

Facebook ramps up efforts to curb 'hateful content' in ads

26 June 2020



Facebook CEO Mark Zuckerberg, seen in an October 2019 picture, said the social network will step up efforts to curb hateful content and also add tags to rule-breaking posts that are "newsworthy"

Facebook said Friday it would ban a "wider category of hateful content" in ads as the embattled social media giant moved to respond to growing protests over its handling of inflammatory posts.

Chief executive Mark Zuckerberg said Facebook also would add tags to posts that are "newsworthy" but violate platform rules—following the lead of Twitter, which has used such labels on tweets from President Donald Trump.

The initiative comes with the leading social network facing a growing boycott by advertisers—with soft drink behemoth Coca-Cola and Anglo-Dutch giant Unilever joining Friday—as activists seek tougher action on content they deem to promote discrimination, hatred or violence.

The new policy on hateful content in ads will "prohibit claims that people from a specific race, ethnicity, national origin, religious affiliation, caste,

sexual orientation, gender identity or immigration status are a threat to the physical safety, health or survival of others," Zuckerberg said.

"We're also expanding our policies to better protect immigrants, migrants, refugees and asylum seekers" from hateful ads, he continued.

Facebook has underscored its moves to stem racism in the wake of civil unrest triggered by the May 25 killing of African American George Floyd at the hands of Minneapolis police.

"We invest billions of dollars each year to keep our community safe and continuously work with outside experts to review and update our policies," a spokesperson said.

"The investments we have made in (artificial intelligence) mean that we find nearly 90 percent of hate speech" and take action before users report it.



Consumer products giant Unilever said it would pause ads on Facebook, Instagram and Twitter through 2020

'Newsworthy' exception

Zuckerberg said the "newsworthy" exemption normally occurs "a handful of times a year," when Facebook decides to leave up a message that would ordinarily be removed for rule violations.

Under the new policy, Zuckerberg said, "we will soon start labeling some of the content we leave up because it is deemed newsworthy, so people can know when this is the case."

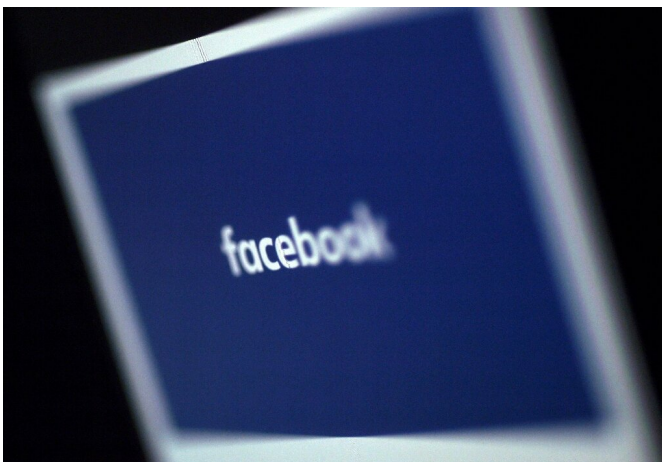
He said users will be allowed to share the content "but we'll add a prompt to tell people that the content they're sharing may violate our policies."

Twitter in recent weeks has labeled at least one Trump tweet misleading and has flagged others as violating platform rules, accessible only when users click through a warning. The move has angered the president and his allies.

Internet platforms have faced intense pressure from activists following Floyd's death.

A coalition including the National Association for the Advancement of Colored People (NAACP) has been urging companies to stop advertising on Facebook, using the #StopHateForProfit hashtag.

At the same time, Trump and his allies have voiced anger over what they claim is biased against conservatives.



Facebook said it revamped some policies based on "feedback from the civil rights community"

Brand boycott accelerates

Zuckerberg made no mention of the ad boycott but said the changes were based on "feedback from the civil rights community and reflect months of work with our civil rights auditors."

Coca-Cola, a major force in global advertising, said it would suspend ads on social media for at least 30 days as it reassesses its policies, though it said the decision was not related to the #StopHateForProfit campaign.

"There is no place for racism in the world and there is no place for racism on social media," James Quincey, chairman and CEO of The Coca-Cola Company, said in a brief statement.

He said social media companies need to provide "greater accountability and transparency."

Unilever, home to brands including Lipton tea and Ben and Jerry's ice cream, said it would stop advertising on Facebook, Twitter and Instagram in the US until the end of 2020 due to the "polarized election period."

American Honda said it would halt ads on Facebook in July, "choosing to stand with people united against hate and racism," adding to a list that includes US telecom giant Verizon and sporting goods makers Patagonia, North Face and REI.

The Facebook move on hate speech in ads "is welcome but (they) account for a small portion of harmful content on the platform," said Graham Brookie, director of the Atlantic Council's Digital Forensics Research Lab, which monitors social media disinformation.

Michelle Amazeen, a Boston University professor of political communication, said details still remain unclear.

"Will Facebook allow independent verification of

which content they tag and the subsequent effects on diffusion?" she asked.

© 2020 AFP

APA citation: Facebook ramps up efforts to curb 'hateful content' in ads (2020, June 26) retrieved 24 November 2020 from <https://techxplore.com/news/2020-06-facebook-ramps-efforts-curb-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.