updated policies to bar specific kinds of "implicit hate speech" such as blackface and "stereotypes about Jewish people controlling the world."

The "proactive detection rate" rose six percent to 95 percent, with the leading social network taking action on 22.5 million pieces of content deemed hateful at Facebook and Instagram in the second quarter of this year, according to the internet giant's latest enforcement report.

"Despite the impact of COVID-19, improvements to our technology enabled us to take action on more content in some areas," the report stated.

Some of the improvement was credited to expanding automated detection to more languages including Spanish and Burmese and to better understanding posts in English.

Automated detection of hate speech at Instagram rose to 84 percent, with the Instagram, with the image-centric social network taking action on a total of 3.3 million pieces of content in the second quarter, according to the report.

"We've made progress in combating hate on our apps, but we know we have more to do to ensure everyone feels comfortable using our services," the report stated.

An Instagram Equity Team and a Facebook Inclusive Product Council were created to help make sure cultural fairness is built into products, according to the report.

Facebook also said it is launching a Diversity Advisory Council to provide input topics and issues.

Progress was also reported in automatically detecting terrorism content, with Facebook taking action on 8.7 million pieces of such content in the [second quarter](#).

Facebook remains under pressure to fight abusive and deceitful content on its platform—amid a boycott by advertisers—while fending off accusations it unfairly stifles politically conservative voices.

© 2020 AFP

Citation: Facebook reports progress catching hate speech (2020, August 11) retrieved 26 April 2024 from <https://techxplore.com/news/2020-08-facebook-speech.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.