

September 2024

# Meta Update: Israel and Palestine Human Rights Due Diligence

TABLE OF CONTENTS

|   |          |
|---|----------|
| <b>Background</b>                                 | <b>2</b> |
| <b>Meta Recommendations Implementation Update</b> | <b>4</b> |
| Overview  | 4        |
| Update to our response                            | 5        |

## Background

Meta<sup>1</sup> commissioned [Business for Social Responsibility \(BSR\)](#) to carry out an independent Human Rights Due Diligence (HRDD) Exercise on Israel and Palestine for the period of May 2021. This work was scoped as rapid human rights due diligence rather than a full Human Rights Impact Assessment in order to enable swift project launch and implementation. Although planning for rapid due diligence began prior to the Oversight Board's [September 2021 recommendation in the Al Jazeera case](#), the exercise seeks to be responsive to the Oversight Board's recommendation to:

Engage an independent entity not associated with either side of the Israeli-Palestinian conflict to conduct a thorough examination to determine whether Facebook's content moderation in Arabic and Hebrew, including its use of automation, have been applied without bias. This examination should review not only the treatment of Palestinian or pro-Palestinian content, but also content that incites violence against any potential targets, no matter their nationality, ethnicity, religion or belief, or political opinion. The review should look at content posted by Facebook users located in and outside of Israel and the Palestinian Occupied Territories. The report and its conclusions should be made public.

The HRDD was conducted by BSR in 2021 and 2022 in line with the United Nations Guiding Principles on Business and Human Rights. BSR is an organization of sustainable business experts that works with a global network of the world's leading companies to build a just and sustainable world. BSR's methodology and findings are published in the Insights and Recommendations Report of the Due Diligence Exercise.

---

<sup>1</sup> On October 28 2021, Facebook, Inc. changed its name to Meta Platforms, Inc. For consistency, this report uses "Meta" to refer to the company both before and after October 28 2021. References to "Facebook" apply only to the social media platform, not the company as a whole. Further, this response references steps taken, or plans to take steps, by Meta as a company regarding a specific entity. Such a statement is not intended to imply that Meta would, or will, take steps regarding all entities. No statement in this response is intended to create — or should be construed as creating — new obligations (legal or otherwise) regarding the application of a policy or procedure to other products or entities. For example (and in contrast to other Meta technologies), WhatsApp is an end-to-end encrypted messaging and calling application with unique human rights touchpoints. This response's discussion of content moderation and related issues on Facebook and Instagram do not apply to WhatsApp. Unless a policy or commitment is specified as applying to WhatsApp, it does not apply to WhatsApp.

The HRDD made 21 prioritized recommendations. In September 2022, Meta committed to implement 10 recommendations and partly implement four others. It committed to assess the feasibility of another six. Meta declined to take further action in relation to one recommendation.

In providing an update on the status of our response to the recommendations made in the HRDD, we used the following categories, which are the same as the ones we use for [reporting](#) on implementation of Oversight Board recommendations:

- **Complete**: We have completed full or partial implementation of our commitments in response to BSR's recommendation. We will not share future updates on these items.
- **In progress**: We are continuing to make progress, and will have further updates in the future.
- **No further updates**: We have already addressed the recommendation through an action we already do, or will not implement the recommendation. We will not share future updates on these items.

In our 2023 update we reported that five recommendations were complete and ten were in progress. We also noted that we would not provide further updates on six recommendations: four were already completed at the time of our [September 2022 response](#); there was one for which we assessed feasibility, but decided that we would not take further action; and one that we declined to implement when responding to BSR's report.

| Implementation status update as of September 2023 |                                   |                            |  |
|---|-----------------------------------|----------------------------|--|
| BSR recommendations                               | Recommendations fully implemented | Implementation in progress | No further updates                               |
| 21  | 5                                 | 10                         | 6 (of which 4 were completed by September 2022). |

# Meta Recommendations Implementation Update

## Overview

This is an update on Meta’s progress in implementing the 10 outstanding recommendations from BSR’s HRDD, during the period June 30, 2023 to June 30, 2024. This update is part of our [broader commitment](#) to meaningful transparency about our human rights due diligence and about our integrity work.<sup>2</sup>

Of the ten remaining recommendations outstanding, we had completed implementation of five of them by June 30 2024. Implementation of five recommendations was still in progress.

| Implementation status update as of September 2024 |                                   |                            |  |
|---|-----------------------------------|----------------------------|--|
| BSR recommendations                               | Recommendations fully implemented | Implementation in progress | No further updates                               |
| 21  | 10                                | 5                          | 6 (of which 4 were completed by September 2022). |

<sup>2</sup> Meta’s publication of this response should not be construed as an admission, agreement with, or acceptance of any of the findings, conclusions, opinions, or viewpoints identified by BSR, or the methodology employed to reach such findings, conclusions, opinions, or viewpoints. Likewise, while Meta references steps it has taken, or plans to take, that may correlate to points BSR raised or recommendations it made, these also should not be construed as an admission, agreement with, or acceptance of any findings, conclusions, opinions, or viewpoints.

## Update to our response

**Recommendation 1:** Review whether Meta should create policy measures for content that praises or glorifies violence (including indiscriminate attacks, such as violence that is not targeted at any particular person or group).

Meta commitment: **Implementing**

Status update: **Complete**

*We conducted a policy development process - including stakeholder engagement - on glorification of violence in different areas of content policy. Acknowledging the complex balance between political expression and safety, we made changes to our at-scale policies on Violence & Incitement on gender-based violence, but otherwise concluded that existing at-scale policy lines covered this challenging area appropriately. We will continue to make decisions on other types of content on-escalation upon adequately weighing the safety risks.*

*In addition, in January 2024, we announced an [update](#) to our Dangerous Organizations and Individuals ('DOI') Policy. This policy is key to addressing some of the most serious threats to the safety of our users and platform. We do not allow organizations or individuals that proclaim a violent or hateful mission, or are engaged in violence, to have a presence on Meta.*

*We heard feedback that our previous definition of "Praise" was too broad. We redefined our policy in order to make it more nuanced and proportionate. We now use the term 'glorification' instead. More details about the policy and how we define entities and relevant terms can be found [here](#).*

*We have also [recently](#) updated our Dangerous Organizations and Individuals policy to allow for more social and political discourse in certain situations, and to ensure users are not unduly penalized for sharing such discourse. These situations include (among others) peace agreements, elections, human rights related issues, and academic discussion. This update was prompted by discussion and feedback from internal and external stakeholders.*

**Recommendation 3:** Review the practice of designating deceased historical individuals under the DOI Policy and assess feasibility of alternative policy approaches to improve transparency and fairness.

Meta commitment: **Implementing**

Status update: **Complete**

*In our January 2024 DOI policy [update](#), we announced that we had revamped our policy for de-designating dangerous organizations and individuals so that it covers all DOI categories and is triggered by an entity's demonstrated behavioral change. To be considered for delisting, an entity must (1) not be designated by the U.S. government as a Specially Designated Narcotics Trafficking Kingpin (SDNTK); Foreign Terrorist Organization (FTO); or Specially Designated Global Terrorist (SDGT); (2) no longer be involved in violence or hate; (3) not be symbolic to violence and hate or be used to incite further violence or spread hateful propaganda. The policy also applies to designations of deceased historical individuals.*

*As noted above, we have also updated our DOI policy to allow for more social and political discourse in certain instances. We believe that this should help mitigate the impacts on speech that arise as a result of designations of historical individuals. We have also committed to [implementing in full](#) the recommendations of the Oversight Board Policy Advisory Opinion on referring to designated dangerous individuals as “Shaheed”.*

**Recommendation 4:** Tier the designation system and strikes for DOI violations to take into account who the organization or individual is and what the violation is (praise, support, or representation) so that the strike is proportional to the violation.

Meta commitment: **Assessing feasibility**

Status update: **Complete**

*As part of our policy review process of our definition of “praise” we examined how our strike system could be more proportionately applied. This is in line with feedback from the Oversight Board and other human rights experts.*

*In an effort to simplify the way in which we designate organizations or individuals under our DOI policy, and in order to ensure more effective and consistent enforcement, we recently announced that we had divided the designations into two tiers.*

- *[Tier 1](#) focuses on entities that engage in serious offline harms including organizing or advocating for violence against civilians, repeatedly dehumanizing or advocating for harm against people based on protected characteristics, or engaging in systematic criminal operations. We do not allow Glorification, Support, and Representation of these Tier 1 entities, their leaders, founders or prominent members, as well as unclear references to them when the user’s intent was not clearly indicated. In addition, we do not allow content that glorifies, supports, or represents events that Meta designates as violating*

*violent events. For example, terrorist attacks, multiple-victim violence or attempted multiple-victim violence, serial murders, or hate crimes.*

*We have improved proportionality by ensuring unclear references to designated Tier 1 events or ideologies are removed without a strike against the user's account.*

- *[Tier 2](#) includes Violent Non-State Actors that engage in violence against state or military actors in an armed conflict but do not intentionally target civilians. It also includes Violence Inducing Entities that are engaged in preparing or advocating for future violence but have not necessarily engaged in violence to date. These are also entities that may repeatedly engage in violations of our Hate Speech or Dangerous Organizations and Individuals policies on or off the platform. We remove Glorification, Material Support, and Representation of these entities, their leaders, founders and prominent members.*

*We have improved proportionality to allow references and non-material support to Tier 2 designated entities, mitigating risks of over-enforcing on speech about this category of entities.*

Recommendation 5: Provide users with a more specific and granular policy rationale when strikes are applied. This should not just include the category of the violation, but how a post was violating, so that users can better understand the justification, submit an informed appeal, and be less likely to post violating content in the future.

Meta commitment: **Implementing in part**

Status update: **In progress**

*In our [September 2023 Israel Palestine Human Rights Due Diligence update](#), we noted that we already provide specific granular reasoning when content is removed and strikes are applied in the vast majority of cases and that we were working to expand this to the limited number of areas where we do not yet provide this specificity.*

*Given the scale of our enforcement and the fact that users sometimes violate multiple policies, we are sometimes limited in the specificity we can provide when applying strikes.*

*In addition to the update provided in September 2023, we have also launched a wide-ranging effort to harmonize our policies into a more standardized structure across Meta surfaces to improve our enforcement accuracy and increase clarity for people on our platforms. This effort includes aligning definitions across policies to create more efficient and clear internal guidance for policy enforcement. These harmonization efforts also help support updated, and in some cases more granular, user messaging. We are also updating our external policy documents for further clarity. We are assessing the*



*feasibility of providing more granular user messaging under our DOI policy by the end of H2 2024.*

**Recommendation 8:** Determine the required market composition (e.g. headcount, language, location) for standby or rapid response capacities for Hebrew and Arabic markets.

Meta commitment: **Implementing in part**

Status update: **Complete**

*In our [September 2023 Israel Palestine Human Rights Due Diligence update](#), we noted our continued commitment and approach to ensuring correct resource investment to address critical events on a sustainable basis. Our commitment and approach remain unchanged.*

*In response to the terrorist attack on October 7th and the war in Gaza, we were able to route Arabic content more appropriately across our systems to enable greater precision. We also increased our content moderation resources in the Hebrew market.*

*We will continue to review the size of this workforce relative to business and regional needs as these evolve.*

**Recommendation 9:** Continue establishing mechanisms to better route potentially violating Arabic content by dialect/region.

Meta commitment: **Assessing feasibility**

Status update: **In progress**

*In our [September 2023 Israel Palestine Human Rights Due Diligence update](#), we said that we were developing mechanisms to efficiently route content by Arabic dialect. We have determined following an extensive internal assessment that creating specialized routing for a variety of Arabic dialects across our systems would contribute to a greater precision in our Arabic content moderation for high severity content. This work is ongoing.*

**Recommendation 14:** Continue plans to disclose the number of formal reports received from government entities (including the Israel State Attorney Office (ISAO) in Israel) about content that is not illegal but potentially violates Meta content policies. This should

take place either quarterly (as part of the Community Standards Enforcement Report) or every six months (as part of the Content Restrictions Report).

Meta commitment: **Implementing**

Status update: **In progress**

*Meta remains committed to implementing this recommendation. As we shared in our [September 2023 Israel/Palestine Human Rights Due Diligence update](#) and most recent [Quarterly Update on the Oversight Board](#), we are still in the process of developing consistent and reliable systems for gathering metrics on the number of pieces of content removed under the Community Standards as a result of government requests. We continue to evaluate approaches to building the necessary internal data logging infrastructure to enable us to publicly report this information across the diversity of request formats we receive.*

*As noted previously, this is a complex, long-term project whose implementation will take time in part due to the need to assess existing reporting and other regulatory reporting requirements. We will endeavor to provide an update on the timeline for public reporting of these metrics in a future Oversight Board Quarterly Update and in our next annual Human Rights Report.*

**Recommendation 16:** Develop a mechanism to track the prevalence of content that attacks based on specific protected characteristics (e.g., antisemitic, Islamophobic, homophobic content). This might involve, for example, prompting users to mark relevant hate speech content with tags.

Meta commitment: **Assessing feasibility**

Status update: **In progress**

We continue to assess the feasibility of this recommendation: We tested a potential method, Race Data Measurement, to assess whether we can measure differences in product experiences across race and ethnicity. Following that testing, we are exploring further options to assess whether product and system differences exist across race and ethnicity and to accurately disaggregate attacks based on specific protected classes, while balancing privacy concerns. We will consult with representative groups and experts as we move forward.

**Recommendation 18:** Increase the capacity of Meta’s special escalation channels via more staff and more resources to enable sufficiently prompt response to escalations from Trusted Partners, governments, and other actors in both normal times and times of crisis.

Meta commitment: **Implementing in part**

Status update: **In progress**

*At the end of 2023 we conducted a comprehensive review of our civil society escalations channels and consulted with external partners about how to improve the program. In line with the findings of our internal review and feedback received from external partners, we are implementing key changes to the program, which include clarifying the scope of these escalation mechanisms and providing improved tools to facilitate partner reporting and increase the efficiency of review processes. We believe that these changes will lead to an improved rapid response to escalations in line with the spirit of the recommendation.*

*Reporting from our Trusted Partners across the region following the October 7th attacks and throughout the war in Gaza has been significant. Trusted Partners play a critical role in reporting content trends and helping us keep our communities safe.*

*We continue onboarding Trusted Partners in the region as needed and in line with our onboarding criteria.*

**Recommendation 21:** Separate and apart from existing data and law enforcement policies, develop new methods or policies to enable Meta to store content where Meta is under no legal obligation to preserve but where the content may hold potential use for a rightsholder in future remedy processes.

Meta commitment: **Assessing feasibility**

Status update: **Complete**

*We’ve worked since 2019 to explore rights-respecting initiatives for evidence retention and disclosure, consulting extensively with civil society, academia, and international prosecutorial experts and bodies.*

*As we [publicly stated](#) in response to recommendations from the Oversight Board, we worked for several years to develop an approach to allow international courts and accountability mechanisms to make requests to us for extended retention of data that is relevant to their ongoing investigations. This is a novel area without established or tested best practices, and there remain significant legal, privacy, and policy considerations*

*inherent to this work. This work, which is distinct from our [longstanding policies for responding to preservation requests from law enforcement](#), is now completed subject to ongoing learning and refinement as appropriate.*

*We will look for opportunities to share further updates on our work in this area in our [Quarterly Updates on the Oversight Board](#) and our annual human rights reporting.*