United States Senate

November 16, 2020

Mark Zuckerberg
Chairman and Chief Executive Officer
Facebook, Inc.
1601 Willow Road
Menlo Park, CA 94025

Dear Mr. Zuckerberg:

We write to express our deep concern regarding anti-Muslim bigotry on Facebook. An independent civil rights audit of the company from July 2020 highlighted disturbing examples of anti-Muslim abuse on the platform ranging "[f]rom the organization of events designed to intimidate members of the Muslim community at gathering places, to the prevalence of content demonizing Islam and Muslims, and the use of Facebook Live during the Christchurch massacre . . . ."[1] These concerns have also prompted current Facebook employees to write a letter demanding action on anti-Muslim bigotry and calling for broader structural changes.[2] As members of Congress who are committed to protecting the Muslim community, we urge you to take immediate action to combat this bigotry on Facebook's platforms.

Facebook is a groundbreaking company that has revolutionized the way we communicate. Unfortunately, the connectivity that can bring people together in many positive ways also has been used to dehumanize and stoke violence against Muslims, Black people, Latinos, immigrants, the Jewish community, Sikhs, Christians, women, and other communities here and across the world. The enabling of hate speech and violence against any group is not acceptable. We appreciate that Facebook has taken certain steps to combat these problems. For instance, you recently reversed a prior decision that had allowed content denying the Holocaust, and you have altered your policies to ban blackface and certain anti-Jewish stereotypes. But much more must be done to protect these vulnerable communities. With regard to the Muslim community in particular, the civil rights audit noted advocates' "alarm that Muslims feel under siege on Facebook" and explained how attacks on Muslims present unique considerations that require separate analysis and response compared to other kinds of attacks.[3] Yet, the auditors noted, "Facebook has not yet publicly studied or acknowledged the particular ways anti-Muslim bigotry manifests on its platform."[4]

---

[1] Facebook, *Facebook's Civil Rights Audit – Final Report*, at 57 (July 8, 2020), https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf ("Civil Rights Audit").
[2] Aditya Kalra & Munsif Vengattil, *Exclusive: Facebook Employees Internally Question Policy After India Content Controversy*, Reuters (Aug. 19, 2020), https://www.reuters.com/article/us-facebook-india-exclusive/exclusive-facebook-employees-internally-question-policy-after-india-content-controversy-sources-memos-idUSKCN25F1VW.
[3] Civil Rights Audit, *supra* note 1, at 57.
[4] *Id.*

Of particular concern is how Facebook has addressed the targeting of mosques and Muslim community events by armed protesters through the platform. In June 2019, Facebook responded to concerns about these practices by creating a "call to arms" policy that prohibits event pages that call for individuals to bring weapons to locations.[5] Yet, in August 2019, when advocates reported to Facebook that a militia group was using an event page to plan an armed protest at the largest Muslim community convention in the country for the second year in a row, it took Facebook more than a full day to remove the content, a delay that Facebook acknowledged was too long and an "enforcement misstep."[6]

Other recent events have demonstrated how Facebook has not taken adequate steps to enforce this call to arms policy. In August 2020, a group called the Kenosha Guard posted an event page titled "Armed Citizens to Protect Our Lives and Property," calling for armed individuals to gather in Kenosha, Wisconsin, following the shooting of Jacob Blake. Notwithstanding multiple reports by users that this page violated Facebook policies, Facebook did not take the page down. An armed 17-year-old traveled from out of state to join this gathering, fatally shot two protestors that night, and is charged with their murder. You stated that the failure to take down the event page and the Kenosha Guard's group page was "largely an operational mistake" because contract content moderators without specialized training failed to detect that the pages violated a new militia policy Facebook had established in August 2020.[7] Your statement was misleading as to the event page, however, because it did not mention that the event page also violated the call to arms policy that had been in place for over a year. Importantly, we understand that the contractors who review user-reported content are not instructed to enforce a core component of the call to arms policy. It is not apparent that Facebook ensures meaningful enforcement of this policy, and that is not acceptable. As the Change the Terms Coalition has explained, that "isn't an operational mistake – that's a design defect."[8]

We have similar concerns about Facebook's efforts to ensure that the platform is not used to enable systematic violence and discrimination against Muslims around the world. A United Nations report concluded that the company played a "determining"[9] role in violence against Rohingya Muslims in Myanmar, and Facebook has similarly acknowledged that the platform was used to "foment division and incite offline violence"[10] against the Rohingya. Unfortunately, this is not an isolated incident. According to a *New York Times* report published a month after anti-Muslim violence erupted in Sri Lanka in March 2018, "Facebook's newsfeed played a

---

[5] Facebook, *Community Standards: Violence and Incitement*, https://www.facebook.com/communitystandards/credible_violence; Sheryl Sandberg, *A Second Update on Our Civil Rights Audit* (June 30, 2019), https://about.fb.com/news/2019/06/second-update-civil-rights-audit/.
[6] Civil Rights Audit, *supra* note 1, at 58.
[7] Ryan Mac & Craig Silverman, *How Facebook Failed Kenosha*, BuzzFeed News (Sept. 3, 2020), https://www.buzzfeednews.com/article/ryanmac/facebook-failed-kenosha.
[8] Letter to Mark Zuckerberg from the Change the Terms Coalition (Sept. 2, 2020), https://www.freepress.net/sites/default/files/2020-09/L-ChangeTheTerms-Facebook-9-2-2020.pdf.
[9] BBC News, *UN: Facebook Has Turned Into a Beast in Myanmar* (Mar. 13, 2018), https://www.bbc.com/news/technology-43385677#:~:text=UN%20investigators%20have%20said%20the,%22turned%20into%20a%20beast.%22.
[10] Facebook, *An Independent Assessment of the Human Rights Impact of Facebook in Myanmar* (Nov. 5, 2018), https://about.fb.com/news/2018/11/myanmar-hria/ ("FB Myanmar Response").

central role in nearly every step from rumor to killing,"[11] despite numerous attempts by Sri Lankan activists and government officials to warn Facebook about potential outbreaks of violence. In an especially horrific episode of anti-Muslim activity on Facebook, in March 2019, a white nationalist gunman broadcasted his 17-minute slaughter of 51 Muslims at two mosques in Christchurch, New Zealand, for the entire world to see using Facebook Live. Reports indicate that the platform has also been used to support the internment of the Uyghurs in China and other human rights violations against this population,[12] that Facebook and WhatsApp have been used to incite violence against Muslims in India,[13] and that Facebook has been used to promote hate and violence in other areas around the world.[14]

The civil rights audit and other reports have documented the shortcomings of Facebook that have led to these results over the years. The United Nations explained in 2018 that Facebook launched its Myanmar-specific services without content moderators who spoke the necessary languages, without adequate technology, and without sufficient transparency and coordination with local organizations. It also documented how speech in clear violation of Facebook's policies remained on the platform notwithstanding multiple reports, and how even after the speech was taken down, re-posts continued to circulate months later. Furthermore, the civil rights audit found that Facebook is not sufficiently attuned to how its algorithms "fuel extreme and polarizing content," and thereby may "driv[e] people toward self-reinforcing echo chambers of extremism," as seen in Myanmar and Sri Lanka.[15] Advocacy groups similarly detailed the extent and persistence of anti-Muslim hate content on Facebook India in multiple reports last year,[16] concerns that have been amplified by recent allegations that some high-ranking employees at Facebook India have enabled hate speech against Muslims and others by applying the platform's content moderation policies in a selective manner.

We recognize that Facebook has announced efforts to address its role in the distribution of anti-Muslim content in some of these areas. These include, for instance, adding country-specific staff and content moderators proficient in certain local languages, investing in proactive detection

---

[11] Amanda Taub & Max Fisher, *Where Countries Are Tinderboxes and Facebook Is a Match*, New York Times (Apr. 21, 2018), https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html.

[12] Ryan Mac, *These New Facebook Ads From Chinese State Media Want You To Believe Xinjiang's Muslim Internment Camps Are Just Great*, BuzzFeed News (Aug. 20, 2019), https://www.buzzfeednews.com/article/ryanmac/chinse-media-facebook-ads-xinjiang-uighur-propaganda.

[13] Newley Purnell & Jeff Horwitz, *Facebook's Hate-Speech Rules Collide With Indian Politics*, Wall Street Journal (Aug. 14, 2020), https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346; Ismat Ara, *"Tear Them Apart": How Hindutva WhatsApp Group Demanded Murder, Rape of Muslims in Delhi Riots*, The Wire (July 6, 2020), https://thewire.in/communalism/delhi-riots-hindutva-whatsapp-muslims-murder-rape.

[14] Muslim Advocates and The Global Project Against Hate and Extremism, *Complicit: The Human Cost of Facebook's Disregard for Human Life* (Oct. 21, 2020), https://muslimadvocates.org/wp-content/uploads/2020/10/Complicit-Report.pdf.

[15] Civil Rights Audit, *supra* note 1, at 56.

[16] Avaaz, *Megaphone for Hate: Disinformation and Hate Speech on Facebook During Assam's Citizenship Count* (Oct. 2019), https://avaazpress.s3.amazonaws.com/FINAL-Facebook%20in%20Assam_Megaphone%20for%20hate%20-%20Compressed%20(1).pdf; Equality Labs, *Facebook India – Towards the Tipping Point of Violence: Caste and Religious Hate Speech* (2019), https://static1.squarespace.com/static/58347d04bebafbb1e66df84c/t/5d0074f67458550001c56af1/1560311033798/Facebook_India_Report_Equality_Labs.pdf.

technologies, strengthening local fact-checking partnerships, and limiting the ability to reshare certain kinds of messages.

Nevertheless, it is not clear that the company is meaningfully better positioned to prevent further human rights abuses and violence against Muslim minorities today. In part, this is because Facebook still does not collect the information needed to evaluate the effectiveness of its responses. For instance, Facebook reported that it took action on 22 million pieces of hate speech content in the second quarter of 2020, up from over 9 million in the first quarter.[17] It is not apparent, however, whether this is a sign of an improving or worsening problem, because this data lacks crucial context: Facebook does not calculate or report on the overall prevalence of hate speech on the platform. It is thus unclear how significant this increase is as a proportion of total hate speech or whether takedowns are increasing only because hate content on the platform is increasing. Facebook recognizes that the statistic it reports "only tells part of the story,"[18] and Facebook does estimate prevalence in other contexts. Its failure to do so as to hate speech is concerning.[19]

In addition, the civil rights audit pointed out that for content that Facebook does remove, the company does not collect data about which protected groups were the target of the removed post. This prevents Facebook and the public from understanding the volume of hate against a particular group, whether attacks against certain groups are consistently not removed, and whether there are gaps in Facebook's policies that result in perpetuating or increasing hate speech and attacks against particular groups. It is difficult to understand how Facebook can effectively combat hate speech without this information.

There is also basic information that Facebook has or could readily make available, but which it has inexplicably declined to make public. For instance, while pointing to its increases in country-specific staff and language-specific content moderators in certain areas, Facebook has declined repeated requests from advocates to provide detailed information about its country-specific staff or language-specific content moderators across the world. Such information is necessary to evaluate Facebook's suggestion that its additions are adequate and to determine whether there are gaps in coverage in other regions that should be addressed proactively before the next violent event. Facebook similarly does not provide information about how the hate speech it has taken down is disaggregated by language or country of origin, information that would help identify volatile areas in need of further attention from content moderators or others at Facebook. That is so even though Facebook has conceded that "[t]hese breakdowns are

---

[17] Facebook, *Community Standards Enforcement Report* (Aug. 11, 2020), https://transparency.facebook.com/community-standards-enforcement.

[18] Facebook, *Understanding the Facebook Community Standards Enforcement Report* (May 2018), https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf.

[19] In addition, Facebook's Data Transparency and Advisory Group (DTAG) specifically recommended calculating prevalence by region. *See* The Justice Collaboratory, Yale Law School, *Report of the Facebook Data Transparency Advisory Group* (Apr. 2019), https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf. Facebook told the DTAG that the prevalence of hate speech "cannot currently be objectively identified across contexts at scale." *Id.* at 22. While we are cognizant that any estimate of prevalence will involve technical challenges, we are confident that a company with Facebook's ingenuity and resources can develop actionable metrics along these lines.

feasible for these count-based metrics" and that it "recognize[s] the value in having different subpopulations of the various metrics."[20] The United Nations 2018 Myanmar report expressed "regret[]" that Facebook did not provide country-specific data about hate speech and deemed it "essential" that such information be disclosed.[21]

Though these concerns have been raised for years, Facebook thus far has not taken the steps required to effectively address hate speech and violence targeting Muslims. In 2018, Facebook acknowledged that it "can and should do better" after its platform fueled violence in Myanmar and outlined steps it would take.[22] In 2020, Facebook "apologize[d] for" the human rights impacts that resulted from misuse of its platform in Sri Lanka and outlined more steps.[23] Despite these experiences, recent reporting suggests that today, Facebook is contributing to the spread of hate speech and violence against ethnic and religious groups in Ethiopia, where Facebook "dominates" the internet.[24] Meanwhile, it announced a call to arms policy to assuage concerns but has failed to adequately enforce it. As members of Congress who are deeply disturbed by the proliferation of this hate speech on your platform, we urge you to do more. We believe Facebook must frankly and openly detail the scope of the problem and take concerted and sustained actions to address this problem fully. We respectfully request that you respond to the questions below by December 16, 2020. As to each question, insofar as Facebook will commit to taking action, please provide details of its plan and expected timing.

1. Will Facebook commit to developing and implementing a plan to ensure robust enforcement of its call to arms policy, including through proactive review of event pages, content moderator review of user reports, and prioritization of highly reported events? If not, why not?

2. Will Facebook commit to collecting and publishing data about the overall amount and prevalence of hate content on the platform and whether hate content is increasing on its platform? If so, please specify whether Facebook will break down this data by country and language. If not, why not?

3. Will Facebook commit to collecting and publishing data about which groups were the subject of the hate speech it removes and enforcement rates across groups? If so, please specify the groups for which Facebook will provide this information. If not, why not?

[20] Radha Iyengar Plumb, *Exploring Feedback From Data and Governance Experts: A Research-Based Response to the Data Transparency Advisory Group Report* (May 23, 2019), https://research.fb.com/exploring-feedback-from-data-and-governance-experts-a-research-based-response-to-the-data-transparency-advisory-group-report/.

[21] United Nations Human Rights Council, *Report of the detailed findings of the Independent International Fact-Finding Mission on Myanmar*, A/HRC/39/CRP.2, 343 (Sept. 17, 2018), https://www.ohchr.org/Documents/HRBodies/HRCouncil/FFM-Myanmar/A_HRC_39_CRP.2.pdf?utm_campaign=The%20Interface&utm_medium=email&utm_source=Revue%20newsletter.

[22] FB Myanmar Response, *supra* note 10.

[23] Joshua Brustein, *Facebook Apologies for Role in Sri Lankan Violence*, Bloomberg (May 12, 2020), https://www.bloomberg.com/news/articles/2020-05-12/facebook-apologizes-for-role-in-sri-lankan-violence.

[24] David Gilbert, *Hate Speech on Facebook Is Pushing Ethiopia Dangerously Close to Genocide*, VICE News (Sept. 14, 2020), https://www.vice.com/en_us/article/xg897a/hate-speech-on-facebook-is-pushing-ethiopia-dangerously-close-to-a-genocide.

4. Will Facebook commit to collecting and publishing country-specific or language-specific data on hate speech that is on or removed from the platform? If not, why not?

5. Will Facebook publish detailed information about the number of country-specific staff and language-specific content moderators it employs? If not, why not?

6. Will Facebook commit to studying regularly its civil rights and human rights impacts and making future human rights impact assessments or rights audits public in their entirety? If not, why not?

7. Will Facebook commit to establishing and publishing criteria that must be met for Facebook to expand or maintain usage of its services in markets at risk of hate content fueling religious and/or ethnic violence to ensure Facebook does not enable human rights violations? If so, please specify the outside input that Facebook will solicit in developing these criteria. If not, why not?

8. Will Facebook conduct an analysis of how it can better design its systems and algorithms to not just identify and take down hate speech, but limit the reach of this content and its ability to cause offline violence? If not, why not?

9. Will Facebook commit to creating a working group led by a senior employee with expertise in anti-Muslim bigotry specifically tasked with monitoring, reviewing, and coordinating efforts to proactively remove anti-Muslim content on the platform? If not, why not?

Thank you for your consideration of our views. We appreciate your prompt attention to this matter.

Sincerely,

CHRISTOPHER A. COONS
United States Senator

RICHARD BLUMENTHAL
United States Senator
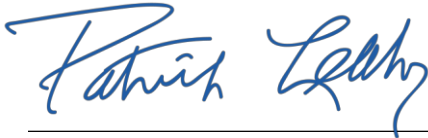

MAZIE K. HIRONO
United States Senator

RICHARD J. DURBIN
United States Senator

MARK R. WARNER
United States Senator

ROBERT MENENDEZ
United States Senator

PATRICK LEAHY
United States Senator

BENJAMIN L. CARDIN
United States Senator

MICHAEL F. BENNET
United States Senator

GARY C. PETERS
United States Senator

AMY KLOBUCHAR
United States Senator

KIRSTEN GILLIBRAND
United States Senator

ELIZABETH WARREN
United States Senator

CHRISTOPHER S. MURPHY
United States Senator

BERNARD SANDERS
United States Senator