September 2022

Meta

Meta Response:
Israel and Palestine Due
Diligence Exercise

TABLE OF CONTENTS

Overview

Meta[1] 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.

This 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.

---

[1] 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.

This disclosure is part of our [broader commitment](#) to meaningful transparency about our human rights due diligence, and about our integrity work.[2]

Acknowledgements

We are deeply grateful to the numerous Israeli, Palestinian and international human rights defenders, civil society organisations, and others who provided input to this human rights due diligence. You have enabled us to make progress on the very important mitigations described in this response, benefiting people in Israel, Palestine and related diasporas around the world.

---

[2] 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.

Meta Response to Recommendations

Overview

The HRDD made 21 prioritized recommendations. At the time of writing, Meta had committed to implement 10 recommendations and partly implement 4 recommendations, and was assessing the feasibility of another 6. Meta declined to take further action in relation to one recommendation. We will provide an update on the status of our implementation efforts in our 2023 Annual Human Rights Report.

For ease of reading, we have categorized our response to HRDD recommendations to match the order in which they appear in the HRDD executive summary. We are categorizing our responses as follows:

> **Implementing:** We have implemented or are implementing steps that are consistent with, or have otherwise satisfied, the recommendation.

> **Implementing in part:** We have implemented or are implementing steps that encompass, or have satisfied, certain of the recommended actions.

> **Assessing feasibility:** We are assessing the feasibility and impact of the recommendation and will provide further updates in the future.

> **No further action:** We will not implement the recommendation, either due to a lack of feasibility or disagreement about how to reach the desired outcome.

Responses

#	BSR Recommendation	Meta Commitment	Considerations
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).	Implementing	Our policies already cover praise of violence related to dangerous organizations and individuals, and we already aim under our Violence & Incitement Policy to take action to prevent potential content-related offline harms. We are also currently exploring refinements to our policies around interpersonal violence and definitions of praise, support, and representation of dangerous organizations. These work streams will likely address some concerns around praise and glorification of violence and will inform future work streams in this area. Any resulting policy updates would be in compliance with our legal obligations in this area.
2	Review whether Meta should limit the Dangerous Individuals and Organizations Policy to "support" or "representation" only.	Implementing	We have started a policy development process to review our definitions of praise, support and representation in our DOI Policy, as well as to explore possible related approaches to social and political discourse. Any resulting policy updates would be in compliance with our legal obligations in this area.

#	BSR Recommendation	Meta Commitment	Considerations
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.	Implementing	We will review the practice of historical designations, and assess the feasibility of alternative approaches. This process will require internal and external consultations with experts in order to understand the offline impact of online praise, support and representation of deceased historical figures.
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.	Assessing Feasibility	We are working on ways to make user experiences of our DOI strikes simpler and more transparent. As noted above, we are also exploring policy revisions, including to our definitions of praise, support, and representation in our DOI Policy. This work is underway.

#	BSR Recommendation	Meta Commitment	Considerations
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.	Implementing in Part	We already provide specific granular reasoning when content is removed and strikes are applied in the majority of cases.

Work is underway 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. |
| 6 | Increase transparency about Meta's enforcement actions—such as feature limiting and search limiting—and communicate enforcement actions clearly to users. | Assessing Feasibility | Currently, when a strike is applied, we provide detail on the specific policy area that was violated, accompanied by relevant examples.

In addition to this, we are assessing the feasibility at scale of providing users with granular notifications of the feature limitations we put in place when strikes are applied. |

#	BSR Recommendation	Meta Commitment	Considerations
7	Publish key elements of internal community operations resources that help content moderators interpret and apply Meta's content policies so that users can better understand and abide by the policies, excepting adversarial content.	Assessing feasibility	We are committed to continued improvement in our transparency efforts. We will review providing additional detailed guidance on how we interpret our policies. We regularly update our Transparency Center so people can better understand and abide by the policies.
8	Determine the required market composition (e.g., headcount, language, location) for standby or rapid response capacities for Hebrew and Arabic markets.	Implementing in Part	We are committed to ensuring correct resource investment to address critical events on a sustainable basis. We review market composition regularly, including incorporating insights from human rights due diligence, with this in mind. We are already assessing how to create a more flexible, diverse and agile workforce, including to reflect that recruiting and training teams with the necessary linguistic skills and cultural competency involves time investments that do not always align with emergency events.

#	BSR Recommendation	Meta Commitment	Considerations
9	Continue establishing mechanisms to better route potentially violating Arabic content by dialect/region.	Assessing Feasibility	We are already exploring routing Arabic content by dialect to reviewers with expertise in that particular dialect. We have conducted research to improve our approach to routing content across the many diverse dialects of Arabic. Based on this research, we're reviewing and testing a range of options to address this recommendation, including hiring more content reviewers with diverse dialect and language capabilities. Our goal is for these steps to assist with routing.
10	Assess whether it is feasible and desirable to create a dialect-specific Arabic classifier, working in partnership with Arab linguists and language model experts.	Implementing	We are committed to updating our classifiers to increase their accuracy and performance, including, where possible, taking dialects into account. Our teams have begun experimentation on building a dialect-specific Arabic classifier. This classifier would help identify the variety of Arabic dialects in which content is written to assist with routing.
11	Continue work on having functioning Hebrew classifiers.	Implementing	We have now launched a Hebrew classifier that proactively detects and actions violating content in Hebrew. As noted above, we are committed to updating our classifiers to regularly improve accuracy and performance.

#	BSR Recommendation	Meta Commitment	Considerations
			We note that it is more challenging to maintain accurate classifiers in less widely spoken languages. This is because large amounts of language data are needed to train classifiers.
12	Adjust the process that allows staff at outsourced providers to add keywords to blocklists to ensure approval by relevant Facebook FTEs.	Implementing	We have now implemented processes for keywords to be raised by outsourced providers to internal expert teams for assessment and approval for addition. We will continuously review and update relevant processes.
13	Develop a vetting / oversight / quality control process for new additions to hashtag / keyword blocklists.	Implementing	We have now implemented this. Internal teams supporting specific policy areas are responsible for updating and maintaining keywords within Meta's tooling features. We will continuously review and update relevant processes.

#	BSR Recommendation	Meta Commitment	Considerations
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 which 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).	Implementing	We are actively seeking ways to expand our transparency reporting around government requests to remove or restrict content. In 2021, in response to a recommendation by the Oversight Board, we committed to reporting five country-level metrics, including the number of: (1) unique requests we receive; (2) pieces of content covered by these requests; (3) such pieces of content removed under the Community Standards; (4) such pieces of content locally restricted based on local law; and (5) such pieces of content where no action is taken. Because this is a large, complex project, we do not anticipate publishing these new metrics in 2022. We will provide updates on our progress in future Quarterly Updates on the Oversight Board and our annual human rights reporting.
15	Assess the review accuracy of the DOI Policy enforcement in Arabic across both internal and outsourced teams, and including both machine and human based review, and address findings	Implementing	We have a robust accuracy program in place, including for the Arabic market, to help ensure the decisions made by both automated systems and human reviewers are correct. As part of this program, we consistently strive to improve the accuracy of the enforcement of our policies, including reviewing decisions made by human and automated review, and adjusting our operations accordingly.

#	BSR Recommendation	Meta Commitment	Considerations
	(BSR notes this is an ongoing effort).		This is a continuous process, and extends across all markets and policy areas.
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.	Assessing Feasibility	We are committed to tracking the overall prevalence of hate speech on our platform, but face challenges measuring prevalence with this level of granularity. We note that subjective assessments run certain profiling risks.
17	Establish a structure, protocol, or team to gauge over and under content policy enforcement in a systematic manner during a crisis.	Implementing	We actively monitor both active content removals and appeals. In a crisis, we may deploy our Integrity Product Operations Centers model to monitor and respond to these trends and other threats in real time. This includes closely monitoring potential over- and under-enforcement of our Community Standards.

#	BSR Recommendation	Meta Commitment	Considerations
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.	Implementing in Part	We are committed to strengthening partnerships with expert civil society organizations that can help us understand content-related risks. Increased resourcing for these specialized escalation channels is underway.
19	Engage in stakeholder engagement and prepare public transparency statement(s) regarding Meta's understanding of its Foreign Terrorist Organization (FTO) and Specially Designated Global Terrorist (SDGT) obligations.	Implementing in Part	We rely on legal counsel and relevant sanctions authorities to understand our compliance obligations. We also regularly review our policies and explore updates to strike an appropriate balance between voice and safety, while complying with our legal obligations. For example, we are reviewing policies as per recommendation 2. As with all policy reviews, we are committed to carrying out broad stakeholder engagement.

#	BSR Recommendation	Meta Commitment	Considerations
20	Fund public research into the optimal relationship between legally required counterterrorism obligations and the policies and practices of social media platforms. This would address questions such as how the concept of material support for terrorism should be interpreted in the context of social media and whether governments should establish different regulations or interpretations for social media companies.	No Further Action	Like other social media platforms, we rely on legal counsel and relevant sanctions authorities to understand the company's legal obligations.

Legal advice is an important foundation to our DOI Policy. As with other legal advice, we do not direct or fund legal guidance for other companies.

As we review our policies, we will continue to consider the insights of diverse experts and related research. We encourage experts to engage with the sanction authorities that administer sanctions regulations for further guidance. |

#	BSR Recommendation	Meta Commitment	Considerations
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.	Assessing Feasibility	We are assessing the feasibility of actions in this area, as reported to the Oversight Board in May 2022. There is complex interplay between identification, preservation, and disclosure of content; international accountability process; and our privacy and legal obligations.