

Annex I: Human Rights Watch Letter to Meta

Human Rights Watch

350 Fifth Avenue, 34th Floor
New York, NY 10118-3299
Tel: +1-212-290-4700
Fax: +1-212-736-1300; 917-591-3452

Tirana Hassan, Executive Director

Deputy Executive Director
Wisla Heneghan, DED/ Chief Operating Officer

Senior Leadership

Sari Bashi, Program Director
Lauren Camilli, General Counsel
Mei Fong, Chief Communications Officer
Abir Ghattas, Chief Information Officer (Acting)
Marusya Lazo, Chief Financial Officer
Joe Lisi, Chief People Officer (Acting)
Valentina Rosa, Chief Development Officer
James Ross, Legal and Policy Director
Bruno Stagno Ugarte, Chief Advocacy Officer
Minjon Tholer, Global Head of Diversity, Equity, and Inclusion

Division and Program Directors

Lama Fakih, Middle East & North Africa
Juanita Goebertus, Americas
Tanya Greene, United States
Elaine Pearson, Asia
Maus Segun, Africa
Hugh Williamson, Europe and Central Asia

Shantha Rau Barniga, Disability Rights
Elizabeth Evenson, International Justice
Bill Frellick, Refugee and Migrant Rights
Arvind Ganesan, Economic Justice and Rights
Steve Goose, Arms
Maria McFarland Sánchez-Moreno, Technology and Human Rights (Acting)
Zama Neff, Children's Rights
Richard Pearshouse, Environment and Human Rights
Graeme Reid, Lesbian, Gay, Bisexual, and Transgender Rights
Macarena Sáez, Women's Rights
Ida Sawyer, Crisis and Conflict

Advocacy Directors

Yasmine Ahmed, United Kingdom
Maria Laura Canineu, Brazil
Louis Charbonneau, United Nations, New York
Philippe Dam, European Union
Farida Delf, Canada
Kanae Doi, Japan
Meenakshi Ganguly, South Asia
Daniela Gavshon, Australia
Bénédictine Jeannerod, France
Iskra Kirova, Europe and Central Asia
Wenzel Michalski, Germany
Måns Mollander, Nordic
Allan Ngari, Africa
Hilary Power, United Nations, Geneva
Sarah Yager, Washington DC

Board of Directors

Amy Rao, Co-Chair
Neil Rimer, Co-Chair
Oki Matsumoto, Vice Chair
Amy Towers, Vice Chair; Treasurer
Catherine Zennström, Vice Chair
Bruce Rabb, Secretary
Akwas Aidoo
Lishan Aklög
George Coelho
Roberto Dañino
Kimberly Marteau Emerson
Louba Freih
Leslie Gilbert-Lurie
Paul Gray
Caitlin Hiesing
David Lakhdhir
Louisa Lee-Reizès
Alicia Mirfana
Gloria Principe
Ambassador Robin Sanders
Bruce Simpson
Joseph Skrzyński AO
Danna Slight, C.M.
Siri Stolt-Nielsen
Marie Warburg
Isabelle de Wismes
Masa Yanagisawa
Andrew Zoll

Nick Clegg
President, Global Affairs
Facebook Inc. (Meta)

Monika Bickert
Vice President, Content Policy
Facebook Inc. (Meta)

November 15, 2023

Dear Mr. Clegg and Ms. Bickert,

We are writing to you on behalf of Human Rights Watch to share the findings of research we have conducted on the censorship and suppression of content related to Israel and Palestine on social media since October 7, and to solicit Meta's perspective by responding to several specific questions that have stemmed from the research.

Human Rights Watch has been documenting online censorship related to Israel and Palestine since October 7 to November 15, a period during which we have observed takedowns or other suppression of content by Palestinians and their supporters, including about human rights abuses, on Instagram and Facebook in significant numbers and even when similar content by other users is not suppressed. In October, Human Rights Watch published a call for evidence of online censorship and suppression of content related to Israel and Palestine on social media since October 7, which we posted from the HRW main accounts on multiple social media platforms, including Instagram, Facebook, X, and TikTok.

We have gathered hundreds of cases which detail various forms of censorship impacting posts and accounts documenting, condemning, and raising awareness about the recent and ongoing hostilities in Israel and Palestine, which involve serious human rights abuses.

Examples include deleting posts, stories, and comments, and suspending accounts, at times without warning the user. We have also documented users' inability to engage with content, such as commenting, sharing, and reposting on stories, on Instagram and Facebook, and restrictions on users' ability to use certain features, such as Instagram/Facebook Live, for prolonged periods. The most widely cited reasons for these restrictions on content and accounts were Community Guidelines or Standards violations, specifically those relating to "dangerous organizations and individuals," "adult nudity and sexual activity," "hate speech," "violent and graphic content," and "spam." In dozens of cases we have documented, users have been unable to appeal content or account removal due to the appeal mechanism malfunctioning, meaning that they have no effective access to a remedy.



HRW.org

While Meta attributed some users' significantly reduced reach to reshared Reels and Feed posts not showing up properly in their Stories on Instagram and inability to go Live on Facebook to a "bug," and [reported](#) having fixed the issue on October 18, we have continued to receive cases of such restrictions after that date.

Another common complaint was "shadowbanning," specifically reports of reduction in the distribution or reach of content relating to developments in Israel and Palestine and disabling searches for accounts that post about Palestine. We note that Meta has [reported](#) taking steps to lower the threshold at which it takes action against potentially violating and borderline content across Instagram and Facebook to [avoid recommending](#) this type of content and to reduce the visibility of potentially offensive comments. We are concerned that these measures increase the margin of error and result in false positives flagging non-violative content. Furthermore, because the user is not informed of any action taken on their account or content, the user is left without access to a remedy. This censorship is ongoing as we are receiving daily cases documenting consistent patterns.

Based on our findings, we are concerned about the number of cases in which it appears that erroneous enforcement of Meta's policies disproportionately targets content in support of Palestine, effectively silencing Palestinian journalists, activists, and others, as well as voices raising awareness about human rights abuses in Palestine during a critical time.

In 2021, Human Rights Watch published a [report](#) which documented Facebook's censorship of the discussion of rights issues pertaining to Israel and Palestine, and called for the company to take up the Facebook Oversight Board's [recommendation](#) to commission an independent investigation. Facebook then commissioned an investigation into its human rights' due diligence in relation to the May 2021 events in Israel and Palestine. In September 2022, the Business for Social Responsibility (BSR) published its findings in the report "[Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021.](#)" Among other findings, the report concluded that Meta's actions in May 2021 "appear to have had an adverse human rights impact 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." The BSR [report](#) identified various instances where Meta policy and practice, combined with broader external dynamics, lead to different human rights impacts on Palestinian as well as other Arabic-speaking users.

We will publish a report on our findings in early December. We acknowledge the documents Meta sent us, specifically its [September 2023 update](#) on Israel and Palestine Human Rights Due Diligence and its [newsroom post](#) on ongoing work in response to the crisis. In order for our report to be as complete as possible, and to provide Meta with an opportunity to comment, we would appreciate your responses to the following questions in writing:

- 1) How does Meta explain the suppression of content related to Palestine detailed above? What steps, if any, does Meta plan to undertake to investigate whether it is arbitrarily censoring content related to Palestine?
- 2) What steps, if any, has Meta undertaken to consider and assess the impact of changes in its recommendation algorithms on Instagram and Facebook referenced in [its October 13 newsroom post](#) on the freedom of expression of its users? Specifically:

- Before adjusting or introducing temporary measures to take swift action on violating content, did Meta conduct an impact assessment of such measures on the free expression of Palestinian and pro-Palestinian voices on its platforms?
 - Can you confirm the [Wall Street Journal's](#) reporting that Meta lowered the threshold to hide comments from users in the Occupied Palestinian Territory if its automated detection systems determined there was at least a 25% chance they violated rules?
 - Given that such measures are expected to lead to erroneous takedowns of speech that doesn't violate Meta's policies, did Meta determine what error rates it deems to be acceptable?
 - Has Meta seen an increase in appeals of its content moderation decisions concerning Palestinian and pro-Palestine content? What percentage of these appeals have been effective?
 - Given that these measures are timebound, how will Meta decide when to lift them?
- 3) Has Meta made temporary changes to its [strikes policy](#) in response to the current hostilities? If so, please provide a revised policy, including which policy violations will result in which account restrictions.
 - 4) Has Meta considered notifying users that their content is being demoted or that its distribution is reduced based on policy considerations, in order to provide the user with the opportunity to appeal these decisions?
 - 5) What steps, if any, has Meta undertaken to determine the percentage of content flagged by automation for violating Meta's policies, including on dangerous organizations and individuals, that has been found to be in error? In these cases, what additional steps has Meta taken to provide users whose content was wrongfully removed with a remedy?
 - 6) Meta has [referred to violations](#) of its "most serious policies" as leading to restricting a violating user's ability to use Live on Instagram and Facebook for prolonged periods from their first offense, which Meta underpinned with the example of "someone who shares a link to a statement from a terrorist group with no context." However, Human Rights Watch has documented dozens of cases where Meta imposed such restrictions when users merely shared posts documenting hostilities in Gaza, sharing on-the-ground developments, or expressing solidarity with Palestinians. In these circumstances:
 - How does Meta explain the prevalence of such disproportionate restrictions?
 - Please provide a reasoned justification to what Meta considers to be its "most serious policies." What are the parameters that determine the seriousness of one policy over another?
 - Has Meta seen an increase in appeals regarding such restrictions since October 7? If so, what percentage of those appeals has been successful, and what types of remedy are being provided to users whose accounts have been wrongfully restricted from accessing features such as Instagram and Facebook Live?
 - 7) How does Meta explain some users' reports of their inability to access appeal mechanisms on Instagram and Facebook?
 - 8) Has Meta adjusted its internal guidelines on enforcing its policies in response to the October 7 hostilities? If so, could you share them with Human Rights Watch?

- 9) The Israeli Cyber Unit, based within the State Attorney's Office, flags and submits requests to social media companies to "voluntarily" remove content. Instead of going through the legal process of filing a court order based on Israeli criminal law to take down online content, the Cyber Unit makes appeals directly to platforms, including Meta, based on their own terms of service. Meta [committed](#) in 2021 to "disclose the number of formal reports received from government entities (including the Israel State Attorney Office) about content that is not illegal but potentially violates Meta content policies." Can you provide Human Rights Watch with those statistics? If not, why not?
- 10) What steps, if any, has Meta undertaken or does Meta plan to undertake to institute any firewalls to prevent undue influence of its public policy staff, including former Israeli and other government officials, over content moderation decisions regarding Israel and Palestine?

We request that you provide us with a response to these questions and any other information that you wish to provide **in writing by December 6** so that we may have the opportunity to review and reflect it as appropriate in our reporting. We may post on our website relevant parts of any response received after that date.

Thank you in advance for your engagement. Should you have any clarificatory questions or comments, please feel free to reach out to my colleague Deborah Brown, Senior Researcher on Technology and Human Rights, at [redacted].

Sincerely,

Tirana Hassan
Executive Director
Human Rights Watch

Cc:

Iain Levine, Senior Human Rights Advisor
Miranda Sissons, Director of Human Rights
Alex Warofka, Human Rights Manager and Emerging Risks Lead