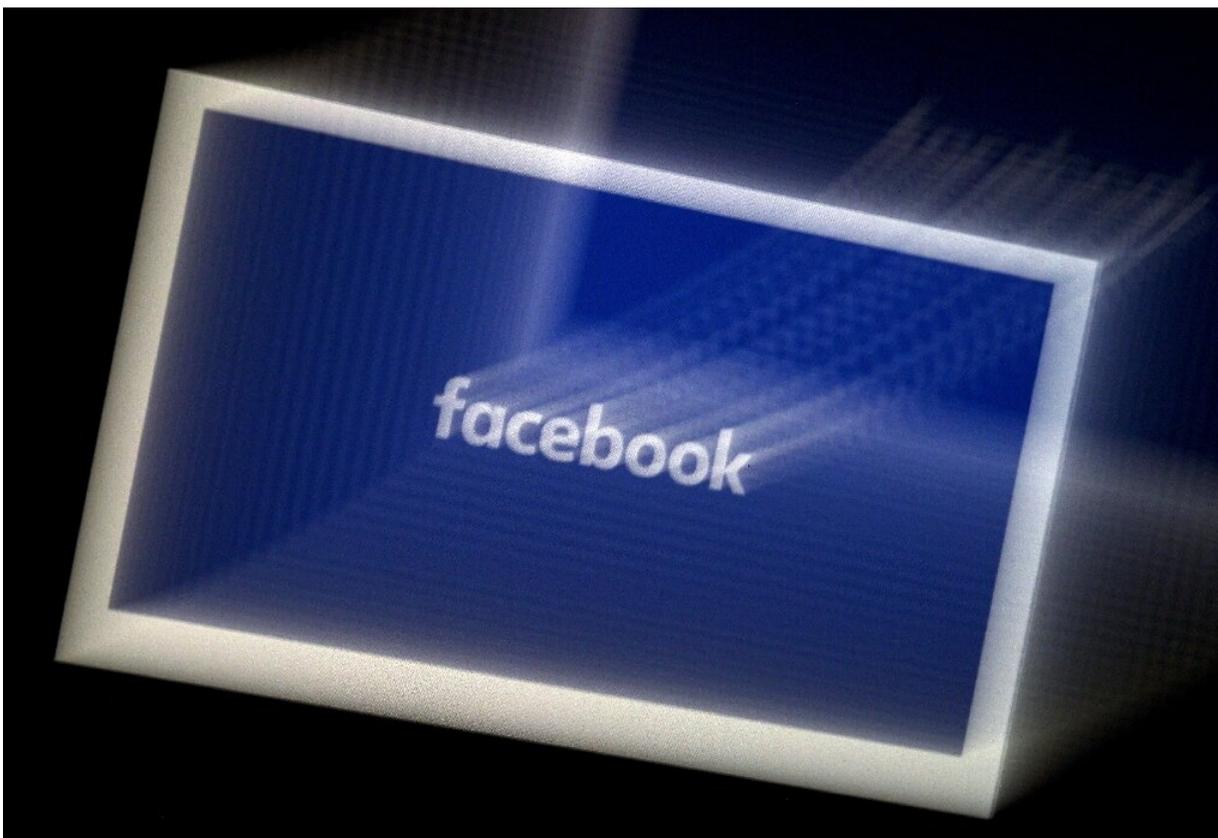


Facebook changing 'race-blind' hate speech policy

December 3 2020



Facebook is reportedly changing its policy on hate speech and shifting away from a "race-blind" approach

Facebook on Thursday said it is revising its systems to prioritize blocking slurs against Black people, gays and other groups historically

targeted by hateful vitriol, no longer automatically filtering out barbs aimed broadly at whites, men, or Americans.

The change in Facebook's algorithm is a shift from the social network's ethnicity and gender-neutral system that removed anti-white comments and comments such as "Men are dumb" or "Americans are stupid."

"We know that hate speech targeted towards underrepresented groups can be the most harmful, which is why we have focused our technology on finding the hate speech that users and experts tell us is the most serious," said Facebook spokeswoman Sally Aldous.

The leading social network said it has also become more precise about the hate speech that its software seeks out.

The changes are an adjustment of the leading social network's automated systems, meaning that hateful posts about whites, men, or Americans that are reported by users will still be deleted if they violate Facebook policies.

Over the past year Facebook has also updated its policies to catch more implicit hate speech, such as depictions of Blackface and stereotypes about Jewish people, Aldous noted.

"Thanks to significant investments in our technology we proactively detect 95 percent of the content we remove and we continue to improve how we enforce our rules as hate speech evolves over time," Aldous said.

The reform is at an early stage and aims to target speech deemed "the worst of the worst," including slurs against Black people, Muslims, people of more than one race, the LGBTQ community and Jews, according to a report in the Washington Post.

The move comes as Facebook faces pressure from civil rights groups who have long complained the social network does too little to police hate speech. Earlier this year, more than 1,000 advertisers joined a boycott of Facebook to protest its handling of hate speech and misinformation.

At the same time, the company and its rival Twitter have also been taken to task on Capitol Hill by Republicans who say the platforms are biased against conservatives.

On Wednesday, Twitter said it was expanding its definition of hateful content to ban language which "dehumanizes" people on the basis of race, ethnicity or national origin.

Twitter said it would remove offending tweets when they are reported, and offered examples such as describing a particular ethnic group as "scum" or "leeches."

© 2020 AFP

Citation: Facebook changing 'race-blind' hate speech policy (2020, December 3) retrieved 23 February 2026 from <https://techxplore.com/news/2020-12-facebook-race-blind-speech-policy.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.