

# Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021

# Insights and Recommendations | September 2022

#### **PURPOSE**

This human rights due diligence exercise reviews the impact of Meta's¹ policies and activities during the May 2021 crisis in Israel and Palestine. The primary purpose is to provide Meta with prioritized, action-oriented, and decision-useful recommendations for policies and practices to fulfill Meta's commitments under its <a href="Corporate">Corporate</a> <a href="Human Rights Policy">Human Rights Policy</a>, and responsibilities under the United Nations Guiding Principles on Business and Human Rights (UNGPs), including Principles 20-22.² To meet this brief, BSR herein:

- Identifies Meta's human rights impacts, 3 including its responsibilities under the UNGPs.
- Reviews Meta's responsibility to address these human rights impacts according to the UNGPs.
- Makes recommendations for policies and practices that adhere to these responsibilities.

This human rights due diligence helps fulfill the <u>recommendation of the Oversight Board</u> that Meta should engage an independent entity not associated with either side of the Israeli-Palestinian conflict to determine whether Meta's content moderation in Arabic and Hebrew has been applied without bias.

BSR has identified good practice, areas for improvement, and lessons learned by Meta during May 2021. It is to Meta's credit that it is undertaking this exercise, to learn what went well and what can be enhanced and addressed. We also note that Meta is already making progress on many of BSR's recommendations.

BSR and Meta would like to thank all stakeholders who took the time to participate in this due diligence.

<sup>&</sup>lt;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 report references actions taken by Meta as a company regarding a specific entity. Such a statement is not intended to imply that Meta took that action regarding all entities. For example, unless a policy is specified as applying to WhatsApp, it does not apply to WhatsApp.

<sup>&</sup>lt;sup>2</sup> For instance, Principle 20 states that companies should track the effectiveness of their response to human rights impacts by engaging with internal and external stakeholders, integrating findings into relevant internal reporting processes, and driving continuous improvement. Principle 21 states that companies should communicate externally when human rights concerns are raised by stakeholders and provide information that is sufficient to evaluate the adequacy of the company's response. Principle 22 states that companies should provide for or cooperate in the remediation of adverse impacts, and this may include seeking to guarantee non-repetition of prior harms.

<sup>&</sup>lt;sup>3</sup> The term "human rights impacts" is used in line with the UNGPs, which defines an "adverse human rights impact" as occurring when an action removes or reduces the ability of an individual to enjoy his or her human rights. The term "adverse human rights impact" is not equivalent to--and does not imply--a legal obligation or breach, or legal causation. See <a href="https://www.ohchr.org/sites/default/files/Documents/Publications/HR.PUB.12.2\_En.pdf">https://www.ohchr.org/sites/default/files/Documents/Publications/HR.PUB.12.2\_En.pdf</a>.

## 2. SCOPE AND LIMITATIONS

The scope of this due diligence review is limited to the May 2021 crisis and does not explore Meta's role in Israel and Palestine more broadly. All relevant Meta family of apps—Facebook, Instagram, and WhatsApp—are within scope. This report summarizes the human rights due diligence exercise that BSR started in September 2021 and completed in April 2022.

# 3. METHODOLOGY

This human rights due diligence review is based upon: publicly available information; interviews with relevant Meta staff and a review of related materials; the perspectives of a wide variety of affected rightsholders and stakeholders in Israel, Palestine, and globally obtained through interviews with BSR and public written communications between affected stakeholders and Meta; relevant data, including analyses of "jobs" (i.e., cases) in Arabic and Hebrew originating from Israel or Palestine; analysis of enforcement actions by human review and automated review; and review of individual cases of content removal and account restrictions.

BSR's methodology included a review of data internal to Meta. However, it is important to note that there is no established "ground truth" for what absolute or relative rates of content enforcement should be in Israel and Palestine, as no data is available for the prevalence of violating content at the country level (as opposed to the market level, such as the Arabic market). The data reviewed by BSR was shared by Meta to help BSR understand the on-platform manifestations of the May 2021 crisis in Israel and Palestine, and validate trends and insights we gathered through qualitative research; however, while we have referred to the use of data to inform our analysis, the data does not meet the quality, confidence, and assurability requirements for detailed and formal external disclosure.

#### 4. CONTEXT

Meta's policies and their implementation took place in the context of a highly complex set of social and historical dynamics. Those dynamics include international and regional politics and power asymmetry on the one hand, and Meta's content policy<sup>4</sup> and content policy enforcement on the other.

Meta's actions during the events of May 2021 cannot be viewed in isolation but must be understood in the context of the ongoing conflict in Israel and Palestine, which is characterized by competing historical narratives of victimization and persecution; the conflation of civilian populations and diaspora communities with the apparatus and actions of state governments and terrorist organizations; modern power asymmetries in which the Israeli state has greater administrative, financial, and military might vis-a-vis Palestinian political institutions; and global ramifications for both the perception and treatment of Jewish communities and the Palestinian diaspora outside the region.

An outbreak of violence in the ongoing Israeli–Palestinian conflict took place in May 2021, triggered by protests in East Jerusalem over the eviction of Palestinian families in the Sheikh Jarrah neighborhood. This outbreak occurred in the context of Israeli occupation of the West Bank and increased tensions relating to the expansion of Israeli settlements and the eviction of Palestinian communities.<sup>5</sup> The violence included some protests, riots, and police riot control inside Israel, indiscriminate rocket attacks by Hamas on Israel, and Israeli airstrikes targeting

<sup>&</sup>lt;sup>4</sup> BSR uses the generic term "content policy" to encompass Facebook Community Standards and the Instagram Community Guidelines.

<sup>&</sup>lt;sup>5</sup> Israel's policy of settling citizens in the occupied Palestinian territories and displacing Palestinian communities has been declared a violation of international law by <u>several UN bodies</u>.

the Gaza Strip. The violence prompted protests around the world, and largely ended on May 21 following a ceasefire agreement between Israel and Hamas the day before.<sup>6</sup>

People on all sides of the conflict, both regionally and globally, have suffered and understand themselves to be victims of different historical events. Meta's role is not to arbitrate this conflict, but rather to generate and enforce policies to mitigate the risk that its platforms aggravate it by silencing voices, reinforcing power asymmetries, or allowing the spread of content that incites violence.

The UNGPs lay out the expectation that Meta should avoid infringing on the human rights of others and should address adverse human rights impacts (as defined in footnote 3) with which it is involved. In a conflict-affected context like Israel and Palestine, this includes understanding how the ongoing conflict intersects with Meta's platforms (e.g., via competing narratives about both history and current events), how the online actions of a range of actors, including Meta, are possibly shaping offline events (e.g., via hate speech and incitement to violence that mobilizes offline harm), and which groups are particularly vulnerable given the context.

Key events and milestones during this period are as follows:

#### **Broader Context**

Mid-April: Sheikh Jarrah property dispute, evictions, and protests

May 6: Rising tensions in West Bank and East Jerusalem with the death of two Palestinians

May 7: Israeli police enter Al-Aqsa Mosque during prayers

May 8: Clashes between protestors and Israeli police

May 10 onwards: Further escalation, including protests, violence, and rocket strikes

May 16: Al-Jazeera building in Gaza struck by an Israeli airstrike

May 20: Israel and Hamas agree to ceasefire

#### Meta

April 23: Jerusalem is made a "Temporary High-Risk Location."<sup>7</sup>

May 5-6: Global technical glitch preventing the posting of Instagram stories containing re-shares of posts (including, but not limited to, Israel and Palestine) is reported and addressed

May 11: Al-Aqsa hashtag block is reported, and work begins on understanding the cause; Meta publishes reason for error on May 12

May 12: Dedicated crisis response team activated

May 13: Temporary High-Risk Location status expanded to all of Israel, the West Bank, and Gaza

May 14-20: Letters received from civil society organizations expressing concern with restrictions on Palestinian speech and increases in antisemitic content, and content contributing to violence on all sides

<sup>&</sup>lt;sup>6</sup> See <u>Israel and Hamas Begin Cease-Fire in Gaza Conflict</u>, <u>Israel-Gaza ceasefire holds despite Jerusalem clash</u>, <u>Timeline of the Israeli—Palestinian conflict in 2021</u>, and description of the <u>2021 Israel—Palestine crisis</u>

<sup>&</sup>lt;sup>7</sup> A "temporary high-risk location" is a designation used by Meta where certain stricter content policies are applied to a defined geographical area, most notably restricting calls to bring armaments to locations where there are temporary signals of a heightened risk of violence or offline harm, such as known protest, a counter-protest, or recent violence. In addition, Facebook's veiled threats policy, which covers coded statements where the method of violence or harm is not clearly articulated, and which requires significant understanding of context to enforce, is more rigorously applied in a "temporary high-risk location".

May 21: Palestinian journalists report blocking of their WhatsApp accounts
May 27: Dedicated crisis response team closed
June 24: Meta publishes new information on how Praise, Support and Representation are defined and interpreted in Dangerous Individuals and Organizations (DOI) policy

#### 5. ANALYSIS

#### **UNGP-Aligned Actions**

Meta took many appropriate actions during the May 2021 crisis, including establishing a special operations
center/crisis response team, prioritizing risks of imminent offline harm, seeking an approach to content
removal and visibility based on necessary and proportionate restrictions consistent with the International
Covenant on Civil and Political Rights (ICCPR) Article 19(3), and overturning enforcement errors in response
to user appeals. Some of BSR's recommendations to Meta build upon important foundations for a human
rights-based approach to content governance, such as these, that have already been established by Meta.

#### Challenges and Insights

- From the interviews conducted and the data reviewed, hate speech and incitement to violence against
  Palestinians, Arab Israelis, Jewish Israelis, and Jewish communities outside the region occurred on Meta
  platforms.
- Based on the data reviewed, examination of individual cases and related materials, and external stakeholder engagement, Meta's actions<sup>8</sup> in May 2021 appear to have had an adverse human rights impact (as defined in footnote 3) on the rights of Palestinian users to freedom of expression, freedom of assembly, political participation, and non-discrimination, and therefore on the ability of Palestinians to share information and insights about their experiences as they occurred. This was reflected in conversations with affected stakeholders, many of whom shared with BSR their view that Meta appears to be another powerful entity repressing their voice that they are helpless to change.
- The data reviewed, examination of individual cases and related materials, and interviews with both internal and external stakeholders all identified both over-enforcement (erroneously removed content and erroneous account penalties) and under-enforcement (failure to remove violating content and failure to apply penalties to offending accounts) of Meta content policies during May 2021, especially the Dangerous Individuals and Organizations (DOI) and Violence and Incitement (V&I) policies.<sup>9</sup>

<sup>&</sup>lt;sup>8</sup> Consistent with UNGPs Principle 13, BSR interprets "actions" as "actions and omissions."

<sup>&</sup>lt;sup>9</sup> As noted above, BSR's methodology did not attempt to identify what should be the absolute or relative rates of content enforcement in Israel and Palestine. The terms over-enforcement and under-enforcement therefore mean only that instances were identified of erroneously

- The data reviewed indicated that Arabic content had greater over-enforcement (e.g., erroneously removing Palestinian voice) on a per user basis (i.e., adjusting for the population size difference between Arabic and Hebrew speakers in Israel and Palestine). Data reviewed by BSR also showed that proactive detection rates of potentially violating Arabic content were significantly higher than proactive detection rates of potentially violating Hebrew content, which can be likely attributed in large part to Meta's policies which incorporate certain legal obligations relating to designated foreign terrorist organizations, and the fact that there was an Arabic hostile speech classifier but not a Hebrew hostile speech classifier. Based on indications in the data and examination of individual cases, under-enforcement also occurred (e.g., incitement to violence against Israelis, praise of Hamas, including by Palestinian political authorities), though BSR notes that underenforcement is more challenging to measure than over-enforcement.
- Meta has no specific metrics for under-enforcement because it is challenging to conclusively measure. However, the materials reviewed and internal interviews indicated that Hebrew content experienced greater under-enforcement, largely due to the lack of a Hebrew classifier and the loss of Hebrew-speaking FTEs and outsourced content moderators in the weeks leading up to May 2021. However, an examination of individual cases showed that over-enforcement also occurred, with Israeli accounts having content erroneously removed and receiving erroneous account restrictions.
- Based on the materials reviewed, BSR identifies the following possible root causes for over-enforcement that Meta should investigate further:
  - Arabic classifiers<sup>10</sup> may have higher error rates for Palestinian Arabic.
  - Potentially violating Arabic content may not have been routed to content reviewers who speak or understand the specific dialect of the content.
- A substantial increase in case volume during May 2021 (up to 10x on peak days was reported by markets staff to BSR and illustrated in the data BSR examined) presented significant challenges to the effective content policy enforcement during the crisis. According to internal stakeholders, Meta did not have sufficient Arabic and Hebrew-speaking reviewers to handle the spike.
- Based on BSR's review of tickets and input from internal stakeholders, a key over-enforcement issue in May 2021 occurred when users accumulated "false" strikes that impacted visibility and engagement after posts were erroneously removed for violating content policies. 11 The human rights impacts (as defined in footnote 3) of these errors were more severe given a context where rights such as freedom of expression, freedom of association, and safety were of heightened significance, especially for activists and journalists, and given the prominence of more severe DOI policy violations. Further, these strikes remain in place for those users that did not appeal erroneous content removals.

removed content, erroneous account penalties, failure to remove content, or failure to apply penalties. BSR did not consider or draw any conclusions with respect to whether Meta's enforcement complied with any legal obligations.

<sup>&</sup>lt;sup>10</sup> A classifier is an algorithm that identifies and sorts content into types of content—for example, proactively identifying content with a high likelihood of violating Meta's policies.

<sup>&</sup>lt;sup>11</sup> In Meta's general approach to content policy enforcement, violations considered more severe (such as violations of the DOI policy) result in longer restrictions or additional restrictions, such as creating ads, while violations considered less severe (such as incitement to violence, hate speech, or bullying and harassment) involve lesser restrictions, such as reduced searchability of the account (i.e., requiring users to enter the exact name of the users account in order to find it rather than a normal keyword search) or reduced content visibility (i.e., placing content lower in feeds). Meta notifies users when their searchability has been impacted, but not when content visibility is reduced. When users successfully appeal a content removal and Meta reverses its decision, the associated strikes and penalties are removed.

- Reviewed materials and input from internal stakeholders revealed a lack of oversight at Meta that allowed content policy errors with significant consequences to occur. One key example during this crisis shared with BSR was that #AlAqsa was added to a hashtag block list by an employee in Meta's Outsourced Services when pulling from an updated list of terms from the US Treasury Department containing the Al Aqsa Brigade, resulting in #AlAqsa being hidden from search results. The hashtag #AlAqsa in fact had been used extensively in posts referring to the Al-Aqsa Mosque, one of the holiest sites in Islam.
- Many stakeholders reported to BSR that users reported experiencing reduced distribution of content during the crisis period. While assessments of content visibility are complex, BSR discussed this question with internal stakeholders and identified the following possible factors that could have contributed to this: 1) enforcement penalties for alleged content policy violations that affected searchability or visibility of content; 2) Meta implemented some content-neutral "break-the-glass" measures for Israel and Palestine during May 2021 designed to mitigate the risk of online-to-offline harm that intentionally reduced the visibility of all repeatedly reshared content; 3) users experienced two Instagram glitches that affected the reach of stories globally on May 5-6.
- The May 2021 crisis brought to the forefront questions around the contours and details of praise for and glorification of violence—as distinct from Meta's Policy on Violence and Incitement and similar policies, which prohibit "language that incites or facilitates serious violence." Specifically, Meta should consider whether its policies sufficiently address praise for and glorification of indiscriminate violence, meaning violence that is not targeted at any particular person or group.
- Stakeholders raised concerns about antisemitic content on Meta platforms. Antisemitic content is a type of hate speech, and so it often falls under Meta's hate speech policy for Instagram and Facebook—however, this policy applies to all types of hate speech, and does not clearly delineate different categories of hate speech, and does not contain a full definition of these. At present, Meta does not track (i.e., label or count) specific types and targets of hate speech, just the "tier", and therefore has no metrics to understand the prevalence of antisemitic content—including whether or not it increased in May 2021. While BSR could not independently verify the data, organizations dedicated to antisemitism tracked clearly antisemitic content that violated Meta's policies that were not detected and removed during this period. BSR's analysis suggests that there was insufficient cultural competency on the part of content moderators and those setting policy had an insufficient linguistic capacity in the range of languages (including small European languages) in which antisemitic content has appeared.<sup>12</sup>
- External stakeholders interviewed by BSR reported cases where WhatsApp was used by right-wing Israelis in Israel to incite violence and coordinate attacks against both Arab and Jewish Israelis, as well as against Israeli journalists.
- Some journalists and academics had their WhatsApp accounts unintentionally disabled after correct
  enforcement against designated terrorist organization groups for violation of WhatsApp policy. When
  WhatsApp was made aware that these accounts had been unintentionally disabled, they were reinstated to
  the service.

6

<sup>&</sup>lt;sup>12</sup> See for example reporting by World Jewish Congress, Fighting Online Antisemitism, and CST

- Stakeholders reported to BSR their view that Meta's DOI and V&I policies are poorly understood by users—for example, that praise of terrorist organizations and what constitutes incitement to violence may be challenging for users to understand.
- These factors all point to broader known themes in social media platform content governance, such as the importance of understanding geographic context and monitoring how it is changing, the need for a sufficient volume of staff with relevant language skills and cultural understanding, and the ongoing tension between the right to freedom of expression and efforts to enhance safety and prevent physical harm both on and off-platform. Measures that enable Meta to moderate content in relevant languages and ensure that both policy and policy enforcement guidance is informed by regional expertise reflecting diverse perspectives will be foundational to addressing these challenges. BSR emphasizes that many of the factors it has identified are likely to be prevalent across the social media industry and in other conflict-affected regions and that industry-wide approaches and multi-stakeholder dialogue to address these challenges will be beneficial.
- Based on all sources reviewed as part of this human rights due diligence exercise, the main adverse human rights impacts (as defined in footnote 3) included freedom of expression (e.g., from the over-enforcement of content policy), freedom of assembly and association (e.g., reducing the ability to organize and associate online), freedom from incitement (e.g., under-enforcement of content inciting violence), bodily security (e.g., under-enforcement of content organizing violence), non-discrimination (e.g., different impacts on Arabic speakers), and access to remedy (e.g., loss of access to content that Meta is under no legal obligation to preserve, but could potentially assist rightsholders in future processes.<sup>13</sup>)

#### Bias

- The Oversight Board sought a determination of whether Meta's content moderation in Arabic and Hebrew, including its use of automation, was applied without bias. The Oversight Board did not define "bias" in its recommendation, nor provide a governing framework.
- For this assessment, BSR considers both intentional bias (where some people are deliberately treated differently than others) and unintentional bias (where policies and processes may be neutral on their face, or in place for reasons of legal compliance, but impact some people differently than others). BSR also considered the distinction between bias at the content policy level and bias at the content moderation system level.
- From the interviews conducted by BSR and the data reviewed, BSR did not identify intentional bias at Meta writ large or among employees individually. BSR found no evidence of racial, ethnic, nationality or religious animus in governing teams and notes that Meta has diverse employees representing different viewpoints, nationalities, races, ethnicities, and religions relevant to this conflict. Further, BSR did not identify evidence that in developing or implementing policies, Meta intentionally sought to benefit or harm any particular group because of their race, religion, nationality, ethnicity, or any other protected characteristic.

7

<sup>&</sup>lt;sup>13</sup> For example, see UN Berkeley School of Law, Digital Lockers, June 2021.

- However, BSR did identify various instances of unintentional bias where Meta policy and practice, combined with broader external dynamics, does lead to different human rights impacts on Palestinian and Arabic speaking users.
- Meta, as a U.S. company, must comply with U.S. laws, including those regarding the provision of "material support" or resources to designated foreign terrorist organizations, where a "foreign terrorist organization" is a foreign organization that is designated by the U.S. Secretary of State in accordance with section 219 of the Immigration and Nationality Act (INA) (18 U.S.C. §2339B). Legal designations of terrorist organizations around the world have a disproportionate focus on individuals and organizations that have identified as Muslim, and thus Meta's DOI policy and the list are more likely to impact Palestinian and Arabic-speaking users, both based upon Meta's interpretation of legal obligations, and in error.<sup>14</sup>
- Palestinians are more likely to violate Meta's DOI policy because of the presence of Hamas as a governing
  entity in Gaza and political candidates affiliated with designated organizations. DOI violations also come with
  particularly steep penalties, which means Palestinians are more likely to face steeper consequences for both
  correct and incorrect enforcement of policy. In contrast to Israelis and others, Palestinians are prevented from
  sharing types of political content because the Meta DOI policy has no exception for the praise of designated
  entities in their governing capacity.
- Further, the contrasting status of Meta's content moderation system for Arabic compared with Hebrew may have resulted in unintentional bias via greater over-enforcement of Arabic content compared to Hebrew content, even when adjusting for population size. (Note that this does not account for potential differences in rates of violation for Hebrew vs. Arabic content, for which there was no data).
- BSR notes the following factors: (1) a possible insufficient routing of Arabic content by dialect or regional expertise, which would ensure that outsourced reviewers clearly understand the dialect of the content they are reviewing, as well as related cultural context that might be important for informing decisions; (2) the use of classifiers for Arabic content, whereas there has been no functional classifier for Hebrew content; and (3) Arabic classifiers are likely less accurate for Palestinian Arabic than other dialects, both because the dialect is less common, and because the training data—which is based on the assessments of human reviewers—likely reproduces the errors of human reviewers due to lack of linguistic and cultural competence. By contrast, Hebrew is a more standardized language and is predominantly spoken in Israel, meaning that reviewers for Hebrew content are both fluent in the language and highly likely to understand the context. As a result of these factors, the content moderation system may not be as precise for Arabic content as for Hebrew content.

#### 6. RECOMMENDATIONS

BSR has discussed observations and points for consideration with Meta, consistent with its responsibility under the UNGPs to take appropriate action to address its adverse human rights impacts. Our recommendations include:

<sup>&</sup>lt;sup>14</sup> We note that civil society organizations raised with BSR instances where Israeli violent extremist groups have not been added to the DOI list, as well as several historical individuals who committed acts of terror, raising some questions as to the methodology about how groups and individuals are added to the list. While this insight does not mean there should be an equal number of Palestinian and Israeli designated groups, the fact that known Israeli groups and individuals are not included may result in discrimination.

- 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).
- 2. Review whether Meta should limit the DOI policy to "support" or "representation" only.
- 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.
- 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.
- 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.
- 6. Increase transparency about Meta's enforcement actions—such as feature limiting and search limiting—and communicate enforcement actions clearly to users.
- 7. Publish key elements of Meta's 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.
- 8. Determine the required market composition (e.g., headcount, language, location) for standby or rapid response capacities for Hebrew and Arabic markets.
- 9. Continue establishing mechanisms to better route potentially violating Arabic content by dialect/region.
- 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.
- 11. Continue work on having functioning Hebrew classifiers.
- 12. Adjust the process that allows staff at outsourced providers to add keywords to blocklists to ensure approval by relevant Facebook FTEs.
- 13. Develop a vetting/oversight quality control process for new additions to hashtag / keyword blocklists.
- 14. Continue plans to disclose the number of formal reports received from government entities (including the 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).
- 15. Assess the review accuracy of the DOI policy enforcement in Arabic across both internal and outsourced teams, including both machine and human-based review, and address findings (BSR notes this is an ongoing effort).
- 16. Develop a mechanism to track the prevalence of content that attacks based on specific protected characteristics (e.g., antisemitic, Islamophobic, anti-Arab, homophobic content). This might involve, for example, prompting users to mark relevant hate speech content with tags.
- 17. Establish a structure, protocol, or team to gauge over and under content policy enforcement in a systematic manner during a crisis.

- 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.
- 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.
- 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.
- 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.

### **DISCLAIMER**

The conclusions presented in this document represent BSR's best professional judgment, based upon the information available and conditions existing as of the date of the review.

In performing its assignment, BSR relies upon publicly available information, information provided by Meta, and information provided by third parties. Accordingly, the conclusions in this document are valid only to the extent that the information provided or available to BSR was accurate and complete, and the strength and accuracy of the conclusions may be impacted by facts, data, and context in which BSR was not privy.

As such, the facts or conclusions referenced in this document should not be considered an audit, certification, or any form of qualification. This document does not constitute and cannot be relied upon as legal advice of any sort and cannot be considered an exhaustive review of legal or regulatory compliance.

BSR makes no representations or warranties, express or implied, about the business or its operations. BSR maintains a policy of not acting as a representative of its membership, nor does it endorse specific policies or standards. The views expressed in this document do not reflect those of BSR member companies.

#### **ABOUT BSR**

BSR is a sustainable business network and consultancy focused on creating a world in which all people can thrive on a healthy planet. With offices in Asia, Europe, and North America, BSR provides its 300+ member companies with insight, advice, and collaborative initiatives to help them see a changing world more clearly, create long-term value, and scale impact.

All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the publisher, except in the case of brief quotations embodied in critical reviews and certain other noncommercial uses permitted by copyright law.

Copyright © 2022 by Business for Social Responsibility (BSR)