From the River to the Sea Oversight Board Comment
The phrase “from the river to the sea, Palestine will be free” and its variations are at the heart of the vigorous debate over the current conflict in Gaza. It is a common refrain both online and offline by supporters of the Palestinian cause while it has been frequently condemned by supporters of Israel, including in notable exchanges between members of Congress and presidents of Ivy League universities. There are several challenges to moderating this term, including interpreting a range of meanings often involving implicit and coded speech, difficulty navigating issues of neutrality and bias, and consistency with existing policies and Board decisions that remove other types of speech that are arguably less harmful. The key takeaway for the Board is that even for content that is viewed as hateful or harmful, there should be a closer and clear connection to actual harms to justify its removal. Otherwise, the Board and Meta risks interjecting bias and uneven policies based on how offensive some may find certain types of speech.

Handling Content with a Range of Meanings
While other experts will likely provide in-depth analysis and history of this phrase, I will merely highlight several meanings to show how a blanket policy approach would be inappropriate.

From a common Israeli and Jewish perspective, “From the river to the sea, Palestine will be free” is a clear call for a “Palestinian state extending from the Jordan River to the Mediterranean Sea, territory that includes the State of Israel, which would mean the dismantling of the Jewish state.” The resulting Palestinian state would deny Israelis self-determination and force them to face repression, genocide, or ethnic cleansing at the hands of a regime that is fundamentally hostile to Israeli and Jewish identities and is empowered by terrorist factions such as Hamas. As such, this view sees this phrase as a call for the destruction of not just the Israeli state, but also the people living in Israel.

On the other hand, a common interpretation from the Palestinian perspective is that the phrase represents an aspirational call for Palestinian political rights in the region. Supporters of this view would point to the Arab rejection of the U.N.’s two-state partition plan and the subsequent war in 1948 in which hundreds of thousands of Palestinians were displaced from their homes. At the time, Arab leaders and Palestinians expressed a desire for a single secular state rather than two states. From the river to the sea thus became a desire for many to return to their homes and to have rights and protection under a single state.

Issues of Neutrality and Bias
Both views are likely correct. Some uses of the phrase certainly reflect an antisemitic call for the destruction of Jews living between the river and the sea. Others certainly use the term as an aspirational desire for political freedom and human rights. And some are likely somewhere in between, desirous of freedom but also open to or supportive of harming Israelis and Jews to achieve their aspirations. Still, others are likely uninformed uses of the term to express general solidarity with the Palestinian cause, with little consideration for what the term actually means.

These multiple meanings and variations in actual intent mean that any at-scale policy cannot allay the concerns of both sides. A policy to restrict the use of the phrase will silence political speech while the status quo will allow those with violent intentions to cloak their words with clever doublespeak.

When multiple meanings are present, there will always be pressure to view one interpretation as the correct one. This, however, invites bias. I have a personal view on the use of this term as do my former colleagues working to set policy at Meta, as do members of the Oversight Board. Our beliefs and experiences all bias us toward interpreting this phrase one way or another. But as I noted, it's likely that the term is used in a multitude of ways with a variety of intentions.

If a user posts this phrase together with other violating content—clear calls for violence against Jewish or Israeli targets, praising Hamas or October 7, etc.— then their intentions are clear and Meta will remove such content under existing policies. Restricting speech in this case, however, will assert that the Board can ascertain the intention and impact of the phrase. In the absence of clear evidence that a given word or phrase is predominantly used as a call for violence against people, the Board should favor greater expression and not remove the term.

The Board may be tempted to look at narrower ways to address the use of this term, such as an on-escalation policy. However, I would similarly advise caution here. The Board in its Drill Rap case noted that policies handled on escalation and with input from local experts or law enforcement to determine when content is a veiled threat, are open to bias and government abuse. This is certainly true in this case.

What external experts could neutrally assess when the use of this phrase was meant as a coded call for violence vs. a political rallying cry? And if no external experts can be trusted to provide their views on how this content is connected to violence, the task would somehow fall on Meta’s teams to understand the specific contexts in which this phrase is being used, the intent of the speakers using this term, and to adjudicate its nexus to harm without external expertise. This would not be feasible.

Consistent Norms that Maximize Speech

The last challenge to consider here is that while I believe this phrase is not sufficiently connected to clear physical harm, it's worth considering how a decision in this case is or is not consistent with the norms around other forms of potentially harmful speech.

Take, for example, one of the Board’s more recent cases involving a French politician who made a comment about “Africa colonizing Europe.” While the Board said the content should be allowed, it apparently was a close call as the rest of the decision was dominated by a minority opinion that argued strongly for removing such speech. The decision cites and discusses several key precedents from prior Oversight Board decisions:

- that implicit attacks against vulnerable groups should be removed;

- that “cumulative harms of hate speech” justify removal “even when the expression does not directly incite violence or discrimination”;
- posts should be assessed “in context according to the way they are likely to be understood, even if their incendiary message is couched in language designed to avoid responsibility”; and
- hate speech should be removed given that the accumulation of such content “creates an environment where acts of violence are more likely to be tolerated.”

Using these same lines of argumentation, there is a strong case for removing the phrase “from the river to the sea.” The phrase can be, and is by many people, understood as a barely or not veiled call for genocidal violence against a vulnerable protected group (Jews, Israelis), which presents a significant cumulative harm even if it is not directly inciting violence or discrimination. According to many, this phrase expresses incendiary messages that are couched in language that avoids responsibility, even as it is directly invoked by Hamas (and other terrorist groups) in its core documents and by its leaders. This phrase, then, is reasonably connected to creating an environment where the violence of October 7 was more than tolerated—it was supported, planned, executed, and defended.

The real challenge the Board faces in this case, then, is why is non-violent but prejudiced speech (what many call lawful but awful speech) considered as or more harmful than a phrase that can be understood as a call for mass violence against Israelis? Can the board meaningfully make the case that Hamas’s prominent use of this phrase is less connected to harm than the existence of Great Replacement theories in the manifestos of the perpetrators in Christchurch or Buffalo? Or that this phrase’s cumulative use is less connected to harm than hate speech in various cases—notably including holocaust denial— where the Board found the cumulative harms of non-violent speech justified the removal of content as a matter of international human rights law?

To be clear, I don’t think “from the river to the sea” should be removed from the platform for the reasons I describe in the prior sections. The Board is free to make its own decisions and Meta is free to choose how it sets its policies. But the case underscores the difficulty the Board faces in setting broadly applicable norms for speech and why before it calls for content to be removed, it should insist on a clearer connection between content and the real, tangible harm it will supposedly inflict. This is especially true in cases where hate speech and violence against protected characteristics are alleged. In such cases, it is easy to protect a sympathetic group or viewpoint but more difficult to hold the same position when defending the noxious speech of unsympathetic actors.

I make these comments in my individual capacity.