

ONE HUNDRED EIGHTEENTH CONGRESS
Congress of the United States
House of Representatives
COMMITTEE ON ENERGY AND COMMERCE
2125 RAYBURN HOUSE OFFICE BUILDING
WASHINGTON, DC 20515-6115

Majority (202) 225-3641
Minority (202) 225-2927

December 13, 2023

Linda Yaccarino
Chief Executive Officer
X
1355 Market Street, Suite 900
San Francisco, CA 94103

Dear Ms. Yaccarino:

I appreciate your company's recent briefing to Committee staff on the use of your platform to proliferate extreme, graphic, false, misleading, or other harmful content in the aftermath of Hamas's October attacks, and Israel's ongoing response in Gaza. However, it did not fully address my questions, nor did it resolve my concerns. I am particularly alarmed by your apparent outsourcing of content moderation to your users through the use of "Community Notes."

Since the attacks on October 7, 2023, numerous reports have described how terrorist organizations and others have used social media platforms to inflict more suffering on civilians, mislead the public about the Israel-Hamas conflict, and spread hate. For instance, Meta products were reportedly used to broadcast videos of violence against Israeli civilians, and Hamas threatened to post videos of hostage executions online in retaliation for Israeli airstrikes.¹

In other cases, users spread doctored content designed to incite outrage among the Israeli and Palestinian public. For example, a video on your platform, X, purported to show a second air assault on Israel but was later debunked as a clip from a video game.² Another video on Meta purported to show a young Israeli woman being beaten by a Palestinian mob but was later revealed to be a 2015 clip depicting gang violence in Guatemala.³ Other articles intentionally mischaracterized the global response, such as a deliberately mis-captioned video falsely

¹ *Hamas Hijacked Victims' Social Media Accounts to Spread Terror*, The New York Times (Oct. 17, 2023); *Parents Urged to Delete Their Kids' Social Media Accounts Ahead of Possible Israeli Hostage Videos*, CNN (Oct. 12, 2023).

² Shayan Sardarizadeh (@Chayan86), X (Oct. 8, 2023, 8:44 PM) (<https://twitter.com/Shayan86/status/1711180889453953238>).

³ *Fact Check: Graphic Video of Girl Burned by Mob is from Guatemala, not Israel*, Reuters (Oct. 13, 2023).

representing that Russian President Vladimir Putin warned the United States to “stay out of the Gaza war.”⁴ More recently, widely viewed images intended to illustrate the human toll of the war on Palestinians were later found to be from other conflict zones or disaster sites, such as Syria or Tajikistan, and some widely circulated videos depicting Israeli victims of Hamas’s massacre were revealed to be deepfakes generated through artificial intelligence.⁵

Against this backdrop, antisemitic and Islamophobic accounts have spread despicable rhetoric against Jews, Israelis, and Palestinians. One account on your platform asserted that Israel does not deserve sympathy because it helped coordinate the 9/11 terrorist attacks, while another posted that when young girls convert to Islam, they end up “in a suitcase or fridge.”⁶

Reports of weakened trust and safety policies and increased automation of content moderation workforces raise additional concerns about your industry’s willingness and ability to adequately handle sudden and complex world events. In December 2022, your company, X, disbanded its Trust and Safety Council—a group designed to provide expert advice about online safety—as part of its reorientation towards “free speech absolutist” ideals.⁷ Four months later, YouTube loosened its policy on election-related disinformation, despite a professed commitment to preventing the proliferation of false information.⁸ Alphabet reduced its workforce tasked with building content moderation tools by one-third, and Meta ended the contracts of about 200 content moderators in early January 2023.⁹ In total, tech companies—including social media giants like your company, X, and Meta—have laid off over 100,000 staff that, in part, helped provide a safer online environment.¹⁰ The recent uptick in graphic, false, and xenophobic hate speech on platforms calls into question your decision to automate content moderation tools and downgrade your trust and safety policies.

To understand your efforts to address extreme, graphic, false, misleading, and otherwise harmful content on your platforms, I request responses to the following questions:

⁴ *Old Videos Do Not Show Putin Warning the US to Stay out of the Latest Gaza War*, Associated Press (Oct. 9, 2023).

⁵ *The Horrifying Images are Real. But They’re Not from the Israel-Gaza War*, The New York Times (Nov. 2, 2023); *TikTok Struggles to take down Deepfake Videos of Hamas’ Victims*, Bloomberg (Dec. 4, 2023).

⁶ Anti-Defamation League, *Hamas Attack Draws Cheers from Extremists, Spurs Antisemitism and Conspiracies Online* (Oct. 7, 2023) (<https://www.adl.org/resources/blog/hamas-attack-draws-cheers-extremists-spurs-antisemitism-and-conspiracies-online>); *Analysis: Why is so Much Anti-Palestinian Disinformation Coming from India?*, Aljazeera (Oct. 16, 2023).

⁷ *Twitter Disbands its ‘Trust and Safety Council’ that Tackled Harassment and Child Exploitation*, CNN (Dec. 13, 2022).

⁸ *Scoop: YouTube Reverses Misinformation Policy to Allow U.S. Election Denialism*, Axios (June 2, 2023).

⁹ *Tech Layoffs Ravage the Teams that Fight Online Misinformation and Hate Speech*, CNBC (May 26, 2023).

¹⁰ *Google, Meta, Amazon and Other Tech Companies Have Laid off More than 104,000 Employees in the Last Year*, CNBC (Mar. 20, 2023).

1. Please produce any company policies, manuals, or documents that were designed to address violative content on your platform related to the current crisis in Israel and Gaza.
2. Please provide the total number of pieces of content (e.g., posts, videos, short clips, or comments) relating to the crisis in Israel and Gaza that have been posted to your platform since October 7, 2023, and the total number of pieces of content that have been removed from your platform for violating your trust and safety terms as applied to the crisis. Of the content that has been removed, please identify which policy was violated (e.g., terrorist link, glorification of violence, graphic content, etc.).
3. Please provide the number of accounts suspended since October 7, 2023, because of a link to a terrorist organization involved in the conflict between Israel and Hamas, including Hamas, Hezbollah, or Palestinian Islamic Jihad. If you currently have company policies that forbid terrorist organizations from maintaining accounts on your platform, please also explain why these accounts were not taken down prior to October 7, 2023.
4. Please provide a list of the hashtags your platform has blocked that are linked to the current crisis in Israel and Gaza since October 7, 2023.
5. For content on your platform related to the crisis in Israel and Gaza, please describe your process for placing interstitial warnings, the types of posts that qualify for an interstitial warning, and how many posts have received an interstitial warning.
6. Please explain any measures you have in place to allow users to report concerns about extreme, graphic, false, misleading, and otherwise harmful content related to the crisis in the Middle East, including content that is antisemitic or Islamophobic.
 - a. How many reports of such content have you received and how have you responded?
 - b. Have you categorized and evaluated each such report?
 - c. Have you responded individually or collectively to such reports and, if so, how have you responded?
 - d. Have you considered or made changes to your platform's policies based on such reports? If so, how so? If not, why not?
7. Please describe the process by which Community Notes appear on content related to the events in Israel and Gaza since October 7, 2023. Please include in your response the criteria for becoming a contributor of Community Notes and whether the presence of a Note has resulted in the reduction of likes or reposts of the content.

8. Do you provide your users with tools to minimize extreme, graphic, false, misleading, and otherwise harmful content related to the crisis in the Middle East, including content that is antisemitic or Islamophobic?
 - a. If so, please describe such tools.
 - b. How do you make your users aware of such tools?
 - c. How long have such tools been in place?
 - d. How many users have taken advantage of those tools since October 7, 2023?
9. What is your policy on allowing misinformation and disinformation to remain on your platform as it relates to events in Israel and Gaza since October 7, 2023? Please describe how your company distinguishes between false content that mandates removal, false content that requires an interstitial warning or additional content, and false content that is deemphasized.
10. Please describe if, and how, your company differentiates its treatment of violent or inciteful content committed by a terrorist group (e.g., glorification of Hamas) and violent or inciteful content committed by other actors (e.g., calls for violence against civilians).
11. Does your company have a “public interest” exception to your trust and safety standards that allows otherwise violative content to remain on your platform? If so, please describe:
 - a. How you have applied that policy in the context of the recent crisis in Israel and Gaza, including any process for determining whether the exception applies and any additional framing you provide users viewing content on your platform under this exception; and
 - b. How many posts and/or users have been granted a “public interest” (or similar) exception related to the violence in Israel and Gaza since October 7, 2023?
12. For each of the following categories of content, please describe the role of human personnel and automated processes in identifying and, if necessary, removing content from your platform as it relates to the current crisis in Israel and Gaza:
 - a. Extreme or terrorist content;
 - b. Fake and doctored content;
 - c. Denial of major historical events (e.g., the Holocaust or the October 7, 2023, terrorist attacks);

- d. Misinformation, disinformation, and false content; and
 - e. Antisemitism, Islamophobia, and anti-Arab hate speech.
13. Please describe in detail the current composition of your content moderation workforce and procedures.
- a. Please describe any modifications your company has made to the number of human personnel reviewing content for violations of your trust and safety criteria since October 7, 2023.
 - b. Please describe any modifications to your automated content review and moderation algorithms that you have made to better address violative content in response to the violence in Israel and Gaza.
14. Does your platform have a list of terms or phrases that it uses to identify content for deemphasis, automatic removal, or secondary review? If so, please provide that list.
15. What specific policies has your company put in place to limit children's access to content that is violative of your trust and safety guidelines as it relates to the crisis in Israel and Gaza? If your company has guidelines specific to children in different age brackets, please explain those guidelines and identify which age bracket they apply to.
16. What systems are in place for the continuous monitoring, evaluation, and, if necessary, revision of your trust and safety guidelines as they relate to the current crisis in Israel and Gaza?
17. Please describe the standards and the mechanisms your platform has in place to detect, and where appropriate, remove deepfakes.
18. Please provide a copy of any response your company has made to the European Commission in connection with their request for information about your company's compliance with the Digital Services Act and how your company is reducing the risk of spreading violent or terrorist content, hate speech, or disinformation related to the current crisis in Israel and Gaza.

Ms. Linda Yaccarino
December 13, 2023
Page 6

Please provide responses to these questions by January 3, 2024. For further questions, please reach out to Constance O'Connor and Harry Samuels with the Committee Democratic staff. I appreciate your urgent attention to this issue.

Sincerely,

A handwritten signature in blue ink that reads "Frank Pallone, Jr." with a stylized flourish at the end.

Frank Pallone, Jr.
Ranking Member

cc: The Honorable Cathy McMorris Rodgers
Chair