

15 September 2020

Dear Special Rapporteur Andrews,

I am responding to the letter from your predecessor addressed to Mark Zuckerberg on December 11, 2019. I do appreciate you raising this letter with us after taking up your mandate, allowing us the opportunity to respond.

The information we're providing in response to your questions is attached to this cover letter. We hope it gives you the information you need.

Our work on Myanmar, including our human rights due diligence, and our efforts to prevent and mitigate human rights harms, is some of the most important work being done at Facebook.

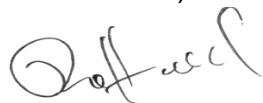
We have made a major effort in recent years to address the challenges we face in Myanmar, which our dedicated product, engineering, partnerships, operations and policy teams continue to work on. This includes our work to root out abuse in the run up to the country's 2020 elections. More than twenty teams are working on preparing for the forthcoming election, and we began our work well over a year ago.

Please do not hesitate to reach out if we can clarify any questions or provide you with any further information. You will appreciate that, for security or other human rights-related reasons, we cannot necessarily disclose all details in a public letter.

Please be assured of my commitment, and that of my colleagues, to continue our dialogue with you on issues relevant to your important mandate, and to listen to your assessment of any additional issues we should consider and address. We would be happy to speak with you at your earliest convenience.

Yours sincerely,

Rafael Frankel, PhD



Director of Public Policy Southeast-Asia

This attachment is a response to the letter from the Special Rapporteur for Myanmar, forwarded to Facebook by OHCHR Special Procedures on December 11, 2019.

1. Please provide information about what bodies of law are used to designate groups as “dangerous” and what Facebook’s factual and legal evaluation is for both state and non-state actors.

Facebook has developed robust definitions of different types of dangerous organizations and individuals who proclaim a violent mission, or are engaged in violence, and thus prohibited from having a presence on Facebook.

Our dangerous organizations policy and the related definitions are [posted publicly](https://www.facebook.com/communitystandards/dangerous_individuals_organizations). They are available at: https://www.facebook.com/communitystandards/dangerous_individuals_organizations

We developed these definitions through a process of extensive outside consultation with experts in a variety of relevant fields, including counterterrorism, human rights, and law, and we use these definitions as a basis for assessing groups for designation. In addition, we also consider whether dangerous individuals and organizations have been designated by relevant sanctions authorities.

2. Please provide information about the procedure followed by which organisations are designated as “dangerous”, including the process for informing organisations or individuals of the designation, whether reasons are given, and any review or appeal process available to designated organisations. Was a human rights impact assessment of any designation or page removal made?

When designating groups as “dangerous organizations”, Facebook engages in a process of structured research and review, focused on determining an organization’s goals and whether there is a track record of offline violence. (See below). The process involves multiple teams across the company. We make findings based on the credible evidence available and the detailed assessments of multiple teams. Users receive a message from Facebook if their accounts or content have been removed under this process.

We conduct routine, ongoing human rights due diligence on our content policy decisions. Our designation process was also referenced as part of our existing HRIA on Myanmar, and will be re-evaluated as part of any future HRIA updates.

3. Please describe the factual basis and the process that was followed to designate the four organisations as “dangerous” in February 2019. Was information provided to them about the process and their designation and any possibility of review or appeal? Please also provide information about any warnings or restrictions considered and applied to the four organisations.

We designate entities under our Dangerous Individuals and Organizations policy after a rigorous process that takes into account both online and offline behavior. During this process, we work to identify an organization’s goals and if it has a track record of offline violence.

We designated these four organizations in February 2019 because there was clear evidence that they had been responsible for attacks against civilians and other practices contrary to the laws of armed conflict, such as (in some cases) the conscription of child soldiers. This analysis was based on multiple external sources, including (but not limited to): media reports; [this report](#)¹ on activity in Chin State; this [June 2016 OHCHR report](#)²; and this recent [2019 OHCHR report](#)³; as well as from the ongoing investigations by our internal team of counterterrorism and country experts.

You asked about warnings or restrictions. Any entity designated as a dangerous organization is banned from having a presence on Facebook. In addition, when we identify content that praises, supports or represents these groups, we remove it. This is a policy through which we seek to prevent or mitigate extremely severe offline harms; you will appreciate we cannot discuss details in a public letter.

However, we recognize the complexities of the local environment and the place these four groups have within Myanmar's diverse, multi-ethnic society. While these organizations remain banned from Facebook, we're taking additional measures to allow for robust community debate about them, in close alignment with the law of armed conflict. We look forward to sharing more soon.

4. Please describe the procedure to review and remove content and posts for inauthentic behavior, and explain if it is different to the removal of content deemed to be "dangerous", including the process of informing organisations or individuals, whether reasons are given, and any review or appeal process made available.

Our Inauthentic Behavior policies are separate from our policies around dangerous organizations and individuals. They are [publicly available](#) in our Community Standards at https://www.facebook.com/communitystandards/inauthentic_behavior.

They focus exclusively on behaviors—not content—seeking to: mislead users about the identity of users engaging on Facebook and Instagram; amplify their content in contravention of our existing limitations; or to evade our enforcement systems.

A subset of our inauthentic behavior enforcement includes our public removals of Coordinated Inauthentic Behavior (CIB), which describes coordinated operations that use fake accounts in order to mislead users.

We publicly disclose our CIB enforcement in a [monthly report](#)⁴ that describes the behaviors we observed; the number of Facebook and Instagram assets removed; and attribution (to the best of our ability) to the networks behind the activity.

Removed users receive in-product notification that their accounts and associated assets were disabled for violations of our Community Standards.

Furthermore, we consistently share information about these networks with third-party independent researchers. This enables further analysis and public reporting about the various Facebook and Instagram assets removed in these processes.

¹ https://www.ecoi.net/en/file/local/1405345/1226_1501750161_597f06624.pdf

² https://www.ecoi.net/en/file/local/1261211/1930_1484061108_g1613541.pdf

³ <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=24089&LangID=E>

⁴ <https://about.fb.com/news/2020/09/august-2020-cib-report/>

We do not allow for appeals by disabled users because every CIB enforcement is vetted at several levels for analytic rigor and narrowness of enforcement scope. CIB is among the most severe policy violations on Facebook; we also take additional steps to ensure that CIB actors do not return to the platform.

5. Please provide information about any other pages, including those of “related praise, support and representation” of the four organisations designated “dangerous”, that were removed.

These four groups currently remain designated. As noted, we do not allow support or representation of designated groups on our platform as part of our efforts to prevent and disrupt real-world violence and thereby protect our users.

That said, we do recognize the prominent role that these groups play in their communities, and we do allow users to talk about the groups. We also allow the news media to report on and discuss these groups.

Going forward, we will continue to review content, individuals and organizations against our policies and legal obligations, and to reassess how our policies apply to ways in which people choose to discuss important regional events and developments.

6. Has Facebook banned any accounts or removed any pages or profiles from Myanmar other than those that were included in the announcements in 2018 and 2019? If yes, what were the reasons for their banning and removal? Why were they not made public?

As you note, we publicly announced multiple takedowns in 2018 and 2019. These were related to three processes: the banning of Myanmar entities; the designation of four groups as dangerous organizations; and routine ongoing Coordinated Inauthentic Behavior takedowns, which take place monthly.

These takedowns, publicly announced, represent processes separate from our day to day enforcement of our Community Standards. Thus, we also regularly remove profiles and pages that repeatedly violate our Community Standards, including our policies on hate speech and impersonation. We have also designated a number of hate figures and organisations in Myanmar, including [REDACTED], [REDACTED], [REDACTED], [REDACTED], and [REDACTED], and will remove content that engages in praise, support, or representation of these designees. Such removals are a regular and ongoing part of our global enforcement policies, and are not publicly announced, but are included in the figures that make up our global [Community Standards Enforcement Report](#)⁵.

We also continue to invest in technology that helps us to remove recidivist accounts deleted for Community Standards violations (such as hate speech violations) to try to ensure that these accounts do not have presence on our platform again.

7. Please outline Facebook’s data preservation policy for content that is removed as a result of moderation and a ban, including whether content is being preserved by Facebook and under what conditions. Does Facebook delete any data that it removes?

⁵ <https://transparency.facebook.com/community-standards-enforcement>

Our data preservation is guided by the requirements of U.S., Irish, European law, as well as by our terms of service. In our [data policy](#)⁶, in the paragraph titled “How Do We Respond to Legal Requests or Prevent Harm”, we explain we retain information from accounts disabled for terms violations for at least a year to prevent repeat abuse or other term violations. In addition, it is important to note that we routinely preserve data at the lawful request of law enforcement authorities. We have a single set of [global guidelines](#)⁷, publicly available in our [Safety Center](#)⁸.

In the case of Myanmar, consistent with our public commitment to work with and provide information to relevant authorities, we have likewise acted on preservation requests from the UN Fact Finding Mission on Myanmar and the Independent Investigative Mechanism on Myanmar. In addition, we preserved data related to a variety of Myanmar-related account takedowns and CIB investigations from 2018 to date. As has been publicly reported, Facebook is seeking to assist international accountability efforts, and begun a process of lawful disclosure to the IIMM in August 2020.

8. Please provide information about Facebook’s plans to align its Community Standards and methods of assessing content with international human rights standards, including the Rabat Plan of Action on the prohibition of advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence.

Facebook’s Community Standards are [broadly aligned](#)⁹ with international human rights norms. They are designed to guide content moderation decisions for many millions of pieces of content daily. They are revised regularly, and we also look forward to learning from the feedback from our new Oversight Board.

Our hate speech policies, available [here](#)¹⁰, are based on extensive legislative and human rights legal analysis, as well as statistical research on the most patterns of hate speech on the platform. They are framed to respect the standards Articles 19 and 20 (2) of the ICCPR, the latter forming an important foundation for effective implementation of the Rabat Principles.

Multiple other Facebook policies are also relevant to Rabat Implementation (see below).

One challenge is that the Rabat Principles articulate a (suitably) high threshold for the criminal prohibition of speech, and ideally require exploration of intent and other factors social media companies are not well-positioned to determine.

Nevertheless, we are incorporating frameworks and insights from the Rabat Principles in other areas of our Community Standards, including our Harmful Stereotypes Policy; our Misinformation and Real World Harm policy; and our Unverifiable Rumors policy (all expanded or launched in 2020). We have also begun to use them as guidance for policy revision elsewhere.

We are currently exploring additional methods to build from the general principles embodied in Rabat’s six part test, and to incorporate them into other policy areas. We recently assisted OHCHR translate the six part test into more than 25 languages, including [Burmese](#) and Bengali¹¹.

⁶ <https://www.facebook.com/policy.php>

⁷ <https://www.facebook.com/safety/groups/law/guidelines/>

⁸ <https://www.facebook.com/safety>

⁹ <https://about.fb.com/news/2019/09/updated-the-values-that-inform-our-community-standards/>

¹⁰ <https://www.facebook.com/communitystandards/>

¹¹ <https://www.ohchr.org/EN/NewsEvents/Pages/Hate-speech-threshold-test.aspx>

We're also incorporating relevant guidance into our [distribution ranking methodologies](#)¹² for inflammatory speech in countries at risk of conflict. You may wish to read the [Facebook Response to the Sri Lanka HRIA](#)¹³ for a more in depth review of relevant conflict prevention and mitigation initiatives, as well as this [newsroom post](#)¹⁴ on our integrity investments related to social media and conflict.

We note that Myanmar's transition from Zawgyi to the global font encoding standard [Unicode](#) is a very significant development. We have invested very significant resources in enabling that transition. Indeed, it has enabled us to build crucial data, research, and integrity measures that were previously impossible to develop or deploy. Likewise, you may wish us to brief you on other product-related steps we're taking to reduce human rights risks in Myanmar and are consistent with the Rabat Principles.

9. Please provide information about how Facebook manages content moderation in Myanmar and the human resources dedicated to it.

There are now more than 35,000 people working on safety and security issues across the company. More than 15,000 people out of the 35,000 are dedicated content reviewers and they include 120 content reviewers who are native Burmese speakers.

We've built a dedicated Myanmar team, including people who are Myanmar nationals and who spend a significant amount of time talking to our partners on the ground—including civil society partners—to try to understand current issues, as well as other issues on the horizon.

We recognize that there are certain types of content, such as hate speech, that could lead to imminent, offline harm but that could also suppress the vote. We have a clear and detailed policy against hate speech, and we remove violating content as soon as we become aware of it.

To do this, we've invested significantly in proactive detection technology to help us catch violating content more quickly. We also use AI to proactively identify hate speech in 45 languages, including Burmese.

We have continued to invest in improving this technology and our overall enforcement against hate speech as the election approaches. In the second quarter of 2020, we took action against 280,000 pieces of content in Myanmar for violations of our Community Standards prohibiting hate speech, of which we detected 97.8% proactively before it was reported to us. This is up significantly from Q1 2020, when we took action against 51,000 pieces of content for hate speech violations, detecting 83% proactively.

Our strong commitment to address issues of concern in Myanmar continues. We look forward to ongoing dialogue with your mandate in doing so, and very much appreciate your engagement and shared experience and expertise.

#END#

¹² <https://about.fb.com/news/2019/06/social-media-and-conflict/>

¹³ <https://about.fb.com/news/2020/05/human-rights-work-in-asia/>

¹⁴ <https://about.fb.com/news/2019/06/social-media-and-conflict/>