Meta's Quarterly Update on the Oversight Board

# Q3/2022 Key Highlights

This Quarterly Update covers Meta's case referrals, Policy Advisory Opinion (PAO) requests, as well as Meta's progress implementing product, policy, and operations recommendations issued by the board from Q1 2021 through Q3 2022. Meta has responded to all 130 board recommendations made through Q3 2022 publicly and we have committed to implement or we are exploring the feasibility of implementing 75% of the recommendations issued to date.

### Meta's Commitment to the Board

On July 22, 2022 Meta approved a new three-year, \$150 million commitment to fund the Oversight Board and its operations. This money was provided to the Oversight Board Trust, which enables the board to continue operating as a separate entity, independent from Meta.

# Quarter by the Numbers

- Meta hosted 16 briefings, 12 subject matter specific and 4 Q&A sessions to help inform the board's deliberations on cases and PAOs
- Meta submitted 5 content referrals and the board accepted 1 concerning sexual harassment in India
- This past quarter, Meta completed taking action on 11 recommendations from the board

## Taking Action on Recommendations:

Because of the board's recommendations, we made significant improvements this quarter across Transparency, Policy & Enforcement. Some highlights include:

• Policy Violation Notifications:

We launched new notifications to inform people which Community Standard their content violated in response to a number of recommendations made by the board<sup>1</sup>. This messaging is available globally in English, and translation work is currently underway to make the messaging available in all languages used on our platforms.

<sup>&</sup>lt;sup>1</sup> Armenians in Azerbaijan Recommendation #1, Breast Cancer Symptoms and Nudity Recommendation #3, Nazi Quote Recommendation #1, Depiction of Zwarte Piet Recommendation #2, South Africa Slur Recommendation #1, Post Discussing a Substance with Psychoactive Properties Recommendation #2, and PAO on Sharing Private Residential Information #17

#### • Updates to our <u>Restricted Goods and Services policy</u>:

We added <u>definitions</u> of the terms "non-medical drugs" and "pharmaceutical drugs" to our Restricted Goods and Services policy in response to the <u>board's recommendation (#1)</u>. Relatedly, we also held a Policy Forum to explore expanding our approach to "positive discussion" of non-medical drugs in traditional or religious contexts in response to a <u>separate recommendation (#3)</u>.

#### • Human Rights Due Diligence:

We published the <u>human rights due diligence</u> that we commissioned to fulfill the <u>Oversight Board's</u> <u>recommendation (#3)</u> that Meta should engage an independent entity to analyze Meta's impact on the Israeli-Palestinian conflict and identify any biases in Meta's moderation of content in Arabic and Hebrew.

#### • Policy Review, Sexual Exploitation & Nudity:

We initiated an in-depth review for our approach to preventing the functional identification of victims of sexual violence in our <u>Child Sexual Exploitation</u>, <u>Abuse and Nudity</u> and <u>Adult Sexual Exploitation</u> policies to respond to the board's <u>recommendations</u>. As part of this review we will gather input from external and internal experts as well as research to inform policy development with the goal to discuss it at a future <u>Policy Forum</u>.

#### • Dangerous Individuals and Organizations Policy:

In response to <u>recommendations from the board</u>, we continued our review of "praise" within the context of this policy with the goal of allowing for more user voice without sacrificing safety in a way that can be equitably operationalized at-scale.

## Scope Expansion

#### • Warning Screens:

The scope of the board has expanded to now include the ability to issue binding judgment on cases involving applying a warning screen to photo and video content. Now, if the board determines that content should be restored or remain on our platforms, it can also issue a binding judgment about whether that content qualifies for the application or removal of a warning screen. Warning screen types include marking content as 'disturbing' or 'sensitive.'

Click here to see the full report

