



OVERTURNED

2021-010-FB-UA



# Colombia protests

The Oversight Board has overturned Facebook's decision to remove a post showing a video of protesters in Colombia criticising the country's president, Ivan Duque.

## Policies and topics

-  Community organizations, Freedom of expression, Protests
-  Hate speech

## Region and countries

-  Latin America and the Caribbean
-  Colombia

## Platform

-  Facebook

## Attachments

[Public Comments 2021-010-FB-UA](#)

## Case summary

The Oversight Board has overturned Facebook’s decision to remove a post showing a video of protesters in Colombia criticizing the country’s president, Ivan Duque. In the video, the protesters use a word designated as a slur under Facebook’s Hate Speech Community Standard. Assessing the public interest value of this content, the Board found that Facebook should have applied the newsworthiness allowance in this case.

### **About the case**

In May 2021, the Facebook page of a regional news outlet in Colombia shared a post by another Facebook page without adding any additional caption. This shared post is the content at issue in this case. The original root post contains a short video showing a protest in Colombia with people marching behind a banner that says “SOS COLOMBIA.”

The protesters are singing in Spanish and address the Colombian president, mentioning the tax reform recently proposed by the Colombian government. As part of their chant, the protesters call the president "hijo de puta" once and say "deja de hacerte el marica en la tv" once. Facebook translated these phrases as "son of a bitch" and "stop being the fag on tv." The video is accompanied by text in Spanish expressing admiration for the protesters. The shared post was viewed around 19,000 times, with fewer than five users reporting it to Facebook.

### **Key findings**

Facebook removed this content as it contained the word “marica” (from here on redacted as “m\*\*ica”). This violated Facebook’s Hate Speech Community Standard which does not allow content that “describes or negatively targets people with slurs” based on protected characteristics such as sexual orientation. Facebook noted that while, in theory, the newsworthiness allowance could apply to such content, the allowance can only be applied if the content moderators who initially review the content decide to escalate it for additional review by Facebook’s content policy team. This did not happen in this case. The Board also notes that Facebook does not make its criteria for escalation publicly available.

The word “m\*\*rica” has been designated as a slur by Facebook on the basis that it is

inherently offensive and used as an insulting and discriminatory label primarily against gay men. While the Board agrees that none of the exceptions currently listed in Facebook's Hate Speech Community Standard permit the slur's use, which can contribute to an environment of intimidation and exclusion for LGBT people, it finds that the company should have applied the newsworthiness allowance in this case.

The newsworthiness allowance requires Facebook to assess the public interest of allowing certain expression against the risk of harm from allowing violating content. As part of this, Facebook considers the nature of the speech as well as country-specific context, such as the political structure of the country and whether it has a free press.

Assessing the public interest value of this content, the Board notes that it was posted during widespread protests against the Colombian government at a significant moment in the country's political history. While participants appear to use the slur term deliberately, it is used once among numerous other utterances and the chant primarily focuses on criticism towards the country's president.

The Board also notes that, in an environment where outlets for political expression are limited, social media has provided a platform for all people, including journalists, to share information about the protests. Applying the newsworthiness allowance in this case means that only exceptional and limited harmful content would be permitted.

## **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:

- Publish illustrative examples from the list of slurs designated as violating under its Hate Speech Community Standard, including borderline cases with words which may be harmful in some contexts but not others.
- Link the short explanation of the newsworthiness allowance provided in the introduction to the Community Standards to the more detailed explanation in the Facebook's Transparency Center of how this policy applies. The company should supplement this explanation with illustrative examples from a range of contexts, including reporting on large scale protests.

contents, including reporting on hate speech protests.

- Develop and publicize clear criteria for content reviewers for escalating for additional review public interest content that potentially violates the

Community Standards but may be eligible for the newsworthiness allowance.

These criteria should cover content depicting large protests on political issues.

- Notify all users who reported content which was assessed as violating but left on the platform for public interest reasons that the newsworthiness allowance was applied to the post.

\*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 a Facebook post showing a video of protesters in Colombia criticizing the Colombian president, Ivan Duque. In the video, protesters used a word which Facebook has designated as a slur that violates its Hate Speech Community Standard for being a direct attack against people based on their sexual orientation. The Board found that, while the removal was prima facie in line with the Hate Speech Community Standard (meaning that on its face the content appeared to violate the Standard), the newsworthiness allowance should have been applied in this case to keep the content on the platform.

### **2. Case description**

In May 2021, the Facebook page of a regional news outlet in Colombia shared a post by another Facebook page, without adding any additional caption – this shared post is the content at issue in this case. The original root post contains a short video (originally shared on TikTok), which shows a protest in Colombia, with people marching behind a banner that says "SOS COLOMBIA." The protesters are singing in Spanish and address the Colombian president, mentioning the tax reform recently proposed by the Colombian government. As part of their chant, the protesters call the president an "hijo de puta" once and say "deja de hacerte el marica en la tv" once. Facebook translated these phrases as "son of a bitch" and "stop being the fag on tv." The video, which is 22 seconds long, is accompanied by text in

Spanish, including a denunciation for the protesters.

Spanish expressing admiration for the protesters.

The shared post was viewed around 19,000 times and shared over 70 times. Fewer than five users reported the content. Following human review, Facebook removed the shared post under its Hate Speech policy. Under its Hate Speech Community Standard, Facebook takes down content that "describes or negatively targets people with slurs, where slurs are defined as words that are inherently offensive and used as insulting labels" on the basis of protected characteristics including sexual orientation. The word "marica" (hereafter redacted as "m\*\*ica") is on Facebook's list of prohibited slur words. The user who posted the shared post appealed Facebook's decision. Following further human review, Facebook upheld its original decision to remove the content. Facebook also removed the original root post from the platform.

### **3. Authority and scope**

The Board has the power 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).

### **4. Relevant standards**

The Oversight Board considered the following standards in its decision:

#### *1. Facebook's Community Standards:*

In the policy rationale for the Hate Speech Community Standard, Facebook states that hate speech is not allowed on the platform "because it creates an environment of intimidation and exclusion and, in some cases, may promote real-world violence."

The Community Standard defines hate speech as "a direct attack against people — rather than concepts or institutions — on the basis of what we call protected characteristics: race,

ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease.” It prohibits content that “describes or negatively targets people with slurs, where slurs are defined as words that are inherently offensive and used as insulting labels for the above-listed characteristics.”

## *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, the relevant one in this case being “Dignity”:

*"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 them.*

## *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. The Board's analysis of Facebook’s human rights responsibilities in this case was informed by the following human rights standards:

- The right to freedom of opinion and 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 non-discrimination: Article 2, para. 1 and Article 26, ICCPR; General Comment No. 18, Human Rights Committee 1989; UN Human Rights Council Resolution 32/2 on the protection against violence and discrimination based on sexual orientation and gender identity, 2016; UN Independent Expert on protection against violence and discrimination based on sexual orientation and gender identity, reports: A/HRC/35/36 (2017) and A/HRC/38/43 (2018).

- The right to peaceful assembly: Article 21, ICCPR, General Comment No. 37, Human Rights Committee, 2020; UN Special Rapporteur on freedom of peaceful assembly and of association, report A/HRC/41/41 (2019).

## 5. User statement

The user, who is the administrator of the page on which the content was posted, submitted their appeal to the Board in Spanish. In the appeal, the user states that they are a journalist reporting on local news from their province. The user claims that the content was posted by another person who took their phone, but that, nevertheless, this content did not intend to cause harm and showed protests in a time of crisis. The user states that they aim to follow Facebook's policies, and claim that this removal led to account penalties.

The user further states that the content shows young people protesting within the framework of freedom of expression and peaceful protest, and that the young people are expressing themselves without violence and demanding rights using typical language. The user also expresses concern about government repression of protest.

## 6. Explanation of Facebook's decision

Facebook removed this content on the basis that it contained the word "m\*\*ica" and therefore violated Facebook's Hate Speech Community Standard, which prohibits "[c]ontent that describes or negatively targets people with slurs, where slurs are defined as words that are inherently offensive and used as insulting labels for the above characteristics [i.e., a protected characteristic]." The word "m\*\*ica" is on Facebook's list of prohibited slur words, on the grounds that it targets people based on their sexual orientation.

Facebook states that there is no exception for using slurs against political leaders or public figures. Furthermore, Facebook notes that "it does not matter if the speaker or the target are members of the protected characteristic group being attacked. Since slurs are inherently offensive terms for a group defined by their protected characteristic, the use of slurs [is] not allowed, unless the user has clearly demonstrated that [the slur] was shared to condemn, to discuss, to raise awareness of the slur, or the slur is used self-referentially or in an empowering way."

With regards to whether the newsworthiness allowance could be applied to this content, Facebook explained that the newsworthiness allowance can only be applied if the content

moderators who initially review the content decide to escalate it for additional review by Facebook’s content policy team – in this case, the content was not escalated for further review. The Board notes that Facebook does not make its criteria for escalation publicly available. It stated that “the newsworthiness allowance, in theory, could apply to such content. In this case, however, the public interest value does not outweigh the risk of harm from allowing content containing an inherently offensive and insulting label to remain on Facebook’s platform.”

## **7. Third-party submissions**

The Oversight Board received 18 public comments related to this case. Five of the comments were submitted from Asia Pacific and Oceania, one from Europe, seven from Latin America and the Caribbean, one from Middle East and North Africa, and four from the United States and Canada.

The submissions covered the following themes: the various meanings and uses of the word “m\*\*ica” in Colombia; concern that Facebook removes journalistic content; censorship of media outlets in Colombia; and analysis of whether the content complied with the Community Standards.

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

## **8. Oversight Board analysis**

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 finds that, although Facebook’s decision to remove the content was prima facie in line with its Hate Speech Community Standard (meaning that on its face the content appeared to violate the Standard), the newsworthiness allowance should have been applied in this case to allow the content to remain on the platform.

The word “m\*\*ica” has been designated as a slur by Facebook on the basis that it is



inherently offensive and used as an insulting and discriminatory label primarily against gay men. As noted in section 6, Facebook explained to the Board that neither the sexual orientation nor the public figure status of the target is relevant to the enforcement of this policy. Since discriminatory slurs are inherently offensive, the use of slurs is not allowed unless a policy exception applies. Those exceptions allow the sharing of slurs to condemn, to discuss, or raise awareness of hate speech, or when used self-referentially or in an empowering way.

The Board sought expert input and public comments that confirmed that the word “m\*\*ica” has multiple meanings and can be used without discriminatory intent. However, there is agreement that its origins are homophobic, principally against gay men, even though its use has evolved to reportedly common usage in Colombia to refer to a person as “friend” or “dude,” and as an insult equivalent to “stupid,” “dumb” or “idiot.” The Board notes that this evolution or normalization does not necessarily mean the term’s usage is less harmful for gay men, as this casual use may continue to marginalize lesbian, gay, bisexual and transgender (LGBT) people and communities by implicitly associating them with negative characteristics.

The Board understands why Facebook designated this word as a slur, and agrees none of the exceptions currently listed in the Hate Speech Community Standard explicitly applies to permit its use on the platform. Nevertheless, the Board finds that the newsworthiness allowance should have been applied to allow this content to remain on the platform.

Facebook has provided more public information about the newsworthiness allowance in response to the Board’s recommendations in case 2021-001-FB-FBR. This allowance requires the company to assess the public interest of expression against the risk of harm from allowing violating content on the platform. Facebook states that it takes into account country-specific circumstances, the nature of the speech, including whether it relates to governance or politics, and the political structure of the country, including whether it has a free press. The allowance is not applied on the basis of the identity of the speaker as a journalist or media outlet, or simply because the subject matter is in the news.

Several contextual factors are relevant to assessing the public interest in this content. It was posted during widespread protests against the Colombian government. The chant in the video was primarily focused on criticism towards the president. While participants appear to use the slur term deliberately, the protest was not discriminatory in its objectives. The slur term is used once, among numerous other utterances. Where it appears that a user shares

footage to raise awareness of the protests and to express support for their cause, and not to insult people on the basis of protected characteristics or to incite discrimination or violence, the newsworthiness exception is particularly applicable.

The Board emphasizes that the application of the newsworthiness allowance in this case should not be understood as endorsement of the language the protesters used. The Board acknowledges that the term used by protesters in this video is offensive to gay men, including in Colombia, and its usage could create a risk of harm. Allowing such slurs on the platform can contribute to an environment of intimidation and exclusion for LGBT people and, in some cases, promote real-world violence. This language is not inherently of public interest value. Rather, the public interest is in allowing expression on the platform that relates to a significant moment in Colombia's political history.

The Board also notes that social media has played an important role in providing a platform for all people, including journalists, to share information about the protests in an environment where public comments and expert reports suggest the media landscape would benefit from greater pluralism. Allowing the content through the application of the newsworthiness allowance means that only exceptional and limited harmful content would be permitted. The newsworthiness exception should not be construed as a broad permission for hate speech to remain up.

## **8.2 Compliance with Facebook's values**

The Board finds that restoring this content is consistent with Facebook's values. Facebook lists "Dignity" as one of its values. The Board shares Facebook's concern that permitting hateful slurs to proliferate on the platform can cause dignitary harm to members of communities targeted by such slurs. The Board also acknowledges that the use of the slur in this specific case may be demeaning and harmful to members of the LGBT community.

At the same time, Facebook has indicated that "Voice" is not just one of its values, but its "paramount" value. The sharing of content that shows widespread protests against a political leader represents the value of "Voice" at its apex, particularly in an environment in which outlets for political expression are limited. Application of the newsworthiness allowance to the slur policy in this setting—the sharing of information about political protests against a national leader—permits Facebook to honor its paramount commitment to "Voice" without sacrificing its legitimate commitment to "Dignity."

### 8.3 Compliance with Facebook's human rights responsibilities

The Board finds 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 that this includes the International Covenant on Civil and Political Rights (ICCPR).

#### *Freedom of expression and freedom of peaceful assembly (Articles 19 and 21 ICCPR)*

Article 19 of the ICCPR provides for broad protection of expression. This protection is “particularly high” for “public debate in a democratic society concerning figures in the public and political domain” ( [General Comment 34](#), para. 34). Article 21 of the ICCPR provides similar protection for freedom of peaceful assembly - assemblies with a political message are accorded heightened protection ( [General Comment No. 37](#), paras 32 and 49), and Article 21 extends to protect associated activities that take place online ( *Ibid.*, paras 6, and 34). The Human Rights Committee has further emphasized the role of journalists, human rights defenders and election monitors and others monitoring or reporting on assemblies, including in respect of the conduct of law enforcement officials ( *Ibid.*, paras 30 and 94). Interference with online communications about assemblies has been interpreted to impede the right to freedom of peaceful assembly ( *Ibid.*, para. 10).

Article 19 requires that where 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).

Facebook has recognized its responsibilities to respect international human rights standards under the UNGPs. Relying on the UNGPs framework, the UN Special Rapporteur on freedom of opinion and expression has called on social media companies to ensure their content rules are guided by the requirements of Article 19, para. 3, ICCPR (see A/HRC/38/35, paras. 45 and 70). The Board examined whether the removal of the post would be justified under the three-part test for restrictions on freedom of expression under Article 19 in accordance with Facebook's human rights commitments.

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

The principle of legality under international human rights law requires rules used to limit expression to be clear, precise, publicly accessible and non-discriminatory ( [General Comment 34](#), para. 25 and para. 26). The Human Rights Committee has further noted that rules “may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution” (General Comment 34, para. 25).

Although Facebook’s Hate Speech Community Standard specifies that slurs related to protected characteristics are prohibited, the specific list of words which Facebook has designated as slurs in different contexts is not publicly available. Given that the word “m\*\*ica” can be used in different ways, it may not be clear to users that this word contravenes Facebook’s prohibition against slurs. Facebook should provide the public with more information on its list of slurs to enable users to regulate their conduct accordingly. The Board has made a policy recommendation below in this regard.

The Board recommended in [case 2021-001-FB-FBR](#) that Facebook should produce more information to help users understand and evaluate the process and criteria for applying the newsworthiness allowance. In [response](#), Facebook published [more information](#) in its Transparency Center and said that from 2022 it would begin providing regular updates about the number of times it applied this allowance in the Community Standards Enforcement Reports. However, the Transparency Center resource is not linked from the more limited explanation of the newsworthiness allowance in the introduction to the Community Standards. While the Board notes the commitment to provide more information in Enforcement Reports, this will not provide information to users who post or view content which is given an allowance.

The Board recommended in [case 2020-003-FB-UA](#) that Facebook should give users more detail on the specific parts of the Hate Speech policy that their content violated, so that users can regulate their behavior accordingly. The Board notes that there is a distinction to be made here. Case 2020-003-FB-UA concerned content originally created by the user themselves that could be easily edited upon notification, whereas the present case concerns content depicting public events. Nevertheless, the Board understands that it is important for users to receive clear information about why their content is removed as a general rule. The Board appreciates the [update](#) Facebook provided in July 2021 on the company’s efforts to implement this recommendation, which when rolled out in all languages should provide more information to users whose content is removed for using slurs. The Board encourages

Facebook to provide clearer timelines for implementing this recommendation in non-English languages.

## *II. Legitimate aim*

Any restriction on expression should pursue one of the legitimate aims listed in the ICCPR, which include the “rights of others.” The policy at issue in this case pursued the legitimate aim of protecting the rights of others (General Comment No. 34, para. 28) to equality, protection against violence and discrimination based on sexual orientation and gender identity (Article 2, para. 1, Article 26 ICCPR; UN Human Rights Committee, *Toonen v. Australia* (1992) and General Comment No. 37, para. 25; UN Human Rights Council Resolution 32/2 on the protection against violence and discrimination based on sexual orientation and gender identity).

## *III. Necessity and proportionality*

Any restrictions on freedom of expression should be appropriate to achieve their protective function and should be the least intrusive instrument among those which might achieve their protective function (General Comment 34, para. 34).

The Board finds that it was not necessary or proportionate to remove the content in this case. As discussed above in section 8.1, the Board recognizes the potential for harms to the rights of LGBT people from allowing homophobic slurs to remain on the platform. However, context is crucial in assessing the proportionality of removal of the content. The UN Special Rapporteur on freedom of expression has stated in relation to hate speech that the “evaluation of context may lead to a decision to make an exception in some instances, when the content must be protected as, for example, political speech” (A/74/486, para. 47(d)).

Taking into account the political context in Colombia, the fact this protest addressed a political figure, and the significant role that social media has played in sharing information about the protests there, the Board finds that removal of this content was not proportionate to achieve the aim of protecting the rights to non-discrimination and equality of LGBT people.

## *Freedom of peaceful assembly*

For a minority of the Board, it is also important to assess the content restriction in this case

for its impact on the right to freedom of peaceful assembly. Journalists and other observers play an important role in amplifying the collective expression and associative power of

protests through disseminating footage of those events online – these acts are protected by Article 21 of the ICCPR (General Comment No. 37, para. 34).

The minority believes that assessing restrictions on the right to peaceful assembly is substantially similar to the test for evaluating restrictions on the right to freedom of expression. Restrictions on the right to freedom of peaceful assembly should be narrowly drawn, meeting the requirements of legality, legitimate aim, and necessity and proportionality ( *ibid.*, paras 8 and 36). The UN Special Rapporteur on freedom of peaceful assembly and of association has also called on companies engaged in content moderation to be guided by international human rights law (see A/HRC/41/41, para. 19), noting “the enormous power of Facebook” ( *Ibid.*, para. 4). The Human Rights Committee has noted that private ownership of communication platforms should inform a contemporary understanding of the legal framework Article 21 of the ICCPR requires ( *op cit.* para. 10 and 34) .

The three-part analysis above, which the minority joins, leads to an additional minority conclusion that Facebook’s removal of the content in this case impaired the right to freedom of peaceful assembly, and that restriction was not justified.

## **9. Oversight Board decision**

The Oversight Board overturns Facebook's decision to take down the content, requiring the post to be restored.

## **10. Policy advisory statement**

The following recommendations are numbered, and the Board requests that Facebook provides an individual response to each as drafted.

### *Content policy*

To further clarify for users its rules on Hate Speech and on how the newsworthiness allowance applies, Facebook should:

1. Publish illustrative examples from the list of slurs it has designated as violating under its Hate Speech Community Standard. These examples should be included in the Community Standard and include edge cases involving words which may be harmful in some contexts but not others, describing when their use would be violating. Facebook should clarify to users that these examples do not constitute a complete list.
2. Link the short explanation of the newsworthiness allowance provided in the introduction to the Community Standards to the more detailed Transparency Center explanation of how this policy applies. The company should supplement this explanation with illustrative examples from a variety of contexts, including reporting on large scale protests.

### *Enforcement*

To safeguard against the wrongful removal of content that is in the public interest, and to ensure provision of adequate information to users who report such content, Facebook should:

3. Develop and publicize clear criteria for content reviewers to escalate for additional review public interest content that potentially violates the Community Standards but may be eligible for the newsworthiness allowance. These criteria should cover content depicting large protests on political issues, in particular in contexts where states are accused of violating human rights and where maintaining a public record of events is of heightened importance.
4. Notify all users who reported content assessed as violating but left on the platform for public interest reasons that the newsworthiness allowance was applied to the post. The notice should link to the Transparency Center explanation of the newsworthiness allowance.

### **\*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.