Press Release

*Issued:* **Friday 25 April 2025 – FOR IMMEDIATE RELEASE**

Peer reviewed: Yes

Evidence type: Observational

Subject: People

Deleted Facebook posts on Middle East conflict broke no rules

Hundreds of posts about an outbreak of violence in Israel and Palestine in 2021 that were removed by a social media company did not violate the platform’s rules, a study suggests.

Analysis of almost 450 deleted Arabic-language Facebook posts referencing the conflict suggests that more than 50 per cent should not have been removed, researchers say.

The study raises important questions about who gets to decide what content remains online – and what gets removed – especially in politically sensitive conflicts, experts say.

A team led by researchers from the University of Edinburgh analysed the content of 448 posts about the 2021 Palestine-Israel conflict – which took place 10 to 21 May 2021 – that were removed by Facebook.

Their findings suggest there are major differences between how the company enforces its community standards and how Arab users perceive that moderation.

As part of the study, more than 100 native Arabic speakers reviewed each of the deleted posts to assess whether they violated Facebook’s guidelines – and whether, in their personal view, they should have been removed. Each post was evaluated by 10 reviewers.

Based on their analysis, 53 per cent of the deleted posts were judged by clear majority – at least seven reviewers agreeing – not to violate any of Facebook’s content rules. Around 30 per cent of all posts had unanimous agreement, with all 10 reviewers judging they did not break any of the platform’s guidelines.

The reviewers agreed the other deleted posts did violate platform rules and were rightly taken down.

The study also found that Facebook’s AI moderation feature often appeared to flag posts expressing support for Palestinians, even when these contained no hate speech or incitement to violence.

The findings highlight broader issues about how online systems built primarily from Western perspectives may lack the cultural and linguistic sensitivity needed for fair global enforcement, the team say. This is of particular concern for marginalised communities, they add.

The researchers urge social media companies to increase the diversity of those involved in setting moderation policies, and to improve transparency about how posts are analysed.

The peer-reviewed study will be presented at the CHI 2025 Conference on Human Factors in Computing Systems. It also involved researchers from HBKU University, Qatar, and the University of Vaasa, Finland.

Dr Walid Magdy, of the University of Edinburgh’s School of Informatics, who led the study, said: “Our analysis highlights a clear disconnect between Facebook’s enforcement and how users from marginalised regions perceive fairness. This is especially important in conflict zones, where digital rights are vulnerable and content visibility can shape global narratives.

“If platforms claim to support free expression and inclusion, they need to rethink how they apply community standards across different languages and cultural contexts. Global platforms can’t rely solely on Western views to moderate global content.”

For further information, please contact: Corin Campbell, Press and PR Office, 07920 404 319,**corin.campbell@ed.ac.uk**