

## Oversight Board Public Comment: From the River to the Sea Input from the Integrity Institute

### Overall summary

We, members of the [Integrity Institute](#), agree with Meta's original decision that **these phrases do not violate Meta's policies** prohibiting content that promotes violence or supports terrorism, and that the identified phrase **does not constitute hate speech, nor is it antisemitic, nor is it a call to abolish the state of Israel.**

We recommend removing the term "from the river to the sea" as a content moderation signal entirely. As with the Oversight Board's [evaluation of the term "shaheed."](#) our concern is that **a policy around this phrase specifically is likely to facilitate significant over-enforcement of content across Meta's diverse platforms.**

From the technical side:

- The current content moderation systems are not equipped to manage an addition to the policy that can properly contextualize the use of this phrase.
- Any automated systems will learn to over-enforce, either based on the term "from the river to the sea" or other non-unique terms on the dangerous individuals list.
- Content moderators are also likely to over-enforce systematically, especially if they cannot confidently identify any names in posts with individuals.
- Over time, any lists of dangerous individuals would get polluted by benign content, as platforms err on the side of recall.

When cultural specificity meets the day-to-day operations of content moderation:

- A policy that still uses "from the river to the sea" as a content moderation signal would lead to scenarios where content critical of violent individuals is over-flagged.
- There remain systemic issues around how we designate "dangerous" individuals and organizations and incitement of violence, and this systematic bias is reflected in the interpretation of "from the river to the sea."
- There are clear mistakes in removing content that is incorrectly conflated with dangerous organizations and individuals. For example, in May 2021, Al-Aqsa was [blocked](#) as a hashtag and was mistakenly linked to the Aqsa Martyrs Brigade.

**Moderating this content impacts freedom of expression on Instagram and Facebook, especially for civil society, journalists, and human rights defenders in regions where the word is commonly used.**

Moderating the content would result in false positive removals of content from news providers, spiritual guidance, or individuals marking moments of cultural, personal or religious importance. This inhibits critical discourse and could be perceived as unfair bias.

Additionally, removing this phrase from Meta's platforms would further exacerbate the global conflict and undermine the global conversation on this critical issue, which could delegitimize Meta's role in this space as a neutral arbiter.

**It is not feasible to prohibit this content, given the complicated historical context and diverse usages.** There are Palestinian Israelis and also Israelis who live and work in the West Bank, with strong differences in use depending on the person's context, and global communities using the terms in diverse, and sometimes competing, ways.

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

The phrase, “From the river to the sea” has multiple origins, from both Palestinian and Jewish Israeli sources. This includes, listed chronologically:

- In the 1960s, the Palestine Liberation Organization (PLO) [used the phrase](#) to call for an Arab state encompassing the entirety of Mandatory Palestine. Then, by 1969, after several revisions, the PLO [used the phrase](#) to call for a single democratic state for Arabs and Jews, that would replace Israel.
- More recently, the dominant Israeli political party Likud has consistently used the framing in their political jargon. Specifically, in 1977, the concept appeared in an [election manifesto](#) of the Likud, which stated that “between the sea and the Jordan (River) there will be only Israeli sovereignty” (more [here](#)).
- Hamas also used the phrase in its [2017 charter](#).
- In recent years, many protestors and activists [have referenced the phrase](#) as a call for peace and equality after decades of Israeli military rule over Palestinians, while for some others [view the phrase](#) as a call for the destruction of Israel.
  - This usage does not intend the phrase as a threat necessarily to all Jewish peoples or a statement of hate. It is a statement recognized in support of the right to belong and stay physically where Palestinian peoples are undergoing and have undergone mass- displacement from their homelands.
  - For example, in a [2018 incident](#), scholar and activist Marc Hill used the phrase to “center Palestinians’ aspirations, not disparage Israelis.”

While the phrase was recently added to Hamas’ political charter, it has a [much longer and diverse history](#). This phrase does not rhyme in the Arabic languages and was not widely adopted in those contexts. There are multiple, competing understandings of this phrase, as [Vox](#), [Jewish Currents](#) and [Al Jazeera](#) research shows.

## **State and institutional (e.g., university) responses to the use of the phrase (e.g., during protests) and the human rights impacts of those responses.**

We caution the privileging of state and institutional responses found in predominantly English written scholarship and news on this topic, as it a) applies to peoples where English is not their native language, and b) these institutions are geographically disparate from the context with their own institutional histories and biases that reinforce structural forms of violence.

Further, historically, institutional responses [have consistently proscribed](#) protest language as a threat or acts of terrorism.

We encourage the Oversight Board to move beyond human rights frameworks and consider how moderating this content will likely contribute to conflict and add to existing structural global inequity. Meta should avoid taking universal policy determinations that privilege a western-centric outlook shared only by a minority of the world’s population, and then applying these positions to the rest of the world.

## **Research into the connection between restricting language on social media and the effective prevention of violence.**

Restricting praise of individuals associated with “dangerous organizations” on social media does not necessarily result in the prevention of violence.

Minority populations often [bear the brunt](#) of sweeping policy changes.

[Research and empirical evidence](#) suggest such solutions tend to **negatively impact everyday users more than bad actors** who adapt to bypass naive content-based filters through coded language; solutions like 'ethical scaling' provide better suggestions.

**There are too few content moderators speaking underserved languages.** For example, Meta had only [one Burmese-speaking content moderator](#) to monitor the posts by 1.2 million active Burmese users in 2014, and action was only taken against "2% of the hate speech on the platform" in 2019.

In this context, it is striking that this policy language, and the [Oversight Board's page](#) for soliciting public comment, are only available in English and Hebrew. By failing to provide important policy content in relevant languages, particularly those represented in the region - namely Arabic - Meta and the Oversight Board are excluding critical and relevant demographics from participating in this important process of "[having your say today and helping \[the Oversight Board\] hold Meta accountable!](#)"

## On Zionism

Zionism is a political movement that [originated in Europe](#).

There are varying definitions of Zionism, [related to](#) the cause of Jewish self-determination and the security of a physical "homeland," though interpretations of this ideology also vary within the Jewish community. For example, on average, young Jewish people [hold different perspectives](#) on Zionism than older Jewish people. One [2022-2023 study](#) of the Jewish community in Greater Portland and Southwest Washington finds that "Twenty-six percent of Jewish adults in Greater Portland explicitly describe themselves as Zionists, 52% explicitly say they are not Zionists, and 22% either do not know or prefer not to say whether they are Zionists."

## On Watermelons

In the second case, the depicted watermelon is reflecting a deep historic symbol of Palestinian symbol of resistance. "The watermelon has long been an emblem of Palestinian solidarity and resistance in the occupied territories where displays of the Palestinian flag are often restricted or banned by Israel," states A Martínez on [NPR](#).

The choice of the watermelon as a symbol of Palestinian resistance not just because the colors resemble the flag, but also because this fruit ties in the significance of agriculture and the [land](#). It is grown throughout Palestine, from Jenin to Gaza.

## Beyond the binary moderation

As one broad reflection on the moderation toolkit, we recognize that there exists a more [expansive set of moderation options](#) than the binary "block" or "no block" decisions that Meta currently utilizes. There are other ways of moderating content, including deamplifying content or removing content from virality.

***A note: This analysis was prepared by members of the [Integrity Institute](#), a community of integrity professionals advancing the theory and practice of the social internet. It does not reflect the views of all members, nor does it reflect the official viewpoint of the Integrity Institute as an organization.***