

December 14, 2023

Mr. Mark Zuckerberg
Chief Executive Officer Meta,
Inc.
1 Hacker Way
Menlo Park, CA 94025

Dear Mr. Zuckerberg,

I write with concerns regarding recent reports that Meta is suppressing Palestinian and Palestinian-related content on Instagram.¹ Over 90 human rights and civil society organizations have raised alarm about Meta's reported suppression, filtering, and mistranslation of Palestine-related content over the past two months, calling on the company to provide additional transparency as to content moderation and to address discriminatory algorithmic design.² Amidst the horrific Hamas terrorist attacks in Israel, a humanitarian catastrophe including the deaths of thousands of civilians in Gaza,² and the killing of dozens of journalists,³ it is more important than ever that social media platforms do not censor truthful and legitimate content, particularly as people around the world turn to online communities to share and find information about developments in the region.

News organizations have reported several instances of Instagram censoring or mistranslating Palestinian and Palestine-related content following Hamas's deadly October 7 attacks. Instagram reportedly labeled the Palestinian flag emoji as "potentially offensive," causing comments

¹ Wall Street Journal, "Inside Meta, Debate Over What's Fair in Suppressing Comments in the Palestinian Territories," Sam Schechner et al., October 21, 2023, <https://www.wsj.com/tech/inside-meta-debate-over-whats-fairin-suppressing-speech-in-the-palestinian-territories-6212aa58>; The Intercept, "Instagram Hid a Comment. It Was Just Three Palestinian Flag Emojis." Sam Biddle, October 28, 2023, <https://theintercept.com/2023/10/28/instagrampalestinian-flag-emoji/>.² 7amleh, "Civil Society Organizations Call for Tech Companies to Respect Palestinian Digital Rights in Times of Crisis," October 13, 2023, <https://7amleh.org/2023/10/13/7amleh-center-along-with-civil-organizations-release-astatement-calling-tech-companies-to-respect-palestinian-digital-rights-in-times-of-crisis>; The Hill, "Palestinian groups accuse Meta of unfairly moderating speech amid Israel-Hamas conflict," Julia Shapero, December 2, 2023, <https://thehill.com/policy/technology/4333648-palestinian-groups-accuse-meta-of-unfairly-moderating-speech-amidisrael-hamas-conflict/>.

² New York Times, "Gaza Civilians, Under Israeli Barrage, Are Being Killed at Historic Pace," Lauren Leatherby, November 30, 2023, <https://www.nytimes.com/2023/11/25/world/middleeast/israel-gaza-death-toll.html>.

³ Committee to Protect Journalists, "Journalist casualties in the Israel-Gaza war," last accessed December 13, 2023, <https://cpj.org/2023/11/journalist-casualties-in-the-israel-gaza-conflict/>.

containing the emoji to be hidden from view.⁴ The platform also translated user bios that contained the Palestinian flag emoji and the words “Palestinian” and “Alhamdulillah” (Arabic for “praise be to God”) as “Palestinian terrorist” or “praise be to God, Palestinian terrorists are fighting for their freedom.”⁵

Even more concerning, Meta reportedly manipulated its content filters to apply stricter standards to content generated in Palestine compared to the standards in place in the rest of the region and the world. Meta generally hides comments designated as hateful only when its content moderation systems are 80% certain that they violate the platform’s policies.⁶ The *Wall Street Journal* reported that after Hamas’s October 7 attacks, Meta tightened those filters, “[a]s part of ‘temporary risk response measures’—emergency calming efforts of the sort that Meta has previously deployed in wars, potential genocides, and the Jan. 6, 2021, U.S. Capitol riot—Meta cut that threshold in half over a swath of the Middle East, hiding any comment deemed 40% likely to be inflammatory.”⁸ Meta also changed the default settings on Facebook to limit who could comment on content created “in the region.”⁷ Soon thereafter, however, Meta lowered the bar further for Palestine only, hiding comments from users in the occupied Palestinian territories if its automated system judged there was at least a 25% chance they violated Meta’s rules.⁸

Unsurprisingly given these changes, Instagram users have reported numerous instances of account suspensions and content suppression linked to communications regarding Palestine. The social media nonprofit 7amleh, also known as the Arab Center for Social Media Advancement, documented more than 200 cases of alleged censorship by Meta in the three weeks following Hamas’s October 7 attack.⁹ Instagram users reported they were restricted from posting or livestreaming and had their content marked as “sensitive.”¹² Other users posted screenshots showing their “Stories” about Palestine received significantly fewer views than other Stories posted on the same day, suggesting that Instagram may be shadow banning material¹³ – that is, limiting

⁴ The Intercept, “Instagram Hid a Comment. It Was Just Three Palestinian Flag Emojis.” Sam Biddle, October 28, 2023, <https://theintercept.com/2023/10/28/instagram-palestinian-flag-emoji/>.

⁵ Wall Street Journal, “Inside Meta, Debate Over What’s Fair in Suppressing Comments in the Palestinian Territories,” Sam Schechner et al., October 21, 2023, <https://www.wsj.com/tech/inside-meta-debate-over-whats-fairin-suppressing-speech-in-the-palestinian-territories-6212aa58>; The Guardian, “Instagram apologises for adding ‘terrorist’ to some Palestinian user profiles,” Josh Taylor, October 19, 2023, <https://www.theguardian.com/technology/2023/oct/20/instagram-palestinian-user-profile-bios-terrorist-addedtranslation-meta-apology>.

⁶ *Id.*

⁸ *Id.*

⁷ Meta, “Meta’s Ongoing Efforts Regarding the Israel-Hamas War,” October 13, 2023, <https://about.fb.com/news/2023/10/metas-efforts-regarding-israel-hamas-war/>.

⁸ Wall Street Journal, “Inside Meta, Debate Over What’s Fair in Suppressing Comments in the Palestinian Territories,” Sam Schechner et al., October 21, 2023, <https://www.wsj.com/tech/inside-meta-debate-over-whats-fairin-suppressing-speech-in-the-palestinian-territories-6212aa58>.

⁹ The Guardian, “Pro-Palestinian Instagram account locked by Meta for ‘security reasons,’” Hilbaq Farah, October 26, 2023, <https://www.theguardian.com/technology/2023/oct/26/pro-palestinian-instagram-account-locked-by-metafor-security-reasons>; Palestinian Observatory of Digital Rights Violations(7or), “Charts,” <https://7or.7amleh.org/charts>.¹²

TechCrunch, “Meta has a moderation bias problem, not just a ‘bug,’ that’s suppressing Palestinian voices,” Morgan Sung, October 19, 2023, <https://techcrunch.com/2023/10/19/meta-instagram-palestine-israel-shadowban-censorshipmoderation-bias/>.¹³ The Guardian, “Instagram users accuse platform of censoring posts supporting Palestine,” Kari Paul, October 18, 2023, <https://www.theguardian.com/technology/2023/oct/18/instagram-palestine-posts-censorship-accusations>; Mashable, “People are accusing Instagram of shadowbanning content about Palestine,” Meera Navlakha, October 16, 2023, <https://mashable.com/article/instagram-shadowbanning-censor-israel-palestine>.

the visibility of users' posts without notifying them.¹⁰ Multiple journalists have also reported issues with their accounts,¹¹ including reports of flagging Arabic posts but not identical English or Hebrew versions.¹²

These problems are not new for Meta. For years, civil society organizations have called on Meta to address systems and practices that result in disproportionate censorship of Palestinians and Palestine-related content. During Israel's attacks on Gaza in 2021, digital rights groups documented hundreds of apparently unwarranted content removals affecting Palestinians.¹⁷ Meta eventually acknowledged its systems were blocking posts related to the Muslim holy site al-Aqsa, which Meta had incorrectly flagged as connected to terrorist groups.¹³ After over 200 Meta employees published a letter demanding the company address potential biases,¹⁹ Meta commissioned a third party to review the human rights impacts of its policies and activities during the 2021 conflict.²⁰

A third-party review of these actions – conducted by Business for Social Responsibility (BSR) on behalf of Meta - highlighted clear issues. BSR “identified a variety of human rights impacts” resulting from “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.”¹⁴ On a per-user basis, Meta over-enforced on Arabic content.¹⁵ BSR concluded that Meta's actions “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.”¹⁶

Meta's content moderation systems and practices have also exacerbated violence and failed to combat hate speech. Facebook, which is owned by Meta, recently approved paid advertisements by a right-wing Israeli group calling for the assassination of a pro-Palestine activist, and subsequent

¹⁰ Electronic Frontier Foundation, “Platforms Must Stop Unjustified Takedowns of Posts By and About Palestinians,” Karen Gullo and Jillian C. York, November 8, 2023, <https://www.eff.org/deeplinks/2023/11/platforms-must-stopunjustified-takedowns-posts-and-about-palestinians>.

¹¹ This includes Pulitzer Prize-winning New York Times reporter Azmat Khan, photojournalist Motaz Azaiza, and Palestinian-American commentator Mariam Barghouti. Twitter, @AzmatZahra, October 15, 2023, 11:13 AM, <https://twitter.com/azmatzahra/status/1713573877664620711?s=46&t=dlpx88UvGvCPTofxCocUsQ>; Daily Dot, “A bombing in Gaza killed an influencer's family—Instagram responded by suspending his account,” Ernie Piper, October 24, 2023, <https://www.dailydot.com/debug/motaz-azaiza-palestine-social-media-instagram/>.

¹² TechCrunch, “Meta has a moderation bias problem, not just a ‘bug,’ that's suppressing Palestinian voices,” Morgan Sung, October 19, 2023, <https://techcrunch.com/2023/10/19/meta-instagram-palestine-israel-shadowban-censorshipmoderation-bias/>.¹⁷ Vox, “Why some Palestinians believe social media companies are suppressing their posts,” A.W. Ohlheiser, October 29, 2023, <https://www.vox.com/technology/23933846/shadowbanning-meta-israel-hamas-war-palestine>.

¹³ *Id.*; The Guardian, “Pro-Palestinian Instagram account locked by Meta for ‘security reasons,’” Hilbaq Farah, October 26, 2023, <https://www.theguardian.com/technology/2023/oct/26/pro-palestinian-instagram-account-lockedby-meta-for-security-reasons>.¹⁹ Al Jazeera, “Facebook employees demand change around Palestine posts,” June 2, 2021, <https://www.aljazeera.com/economy/2021/6/2/facebook-employees-demand-change-around-palestine-posts-report>.²⁰ BSR, “Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021,” September 22, 2022, <https://www.bsr.org/en/blog/human-rights-due-diligence-of-meta-impacts-in-israel-and-palestine-may-2021>.

¹⁴ *Id.*

¹⁵ BSR, “Human Rights Due Diligence of Meta's Impacts in Israel and Palestine in May 2021,” September 2022, p. 5, https://www.bsr.org/reports/BSR_Meta_Human_Rights_Israel_Palestine_English.pdf.

¹⁶ *Id.* at p. 4.

test ads submitted by 7amleh calling for the murder of Palestinian civilians.¹⁷ Meta had ample notice of these types of issues: for example, Amnesty International found that “Meta’s dangerous algorithms and reckless pursuit of profit substantially contributed to the atrocities perpetrated by the Myanmar military against the Rohingya people in 2017.”²⁵ The organization explained that “Meta’s business model, based on invasive profiling and targeted advertising, fuels the spread of harmful content, including incitement to violence.”¹⁸ Meta also approved test ads containing hate speech against the Rohingya last year.¹⁹

Indeed, internal documents released by a Facebook whistleblower in 2020 identified a range of content moderation issues affecting Facebook’s operations in the Middle East.²⁸ Specifically, Facebook’s “lack of local language expertise and cultural knowledge made it difficult, if not impossible to crack down on online sectarianism and other forms of harmful content aimed at local vulnerable groups.”²⁰ And “[i]n documents published in late 2020, [Facebook’s] engineers discovered that more than three-quarters of Arabic-language content automatically removed from the platform for allegedly promoting terrorism has been mistakenly labeled as harmful material.”³⁰

Meta must not make the same mistakes again. Social media communications are vital sources of news from the ground, particularly given widespread communications blackouts,²¹ ongoing misinformation and information war efforts,²² and the killing of over 60 journalists and media workers in Gaza.²³ Social media users also play a critical role in recording human rights violations

¹⁷ The Intercept, “Facebook Approved an Israeli Ad Calling for Assassination of Pro-Palestine Activist,” Sam Biddle, November 21, 2023, <https://theintercept.com/2023/11/21/facebook-ad-israel-palestine-violence/7amleh>, “Meta Should Stop Profiting from Hate,” November 21, 2023, <https://7amleh.org/2023/11/21/meta-should-stop-profiting-from-hate>. ²⁵ Amnesty International, “Myanmar: Facebook’s Systems Promoted Violence Against Rohingya; Meta Owes Reparations,” September 29, 2022, <https://www.amnesty.org/en/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohingya-meta-owes-reparations-new-report/>.

¹⁸ Amnesty International, “The Social Atrocity: Meta and the Right to Remedy for the Rohingya,” report at p. 7, September 29, 2022, <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/>.

¹⁹ Global Witness, “Facebook approves adverts containing hate speech inciting violence and genocide against the Rohingya,” March 20, 2022, <https://www.globalwitness.org/en/campaigns/digital-threats/rohingya-facebook-hatespeech/>; The Intercept, “Facebook Approved an Israeli Ad Calling for Assassination of Pro-Palestine Activist,” Sam Biddle, November 21, 2023, <https://theintercept.com/2023/11/21/facebook-ad-israel-palestine-violence/>. ²⁸ Politico, “Facebook did little to moderate posts in the world’s most violent countries,” Mark Scott, October 25, 2021, <https://www.politico.eu/article/facebook-content-moderation-posts-wars-afghanistan-middle-east-arabic>.

²⁰ *Id.*

³⁰ *Id.*

²¹ Washington Post, “Communication blackouts are sweeping Gaza. Here’s what to know.” Leo Sands & Bryan Pietsch, November 16, 2023, <https://www.washingtonpost.com/world/2023/11/06/gaza-internet-communications-blackout-explainer/>.

²² RAND Blog, “Lies, Misinformation Play Key Role in Israel-Hamas Fight,” Todd C. Helmus & William Marcellino, Oct. 31, 2023, <https://www.rand.org/pubs/commentary/2023/10/lie-misinformation-play-key-role-in-israel-hamasfight.html>.

²³ Committee to Protect Journalists, “Journalist casualties in the Israel-Gaza war,” last accessed December 13, 2023, <https://cpj.org/2023/11/journalist-casualties-in-the-israel-gaza-conflict/>.

and war crimes.²⁴ Human rights organizations have called on social media platforms to ensure materials from war zones are properly archived, even if they are removed from visibility, so that they can be used to investigate crimes.³⁵ Human Rights Watch has explained that preserving digital evidence can help “hold individuals accountable,” “deter future violations,” “and also help restore dignity to victims by acknowledging their suffering and helping to create a historical record that protects against revision by those who denied that atrocities occurred.”³⁶

Reports of Meta’s suppression of Palestinian voices raise serious questions about Meta’s content moderation practices and anti-discrimination protections. Social media users deserve to know when and why their accounts and posts are restricted, particularly on the largest platforms where vital information-sharing occurs. Users also deserve protection against discrimination based on their national origin, religion, and other protected characteristics. Legislative proposals like the bipartisan *Digital Consumer Protection Commission Act*, the *Digital Platform Commission Act*, and the *Algorithmic Accountability Act* would increase transparency around algorithmic decision making on major online platforms. In order to further understand what legislative action might be necessary to address these issues, we ask for additional information about Meta’s application of its policies. Please answer the below questions no later than January 5, 2024.

1. What steps has Meta taken to implement the recommendations in BSR’s September 2022 report titled “Human Rights Due Diligence of Meta’s Impacts in Israel and Palestine”? For each recommendation, please specify:
 - a. Whether Meta has pursued or is pursuing actions related to the recommendation. If yes, please specify the nature of those actions.
 - b. What resources, in terms of staff time and expenses, Meta has allocated toward addressing the recommendation.
 - c. What metric(s) Meta is using to define success with regards to the recommendation.
 - d. When Meta will complete action related to the recommendation.

2. Please list each instance, in the past five years, in which Meta has changed the content moderation threshold for a particular nation or occupied territory in the manner described in the October 21, 2023 *Wall Street Journal* article titled “Inside Meta, Debate Over What’s Fair in Suppressing Comments in the Palestinian Territories.”
 - a. For each instance, please list the location, the timespan during which the threshold was changed, the language to which the threshold change applied, the reason for the threshold change, and the level to which the threshold was set.
 - b. For each instance, please list the number of posts that were flagged or otherwise affected as a result of the new threshold.

²⁴ For example, the International Criminal Court has issued a war crimes indictment based on videos from Libya posted on social media. Bellingcat, “Geolocating Libya’s Social Media Executioner,” September 4, 2017, <https://www.bellingcat.com/news/mena/2017/09/04/geolocating-libyas-social-media-executioner/>. The Syrian Archive has documented digital evidence related to the Syrian Civil War, with a project dedicated to preserving materials that social media platforms might remove. Syrian Archive, “About,” <https://syrianarchive.org/en/about>.

³⁵ Human Rights Watch, “Social Media Platforms Remove War Crimes Evidence,” September 10, 2020, <https://www.hrw.org/news/2020/09/10/social-media-platforms-remove-war-crimes-evidence>; Time, “These Tech Companies Managed to Eradicate ISIS Content. But They’re Also Erasing Crucial Evidence of War Crimes,” Billy Perrigo, April 11, 2020, <https://time.com/5798001/facebook-youtube-algorithms-extremism/>. ³⁶ Human Rights Watch, “Social Media Platforms Remove War Crimes Evidence,” September 10, 2020, <https://www.hrw.org/news/2020/09/10/social-media-platforms-remove-war-crimes-evidence>.

3. During the time period spanning October 7, 2023 to the present:
 - a. How many Arabic language posts originating from Palestine have been removed?
 - b. What percentage of total Arabic language posts originating from Palestine does the above number represent?
 - c. What percentage of the removed posts were removed due to automated systems versus human moderation?
 - d. How often did Meta limit the reachability of Arabic language posts originating from Palestine without notifying the user?
 - e. How often did Meta limit the reachability of Arabic language posts originating from Palestine while notifying the user?

4. During the time period spanning October 7, 2023 to the present:
 - a. How many English language posts originating from Palestine have been removed?
 - b. What percentage of total English language posts originating from Palestine does the above number represent?
 - c. What percentage of the removed posts were removed due to automated systems versus human moderation?
 - d. How often did Meta limit the reachability of English language posts originating from Palestine without notifying the user?
 - e. How often did Meta limit the reachability of English language posts originating from Palestine while notifying the user?

5. During the time period spanning October 7, 2023, to the present:
 - a. How many Hebrew language posts originating from Israel have been removed?
 - b. What percentage of total Hebrew language posts originating from Israel does the above number represent?
 - c. What percentage of the removed posts were removed due to automated systems versus human moderation?
 - d. How often did Meta limit the reachability of Hebrew language posts originating from Israel without notifying the user?
 - e. How often did Meta limit the reachability of Hebrew language posts originating from Israel while notifying the user?

6. During the time period spanning October 7, 2023, to the present:
 - a. How many English language posts originating from Israel have been removed?
 - b. What percentage of total English language posts originating from Israel does the above number represent?
 - c. What percentage of the removed posts were removed due to automated systems versus human moderation?
 - d. How often did Meta limit the reachability of English language posts originating from Israel without notifying the user?
 - e. How often did Meta limit the reachability of English language posts originating from Israel while notifying the user?

7. During the time period spanning October 7, 2023, to the present:
 - a. How many posts globally have been removed?
 - b. What percentage of global posts does the above number represent?

- c. What percentage of those removed posts were removed due to automated systems versus human moderations?
 - d. How often did Meta limit the reachability of posts globally without notifying the user?
 - e. How often did Meta limit the reachability of posts globally while notifying the user?
8. During the time period spanning October 7, 2023 to the present:
 - a. How many appeals did users submit regarding content decisions related to Arabic language posts originating from Palestine?
 - b. How often were content decisions regarding Arabic language posts originating from Palestine appealed?
9. During the time period spanning October 7, 2023 to the present:
 - a. How many appeals did users submit regarding content decisions related to English language posts originating from Palestine?
 - b. How often were content decisions regarding English language posts originating from Palestine appealed?
10. During the time period spanning October 7, 2023 to the present:
 - a. How many appeals did users submit regarding content decisions related to Hebrew language posts originating from Israel?
 - b. How often were content decision regarding Hebrew language posts originating from Israel appealed?
11. During the time period spanning October 7, 2023 to the present:
 - a. How many appeals did users submit regarding content decisions related English language posts originating from Israel?
 - b. How often were content decisions regarding English language posts originating from Israel appealed?
12. During the time period spanning October 7, 2023 to the present:
 - a. How many appeals did users submit globally regarding content decisions?
 - b. How often were content decisions appealed globally?
13. What is the average response time a user can typically expect after appealing a content moderation decision from Meta?
14. During the time period spanning October 7, 2023 to the present:
 - a. What was the average response time for a user appeal of a content moderation decision for Arabic language posts originating from Palestine?
 - b. Based on data from user appeals, what percentage of Arabic language posts originating from Palestine were found to have been wrongfully taken down (i.e. false positives) and then reinstated?
15. During the time period spanning October 7, 2023 to the present:
 - a. What was the average response time for a user appeal of a content moderation decision for English language posts originating from Palestine?

- b. Based on data from user appeals, what percentage of English language posts originating from Palestine were found to have been wrongfully taken down (i.e. false positives) and then reinstated?
16. During the time period spanning October 7, 2023 to the present:
 - a. What was the average response time for a user appeal of a content moderation decision for Hebrew language posts originating from Israel?
 - b. Based on data from user appeals, what percentage of Hebrew language posts originating from Israel were found to have been wrongfully taken down (i.e. false positives) and then reinstated?
17. During the time period spanning October 7, 2023 to the present:
 - a. What was the average response time for a user appeal of a content moderation decision for Hebrew language posts originating from Israel?
 - b. Based on data from user appeals, what percentage of English language posts originating from Israel were found to have been wrongfully taken down (i.e. false positives) and then reinstated?
18. A Meta spokesperson stated that Instagram hid comments containing the Palestinian flag emoji due to Meta’s Dangerous Organizations and Individuals policy and community standards.²⁵ However, the *Intercept* reportedly reviewed several hidden comments containing the Palestinian flag emoji that had no reference to dangerous organizations or individuals and did not otherwise appear to violate Meta’s community standards.²⁶ Is the Intercept report accurate? If not, what is the explanation for the reported hiding of comments?
19. During the time period spanning October 7, 2023 to the present:
 - a. How many staffers does Meta employ to review Arabic language content? Of that number, how many staffers possess native language skills?
 - b. How many staffers does Meta employ to review English language content? Of that number, how many staffers possess native language skills?
 - c. How many staffers does Meta employ to review Hebrew language content? Of that number, how many staffers possess native language skills?
20. What is Meta’s current policy for retention of content that may contain evidence of human rights abuses?
 - a. What are the criteria for initiating and terminating preservation of content that may contain evidence of human rights abuses?
 - b. What is the period of data retention for content that may contain evidence of human rights abuses?
 - c. What are Meta’s policies for researcher, civil society, and governmental access to content that may contain evidence of human rights abuses?

Thank you for your attention to this important matter.

²⁵ The Intercept, “Instagram Hid a Comment. It Was Just Three Palestinian Flag Emojis.” Sam Biddle, October 28, 2023, <https://theintercept.com/2023/10/28/instagram-palestinian-flag-emoji/>.

²⁶ *Id.*

Sincerely,

A handwritten signature in black ink, appearing to read "Elizabeth Warren", with a long horizontal flourish extending to the right.

Elizabeth Warren
United States Senator