



Weapons Post Linked to Sudan's Conflict

2023-028-FB-UA

Summary

The Oversight Board has upheld Meta's decision to remove a post containing a graphic of a gun cartridge, accompanied by a caption providing instructions on how to create and throw a Molotov cocktail. The Board finds the post violated Facebook's Violence and Incitement Community Standard, posing an imminent risk of harm that could exacerbate ongoing violence in Sudan. This case has raised broader concerns about Meta's human rights responsibilities for content containing instructions for weapons shared during armed conflicts. To meet these responsibilities, Meta should ensure exceptions to its violence and incitement rules are clearer. Additionally, Meta should develop tools to correct its own mistakes when it has sent the wrong notification to users about which Community Standard their content violated.

About the Case

In June 2023, a Facebook user posted an illustration of a gun cartridge, with the components identified in Arabic. The post's caption provides instructions on how to create a Molotov cocktail using the components and advises wearing a helmet when throwing the incendiary device. It concludes with a call for victory for the Sudanese people and the Sudanese Armed Forces (SAF).

Two months before the content was posted, fighting broke out in Sudan between the SAF and the Rapid Support Forces (RSF), a paramilitary group designated as dangerous by Meta in August 2023. Sudan's armed conflict is ongoing and has spread across the country, with both sides having used explosive weapons in areas densely populated by civilians.

Meta's automated systems detected the content, determining that it violated Facebook's Violence and Incitement Community Standard. Meta removed the post, applying a standard strike to the user's profile. The user immediately appealed. This led to one of Meta's human



reviewers finding that the post violated the Restricted Goods and Services policy.

The user then appealed to the Board, after which Meta determined the content should have been removed but, as per its original decision, under the Violence and Incitement Community Standard.

Key Findings

The Board finds the post violated the Violence and Incitement policy in two ways. First, the combined effect of the image and caption violated the rule that prohibits “instructions on how to make or use weapons where there is language explicitly stating the goal to seriously injure or kill people.” Regardless of the intent of the person who created the post, the step-by-step guide on how to build a Molotov cocktail and the advice to “use a helmet” indicates the content is calling on people to act on the instructions. Second, resorting to violence in support of the SAF during the ongoing armed conflict does not relate to a non-violent purpose. The Violence and Incitement policy prohibits instructions on how to make weapons, unless there is “context that the content is for a non-violent purpose.”

The rule that prohibits instructions on making and using weapons does include an exception for content when it is shared for “recreational self-defense, military training purposes, commercial video games or news coverage.” Stakeholders consulted by the Board as well as news reports have claimed that Meta allows such instructions in exercise of self-defense for some armed conflicts. Meta has denied this is true. The Board is not in a position to determine the truth of these competing claims.

What is essential, however, is that Meta’s rules on such an important issue are clear, and enforced consistently and rigorously. Given the use of Meta’s platforms by combatants and civilians during conflicts to share information on the use of weapons, or violent content for self-defense, Meta should clarify what the “recreational self-defense” and “military training” exceptions mean. The Board disagrees with Meta that these terms in the public language of the Violence and Incitement Community Standard have “a plain meaning.” To improve clarity, Meta should clarify which actors can benefit from “recreational self-defense” and in which settings this exception applies. Additionally, the public language of the policy on instructions to make or use weapons or explosives fails to expressly state that self-defense contexts are not considered during armed conflicts.



This case also highlights another unclear exception to the Violence and Incitement Community Standard, which allows threats directed at terrorists and other violent actors. This is insufficiently clear because Meta does not clarify whether this applies to all organizations and individuals it designates under its separate Dangerous Organizations and Individuals policy. This is relevant to this case since the RSF was designated for a relevant period in 2023. However, it is impossible for users to know whether their post could be removed or not on this basis since the list of designated organizations and individuals is not available publicly. The Board has already raised concerns about such lack of clarity in our [Haitian Police Station Video](#) decision.

The Board is also concerned that Meta’s notification system does not allow the company to rectify its own mistakes when it does not correctly communicate which Community Standard a user has violated. Being able to correctly inform users of their violation is crucial, guaranteeing fairness. Incorrect notifications undermine the user’s ability to appeal and access remedy. In this case, the user was informed in error that their post was removed for hate speech, even though it had been taken down for violating the Violence and Incitement Community Standard. Therefore, the Board encourages Meta to explore technically feasible ways in which it can make corrections to user notifications.

The Oversight Board’s Decision

The Oversight Board has upheld Meta’s decision to remove the post.

The Board recommends that Meta:

- Amend its Violence and Incitement policy to include a definition of “recreational self-defense” and “military training” as exceptions to its rules prohibiting users from providing instructions on making or using weapons, and clarify that it does not allow any self-defense exception for such instructions in an armed conflict.
- Develop tools to rectify its own mistakes when sending users messages that notify them about the Community Standard they violated, so that users can correctly understand which policies their content violated.



* Case summaries provide an overview of cases and do not have precedential value.

Full Case Decision

1. Decision Summary

The Oversight Board upholds Meta’s decision to remove a post containing a graphic of a gun cartridge, with notes in Arabic identifying its different components. The post was accompanied by a caption in Arabic providing instructions on how to empty a shotgun shell of its pellets, and use the components to create a Molotov cocktail. The post also advises people throwing such an incendiary device to use a helmet to avoid injury. The caption ends with the call, “Victory for the Sudanese people / Victory for the Sudanese Armed Forces / Step forward O, my country.” A hostile speech classifier detected the content and found it violated Facebook’s Violence and Incitement Community Standard.

The Board finds that the post did violate the Violence and Incitement Community Standard, which prohibits providing instructions on how to make or use weapons where there is language explicitly stating the goal to seriously injure or kill people. The Board also concludes that the post violates another policy line under the Violence and Incitement Community Standard, which prohibits instructions on how to make or use explosives unless there is context that the content is for a non-violent purpose. The Board finds that the content poses an imminent risk of harm that could exacerbate ongoing violence in Sudan.

The case raises broader concerns about instructions for weapons that may be shared during an armed conflict and Meta’s human-rights responsibilities in this context. Implementing those responsibilities requires Meta to ensure greater coherence of the rules by clearly defining exceptions to the policy lines on making or using weapons or explosives under the Violence and Incitement Community Standard. Moreover, Meta should develop tools to enable the company to correct mistakes when informing users about which Community Standard they violated.

2. Case Description and Background

In June 2023, a Facebook user posted an illustration of a gun cartridge. The different components of the cartridge are identified in Arabic. The caption for the post, also in Arabic, provides instructions on how to empty a shotgun shell of its pellets and use the components



to create a Molotov cocktail – an incendiary device, typically in a bottle, which is easy to make. The caption also advises using a helmet when throwing the device to protect the person who is throwing the incendiary, and concludes, “Victory for the Sudanese people,” “Victory for the Sudanese Armed Forces,” and “Step forward O, my country.” Linguistic experts the Board consulted said that these phrases did not, in isolation, call for civilians to engage in violence. The content had only a few views before Meta removed it without human review, seven minutes after it was posted. At the time the content was posted in June 2023, the Sudanese Armed Forces (SAF) and the paramilitary group, the Rapid Support Forces (RSF), had been engaged in an armed conflict since mid-April, which continues to the present day. The RSF was designated as a dangerous organization under Meta’s [Dangerous Organizations and Individuals policy](#) on August 11, 2023, months after the conflict escalated.

A hostile speech classifier, an algorithm Meta uses to identify potential violations to the [Hate Speech, Violence and Incitement](#) and [Bullying and Harassment](#) Community Standards, detected the content and determined that it violated the Violence and Incitement Community Standard. Meta removed the content and applied a standard strike to the content creator’s profile, which prevented them from interacting with groups and from creating or joining any messenger rooms for three days. The user immediately appealed Meta’s decision. This led to a human reviewer finding that the post violated the [Restricted Goods and Services policy](#). The user then appealed to the Board. After the Board brought the case to Meta’s attention, the company determined that its original decision to remove the content under the Violence and Incitement Community Standard was correct, and that the post did not violate the Restricted Goods and Services policy.

The Board has considered the following context in reaching its decision on this case.

In April 2023, fighting broke out in Sudan’s capital between the SAF and the RSF. The user who posted the content in this case appears to support the SAF. While fighting initially centered on Khartoum, the capital of Sudan, the conflict then spread across the country including to [Darfur and Kordofan](#). Both groups [have used](#) explosive weapons, including aerial bombs, artillery and mortar projectiles, and rockets and missiles, in areas densely populated by civilians. [According to the United Nations](#), as of January 2024, more than 7 million people have been displaced since mid-April and more than 1.2 million people have fled Sudan. Up to 9,000 people have [reportedly been killed](#). Some of the [attacks](#) against civilians have been [ethnically motivated](#). In October 2023, the UN Special Rapporteur on trafficking in persons [expressed concern](#) over the increased risk of recruitment and use of child soldiers by armed



forces and groups. As fighting continues throughout the country, experts have also [noted](#) the growing [involvement](#) of other armed groups.

According to experts the Board consulted about the conflict in Sudan, both the SAF and the RSF rely on social media to “disseminate information and propaganda” about their respective agendas. While internet penetration [remains low](#) in Sudan, [news organizations](#) and [civil society groups](#) reported that both the SAF and the RSF use social media in an attempt to control narratives surrounding the conflict. For instance, both parties have posted proclamations of victory in areas where fighting is ongoing, thereby putting returning civilians who relied on inaccurate information at risk.

3. Oversight Board Authority and Scope

The Board has authority to review Meta’s decision following an appeal from the person whose content was removed (Charter Article 2, Section 1; Bylaws Article 3, Section 1).

The Board may uphold or overturn Meta’s decision (Charter Article 3, Section 5), and this decision is binding on the company (Charter Article 4). Meta must also assess the feasibility of applying its decision in respect of identical content with parallel context (Charter Article 4). The Board’s decisions may include non-binding recommendations that Meta must respond to (Charter Article 3, Section 4; Article 4). When Meta commits to act on recommendations, the Board monitors their implementation.

4. Sources of Authority and Guidance

The following standards and precedents informed the Board’s analysis in this case:

I. Oversight Board Decisions

The most relevant previous decisions of the Oversight Board include:

- [Haitian Police Station Video](#)
- [Armenian Prisoners of War Video](#)
- [Brazilian General’s Speech](#)
- [Russian Poem](#)
- [Tigray Communication Affairs Bureau](#)
- [Sudan Graphic Video](#)



- [Alleged Crimes in Raya Kobo](#)
- [Ayahuasca Brew](#)
- [Nazi Quote](#)
- [Armenians in Azerbaijan](#)

II. *Meta's Content Policies*

Meta's [Violence and Incitement Community Standard](#) aims to “prevent potential offline violence that may be related to content on our platforms.” Meta states it removes “language that incites or facilitates serious violence” and “threats to public or personal safety.” Meta distinguishes between casual statements, allowed under the policy, and those that pose a “genuine risk of physical harm or direct threats to public safety.”

The Violence and Incitement Community Standard prohibits “threats that could lead to death (or other forms of high-severity violence).” It also states that Meta allows threats “directed against certain violent actors, like terrorist groups.” Meta updated the policy on December 6, 2023 to say that Meta does not prohibit threats when shared to raise awareness, in line with the Board’s recommendation in the [Russian Poem case](#). In exchanges with the Board, Meta clarified that “calls for violence, aspirational threats and conditional threats of high or mid severity violence are all allowed if the target is a designated DOI [Dangerous Organization or Individual] entity. Statements of intent, however, always violate the policy.”

The Violence and Incitement Community Standard also has two rules related to instructions on making and using weapons.

The first rule prohibits content providing “instructions on how to make or use weapons where there is language explicitly stating the goal to seriously injure or kill people” or “imagery that shows or simulates the end result.” Such content is allowed only when shared in a context of “recreational self-defense, training by a country’s military, commercial video games, or news coverage (posted by a Page or with a news logo).”

The second rule prohibits content providing instructions on how to make or use explosives, “unless with context that the content is for a non-violent purpose.” Examples of a non-violent purpose include “commercial video games, clear scientific/educational purpose, fireworks or specifically for fishing.”



The Board's analysis was informed by Meta's [commitment to voice](#), which the company describes as "paramount," and its value of safety.

III. Meta's Human Rights Responsibilities

The UN Guiding Principles on Business and Human Rights (UNGPs), endorsed by the UN Human Rights Council in 2011, establish a voluntary framework for the human rights responsibilities of private businesses. In 2021, Meta [announced](#) its [Corporate Human Rights Policy](#), in which it reaffirmed its commitment to respecting human rights in accordance with the UNGPs. As per the UNGPs, the human rights responsibilities of businesses operating in a conflict setting are heightened ("Business, human rights and conflict-affected regions: towards heightened action," [A/75/212](#)).

The Board's analysis of Meta's human rights responsibilities in this case was informed by the following international standards:

- [The right to freedom of expression](#): Article 19, International Covenant on Civil and Political Rights (ICCPR), General Comment No. 34, Human Rights Committee, 2011; UN Special Rapporteur on freedom of opinion and expression, reports: [A/HRC/38/35](#) (2018) and [A/74/486](#) (2019).
- [The right to life](#): Article 6, ICCPR.
- [The right to effective remedy](#): Article 2, ICCPR; [General Comment No. 31](#), Human Rights Committee, (2004).
- [Protection of victims of non-international armed conflicts](#): [Additional Protocol II, Geneva Conventions](#); [Rule 70, Rules of Customary International Humanitarian Law](#)
- [Restrictions on the use of incendiary weapons during an armed conflict](#): [Convention on Certain Conventional Weapons](#) (1980); [Protocol on Prohibitions or Restrictions on the Use of Incendiary Weapons \(Protocol III\)](#)

5. User Submissions

In their appeal to the Board, the author of the content stated that Meta misunderstood the post and that they were only sharing facts.

6. Meta's Submissions

Meta claims that the post violated two policy lines in the Violence and Incitement Community Standard. First, it violated the policy line that prohibits content providing instructions on how



to make or use weapons if there is evidence of a goal to seriously injure or kill people. Meta said the post was violating because its “caption suggests that the intended use is for throwing the Molotov cocktail against a target..” Moreover, the instructions are shared “in the context of armed conflict and apparently in support of the SAF.” Meta found the post did not fall under any of the exceptions under the policy line, such as for “recreational self-defense,” “military training” or “news coverage.” When asked by the Board about the meaning of “recreational self-defense” and “military training,” Meta stated it does not have a definition of these terms “beyond the plain meaning of those words.”

The content also violated the policy line prohibiting instructions on how to make or use explosives, without clear context that the content is for a non-violent purpose. Meta considers Molotov cocktails to be explosives under the meaning of the policy. Moreover, Meta assessed that the content “was shared with the intention of furthering the violent conflict.”

According to the internal guidelines on how to apply the Violence and Incitement policy in place when the content was posted, Meta allows content that violates the policy “when shared in awareness-raising or condemning context.” Following updates to the public-facing Community Standards on December 6, 2023, the policy now reflects this guidance: “[Meta] do[es] not prohibit threats when shared in awareness-raising or condemning context.” However, Meta says the caption “makes clear that the intention is not to raise awareness but to enable violent action.” When asked whether Meta has exempted any countries or conflicts from the application of the Violence and Incitement policy lines prohibiting instructions on making or using weapons or explosives, Meta told the Board it has not applied any country-specific policy exceptions or allowances, “regardless of active conflicts.”

Based on Meta’s updated Violence and Incitement Community Standard, Meta does not prohibit threats directed against “certain violent actors, like terrorist groups.” This means that some threats targeting designated dangerous organization or individual entities, such as the RSF, are allowed on Meta’s platforms. Meta explained, however, that this exception does not apply to the two policy lines under the Violence and Incitement Community Standard on instructions on how to make or use weapons or explosives. In other words, content explaining how to create or use weapons is prohibited even if it targets a designated dangerous organization or individual entity.

Meta did not set up an Integrity Product Operations Center, which is used to respond to threats in real-time, to address the outbreak of violence in Sudan in April 2023. According to Meta, it was able to “handle the identified content risks through the current processes.” The company’s



efforts to respond to the current conflict continue and build on work first described in the Board’s [Sudan Graphic Video](#) case. In response to the military coup in Sudan in October 2021, Meta created “a crisis response cross-functional team to monitor the situation and communicate emerging trends and risks,” which is ongoing. Additionally, Meta took the following steps, among others, to address potential content risks related to the 2023 Sudan conflict: removed pages and accounts representing the RSF, following Meta’s designation of the group as a dangerous organization; investigated potential fake accounts that could mislead public discourse surrounding the conflict; and designated Sudan as a Temporary High-Risk Location (for a description of the THRL designation and its relationship to the Violence and Incitement Community Standard, see [Brazilian General’s Speech](#) decision, Section 8.1). Meta informed the Board that it is working to establish longer-term crisis coordination “to provide dedicated operations oversight throughout the lifecycle of imminent and emerging crises,” following on from the Board’s recommendation in the [Tigray Communication Affairs Bureau](#) case.

As of May 30, 2023, Sudan reached Meta’s highest internal crisis designation. Since then, Meta has been maintaining a heightened risk management level and are monitoring the situation for content risks as part of that work. The Crisis Policy Protocol is the framework Meta adopted for developing time-bound policy-specific responses to an emerging crisis. There are three crisis categories under the Crisis Policy Protocol – Category 1 being the least severe and Category 3 being the most severe. The Category 3 crisis designation in Sudan was a result of the escalating crisis meeting additional entry criteria, such as the existence of a “major internal conflict” and “military intervention.”

The Board asked Meta sixteen questions in writing. Questions related to Meta’s hostile speech classifier; how Meta understands the concept of self-defense in relation to the Violence and Incitement Community Standard; measures taken in response to the conflict in Sudan; and the enforcement of the weapons-related policy lines of the Violence and Incitement Community Standard in armed conflicts. Meta answered all questions.

7. Public Comments

The Oversight Board received 10 public comments relevant to this case. Three of the comments were submitted from the United States and Canada, two from Asia Pacific and Oceania, two from Europe, one from Latin American and Caribbean, one from Middle East and North Africa and one from Sub-Saharan Africa. This total includes public comments that were either duplicates, were submitted without consent to publish or were submitted with



consent to publish but did not meet the Board’s conditions for publication. Public comments can be submitted to the Board with or without consent to publish, and with or without attribution. The submissions covered the following themes: conflict dynamics in Sudan; Meta’s human rights responsibilities in situations of armed conflict, particularly in the preservation of online content for human rights accountability; and the impact of Meta’s classifier design on the moderation of conflict-related content.

To read public comments submitted for this case, please click [here](#).

8. Oversight Board Analysis

The Board selected this case to assess Meta’s policies on weapons-related content and the company’s enforcement practices in the context of armed conflicts. The case falls within the Board’s Crisis and Conflict Situations strategic priority.

8.1 Compliance With Meta’s Content Policies

I. Content Rules

The Board finds that the post violates Meta’s Violence and Incitement policy. The combined effect of the image and caption in the post meets the requirements of “language explicitly stating the goal” in the line that prohibits “instructions on how to make or use weapons where there is language explicitly stating the goal to seriously injure or kill people.” Meta considers Molotov cocktails as weapons prohibited under the Violence and Incitement policy. The post provides a step-by-step guide on how to build and use a Molotov cocktail. Intent to seriously injure or kill people can be inferred from this step-by-step guide as well as the advice to “use a helmet” to protect the person who throws the incendiary, which means the post is calling on people to act on the instructions. According to experts consulted by the Board, the calls for victory at the end of the caption clearly articulate support for one of the sides of the armed conflict.

The content further violates the Violence and Incitement prohibition on “instructions on how to make or use explosives, unless with context that the content is for a non-violent purpose.” Resorting to violence in support of the SAF does not relate to a non-violent purpose; such purposes, as outlined in the policy, are limited to “commercial video games, clear scientific/educational purpose, fireworks or specifically for fishing.” The Board notes that,



according to Meta, this prohibition applies whether or not Meta has designated the entity targeted by the content as a dangerous organization or individual under the [Dangerous Organizations and Individuals Community Standard](#). Meta explained to the Board this is because “part of the harm of sharing these instructions is that they can be used by other people intending to harm other targets.” The Board finds that this rule was applied in accordance with Meta’s content policies when it came to removal of the content in this case.

II. Enforcement Action

Although the hostile speech classifier correctly identified the content as a violation of the Violence and Incitement Community Standard, the user was informed in error that their post was removed for hate speech. According to Meta, this was due to a bug in the company’s systems. Meta informed the Board that it is unable to send new messages to the same support inbox thread when it realizes a mistake was made. The Board is concerned that Meta’s user-notification system does not allow the company to rectify its own mistakes when it does not correctly communicate to the user which Community Standard they violated. This prevents users from learning about the actual reason their content was removed. As the Board previously highlighted in several cases (e.g., [Armenians in Azerbaijan](#), [Ayahuasca Brew](#), [Nazi Quote](#)), Meta should transparently inform users about the content policies they violated.

8.2 Compliance With Meta’s Human Rights Responsibilities

The Board finds that Meta’s decision to remove the post was consistent with the company’s human rights responsibilities.

Freedom of Expression (Article 19 ICCPR)

Article 19 of the ICCPR provides for broad protection of expression, including political expression. This right includes the “freedom to seek, receive and impart information and ideas of all kinds.” These protections remain active during armed conflicts, and should continue to inform Meta’s human rights responsibilities, alongside the mutually reinforcing and [complementary rules](#) of international humanitarian law that apply during such conflicts ([General Comment 31](#), Human Rights Committee, 2004, para. 11; Commentary to UNGPs, [Principle 12](#); see also UN Special Rapporteur’s report on Disinformation and freedom of opinion and expression during armed conflicts,



Report [A/77/288](#), paras. 33-35 (2022); and OHCHR [report](#) on International legal protection of human rights in armed conflict (2011) at page 59).

The UN Special Rapporteur on freedom of expression has stated that “[d]uring armed conflict, people are at their most vulnerable and in the greatest need of accurate, trustworthy information to ensure their own safety and well-being. Yet, it is precisely in those situations that their freedom of opinion and expression, which includes ‘the freedom to seek, receive and impart information and ideas of all kinds,’ is most constrained by the circumstances of war and the actions of the parties to the conflict and other actors to manipulate and restrict information for political, military and strategic objectives,” (Report A/77/288, para. 1). The Board recognizes the importance of ensuring that people can freely share information about conflicts, especially when social media is the ultimate source of information, while simultaneously ensuring content that is likely to fuel further offline violence does not go viral.

When restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimate aim, and necessity and proportionality (Article 19, para. 3, ICCPR). These requirements are often referred to as the “three-part test.” The Board uses this framework to interpret Meta’s voluntary human rights commitments, both in relation to the individual content decision under review and what this says about Meta’s broader approach to content governance. As in previous cases (e.g., [Armenians in Azerbaijan](#), [Armenian Prisoners of War Video](#)), the Board agrees with the UN Special Rapporteur on freedom of expression that, although “companies do not have the obligations of Governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users’ right to freedom of expression,” ([A/74/486](#), para. 41). In doing so, the Board attempts to be sensitive to ways in which the human rights responsibilities of a private social media company may differ from a government implementing its human rights obligations.

I. Legality (Clarity and Accessibility of the Rules)

The principle of legality requires rules limiting expression to be accessible and clear, both to those enforcing the rules and those impacted by them (General Comment No. 34, para. 25). Users should be able to predict the consequences of posting content on Facebook and Instagram. The UN Special Rapporteur on freedom of expression has highlighted the need for “clarity and specificity” in content-moderation policies ([A/HRC/38/35](#), para. 46).



The Board finds that the general rule prohibiting instructions on making or using weapons or explosives under certain circumstances is sufficiently clear, meeting the requirements of legality. The Board also notes, however, that Meta could further improve clarity around the policy's exceptions by explaining concepts such as "recreational self-defense" and "training by a country's military" in the public-facing language of the Violence and Incitement Community Standard. The Board disagrees with Meta's claim that these terms have a "plain meaning."

With respect to the term "recreational self-defense," the Board believes Meta should clarify the actors that can benefit from it, and in which settings the exceptions apply. Moreover, it is not expressly stated in the public-facing Violence and Incitement Community Standard that the term does not contemplate self-defense contexts in armed conflict settings. With respect to the term "training by a country's military," Meta does not clarify whether it is limited to militaries of recognized states nor how the company treats armies of *de facto* governments.

Stakeholders consulted by the Board as well as [public reporting](#) have claimed that Meta allows instructions on making or using weapons in exercise of self-defense for certain armed conflicts. In response to the Board's questions, Meta denied that these reports are true. The Board is not in a position to determine the truth of these conflicting claims. In any event, it is essential that Meta's rules on as important an issue as this be enforced consistently and rigorously. Given the use of Meta's platforms to exchange information during armed conflicts when both combatants and civilians may be sharing information on the use of weapons, or violent content invoking self-defense, Meta should clarify what the "recreational self-defense" and "military training" exceptions mean in the Violence and Incitement Community Standard.

Additionally, the Board finds Meta's policy exception to the Violence and Incitement Community Standard, which allows threats "directed against certain violent actors, like terrorist groups," insufficiently clear, thus failing to meet the legality requirement. It does not clarify whether this policy line applies to all dangerous individuals and organizations designated under the Dangerous Organizations and Individuals Community Standard. Moreover, the list of designated organizations and individuals under the Dangerous Organizations and Individuals policy is not public. This exacerbates the lack of clarity to users on which posts will be removed or kept up depending on whether or not the entity referred to in their post is included in Meta's hidden list of dangerous organizations. The Board repeats the concerns raised in the [Haitian Police Station Video](#) decision on this policy exception.



II. *Legitimate Aim*

Restrictions on freedom of expression (Article 19, ICCPR) must pursue a legitimate aim. The Violence and Incitement policy aims to “prevent potential offline harm” by removing content that poses “a genuine risk of physical harm or direct threats to public safety.” As previously concluded by the Board in the [Alleged Crimes in Raya Kobo case](#), this policy serves the legitimate aim of protecting the rights of others, such as the right to life (Article 6, ICCPR).

III. *Necessity and Proportionality*

The principles of necessity and proportionality provide that restrictions on freedom of expression “must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; [and] they must be proportionate to the interest to be protected,” ([General Comment No. 34, para. 34](#)). The Board also considered the Rabat Plan of Action factors on what constitutes incitement to violence ([Rabat Plan of Action, OHCHR, A/HRC/22/17/Add.4,2013](#)), while considering the differences between the legal obligations of states and the human rights responsibilities of businesses. Although the post in this case does not advocate hatred on the basis of nationality, race or religion, the Rabat Plan of Action nonetheless offers a useful framework in assessing whether or not the content incites others to violence.

In this case, the Board finds that Meta removing the content from Facebook complied with the requirements of necessity and proportionality. Using the Rabat Plan of Action’s six-part test to inform its analysis, the Board finds support for the removal of this post, as explained below.

Regardless of the intent of the content creator when posting, the step-by-step guide to making and using a Molotov cocktail created a genuine risk of imminent harm in an already volatile security situation. The incendiary weapon referred to in the post is prohibited under the [Convention on Certain Conventional Weapons](#) for being both excessively injurious and indiscriminate as a means of attack. The impact of an explosion not only poses a high risk of wounding civilians, it can also lead to “unnecessary suffering” or “superfluous injury” on combatants prohibited by [customary international humanitarian law](#). Encouraging civilians with no military training to deploy and use incendiary weapons further increases these risks.

The genuine risk of imminent harm exists despite the user being a private individual who is not influential, with a limited number of friends and followers. The Board notes that Meta’s



hostile speech classifier was able to detect and remove the content within minutes of the content being posted.

The Board also notes that the content was posted in the context of an ongoing armed conflict. At the time it was posted, two months into the armed conflict, [reporting](#) and [expert analysis](#) showed widespread human rights abuses committed by both the SAF and RSF. According to the UN, human rights groups, experts consulted by the Board and public comment submissions, including from Genocide Watch (PC-19006, PC-19001), both parties to the armed conflict have engaged in various abuses of international humanitarian and human rights laws, leading to millions of people being displaced, arbitrarily arrested, sexually violated or killed. The conflict is still ongoing and shows no signs of ending despite condemnation by the [UN Security Council](#), civil society groups and human rights organizations. In this context, the Board finds that the content in this case incited violence, posing an imminent risk of civilians directly taking part in hostilities by using a particularly pernicious and outlawed weapon and further escalating the conflict.

Meta does not allow content like the post under review in the context of self-defense when enforcing the Violence and Incitement policy. The Board finds this to be a sensible approach and urges Meta to enforce it consistently. Under the UNGPs, Meta’s human rights responsibilities include respecting “the standards of international humanitarian law in an armed conflict,” (Commentary to Principle 12, UNGPs). International humanitarian law provides standards for parties engaging in armed conflicts to maximize civilian protection (e.g., Additional Protocol II of the Geneva Conventions protecting civilians during armed conflict; Article 2, Protocol III of the Convention on Certain Conventional Weapons prohibiting the use of incendiary weapons). The Board believes that these standards can also be helpful for social media companies to achieve that aim when their platforms are used in armed conflict settings. In line with these standards, Meta should aim for a policy that results in the widest protection for civilians and civilian property in conflict settings. When applied to the Violence and Incitement policy, this means prohibiting credible threats regardless of the target.

Access to Remedy

The Board is concerned at Meta’s technical inability to correct mistakes in its notifications to users when it informs them of which rule they violated with their content. Being able to correctly inform users of the violation is a crucial component of enforcing Meta’s Community Standards and guarantees fairness and due process to the user. When a notification provided



to users is incorrectly identified, the user’s ability to appeal and access remedy on Meta’s platforms is undermined. The Board encourages Meta to explore technically feasible ways in which it can make corrections to user notifications.

9. Oversight Board Decision

The Oversight Board upholds Meta’s decision to take down the content.

10. Recommendations

Content Policy

1. To better inform users of what content is prohibited on its platforms, Meta should amend its Violence and Incitement policy to include a definition of “recreational self-defense” and “military training” as exceptions to its rules prohibiting users from providing instructions on making or using weapons, and clarify that it does not allow any self-defense exception for instructions on how to make or use weapons in the context of an armed conflict.

The Board will consider this implemented when the Violence and Incitement Community Standard is updated to reflect these changes.

Enforcement

2. To make sure users are able to understand which policies their content was enforced against, Meta should develop tools to rectify mistakes in its user messaging notifying the user about the Community Standard they violated.

The Board will consider this implemented when the related review and notification systems are updated accordingly.

***Procedural Note:**

The Oversight Board’s decisions are prepared by panels of five Members and approved by the majority of the Board. Board decisions do not necessarily represent the personal views of all Members.



For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by an independent research institute headquartered at the University of Gothenburg, which draws on a team of more than 50 social scientists on six continents, as well as more than 3,200 country experts from around the world. The Board was also assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology.* Memetica, an organization that engages in open-source research on social media trends, also provided analysis. Linguistic expertise was provided by Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world.