**Summary and Outcome**

On December 14, 2021, the Oversight Board upheld Meta’s original decision to remove a post alleging the involvement of ethnic Tigrayan civilians in atrocities in Ethiopia’s Amhara region. Meta initially applied the Hate Speech Community Standard to remove the post from Facebook but restored it after the Board selected the case. The Board found Meta’s explanation for restoration lacked detail and deemed it incorrect. The Board determined that the content violated the prohibition on unverified rumors under the Violence and Incitement Community Standard.

\*The Oversight Board is a separate entity from Meta that provides its independent judgment on individual cases and policy questions. An independent trust funds both the Board and its administration. The Board has the authority to decide whether Facebook and Instagram should allow or remove content. These decisions are binding unless implementing them could violate the law. The Board can also choose to issue company content policy recommendations.

**Facts**

The case concerns allegations made during the ongoing civil war in Ethiopia that erupted in 2020 between the Tigray region’s forces, Ethiopian federal government forces, military, and its allies.

In July 2021, a Facebook user posted in Amharic on his timeline that the Tigray People's Liberation Front (TPLF), alongside ethnic Tigrayan civilians, killed and raped women and children and looted the properties of civilians in Raya Kobo and other towns in Ethiopia's Amhara region. The user ended the post with the following words: "we will ensure our freedom through our struggle." [p. 1] The user claimed that he received the information from the residents of Raya Kobo. The post remained on Facebook for approximately one day. However, during that period, it was viewed nearly 5,000 times, received fewer than 35 comments and more than 140 reactions, and was shared over 30 times. Among the comments in Amharic were statements that read: "[o]ur only option is to stand together for revenge” and “are you ready, brothers and sisters, to settle this matter?” [p. 4]

The post was identified by Meta’s Amharic language automated system as potentially violating its policies. A content moderator from the Amharic content review team determined that the post violated Facebook’s Hate Speech Community Standard and removed it. The user then appealed the decision to Meta, and, following a second review by another moderator from the Amharic content review team, Meta confirmed that the post violated Facebook’s policies. As a result, the user then submitted an appeal to the Oversight Board.

On August 27, after the Board had selected the case for review, Meta identified the post’s removal as an “enforcement error” and restored it.

**Decision Overview**

The main issue for the Board was whether Meta's decision to remove a post alleging the involvement of ethnic Tigrayan civilians in atrocities in Ethiopia's Amhara regions violated  Facebook’s Community Standard on Violence and Incitement. The Board also examined if the post's removal aligned with the company's values and human rights responsibilities.

The user stated in his appeal to the Board that he posted this content to protect his community which was in danger. He also said the post did not constitute hate speech but the truth. He also noted that the TPLF targeted his community of one million people and left them without food, water, and other necessities.

Meta explained in its rationale that it initially removed the content for violating its policy prohibiting "violent speech" targeted at Tigrayan people based on ethnicity. However, Meta explained that its moderators did not record their reasons for removing content beyond indicating the infringed Community Standard. Thus, it could not confirm if the moderators who reviewed the post initially and on appeal applied the same criteria within Facebook's Hate Speech policy to remove the post.

The company further noted that as a result of the Board selecting the case, it had determined that its "decision was an error" and restored the post. Upon reexamination, it found that the content did not violate its rules since it did not target the Tigray ethnicity, and the user's allegations about the TPLF or Tigrayans did not rise to the level of hate speech. Additionally, Meta corroborated that its automated Amharic systems are in place, audited, and refreshed every six months. It also explained that the original text in Amharic led the automated system to identify the content as potentially violating. Similarly, the company confirmed that the two content moderators were Amharic speakers and based their review on the original text in Amharic.

The Board requested and received an additional English translation of the text from its linguistic experts, and Meta provided a further translation of the text by its external linguistic vendor. The two versions confirmed that the prevailing meaning of the text indicates that Tigrayan civilians assisted in the atrocities of the TPLF.

Moreover, Meta established that Ethiopia was "designated as a Tier 1 At-Risk Country" [p. 11], the highest risk level. Further, it stated that it had designated Ethiopia as "a crisis location" for its content policy and integrity work.

After explaining the arguments in the parties' submissions, the Board noted that the case concerned allegations made during an ongoing civil and ethnic war in a region with a history of lethal ethnic conflict. Thus, it explained that it was aware of the tension between protecting freedom of expression and reducing the threat of sectarian strife. According to the Board, such tension could only be resolved by attending to the specifics of a given conflict.

The Board stressed its understanding of civilian involvement in the atrocities in various parts of Ethiopia, though not in Raya Kobo, and that at the time of the posting, Meta could not and did not proactively verify the allegations given the communication blackout in the Amhara region. The Board also said it was aware that accurate reports on atrocities could be lifesaving in conflict zones by putting potential victims on notice of possible perpetrators. Yet, it underscored that unsubstantiated claims regarding civilian perpetrators in an ongoing heated conflict likely pose heightened risks of near-term violence.

*Compliance with Community Standards*

To examine whether Meta had complied with the Community Standards by restoring the content, the Board noted that rumors alleging the complicity of an ethnic group in mass atrocities are dangerous and significantly increase the risk of imminent violence during an ongoing violent conflict such as presently in Ethiopia. It held that the post violated the Violence and Incitement policy's prohibition on misinformation and unverifiable rumors that contribute to the risk of imminent violence or physical harm. Thus the Board considered that the content fell within Meta's Internal Implementation Standards definition of an unverifiable rumor.

*Compliance with Meta’s values*

Regarding whether the company's action had complied with Metas values, the Board found that the decision to restore and allow the content was inconsistent with its values of "Dignity" and "Safety".

*Compliance with Meta’s human rights responsibilities*

The Board then proceeded to analyze if removing the content in the present case was consistent with Meta's human rights responsibilities as a business. It stressed that unverifiable rumors in a heated and ongoing conflict might lead to grave atrocities, which the experience in Myanmar has indicated. Therefore, the Board underscored the importance of a transparent system of moderating content in conflict zones, including a policy regarding unverifiable rumours necessary to mitigate such a risk.

To assess the impact of Meta's decision concerning Freedom of Expression, the Board relied on Article 19 of the International Covenant on Civil and Political Rights (ICCPR), which provides broad protection for freedom of expression through any media and regardless of frontiers. However, the Board stressed that the Article allows this right to be restricted under certain narrow and limited conditions. The Board employed the three-part test to assess whether the company's decision to restore the content was in line with such standards.

*I. Legality (clarity and accessibility of the rules)*

First, the Board recalled that the requirement of legality demands that any restriction on freedom of expression is adequately accessible so that individuals can clearly indicate how the law limits their rights. The law must also be formulated with sufficient precision so that individuals can regulate their conduct. In the Board's view, the Community Standards did not define the term "unverifiable rumor." As applied to the facts of this case in which an unverified allegation was made during an ongoing violent conflict, the Board found that the term "unverifiable rumor" provided sufficient clarity. Given the continuing conflict and the communications blackout, it deemed that the rumor was not verifiable for Meta or the user who was not present in Raya Kobo. Thus considering such circumstances, the Board determined it was foreseeable for users that such a post falls within the prohibition.

*I. Legitimate aim*

The Board then recalled that restrictions on freedom of expression must pursue a legitimate aim, which includes the protection of the rights of others. It explained The Facebook Community Standard on Violence and Incitement existed to prevent offline harm that could be related to content on Facebook. Therefore, the Board considered that restrictions based on this policy served the legitimate aim of protecting the right to life.

*III. Necessity and proportionality*

Lastly, the Board specified that International Human Rights law requires that restrictions on the expression should be appropriate to achieve their protective function. They must be the least intrusive instrument amongst those which might achieve their protective function and must be proportionate to the interest to be protected. Moreover, it underscored that the principle of proportionality demanded consideration for the form of expression at issue. The Board stressed the importance of shedding light on human rights violations in a conflict situation. It noted that information on atrocities could save lives, especially where social media are the ultimate source of information. However, the Board highlighted that the user's post did not contain information about an actual threat to life and did not contain specific information that could have been used to document the human rights violation.

The Board then highlighted that on its own, unverifiable rumors could not directly and immediately cause harm. However, when such content appears on an important, influential, and popular social media platform during an ongoing conflict, the risk and likelihood of harm could become more pronounced. Given that the content was posted during an armed conflict, the Board considered that Meta had to exercise heightened due diligence to protect the right to life. The Board believed that removing the content was necessary in the immediate case to prevent innumerable posts from feeding into a hateful narrative through unverified rumors during an ongoing violent ethnic conflict.

*Policy advisory statement:*

The Board recommended that Meta rewrite its value of "Safety" to reflect that online speech may pose a risk to the physical security of persons and the right to life, in addition to intimidation, exclusion, and silencing. Likewise, it urged the company to reflect in the Facebook Community Standards that unverified rumors pose a higher risk to the rights of life and security of persons in war and violent conflict. Finally, the Board urged Meta to commission independent human rights due diligence assessment on how Facebook and Instagram have been used to spread hate speech and unverified rumors that heighten the risk of violence in Ethiopia.

*Dissenting or Concurring Opinions:*

A minority of the Board highlighted its understanding of the limited nature of this decision. They stated that in the context of an ongoing violent conflict, a post constituting an unverified rumor of ethnically-motivated violence by civilians against other civilians posed severe risks of escalating an already violent situation. Particularly where Meta could not verify the rumor in real time.

**Decision Direction**

The Board decision contracted expression by upholding Facebook's decision to remove the content. However, the Board explained that the company did so under a justified and recognized limitation to freedom of expression. It considered that the post in question did not contain information about an actual threat to life and did not include specific details that could be used to document human rights violations. Thus, the Board deemed that unsubstantiated claims regarding civilian perpetrators are likely to pose heightened risks of near-term violence in an ongoing heated conflict.

**Related International and/or regional laws**

The Board interpreted the right to life contained in this article as interpreted by General Comment No. 36 of the Human Rights Committee (2018).

The Board analyzed Facebook's human rights responsibilities through this precept on freedom of expression and employed the three-part test established.

While employing the three-part test to assess if Facebook’s actions allowed expression to be limited, the Board referred to the General Comment for guidance.

Significance

Article 2 of the [Oversight Board Charter](https://www.oversightboard.com/governance/) states, "For each decision, any prior board decisions will have precedential value and should be viewed as highly persuasive when the facts, applicable policies, or other factors are substantially similar. In addition, Article 4 of the Oversight Board Charter establishes, "The Board's resolution of each case will be binding, and Facebook (now Meta) will implement it promptly unless implementation of a resolution could violate the law. In instances where Facebook identifies that identical content with similar context – which the Board has already decided upon – remains on Facebook (now Meta), it will take action by analyzing whether it is technically and operationally feasible to apply the Board's decision to that content as well. When a decision includes policy guidance or an advisory policy opinion, Facebook (now Meta) will take further action by analyzing the operational procedures required to implement the guidance. Considering it in the formal policy development process of Facebook (now Meta), and transparently communicating about actions taken as a result."