

OVERTURNED

2021-007-FB-UA

Myanmar bot

The Oversight Board has overturned Facebook's decision to remove a post in Burmese under its Hate Speech Community Standard.

Policies and topics

-  Freedom of expression, Politics
-  Hate speech

Region and countries

-  Central and South Asia
-  Myanmar

Platform

-  Facebook

Attachments

[2021-007-FB-UA Public Comments](#)

Case summary

The Oversight Board has overturned Facebook's decision to remove a post in Burmese under its Hate Speech Community Standard. The Board found that the post did not target Chinese people, but the Chinese state. Specifically, it used profanity to reference Chinese governmental policy in Hong Kong as part of a political discussion on the Chinese government's role in Myanmar.

About the case

In April 2021, a Facebook user who appeared to be in Myanmar posted in Burmese on their timeline. The post discussed ways to limit financing to the Myanmar military following the coup in Myanmar on February 1, 2021. It proposed that tax revenue be given to the Committee Representing Pyidaungsu Hlutaw (CRPH), a group of legislators opposed to the coup. The post received about half a million views and no Facebook users reported it.

Facebook translated the supposedly violating part of the user's post as "Hong Kong people, because the fucking Chinese tortured them, changed their banking to UK, and now (the Chinese) they cannot touch them." Facebook removed the post under its Hate Speech Community

Standard. This prohibits content targeting a person or group of people based on their race, ethnicity or national origin with “profane terms or phrases with the intent to insult.”

The four content reviewers who examined the post all agreed that it violated Facebook’s rules. In their appeal to the Board, the user stated that they posted the content to “stop the brutal military regime.”

Key findings

This case highlights the importance of considering context when enforcing hate speech policies, as well as the importance of protecting political speech. This is particularly relevant in Myanmar given the February 2021 coup and Facebook’s key role as a communications medium in the country.

The post used the Burmese phrase “တရုတ်,” which Facebook translated as “fucking Chinese” (or “sout ta-yote”). According to Facebook, the word “ta-yote” “is perceived culturally and linguistically as an overlap of identities/meanings between China the country and the Chinese people.” Facebook stated that given the nature of this word and the fact that the user did not “clearly indicate that the term refers to the country/government of China,” it determined that “the user is, at a minimum, referring to Chinese people.” As such, Facebook removed the post under its Hate Speech Community Standard.

As the same word is used in Burmese to refer to a state and people from that state, context is key to understanding the intended meaning. A number of factors convinced the Board that the user was not targeting Chinese people, but the Chinese state.

The part of the post which supposedly violated Facebook’s rules refers to China’s financial policies in Hong Kong as “torture” or “persecution,” and not the actions of individuals or Chinese people in Myanmar. Both of the Board’s translators indicated that, in this case, the word “ta-yote” referred to a state. When questioned on whether there could be any possible ambiguity in this reference, the translators did not indicate any doubt. The Board’s translators also stated that the post contains terms commonly used by Myanmar’s government and the Chinese embassy to address each other. In addition, while half a million people viewed the post and over 6,000 people shared it, no users reported it. Public comments also described the overall tone of the post as a political discussion.

Given that the post did not target people based on race, ethnicity, or national origin, but was aimed at a state, the Board found it did not violate Facebook’s Hate Speech Community Standard.

The Oversight Board’s decision

The Oversight Board overturns Facebook’s decision to remove the content, requiring the post to be restored.

In a policy advisory statement, the Board recommends that Facebook:

- Ensure that its Internal Implementation Standards are available in the language in which content moderators review content. If necessary to prioritize, Facebook should focus first on contexts where the risks to human rights are more severe.

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

Full case decision

1. Decision summary

The Oversight Board has overturned Facebook’s decision to remove content under its Hate Speech Community Standard. The Board found that the post was not hate speech.

2. Case description

In April 2021, a Facebook user who appeared to be in Myanmar posted in Burmese on their timeline. The post discussed ways to limit financing to the Myanmar military following the coup in Myanmar on February 1, 2021. It proposed that tax revenue be given to the Committee

Representing Pyidaungsu Hlutaw (CRPH), a group of legislators opposed to the coup. The post received about 500,000 views, about 6,000 reactions and was shared about 6,000 times. No Facebook users reported the post.

Facebook translated the supposedly violating part of the user's post as "Hong Kong people, because the fucking Chinese tortured them, changed their banking to UK, and now (the Chinese) they cannot touch them." Facebook removed the post as "Tier 2" Hate Speech under its [Hate Speech Community Standard](#) the day after it was posted. This prohibits content targeting a person or group of people based on their race, ethnicity or national origin with "profane terms or phrases with the intent to insult."

A reshare of the post was, according to Facebook, "automatically selected as a part of a sample and sent to a human reviewer to be used for classifier training." This involves Facebook creating data sets of examples of violating and non-violating content to train its automated detection and enforcement processes to predict whether content violates Facebook policies. The reviewer determined that the shared post violated the Hate Speech Community Standard. While the purpose of the process was to create sets of content to train the classifier, once the shared post was found to be violating it was deleted.

Because the shared post was found to have violated Facebook's rules, an "Administrative Action Bot" automatically identified the original post for review. Facebook explained that the Administrative Action Bot is an internal Facebook account that does not make any assessment of content but carries out "a variety of actions throughout the enforcement system based on decisions made by humans or automation." Two human reviewers then analyzed the original post, and both determined it was "Tier 2" Hate Speech. The content was removed. The user appealed the removal to Facebook, where a fourth human reviewer upheld the removal. According to Facebook, "[t]he content reviewers in this case were all members of a Burmese content review team at Facebook." The user then submitted their appeal to the Oversight Board.

3. Authority and scope

The Board has authority to review Facebook's decision following an appeal from the user whose post was removed (Charter Article 2, Section 1; Bylaws Article 2, Section 2.1). The Board may uphold or reverse that decision (Charter Article 3, Section 5). The Board's decisions are binding and may include policy advisory statements with recommendations. These recommendations are non-binding, but Facebook must respond to them (Charter Article 3, Section 4). The Board is an independent grievance mechanism to address disputes in a transparent and principled manner.

4. Relevant standards

The Oversight Board considered the following standards in its decision:

I. Facebook's Community Standards

Facebook's Community Standards define hate speech as "a direct attack on people based on what we call protected characteristics – race, ethnicity, national origin, religious affiliation, sexual orientation, caste, sex, gender, gender identity, and serious disease or disability." Under "Tier 2," prohibited content includes cursing, defined as "[p]rofan terms or phrases with the intent to insult, including, but not limited to: fuck, bitch, motherfucker."

II. Facebook's values

Facebook's values are outlined in the introduction to the Community Standards. The value of "Voice" is described as "paramount":

The goal of our Community Standards has always been to create a place for expression and give people a voice. [...] We want people to be able to talk openly about the issues that matter to them, even if some may disagree or find them objectionable.

Facebook limits "Voice" in service of four values, and two are relevant here:

"Safety": We are committed to making Facebook a safe place. Expression that threatens people has the potential to intimidate, exclude or silence others and isn't allowed on Facebook.

"Dignity": We believe that all people are equal in dignity and rights. We expect that people will respect the dignity of others and not harass or degrade others.

III. Human rights standards

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, [Facebook announced its Corporate Human Rights Policy](#), where it commemorated its commitment to respecting human rights in accordance with the UNGPs. The Board's analysis in this case was informed by the following human rights standards:

1. [Freedom of expression](#): Article 19, International Covenant on Civil and Political Rights ([ICCPR](#)), [General Comment No. 34](#), Human Rights Committee, 2011
2. [Responsibilities of businesses](#): Business, human rights and conflict-affected regions: towards heightened action report ([A/75/212](#)), UN Working Group on the issue of human rights and transnational corporations and other business enterprises

5. User statement

The user stated in their appeal to the Board that they posted this content to “stop the brutal military regime” and provide advice to democratic leaders in Myanmar. The user also reiterated the need to limit the Myanmar military regime’s funding. The user self-identified as an “activist” and speculated that the Myanmar military regime’s informants reported their post. The user also stated that “someone who understands Myanmar Language” should review their post.

6. Explanation of Facebook’s decision

Facebook removed the content as a “Tier 2” attack under the Hate Speech Community Standard, specifically for violating its policy prohibiting profane curse words targeted at people based on their race, ethnicity and/or national origin. According to Facebook, the allegedly violating content was considered to be an attack on Chinese people.

The content included the phrase in Burmese “\$တရုတ်,” which Facebook’s regional team translated as “fucking Chinese” (or “sout ta-yote”). Facebook’s regional team further specified that “\$” can be used as an abbreviation for “ဆော့” or “sout,” which translates to “fucking.” According to Facebook’s team, the word “ta-yote” “is perceived culturally and linguistically as an overlap of identities/meanings between China the country and the Chinese people.” Facebook provided the Board with the relevant confidential internal guidance it provides to its moderators, or Internal Implementation Standards, on distinguishing language that targets people based on protected characteristics and concepts related to protected characteristics.

Facebook also noted in its decision rationale that following the February 2021 coup “there were reports of increasing anti-Chinese sentiment” in Myanmar and that “several Chinese people were injured, trapped, or killed in an alleged arson attack on a Chinese-financed garment factory in Yangon, Myanmar.” In response to a question from the Board, Facebook stated that it did not have any contact with the Myanmar military regime about this post.

Facebook stated that given the nature of the word “ta-yote” and the fact that the user did not “clearly indicate that the term refers to the country/government of China,” Facebook determined that “the user is, at a minimum, referring to Chinese people.” As such, Facebook stated that the removal of the post was consistent with its Hate Speech Community Standard.

Facebook also stated that its removal was consistent with its values of “Dignity” and “Safety,” when balanced against the value of “Voice.” According to Facebook, profane cursing directed at Chinese people “may result in harm to those people” and is “demeaning, dehumanizing, and belittling of their individual dignity.”

Facebook argued that its decision was consistent with international human rights standards. Facebook stated that its decision complied with the international human rights law requirements of legality, legitimate aim, and necessity and proportionality. According to Facebook, its policy was “easily accessible” in the Community Standards and “the user’s choice of words fell squarely within the prohibition on profane terms.” Additionally, the decision to remove the content was legitimate to protect “the rights of others from harm and discrimination.” Finally, its decision to remove the content was “necessary and proportionate” as “the accumulation of content containing profanity directed against Chinese people ‘creates an environment where acts of violence are more likely to be tolerated and reproduce discrimination in a society,’” citing the Board’s decision [2021-002-FB-UA](#) related to Zwarte Piet. Facebook stated it was similar because “both cases involve hate speech directed at people on the basis of their race or ethnicity.”

7. Third-party submissions

The Oversight Board received 10 public comments related to this case. Five of the comments were from Asia Pacific and Oceania, specifically Myanmar, and five were from the United States and Canada. The Board received comments from stakeholders including human rights defenders and civil society organizations focusing on freedom of expression and hate speech in Myanmar.

The submissions covered themes including translation and analysis of the word “sout ta-yote;” whether the content was an attack on China or Chinese people; whether the post was political speech that should be protected in context of the conflict in Myanmar; whether there was an increase in anti-Chinese sentiment in Myanmar following the February 2021 coup; the relations between China and Myanmar’s military regime; and Facebook’s content moderation practices, particularly the use, training and audit of Facebook’s automation tools for Burmese language content.

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

8. Oversight Board analysis

This case highlights the importance of context when enforcing content policies designed to protect users from hate speech, while also respecting political speech. This is particularly relevant in Myanmar due to the February 2021 coup and Facebook’s importance as a medium for communication. The Board looked at the question of whether this content should be restored through three lenses: Facebook’s Community Standards; the company’s values; and its human rights responsibilities.

8.1 Compliance with Community Standards

The Board found that restoring this content is consistent with Facebook’s Community Standard on Hate Speech. Facebook’s policy prohibits “profane terms with the intent to insult” that targets a person or people based on race, ethnicity, or national origin. The Board concludes that the post did not target people, but rather was aimed at Chinese governmental policy in Hong Kong, made in the context of discussing the Chinese government’s role in Myanmar.

In addition to public comments, the Board also sought two translations of the text. These included translations from a Burmese speaker located within Myanmar and another Burmese speaker located outside of Myanmar. Public comments and the Board’s translators noted that in Burmese, the same word is used to refer to states and people from that state. Therefore, context is key to understanding the intended meaning. This is particularly relevant for applying Facebook’s Hate Speech policy. At the time the content was removed, the Hate Speech Community Standard stated it prohibits attacks against people based on national origin but does not prohibit attacks against countries.

The Board considered various factors in deciding this post did not target Chinese people based on their ethnicity, race, or national origin. First, the broader post suggests ways to limit financial engagement with the military regime and provide financial support for the CRPH. Second, the supposedly violating part of the post refers to China’s financial policies in Hong Kong as “torture” or “persecution,” and not the actions of individuals or Chinese people in Myanmar. Third, while the absence of reporting of a widely shared post does not always indicate it is not violating, more than 500,000 people viewed, and more than 6,000 people shared the post and no users reported it. Fourth, both translators consulted by the Board indicated that, while the same term is used to refer to both a state and its people, here it referred to the state. When questioned on any possible ambiguity in this reference, the translators did not indicate any doubt. Fifth, both translators stated that the post contains terms commonly used by the Myanmar government and the Chinese embassy to address each other. Lastly, public comments generally noted the overall tenor of the post as largely a political discussion.

Therefore, given that the profanity did not target people based on race, ethnicity, or national origin, but targeted a state, the Board concludes it does not violate Facebook’s Hate Speech Community Standard. It is crucial to ensure that prohibitions on targeting people based on protected characteristics not be construed in a manner that shields governments or institutions from criticism. The Board recognizes that anti-Chinese hate speech is a serious concern, but this post references the Chinese state.

The Board disagrees with Facebook’s argument that its decision to remove this content followed the Board’s rationale in case decision [2021-002-FB-UA](#) (where the Board upheld the removal of depictions of people in blackface). In that case, Facebook had a rule against depictions of people in blackface, and the Board permitted Facebook to apply that rule to content that included blackface depictions of Zwarte Piet. Here, by contrast, the context of the post indicates that the language used did not violate Facebook’s rules at all.

During the Board’s deliberation regarding this case, Facebook updated its Hate Speech Community Standard to provide more information on how it prohibits “concepts” related to protected characteristics in certain circumstances. This new rule states Facebook “require[s] additional

information and/or context” for enforcement and that users should not post “Content attacking concepts, institutions, ideas, practices, or beliefs associated with protected characteristics, which are likely to contribute to imminent physical harm, intimidation or discrimination against the people associated with that protected characteristic.”

As it was not part of the Community Standard when Facebook removed this content, and Facebook did not argue it removed the content under this updated Standard to the Board, the Board did not analyze the application of this policy to this case. However, the Board notes that “concepts, institutions, ideas, practices, or beliefs” could cover a very wide range of expression, including political speech.

8.2 Compliance with Facebook’s values

The Board concludes that restoring this content is consistent with Facebook’s values. Although Facebook’s values of “Dignity” and “Safety” are important, particularly in the context of the February 2021 coup in Myanmar, this content did not pose a risk to these values such that it would justify displacing “Voice.” The Board also found that the post contains political speech that is central to the value of “Voice.”

8.3 Compliance with Facebook’s human rights responsibilities

The Board concludes that restoring the content is consistent with Facebook’s human rights responsibilities as a business. Facebook has committed itself to respect human rights under the UN Guiding Principles on Business and Human Rights (UNGPs). Its [Corporate Human Rights Policy](#) states this includes the International Covenant on Civil and Political Rights (ICCPR).

Article 19 of the ICCPR provides for broad protection of expression. This protection is “particularly high” for political expression and debate, including about public institutions ([General Comment 34](#), para. 38). Article 19 requires state restrictions on expression to satisfy the three-part test of legality, legitimacy, and necessity and proportionality. The Board concludes that Facebook’s actions did not satisfy its responsibilities as a business under this test.

I. Legality (clarity and accessibility of the rules)

The principle of legality under international human rights law requires rules used by states to limit expression to be clear and accessible ([General Comment 34](#), para. 25). Rules restricting expression must also “provide sufficient guidance to those charged with their execution to enable them to ascertain what sorts of expression are properly restricted and what sorts are not” ([General Comment 34](#), para. 25).

The Hate Speech Community Standard prohibits profanity that targets people based on race, ethnicity, or national origin. Facebook told the Board that because of the difficulties in “determining intent at scale, Facebook considers the phrase ‘fucking Chinese’ as referring to both Chinese people and the Chinese country or government, unless the user provides additional context that it refers solely to the country or government.” The policy of defaulting towards removal is not stated in the Community Standard.

The Board concludes that the user provided additional context that the post referred to a state or country, as noted in the Board’s analysis of the Hate Speech Community Standard (Section 8.1 above). Multiple Facebook reviewers reached a different conclusion than the Board’s translators, people who submitted public comments, and presumably many of the more than 500,000 users who viewed the post and did not report it. Given this divergence, the Board questions the adequacy of Facebook’s internal guidance, resources and training provided to content moderators.

Given the Board’s finding that the user did not violate Facebook’s Hate Speech policy, the Board does not decide whether the non-public policy of defaulting to removal violates the principle of legality. However, the Board is concerned that the policy of defaulting to removal when profanity may be interpreted as directed either to a people or to a state is not clear from the Community Standards. In general, Facebook should make public internal guidance that alters the interpretation of its public-facing Community Standards.

II. Legitimate aim

Any state restriction on expression should pursue one of the legitimate aims listed in the ICCPR. These include the “rights of others.” According to Facebook, its Hate Speech policy aims to protect users from discrimination. The Board agrees that this is a legitimate aim.

III. Necessity and proportionality

The principle of necessity and proportionality under international human rights law requires that restrictions on expression "must be appropriate to achieve their protective function; they must be the least intrusive instrument amongst those which might achieve their protective function; they must be proportionate to the interest to be protected" ([General Comment 34](#), para. 34). In this case, based on its interpretation of the content, the Board determined that restricting this post would not achieve a protective function.

The UNGPs state that businesses should perform ongoing human rights due diligence to assess the impacts of their activities (UNGP 17) and acknowledge that the risk of human rights harms is heightened in conflict-affected contexts (UNGP 7). The UN Working Group on the issue of human rights and transnational corporations and other business enterprises noted that businesses' diligence responsibilities should reflect the greater complexity and risk for harm in some scenarios ([A/75/212](#), paras. 41-49). Similarly, in case decision [2021-001-FB-FBR](#) the Board recommended that Facebook "ensure adequate resourcing and expertise to assess risks of harm from influential accounts globally," recognizing that Facebook should devote attention to regions with greater risks.

In this case, the Board found that these heightened responsibilities should not lead to default removal, as the stakes are high in both leaving up harmful content and removing content that poses little or no risk of harm. While Facebook's concern about hate speech in Myanmar is well founded, it also must take particular care to not remove political criticism and expression, in this case supporting democratic governance.

The Board noted that Facebook's policy of presuming profanity mentioning national origin (in this case "ဝေရ်ဝေရ်") refers to states *and* people may lead to disproportionate enforcement in some linguistic contexts, such as this one, where the same word is used for both. The Board also noted that the impact of this removal extended beyond the case, as Facebook indicated it was used in classifier training as an example of content that violated the Hate Speech Community Standard.

Given the above, international human rights standards support restoring the content to Facebook.

9. Oversight Board decision

The Oversight Board overturns Facebook's decision to remove the content and requires the content to be restored. Facebook is obligated under the Board's Charter to apply this decision to parallel contexts, and should mark this content as non-violating if used in classifier training.

10. Policy recommendation

Facebook should ensure that its Internal Implementation Standards are available in the language in which content moderators review content. If necessary to prioritize, Facebook should focus first on contexts where the risks to human rights are more severe.

*Procedural note:

The Oversight Board's decisions are prepared by panels of five Members and approved by a 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. An independent research institute headquartered at the University of Gothenburg and drawing on a team of over 50 social scientists on six continents, as well as more than 3,200 country experts from around the world, provided expertise on socio-political and cultural context. The company Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world, provided linguistic expertise.