Meta's success in suppressing misinformation on Facebook is patchy at best, finds study

March 22 2024

The content moderation policy adopted by Meta at the time of the COVID-19 pandemic to rein in misinformation on Facebook has proved
no great obstacle to users capable to finding work arounds according to a new study by digital and social media researchers from the University of Technology Sydney and the University of Sydney.

Published recently in the journal *Media International Australia*, the study looked at the effectiveness of strategies such as content labeling and shadowbanning during 2020 and 2021, shadowbanning involving the algorithmic reduction of problematic content in users' newsfeed, search and recommendations.

Lead author UTS Associate Professor Amelia Johns said the analysis found that far-right and anti-vaccination accounts in some cases enjoyed increased engagement and followers after Meta's content policy announcements.

"This calls in question just how serious Meta has been about removing harmful content," Associate Professor Johns said.

"The company has invested in content moderation policies that err on the side of free expression, preferring content labeling and algorithm-driven suppression over removal.

"The company points to internal modeling which shows that users will try to find work arounds to content that is removed, which is why, it asserts, removal is not effective.

"However our research shows that shadowbans and content labeling are only partially effective and likewise incentivize work arounds by users dedicated to overcoming platform interventions and spreading misinformation.

"It was clear far-right and anti-vaccination communities were not deterred by Meta's policies to suppress rather than remove dangerous
misinformation during the pandemic, employing tactics that disproved Meta's internal modeling.

"In essence users came together as a community to game the algorithm rather than allowing the algorithm to determine what content they were able to access, and how.

"This demonstrates that the success of Meta's policy to suppress rather than remove misinformation is piecemeal, inconsistent and seemingly unconcerned about susceptible communities and users encountering misinformation."


Provided by University of Technology, Sydney


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.