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

HUMAN RIGHTS WATCH

HRW.org

June 28, 2021

██████████████

Facebook
Via email to: ███████████

██████████████

Facebook
Via email to: ███████████

## RE: Censorship of Palestinian Content on Facebook and Instagram

Dear ███████████

Thank you and your colleagues for taking the time to speak with us on May 27. We are following up to request information regarding what appears to be censorship of content on Facebook and Instagram. Specifically, we have observed apparent censorship of content in May and June 2021 that documents, condemns, and raises awareness about the recent events in Gaza and Jerusalem, some of which constitute grave human rights abuses. We are also concerned about apparent censorship of Palestinian voices more generally over the past several years.

Examples include Facebook or Instagram deleting posts and suspending accounts and groups, removing or blocking hashtags, preventing certain accounts from using Facebook/Instagram Live, affixing sensitive and graphic-warning labels to text-only posts, and disabling the ability to search for accounts that post about Palestine or the ability to engage with such posts. There have also

been reports of reduction in the distribution or reach of content relating to developments in Gaza and Jerusalem.

We request that you respond to the following questions by **July 9, 2021** so that we can incorporate Facebook's response and perspectives in our reporting. Please feel free to include any other information you consider relevant.

1) Has Facebook deployed any special temporary measures to limit the spread of possibly dangerous or problematic content, sometimes referred to as "break glass" options, in response to the recent escalation in Israel and Palestine? Specifically:

   a) Has Facebook implemented changes to its algorithms on Facebook and Instagram to demote or reduce the distribution of content that Facebook proactively detects as likely violating its policies on hate speech, violence and incitement, or dangerous individuals and organizations before that determination is made?

   b) What terms and classifiers are being used to flag content for potential hate speech or violence and incitement in connection with Israel and Palestine, including Gaza, allowing them to automatically be flagged for demotion and/or removal? What percentage of posts flagged by automation was found to be in error?  Can you provide any statistics about the amount of content demoted or removed?

   c) Does Facebook have a list of hashtags that are banned? If so, please provide us with a list of hashtags that relate to Israel and Palestine.

   d) Has Facebook put in place specific measures to temporarily reduce the distribution of content from accounts that have recently and repeatedly violated its Community Standards? If so, what are those measures, and what number of violations would trigger those measures, and within what time frame?

   e) Has Facebook put in place limitations on the use of live video? If so, what are those measures and which accounts do they apply to?

2) Please explain how Facebook determines which posts are considered sensitive and why sensitive warning labels are being affixed to stories when the content should not necessarily constitute sensitive content. For example, on May 24,

Instagram affixed a sensitive warning label to multiple stories Mohammed el-Kurd, a Palestinian activist and resident of Sheikh Jarrah, posted, including a story that contained a reposted image of an Israeli police truck and another truck with Hebrew writing on it. On June 21, Instagram did the same to a video post Muna el-Kurd, Mohammed's sister and an activist as well, shared in which an Israeli soldier is seen pushing her in the street near her house. Both examples are included in the appendix to this letter.

3) As Adalah and the Association for Civil Rights in Israel (ACRI)'s [petition to the Israeli Supreme Court documents](), Israeli officials have indicated that removal of Palestinian speech from Facebook has occurred in response to "voluntary" requests by Israel's Cyber Unit to social media companies for content removal, not based on legal orders. Citing a 2018 report by the State Attorney's office, the petition notes that such voluntary requests are met with an extremely high compliance rate—90 percent across all platforms. Does Facebook plan to disclose or report on the number and nature of requests for content removals by the Israeli Government's Cyber Unit? If so, can you provide us with those statistics or a timeline for their publication?

4) Does Facebook plan to publish its Dangerous Organizations and Individuals list? If so, please provide us with a timeline. Please also explain Facebook's process for adding organizations and individuals to the list, including how it uses the US State Department's Foreign Terrorist Organization list and the Office of Foreign Assets Control (OFAC) of the US Department of the Treasury sanctions lists, and what channels organizations and individuals have for challenging Facebook's decision to include them on the Dangerous Organizations and Individuals list.

5) The role of public policy staff and content moderators are distinct and can be in conflict. The former includes representing Facebook and its business interests to the government, while the latter includes evaluating content that is critical of a government, that documents human rights abuses committed by state actors, and the speech of politicians or public officials, which may violate Facebook's Community Guidelines. Therefore, firewalls need to be put in place to prevent content moderation decisions from being influenced by public policy staff. Yet, as [has been](] [documented](), public policy staff have influence over content

moderation decisions. Does Facebook have any firewalls in place to prevent undue influence of public policy staff, including [former Israeli](#) and other government officials, over content moderation decisions with regard to Israel and Palestine?[1] If so, please describe them.

6) What measures, if any, does Facebook have in place to detect anti-Palestinian or anti-Arab bias in its moderation of content?

7) *Buzzfeed News* [reported](#) that during the week of May 10, when the latest spate of violence broke out, Facebook received nearly 155,000 reports of content flagged under the company's terrorism rules, the highest number of reports it received from any country under this content category. It was also reported that Israel was third in the world for flagging content under Facebook's policies for hate speech, violence and incitement, outstripping more populous countries like the US, India, and Brazil. According to [NBC News reporting](#), between May 8 and May 18, users in Israel reported 494,463 cases of hate speech while Palestinian users reported 58,618 cases. [Additionally, Act.IL](#), a smartphone app and website [that seeks to influence public perceptions](#) on Israel and Palestine sends its users on missions, including by asking them to report offensive content on social media that appears to endorse violence. Does Facebook have plans to investigate whether this mass reporting constitutes a breach of its [coordinated inauthentic behavior](#) policy by seeking to manipulate public debate? If so, what measures is Facebook taking to detect and disrupt this behavior on its platforms?

8) Facebook announced that it [reversed the change to Instagram's algorithm](#) that prioritized original content over shared content partially in response to concerns over perceived suppression of pro-Palestinian content, as well as other points of view or topics. Can Facebook provide information on what type of analysis or assessment was made before it changed Instagram's algorithm to prioritize original content?
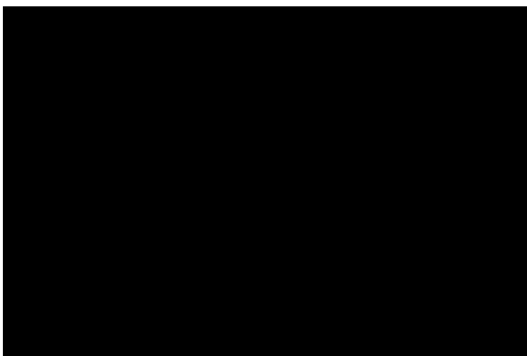
---

[1] A September 2020 article quoted Facebook's head of policy for Israel describing her role as "to represent Facebook to Israel, and represent Israel to Facebook" and saying that she "represent[s] Israel" in weekly meetings "about everything from spam to pornography to hate speech and bullying and violence, and how they relate to [Facebook's] community standards." See: Zev Stub, "Head of policy helps Jews in Israel, diaspora connect over Facebook," *Jerusalem Post*, September 20, 2020, [https://www.jpost.com/50-most-influential-jews/jordana-cutler-642279](https://www.jpost.com/50-most-influential-jews/jordana-cutler-642279)

9) Facebook [announced](#) on May 19 that it established a "special operations center" to monitor content on its platforms regarding the recent escalation in Israel and Palestine. Can you please provide the following information:
   a) The overall mandate of that center and whether Facebook has deployed comparable models in other circumstances.
   b) How many days the "special operations center" has operated.
   c) What factors went into the decision to open the center, and what factors go into deciding when the center should close down.
   d) How many Arabic and Hebrew speaking moderators are/were working with the center.
   e) What data is available about the work the special operations center has done while active.

10) Does Facebook have any plans to investigate whether it may be disproportionately removing Palestinian content? If so, can you detail any plans or timelines for such an investigation?

Thanks in advance for your attention to our questions and feel free to provide any other information you wish.

For any comments or questions, please feel free to contact me ([ganesaa@hrw.org](mailto:ganesaa@hrw.org)) or my colleagues Deborah Brown ([brownd@hrw.org](mailto:brownd@hrw.org)), senior researcher and advocate on digital rights, or Omar Shakir ([shakiro@hrw.org](mailto:shakiro@hrw.org)), Israel and Palestine director.
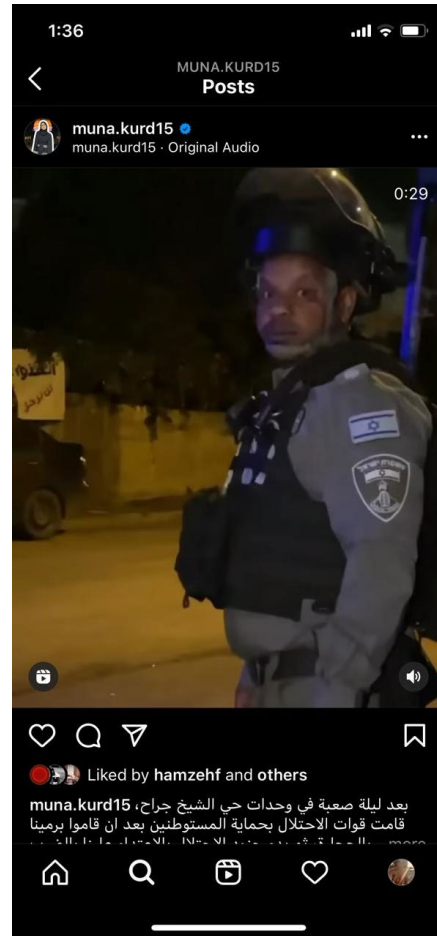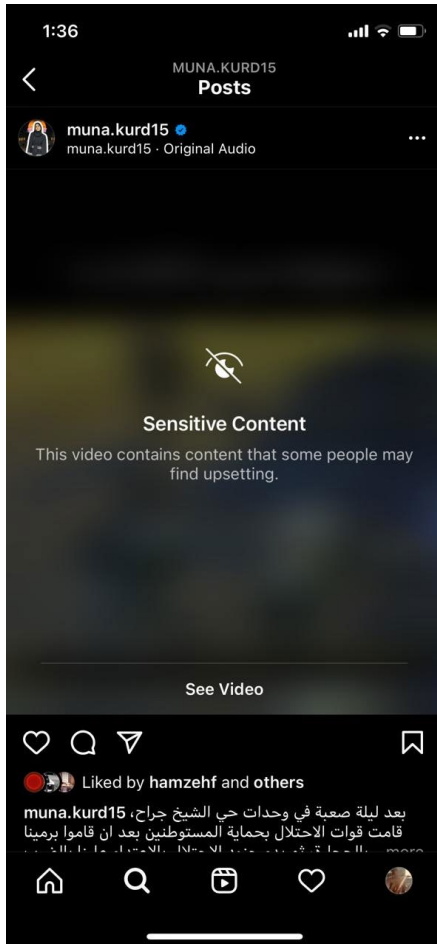
Sincerely,



Human Rights Watch

## APPENDIX

Appendix I – Examples of the apparently arbitrary use of sensitive content labels on Instagram

### E.g. (1) Muna el-Kurd – June 21, 2021

**E.g. (2) Mohammed el-Kurd – May 24, 2021**