

Summary of Stakeholder Engagement Roundtables
Referring to Designated Dangerous Individuals as “Shaheed”
Policy Advisory Opinion

In addition to an open call for public comments for this policy advisory opinion, the Oversight Board conducted additional stakeholder engagement through roundtables.

Three regional roundtables were held, prioritizing geographies where “shaheed” and its variants are commonly used, either in Arabic or as a loanword:

- Southwest Asia and North Africa roundtable: April 13, 2023
- Sub-Saharan Africa roundtable: April 27, 2023
- South/Southeast Asia roundtable: May 9, 2023

Two thematic roundtables were held:

- Counterterrorism and human rights roundtable: May 2, 2023
- Automation and risk assessments roundtable: May 4, 2023

Approximately 140 people participated in the roundtables. Attendees included linguistic and religious scholars, representatives of human rights, digital rights and media organizations, counterterrorism and national security experts, and representatives of governments and inter-governmental organizations. The Board sought expertise from impacted communities, in particular people from places where “shaheed” or its variants are commonly used and who are impacted by content moderation of the term, and people impacted by terrorist violence and the effects of counterterrorism measures. The roundtables were held on the understanding that views shared would not be attributed to individuals.

Southwest Asia and North Africa (SWANA)

This roundtable focused on the impact of Meta’s prohibition of “shaheed” referring to designated individuals, how the term is used and understood across the SWANA region, if Meta’s policy requires reform and how, and what information the company should provide to civil society and the general public to understand and measure the accuracy of policy enforcement in this area. There were about 35 participants.

Participants emphasized the negative impact of the prohibition on free expression, including on political speech and everyday discourse. They noted that “shaheed” may be used to refer to deaths in many contexts including state violence, war, conflict and even natural disasters. For

example, people who died in the February 2023 earthquakes in Turkey and Syria could be called “shaheed.” One participant noted that for the average user, the effects of moderation can include self-censorship. Several participants emphasized that Meta’s moderation was replicating human rights-violating government censorship in the region.

Participants also emphasized that the policy’s impacts were amplified by the way designated organizations and individuals are often enmeshed in the social and political fabric of countries in the region. For example, in Lebanon, Hezbollah is a part of everyday political discourse, yet even coverage critical of the group may get removed. Many participants mentioned moderation of the term “shaheed” in the context of discussions about Israel and Palestine. One participant noted that when they reached out to ask their network about this issue before the roundtable, every Palestinian reported they had some of their posts removed, usually in the context of reporting information.

Many participants’ comments reflected the perception that Meta’s reliance on the U.S. government’s terrorist designations resulted in groups and individuals in the Southwest Asia and North Africa region being extensively designated and potentially over-represented in global lists; many suggested this over-representation showed bias. There was some discussion about the need to publish this list.

Sub-Saharan Africa

Participants discussed the variety of ways in which the term “shaheed” is used, for example to refer to a family member killed in an accident. One participant emphasized that different forms of the word, such as “shahidi” in the Hausa language, must be considered as well. This roundtable was organized in partnership with a civil-society organization that monitors social-media use and misuse in Africa. There were approximately 20 participants.

Participants noted that in 2020 the cousin of the Nigerian prime minister died and he was widely referred to as a “shaheed” on social media. In another example, the term was used online to discuss an individual who killed a politician. Some posts called the murderer “shaheed,” but others expressed disgust that he was described this way. Posts on both sides were removed.

In the context of Sudan, participants discussed how anyone who was killed in the 2019 protests was referred to as a “shaheed.” Facebook was a critical outlet for human-rights documentation, yet people struggled with Meta’s over-enforcement. There were concerns that people’s ability to communicate about ongoing violence in Sudan was being impacted by the policy.

A participant with counterterrorism expertise acknowledged the complexities of the many meanings of the word, while also stating that Meta should not completely cease screening it. In

their opinion, Meta has already pushed many terrorist groups off Instagram and Facebook. They recommended that if Meta wanted to consider the word “shaheed,” it should also look for additional signifiers such as guns or blood. This was echoed by a participant who is a content moderator and who noted that the current policy is too strict, with moderators needing context to enforce it correctly.

South and Southeast Asia

Participants reinforced many of the observations of participants in the Sub-Saharan Africa and SWANA roundtables. They emphasized the complexity around the meaning of the word “shaheed” and the many ways in which it is used, as well as the need to remove content when it is genuinely connected to offline harms, but not when used in everyday life without posing such risks. There were approximately 30 participants for this roundtable.

Participants discussed how “shaheed” can be used neutrally in the South Asian context, for example as part of a title attached to a name. One participant focused on the use of the word in Indonesia where it is translated as “mati syahid,” with the plural form “mati syahid.” They pointed out there are eight ways to die as “mati syahid” and only one is used to refer to dying in a battlefield, whereas other examples include dying during childbirth or from plague.

In the Indian context, a participant observed the word “shaheed” is also used in Hindi, emphasizing that India has the second-largest Muslim population in the world, so loanwords are common. They shared uses of “shaheed” in Hindi that were similar to other contexts. For example, they noted the word is culturally and politically associated with the movement for independence from British colonial rule and that people who died in that struggle are often referred to as “shaheed.”

Several participants raised the Indian state of Kashmir, which they noted is a highly controlled information environment. They shared [the 2016 example](#) of how human-rights defenders, newspapers and everyday users had their accounts suspended and content removed on Facebook for discussing the death of Hizbul Mujahideen commander Burhan Wani in Kashmir. Participants noted that social media companies removed content that discussed his death in sympathetic terms, but also news reporting about Burhan Wani that was not praise.

Similarly to the SWANA roundtable, participants pointed out that by moderating content that uses the word “shaheed” to refer to designated individuals who regard themselves as involved in struggles for political independence, Meta may replicate government censorship. They emphasized that “shaheed” should not be the sole reason for removal, which should only take place if the context and remaining content violates other rules of the Community Standards. Many participants emphasized the need for greater transparency from Meta on its content

moderation decision-making and noted that “shaheed” should not be used as a flag at all until that happens.

Counterterrorism and Human Rights

The Board invited two different expert presentations from an international civil society organization that researches violent extremism and provides policy advice to governments and institutions. Their presentations focused on social and cultural context for the term “shaheed” and a review of its use among Al-Qaeda and Islamic State supporters.

One presentation described how people utilize “shaheed” in myriad contexts. Since the term is used throughout the Quran, over-regulation of it can lead to removals of content reciting the Quran. The expert noted that in their opinion, it is almost impossible to automate this kind of regulation. The other expert presented the results of a review of a small sample of social-media content indicating that while the word “shaheed” was sometimes present, it was not central to terrorist propaganda on Facebook. However, there was a notable lapse in content moderation. For example, an entity that was on a version of Meta’s leaked designation list had an ongoing presence on the platform. The expert recommended that Meta publish its Designated Organizations and Individuals list, provide researchers and civil society with access to anonymized data on related takedowns and, finally, commission research into whether there is a link between “shaheed” content on the platform and offline harm.

In the discussion about the impact of the policy, one participant emphasized the broad and deep misuse of counterterrorism law and practice to target civil-society and human-rights defenders, urging the Board not to compound these problems. Another noted they were concerned about the potential danger of publishing the Designated Organizations and Individuals list, but thought there might be ways to release a limited set of examples. Another participant observed the term is used to describe Muslims killed in a 2019 terrorist attack on a mosque in New Zealand.

The discussion also focused on the mechanics of enforcing the policy. Both expert presenters and participants pointed out that Dangerous Organizations and Individuals have a sophisticated understanding of platform tactics and might try to “game” the rules. Participants wanted to know how Meta uses keyword-based filters, machine-learning classifiers and other technology. They expressed concern these tools were not in line with human rights standards and noted that dangerous content could likely come down under other policies. An expert in content moderation pointed out that, especially with decreasing budgets and capacity at social media platforms, automation will continue to play a central role in this area. Therefore, they noted that asking for better tests before the technology is deployed, building in nuance and ensuring policy exceptions are correctly enforced is essential to ensuring maximum accuracy of these systems.

Automation and Risk Assessments

There were around 30 attendees for this roundtable, organized in collaboration with two civil society organizations from Arabic-speaking countries. Both organizations have a long record of working on human rights and technology issues.

They presented suggestions for recommendations, which they drafted as input to the Board's stakeholder engagement, with participants discussing and refining these recommendations. The recommendations suggested that in consideration of the current perceived failings in automated content moderation, Meta should stop removing all "praise" of designated entities under the Dangerous Organizations and Individuals policy, focusing only on support and representation. That would allow Meta to meet its legal obligations while avoiding the collateral damage on important expression the current policy is, in their view, inflicting. Recommendations acknowledged that at a pragmatic level, it is unlikely Meta will stop deploying automated tools to detect and remove violating content, and suggested ways these tools could be improved, including auditing and co-design between civil society and the engineers designing automated enforcement methods.

Another recommendation called on Meta to better explain how it ensures its automated systems moderating content at-scale are designed and deployed to respect human rights. The suggested recommendations asked Meta to conduct human rights due diligence both before implementing new technology and periodically on existing automated enforcement.

Some of the recommendations focused on addressing how Meta's automated enforcement of its policies plays out in global majority countries, and called on Meta to employ content moderators with sufficient knowledge of the contexts, languages and dialects they are meant to moderate. Others encouraged Meta to ensure its policies, including "spirit of the policy" or other exceptions, are publicly explained.

Several recommendations focused on the need for transparency and researcher access to data, including on automated enforcement. They reiterated the importance of complying with the [Santa Clara Principles](#) on Accountability and Transparency in Content Moderation, and called on Meta to allow researcher access to data about automated systems in a privacy protecting way. Finally, they called for meaningful third-party auditing of automated systems, and one participant noted this could potentially happen in line with the auditing and data access provisions of the Digital Services Act.

Further suggestions built upon recommendations in previous Board decisions, such as pushing for greater transparency of the Designated Organizations and Individuals list, better explanations

for users when action is being taken against them, clarifying the strikes system and establishing an efficient system for appeal.