



**Human Rights Watch’s Public Comment
2024-004-FB-UA, 2024-005-FB-UA, 2024-006-FB-UA**

May 21, 2023

The origin and current uses of the phrase: “From the river to the sea”

The slogan “From the river to the sea, Palestine will be free” has reverberated at protests in solidarity with Palestinians around the world since October 7, 2023.

The phrase "From the river to the sea" has been used to mean different things, ranging from a demand that Palestinians, wherever they live, including in Israel, be free, to a rejection of the state of Israel. For many Palestinians, the phrase reflects a desire for unity of land and people and their aspiration for self-determination. For others, it's a call to replace the state of Israel with Palestine. While the phrase itself can have different political meanings with which different people may agree or disagree, taken on its own it is protected speech under international human rights law.

Research into online trends in content using the phrase.

In a December 2023 [report](#), Human Rights Watch found that Meta’s content moderation policies and systems have increasingly silenced voices in support of Palestine on Instagram and Facebook in the wake of the hostilities between Israeli forces and Palestinian armed groups. The report, “Meta’s Broken Promises: Systemic Censorship of Palestine Content on Instagram and Facebook,” documents a pattern of undue removal and suppression of protected speech including peaceful expression in support of Palestine and public debate about Palestinian human rights. Human Rights Watch found that the problem stems from flawed Meta policies and their inconsistent and erroneous implementation, overreliance on automated tools to moderate content, and undue government influence over content removals.

The complaints under consideration challenge Meta’s decision to not take down content that included the slogan “From the river to the sea, Palestine will be free” on the basis that the content violates Meta’s rules on Hate Speech, Violence and Incitement or Dangerous Organizations and Individuals. Meta made the correct determinations in leaving this content up since the context provided indicates that the phrase was used to express peaceful support for Palestinians.

Human Rights Watch has documented other instances where use of this slogan or related variations was improperly suppressed by Meta for other reasons, including for violating its “spam” policy, for violating unspecified “community guidelines”, and, at times, without any explanation provided at all. Suppression included removal of comments, significantly lower circulation of posts, and temporary restriction from posting.

These examples of suppression are consistent with other treatment of Palestinian content documented by Human Rights Watch. Examples were also frequently accompanied by a lack of transparency and functioning appeal mechanisms. Instances of suppression of this slogan on Meta compound real-world attempts to prohibit its use, which are inconsistent with human rights. These and other aggressive restrictions of Palestinian content contribute to a chilling effect already taking root on Meta platforms.

Meta’s human rights responsibilities in relation to content using the phrase including with regard to freedom of expression, freedom of association, and equality and non-discrimination.

Under the United Nations Guiding Principles on Business and Human Rights (UNGPs), Meta has a responsibility to respect human rights by avoiding infringing on human rights, identifying and addressing the human rights impacts of their operations, and providing meaningful access to a remedy. This responsibility includes aligning their content moderation policies and practices with international human rights standards, ensuring that decisions to take content down are not overly broad or biased, being transparent and accountable in their actions, and enforcing their policies in a consistent manner. Meta’s own Community Standards state that their goal is to create a place for expression and give people a voice. Meta says it wants “people to be able to talk openly about the issues that matter to them...even if some may disagree or find them objectionable”, and in some cases, allows content—which would otherwise go against its standards—if it’s newsworthy and in the public interest.

With respect to the phrase “From the river to the sea, Palestine will be free,” Meta should treat it as protected expression and should not have a blanket policy restricting the use of the phrase. As with other protected expression, posts using this phrase should only be restricted in specific contexts where the speaker is inciting violence, discrimination or hostility.

Unduly restricting or suppressing peaceful content that supports Palestine and Palestinians impermissibly infringes on people’s rights to freedom of expression. Given that social media has become the digital public sphere and the site of social movements, undue restrictions on content and the ability to engage with other users on social media also undermine the rights to freedom of peaceful assembly and association, as well as participation in public affairs. The enforcement of content removal policies and adjustments to recommender algorithms, which determine what content people see in their feeds, to significantly limit circulation of certain types of content may be perceived as biased or selectively targeting specific viewpoints and could undermine the right to non-discrimination and the universality of rights as well as the right to due process.

Removing or suppressing online content can hinder the ability of individuals and organizations to advocate for human rights of Palestinians and raise awareness about the situation in Israel and Palestine. Content removal that is carried out using automated systems, such as on Instagram and Facebook, raises concerns about algorithmic bias. As HRW’s report documents,

these systems may result in the erroneous suppression of content, leading to discriminatory consequences without opportunity for redress.

Users reported that their ability to express opinions, including dissenting or unpopular views about the escalation of violence since October 7, is being restricted repeatedly and increasingly over time. As a result of comment removal or restriction, users reported altering their behavior on Instagram and Facebook to avoid their comments being removed.

Engaging with content, such as posting or reading comments, is a crucial aspect of social media interaction, especially when open discussion is prohibited and contested in offline spaces. Social media can be a vital communications tool in crises and conflicts. However, users experiencing account restrictions, or their mere possibility, may refrain from engaging on social platforms to avoid losing access to their accounts and vital information, resulting in chilling effects and self-censoring behaviors.

Implications of Meta's Dangerous Organizations and Individuals Policy

Should the Oversight Board reverse Meta's decisions on these three cases and link use of the phrase with support for Hamas, the Oversight Board risks establishing use of this slogan as a violation of Meta's Dangerous Organizations and Individuals (DOI) policy. Human Rights Watch recommends against doing so. Prohibiting the use of this slogan under Meta's DOI policy would greatly exacerbate the policy's propensity to stifle peaceful content that supports Palestine and Palestinians in a manner that impermissibly infringes on people's rights to freedom of expression. First, the phrase cannot be understood as support for Hamas without additional context. Second, instead of relying primarily on a definition of terrorist entities or dangerous organizations based on designations set by particular governments, Meta should refocus the policy on prohibiting incitement to terrorism, drawing on the [model definition](#) advanced by the mandate of the Special Rapporteur on the promotion and protection of human rights and fundamental freedoms while countering terrorism.

Meta should also be more transparent about voluntary requests by governments, including internet referral units, to remove content based on Community Standards and Community Guidelines. For example, it should notify users if a government was involved in their content being taken down based on a policy violation, as the Oversight Board recommended in the Öcalan Isolation case, and provide a transparent appeal process for the decision.

Implications of Meta's reliance on automation

Given that the phrase "From the river to the sea" is widely used to engage in protected speech, any guidance from the Oversight Board regarding the moderation of this phrase should take into account Meta's heavy reliance on automated assessment tools, which are notoriously poor at interpreting contextual factors. Specifically, any such guidance should emphasize the challenges automated tools have historically faced when attempting to classify comments advocating for the rights of Palestinians, particularly during times of ongoing hostilities.

Human Rights Watch's research found that Meta's reliance on automation for content moderation is a significant factor in the erroneous enforcement of its policies, which has resulted in the removal of non-violative content in support of Palestine on Instagram and Facebook.

Meta reported on October 13, 2023, that it was taking temporary steps to lower the threshold at which it takes action against potentially violating and borderline content across Instagram and Facebook to avoid recommending this type of content to users in their feeds. However, these measures increase the margin of error and result in false positives flagging non-violative content.

In reviewing hundreds of cases of content removal and the inability of certain users to post comments on Instagram and Facebook, Human Rights Watch found Meta's automated moderation tools failed to accurately distinguish between peaceful and violent comments. After multiple experiences with seemingly automated comment removal, users reported being less likely to engage with content, express their opinions, or participate in discussions about Israel and Palestine.

Meta should improve transparency about where and how automation and machine learning algorithms are used to moderate or translate Palestine-related content, including sharing information on the classifiers programmed and used, and their error rates.

It should also conduct due diligence to assess the human rights impact of temporary changes in Meta's recommendation algorithms in response to the October 7, 2023, attack, and share those findings publicly. This assessment and reporting should become standard practice whenever Meta applies temporary measures in crisis situations.

Implications of Meta's Spam Policies

Human Rights Watch has documented numerous cases where this slogan, as well as comments such as "Free Palestine," "Ceasefire Now," and "Stop the Genocide," were repeatedly removed or restricted by Instagram and Facebook under "spam" Community Guidelines or Standards without appearing to take into account the context of these comments. In many instances, appeal buttons did not work or were not available. An analysis by The Markup of Meta's treatment of Palestinian content under its spam policy similarly concluded that Meta "heavily demoted" legitimate content and "erratically suppressed hashtags," among other problems. These statements and the context in which they are used are clearly not spam, nor do they appear to violate any other Facebook or Instagram Community Guidelines or Standards. For instance, the words in each of these statements on their face do not constitute incitement to violence, discrimination, or hostility.

Repeated posting of slogans and higher post rates is behavior that may share attributes to spam. However, it is precisely what activism and protest look like in the digital sphere. Meta should review its spam policy to ensure that it is not being enforced in a manner that interferes with people's ability to peacefully express themselves and participate in social movements. It should particularly consider including more safeguards against misuse of its spam policies against political expression during times of crisis.

In order to prevent the suppression of political advocacy using "From the river to the sea", the Oversight Board's guidance to Meta should take into account the impact of its spam policy on viral political engagement that uses this and related slogans, particularly at times of crisis.