



OVERTURNED

2023-058-IG-UA

# Heritage of Pride

A user appealed Meta's decision to remove an Instagram post that was celebrating Pride month by reclaiming a slur that has traditionally been used against gay people.

## Policies and topics

-  LGBT, Marginalised communities, Protests
-  Hate speech

## Region and countries

-  United States and Canada
-  United States

## Platform

-  Instagram

*This is a summary decision. Summary decisions examine cases in which Meta has reversed its original decision on a piece of content after the Board brought it to the company's attention. These decisions include information about Meta's acknowledged errors and inform the public about the impact of the Board's work. They are approved by a Board Member panel, not the full Board. They do not involve a public comments process and do not have precedential value for the Board. Summary decisions provide transparency on Meta's corrections and highlight areas in which the company could improve its policy enforcement.*

## **Case summary**

A user appealed Meta's decision to remove an Instagram post that was celebrating Pride month by reclaiming a slur that has traditionally been used against gay people. After the Board brought the appeal to Meta's attention, the company reversed its original decision and restored the post.

## **Case description and background**

In January 2022, an Instagram user posted an image with a caption that includes a quote by writer and civil rights activist James Baldwin, which speaks of the power of love to unite humanity. The caption also states the user's hope for a year of rest, community and revolution, and calls for the continuous affirmation of queer beauty. The image in the post shows a man holding a sign that says, "That's Mr Faggot to you", with the original photographer credited in the caption. The post was viewed approximately 37,000 times.

Under [Meta's Hate Speech policy](#), the company prohibits the use of certain words that it considers to be slurs. The company recognises, however, that "speech, including slurs, that might otherwise violate our standards can be used self-referentially or in an empowering way". Meta explains that its "policies are designed to allow room for these types of speech", but the company requires people to "clearly indicate their intent". If the intention is unclear, Meta may remove content.

Meta initially removed the content from Instagram. The user, a verified Instagram account based in the United States, appealed Meta's decision to remove the post to the Board. After the Board brought this case to Meta's attention, the company determined that the content did not violate the Hate Speech Community Standard and that its original decision was incorrect. The company then restored the content to Instagram.

## **Board authority and scope**

The Board has authority to review Meta's decision following an appeal from the person whose content was removed (Charter Article 2, Section 1; Bylaws Article 3, Section 1).

When Meta acknowledges that it made an error and reverses its decision on a case under consideration for Board review, the Board may select that case for a summary decision

(Bylaws Article 2, Section 2.1.3). The Board reviews the original decision to increase understanding of the content moderation processes involved, reduce errors and increase fairness for Facebook and Instagram users.

## **Case significance**

This case highlights challenges in Meta's ability to enforce exceptions to its Hate Speech policy, as well as shortcomings of the company's cross-check programme. The content in this case was posted by a verified Instagram account eligible for review under the cross-check system. Therefore, the account, which is dedicated to educating users about the LGBTQIA+ movement, should have had additional levels of review. As the caption mentions "infinite queer beauty" and makes references to community and solidarity with LGBTQIA+ people, Meta's moderation systems should have recognised that the slur was used here in an empowering way, rather than to condemn or disparage the LGBTQIA+ community.

Previously, the Board has issued several recommendations relevant to this case. The Board has recommended that "Meta should help moderators better assess when exceptions for content containing slurs are warranted", ( [Reclaiming Arabic Words](#) decision, recommendation no. 1) and that Meta should "let users indicate in their appeal that their content falls into one of the exceptions to the Hate Speech policy. This includes where users share hateful content to condemn it or raise awareness", ( [Two buttons meme](#) decision, recommendation no. 4). Meta has taken no further action on the first recommendation and has implemented in part the second recommendation. Additionally, the Board has recommended that Meta "conduct accuracy assessments focused on Hate Speech policy allowances that cover expression about human rights violations (e.g. condemnation, awareness raising)", ( [Wampus belt](#) decision, recommendation no. 3). Meta has implemented this recommendation in part. Finally, since the content was posted by an account that is part of Meta's cross-check programme, relevant recommendations include encouraging Meta to identify "'historically over-enforced entities' to inform how to improve its enforcement practices at scale", ( [policy advisory opinion on cross-check programme](#), recommendation no. 26) and to establish "a process for users to apply for over-enforcement mistake-prevention protections", ( [policy advisory opinion on cross-check programme](#), recommendation no. 5). Meta is currently fully implementing the first recommendation and declined to take further action on the second recommendation. The Board underlines the need for Meta to address these concerns to reduce the error rate in moderating hate speech content.

## **Decision**

The Board overturns Meta's original decision to remove the content. The Board acknowledges Meta's correction of its initial error once the Board brought the case to Meta's attention.