UPHELD

2021-011-FB-UA

South Africa slurs

The Oversight Board has upheld Facebook’s decision to remove a post discussing South African society under its Hate Speech Community Standard.

Case summary

The Oversight Board has upheld Facebook’s decision to remove a post discussing South African society under its Hate Speech Community Standard. The Board found that the post contained a slur which, in the South African context, was degrading, excluding and harmful to the people it targeted.

About the case

In May 2021, a Facebook user posted in English in a public group that described itself as focused on unlocking minds. The user’s Facebook profile picture and banner photo each depict a black person. The post discussed “multi-racialism” in South Africa, and argued that poverty, homelessness, and landlessness have increased for black people in the country since 1994.

It stated that white people hold and control the majority of the wealth, and that wealthy black people may have ownership of some companies, but not control. It also stated that if “you think” sharing neighborhoods, language, and schools with white people makes you “deputy-white”
then "you need to have your head examined." The post then concluded with "[y]ou are a "sophisticated slave," "a clever black," "n goeie kaffir" or "House nigger" (hereafter redacted as "k***ir" and "n***er").

**Key findings**

Facebook removed the content under its Hate Speech Community Standard for violating its policy prohibiting the use of slurs targeted at people based on their race, ethnicity and/or national origin. The company noted that both "k***ir" and "n***er" are on Facebook’s list of prohibited slurs for the Sub-Saharan market.

The Board found removing this content to be consistent with Facebook’s Community Standards. The Board evaluated public comments and expert research in finding that both "k***ir" and "n***er" have discriminatory uses, and that "k***ir" is a particularly hateful and harmful word in the South African context.

The Board agreed with Facebook that the content did not condemn or raise awareness of the use of "k***ir," and did not use the word in a self-referential or empowering manner. As such, no exception to the company’s Hate Speech Community Standard applied in this case.

While the user’s post discussed relevant and challenging socio-economic and political issues in South Africa, the user racialized this critique by choosing the most severe terminology possible in the country.

In the South African context, the slur "k***ir" is degrading, excluding and harmful to the people it targets. Particularly in a country still dealing with the legacy of apartheid, the use of racial slurs on the platform should be taken seriously by Facebook.

The Board supports greater transparency around Facebook’s slur list. The company should provide more information about the list, including how it is enforced in different markets and why it remains confidential.

The Board also urged Facebook to improve procedural fairness in enforcing its Hate Speech policy, issuing the recommendation below. This would help users understand why Facebook removed their content and allow them to change their behavior in the future.

**The Oversight Board’s decision**

The Oversight Board upholds Facebook’s decision to remove the post.

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

- Notify users of the specific rule within the Hate Speech Community Standard that has been violated in the language in which they use Facebook, as recommended in case decision 2020-003-FB-UA (Armenians in Azerbaijan) and case decision 2021-002-FB-UA (Depiction of Zwarte Piet). In this case, for example, the user should have been notified they violated the slurs prohibition. The Board has noted Facebook’s response to Recommendation No. 2 in case decision 2021-002-FB-UA, which describes a new classifier that should be able to notify English-language Facebook users their content has violated the slur rule. The Board looks forward to Facebook providing information that confirms implementation for English-language users and information about the timeframe for implementation for other language users.

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

**Full case decision**

1. Decision summary

The Oversight Board has upheld Facebook’s decision to remove a post discussing South African society under its Hate Speech Community Standard which prohibits the use of slurs.

2. Case description
In May 2021, a Facebook user posted in English in a public group that described itself as focused on unlocking minds. The user’s Facebook profile picture and banner photo each depict a black person. The post discussed “multi-racialism” in South Africa, and argued that poverty, homelessness, and landlessness have increased for black people in South Africa since 1994. It stated that white people hold and control the majority of wealth, and that wealthy black people may have ownership of some companies, but not control. It also stated that if “you think” sharing neighborhoods, language, and schools with white people makes you “deputy-white” then “you need to have your head examined.” The post then concluded with “[y]ou are” a “sophisticated slave,” “a clever black,” “’n goeie kaffir” or “House nigger” (hereafter redacted as “k***r” and “n***er”).

The post was viewed more than 1,000 times, receiving fewer than five comments and more than 10 reactions. It was shared over 40 times. The post was reported by a Facebook user for violating Facebook’s Hate Speech Community Standard. According to Facebook, the user who posted the content, the user who reported the content, and “all users who reacted to, commented on and/or shared the content” have accounts located in South Africa.

The post remained on the platform for approximately one day. Following review by a moderator, Facebook removed the post under its Hate Speech policy. Facebook’s Hate Speech Community Standard 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” based on their race, ethnicity and/or national origin. Facebook noted that while its prohibition against slurs is global, the designation of slurs on its internal slurs list is market oriented. Both “k***r” and “n***er” are on Facebook’s list of prohibited slurs for the Sub-Saharan market.

Facebook notified the user that their post violated Facebook’s Hate Speech Community Standard. Facebook stated that the notice to the user explained that this Standard prohibits, for example, hateful language, slurs, and claims about the coronavirus. The user appealed the decision to Facebook, and, following a second review by a moderator, Facebook confirmed the post was violating. The user then submitted an 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, and its decision is binding on Facebook (Charter Article 3, Section 5). The Board’s decisions may include policy advisory statements with non-binding recommendations that Facebook must respond to (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 3,” prohibited content includes 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 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, 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.

https://www.oversightboard.com/decision/FB-TYE2766G/
“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 reaffirmed 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:

- **Equality and non-discrimination**: Article 2, para. 1 and Article 26 (ICCPR); Article 2, ICERD; General Recommendation 35, Committee on the Elimination of Racial Discrimination, 2013.

5. User statement

The user stated in their appeal to the Board that people should be allowed to share different views on the platform and “engage in a civil and healthy debate.” The user also stated that they “did not write about any group to be targeted for hatred or for its members to be ill-treated in any way by members of a different group.” The user argued that their post instead “encouraged members of a certain group to do introspection and re-evaluate their priorities and attitudes.” They also stated that there is nothing in the post or “in its spirit or intent” that would promote hate speech, and that it is unfortunate that Facebook is unable to tell them what part of their post is hate speech.

6. Explanation of Facebook’s decision

Facebook removed the content under the Hate Speech Community Standard, specifically for violating its policy prohibiting the use of slurs targeted at people based on their race, ethnicity and/or national origin. Facebook noted in its decision rationale that it prohibits content containing slurs, which are inherently offensive and used as insulting labels, unless the user clearly demonstrates that that content “was shared to condemn, to discuss, to raise awareness of the slur, or the slur is used self-referentially or in an empowering way.” Facebook argued that these exceptions did not apply in this case.

Facebook argued the post addressed itself to “Clever Blacks” and that this phrase “has been used to criticize Black South Africans who are perceived to be ‘excessively anxious to appear impressively clever or intelligent.’” Facebook also noted that the post used the words “k***ir” and “n***er,” both of which are on its confidential list of prohibited slurs. According to Facebook, the word “k***ir” is deemed as “South Africa’s most charged epithet” and historically used by white people in South Africa “as a derogatory term to refer to black people.” Facebook added that this term “has never been reclaimed by the Black community.” Facebook stated that the word “n***er” is also “highly offensive in South Africa” but that it “has been reclaimed by the Black community for use in a positive sense.”

Facebook also noted that, as part of the process for determining whether a word or phrase constitutes a slur, it must be recommended by its internal or external stakeholders. Facebook specified that it recently held consultations with stakeholders that confirmed the need for the exception of the Hate Speech policy that allows the use of slurs when “used self-referentially or in an empowering way.” According to Facebook, external stakeholders generally agreed that it is important “to allow people to use a reclaimed slur in an empowering way,” but it is also critical that Facebook does not “guess, decide, or gather data about users’ membership in a protected characteristic” to decide whether the use of a slur violates its policies. Facebook confirmed in its response to the Board that the external stakeholders included seven experts/organizations in North America, 16 from Europe, 30 from Middle East, two from Africa, six in Latin America and one in the Asia Pacific/India region.

Facebook concluded that while the user’s profile picture depicts a black person, the user “does not identify themselves with the slurs or argue that they should be reconsidered or reclaimed.” According to Facebook, “the slurs in this post are being used in an offensive manner to attack” black people who live among white 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, the slurs in the post were used "to attack other people in a harmful manner antithetical to Facebook's values." In this regard, Facebook referred to the Board's case decision [2020-003-FB-UA](https://www.oversightboard.com/decision/FB-TYE2766G/).

Facebook argued that its decision was consistent with international human rights standards. It stated that its decision complied with the international human rights law requirements that restrictions on freedom of expression respect the principles 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 slurs." Additionally, the decision to remove the content was legitimate to protect "the rights of others from harm and discrimination," and consistent with the requirement under Article 20, para. 2 of the ICCPR to prohibit speech that advocates "national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence." Finally, Facebook argued that its decision to remove the content was "necessary and proportionate to limit harm" against members of the black community and "to other viewers of seeing hate speech," referring to the Israel Democracy Institute and Yad Vashem’s "Recommendations for Reducing Online Hate Speech," and Richard Delgado’s "Words That Wound: A Tort Action for Racial Insults, Epithets, and Name-Calling."

7. Third-party submissions

The Oversight Board received six public comments related to this case. Three of the comments were from Sub-Saharan Africa, specifically South Africa, one was from Middle East and North Africa, one was from Asia Pacific and Oceania, and one was from the United States and Canada. The Board received comments from stakeholders including academia and civil society organizations focusing on freedom of expression and hate speech in South Africa.

The submissions covered themes including the analysis of the words “clever blacks,” “n***er” and “k***ir;” whether the words “n***er” and “k***ir” qualify as hate speech; the user’s and reporter’s identity and its impact on how the post was perceived; and the applicability of Facebook’s Hate Speech policy exceptions.

To read public comments submitted for this case, please click [here](https://www.oversightboard.com/decision/FB-TYE2766G/).

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 removing this content is consistent with Facebook’s Community Standards. The use of the word "k***ir" in the user’s post violated the Hate Speech Community Standard, and no policy exception applied.

The Hate Speech Community Standard prohibits attacks based on protected characteristics. This includes “[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.” Facebook considers “k***ir” and “n***er” racial slurs. The Board evaluated public comments and expert research in finding that both slurs have discriminatory uses, and that “k***ir” is a particularly hateful and harmful word in the South African context.

The internet is a global network and content that is posted on Facebook by a user in one context may circulate and cause damage in other contexts. At the same time, Facebook’s confidential slur list is divided by markets in recognition that words carry different meaning and may cause different impacts in some situations. The Board notes that it has previously dealt with the use of the word “kafir” in case decision [2020-007-FB-FBR](https://www.oversightboard.com/decision/FB-TYE2766G/), where the Board ordered the restoration of the content. In that case, Facebook did not treat the term as a slur, but rather meaning “non-believers” as the target group of an alleged “veiled threat” under the Violence and Incitement policy. The term with one “f,” used in that case in India, has the same origins in Arabic as the South African term with two. This demonstrates the difficulty for Facebook of enforcing a blanket prohibition on certain words globally, where similar or identical terms in the same or different languages can hold different meanings and pose different risks depending on their contextual use.

The Board notes that the post was targeted at a group of black South Africans. The Board further notes that the user’s critique discussed this group’s presumed economic, educational and professional status and privilege. The user argued in their statement to the Board that they were not targeting or inciting hate or discrimination against persons on account of their race. A few Board Members found this argument compelling.
However, the user chose the most severe terminology possible in South Africa to racialize this critique. The use of the “k***ir” term, with the prefix “good” in Afrikaans, has a clear historical association that carries significant weight in South Africa. The Board finds that the use of the “k***ir” term in this context cannot be separated from its harmful and discriminatory meaning.

Facebook told the Board that it reviews its slur list annually. About the designation of “k***ir” on the list, Facebook shared that in 2019 it held a consultation with civil society organizations in South Africa. In that meeting stakeholders told Facebook that “k***ir” “is used in a way to denigrate and demean a Black person as inferior and worthy of contempt.” To meet its human rights responsibilities when developing and reviewing policies, including the slur list, Facebook should consult potentially affected groups and other relevant stakeholders, including human rights experts.

Facebook has four exceptions to its slur policy that are referenced in the policy rationale of the Hate Speech Community Standard: “We recognize that people sometimes share content that includes someone else’s hate speech to condemn it or raise awareness. In other cases, speech that might otherwise violate our standards can be used self-referentially or in an empowering way.” The majority of the Board is of the view that Facebook’s exceptions did not apply in this case. This is because the content did not condemn the use of the word “k***ir,” it did not raise awareness, and it was not used in an empowering manner. The Board also found this content was not self-referential, despite a few members considering this exception should have applied because it expresses criticism against some privileged members of the targeted group. However, the Board found that nothing in the post suggests the user considers themself to be in that targeted group. Further, the user’s reference to “you” and “your” in the post distanced the user from the targeted group.

Therefore, the Board finds that Facebook was acting according to its Community Standard on Hate Speech when it decided to remove this content.

8.2 Compliance with Facebook’s values

The Board recognizes that “Voice” is Facebook’s paramount value, and that Facebook wants users of the platform to be able to express themselves freely. However, Facebook’s values also include “Dignity” and “Safety.”

The Board finds that value of “Voice” to be of particular importance to political discourse about racial and socio-economic equality in South Africa. Arguments about the distribution of wealth, racial division and inequality are highly relevant, especially in a society that many argue is still undergoing transition from apartheid towards greater equality. Those targeted by slurs also may see “Voice” impacted as their use may have a silencing impact on those targeted and inhibit their participation on Facebook.

The Board also considers the values of “Dignity” and “Safety” to be of vital concern in this context. The Board found that the use of slur “k***ir” in the context of South Africa can be degrading, excluding and harmful to the people targeted by the slur (see, for example, 2019 PeaceTech Lab and Media Monitoring Africa’s Lexicon of Hateful Terms, pages 12 and 13). Particularly in a country still dealing with the legacy of apartheid, the mention of racial slurs on the platform should be taken seriously by Facebook.

It is relevant that in this context the user opted to deploy a slur term that is particularly incendiary in South Africa. It was possible for the user to engage in political and socio-economic discussions on Facebook in ways that appealed to the emotions of their audience without referencing this slur. This justified displacing the user’s “Voice” to protect the “Voice,” “Dignity” and “Safety” of others.

8.3 Compliance with Facebook’s human rights responsibilities

The Board concludes that removing 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. While protection is “particularly high” for political expression and debate (General Comment 34, para. 38). The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD), also provides protection to freedom of expression (Article 5), and the Committee tasked with monitoring states’ compliance has emphasized the importance of the right to assist “vulnerable groups in redressing the balance of power among the components of society” and to offer “alternative views and counterpoints” in discussions (CERD Committee, General Recommendation 35, para. 29). At the same time, the Board has upheld Facebook’s decisions to restrict content that meet the Article 19 ICCPR three-part test of legality, legitimacy, and necessity and proportionality. The Board concluded that Facebook’s actions satisfied its responsibilities 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). 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). In some situations, Facebook’s concepts of “inherently offensive” and “insulting” may be too subjective and raise concerns for legality (A/74/486, para. 46, see also A/HRC/38/35, para. 26). Additionally, there may be situations where a slur has multiple meanings or can be used in ways that would not be considered an “attack.”

The Board asked Facebook how its market-specific slur list is enforced, and if a slur’s appearance on any market list means it cannot be used globally. Facebook responded that its “prohibition against slurs is global, but the designation of slurs is market-specific, as Facebook recognizes that cultural and linguistic variations mean that words that are slurs in some places may not be in others.” The Board reiterated its initial question. Facebook then responded “[i]f a term appears on a market slur list, the hate speech policy prohibits its use in that market. The term could be used elsewhere with a different meaning; therefore, Facebook would independently evaluate whether to add it to the other market’s slur list.” It remains unclear to the Board how Facebook enforces the slur prohibition in practice and at scale. The Board does not know how Facebook’s enforcement processes to identify and remove violating content operate globally for market-specific terms, how markets are defined, and when and how this independent evaluation occurs.

In this case, as noted above, the sources consulted by the Board concur that “k***ir” is widely understood as South Africa’s most charged racial epithet. As the expression fell unambiguously within the prohibition, Facebook met its responsibility of legality in this case.

The Board notes its decision in case 2021-010-FB-UA and its recommendation that Facebook provide illustrative examples from the slurs policy in the public-facing Community Standards (Recommendation No. 1). The Board supports greater transparency around the slur list and continues to discuss how Facebook could provide users with sufficient clarity while respecting the rights to equality and non-discrimination. A minority of the Board believes Facebook should make its slur list public, so it is available to all users. A majority believes the Board should better understand the procedure and criteria for building the list and how specifically it is enforced, as well as possible risks in publication, including strategic behavior to evade slur violations and whether certain words accumulate with harmful effect. Facebook should contribute to this discussion by publishing more information about the slur list, designation and review processes, its enforcement and application globally and/or by market or language, and why it remains confidential.

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.” Previously the Board has stated that the slur prohibition “seeks to protect people’s rights to equality and non-discrimination (Article 2, para. 1, ICCPR [and] to exercise their freedom of expression on the platform without being harassed or threatened (Article 19, ICCPR),” among other rights (case decision 2020-003-FB-UA). The Board reiterates that these are legitimate aims.

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, the Board decides that removing the content was appropriate to achieve a protective function. The Board also issues a policy recommendation to Facebook on improving the enforcement of its Hate Speech Community Standard.

Facebook’s Hate Speech Community Standard prohibits some discriminatory expression including slurs, absent any requirement that the expression incite violence or discriminatory acts. While such prohibitions would raise concerns if imposed by a government at a broader level (A/74/486, para. 48), particularly if enforced through criminal or civil sanctions, the Special Rapporteur indicates that entities engaged in content moderation like Facebook can regulate such speech:

The scale and complexity of addressing hateful expression presents long-term challenges and may lead companies to restrict such expression even if it is not clearly linked to adverse outcomes (as hateful advocacy is connected to incitement in Article 20(2) of the ICCPR). Companies should articulate the bases for such restrictions, however, and demonstrate the necessity and proportionality of any content actions (A/HRC/38/35, para. 28).
In this case, the historical and social context was crucial, as the Board notes the use of the word “k***ir” is closely linked with discrimination and the history of apartheid in South Africa. The Board also discussed the status of the speaker and their intent. The Board acknowledges that there may be instances in which the racial identity of the speaker is relevant to analysis of the content’s impact. The Board notes the Special Rapporteur’s concerns that inconsistent Hate Speech policy enforcement may “penaliz[e] minorities while reinforcing the status of dominant or powerful groups” to the extent that harassment and abuse remains online while “critiques of racist phenomena and power structures” may be removed (A/HRC/38/35, para. 27). While a profile photo may lead to inferences about the user, the Board notes it is generally not possible to confirm if profile photos depict those responsible for content. Additionally, the Board discussed concerns Facebook said stakeholders raised about it attempting to determine users’ racial identities. The Board agreed that Facebook gathering or maintaining data on users’ perceived racial identities presents serious privacy concerns. In relation to intent, while the user stated they wished to encourage introspection, the post invoked a racial slur with charged historical implications to criticize some black South Africans.

This was a complex decision for the Board. It results in the removal of expression that discusses relevant and challenging socio-economic and political issues in South Africa. Such discussions are important, and a certain degree of provocation should be tolerated when discussing such matters on Facebook. However, the Board finds that given the information analyzed in the previous paragraphs, Facebook’s decision to remove the content was appropriate. The Board also issues a policy recommendation that Facebook prioritize improving procedural fairness to users about its hate speech policy enforcement, so that users can understand with greater clarity the reasons for content removal where it occurs and have the possibility to consider changing their behavior.

9. **Oversight Board decision**

The Oversight Board upholds Facebook’s decision to remove the content.

10. **Policy recommendation**

**Enforcement**

To ensure procedural fairness for users, Facebook should:

1. Notify users of the specific rule within the Hate Speech Community Standard that has been violated in the language in which they use Facebook, as recommended in case decision 2020-003-FB-UA (Armenians in Azerbaijan) and case decision 2021-002-FB-UA (Depiction of Zwarte Piet). In this case, for example, the user should have been notified they violated the slurs prohibition. The Board has noted Facebook’s response to Recommendation No. 2 in case decision 2021-002-FB-UA, which describes a new classifier that should be able to notify English-language Facebook users their content has violated the slur rule. The Board looks forward to Facebook providing information that confirms implementation for English-language users and information about the timeframe for implementation for other language users.

*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.