



## **Comment on Kenyan Politics Using a Designated Slur**

**2025-042-FB-UA**

### **Summary**

The Oversight Board has overturned Meta’s original decision to remove a comment on Kenyan politics including the term “tugeges.” While the comment negatively describes a group of Kenyan voters, it does not inherently create an atmosphere of discriminatory exclusion and intimidation. When this content was posted, the term should not have qualified as a slur. This case shows how’s Meta overbroad response led to political speech and public debate being suppressed.

### **About the Case**

Kenya’s former Prime Minister Raila Odinga stood as a candidate for the African Union’s chairperson in February 2025. Ahead of that election, a Facebook user posted an image of Kenya’s former Deputy President Rigathi Gachagua. Text on the image describes Gachagua’s support for Odinga’s nomination. In the caption, the user states that Gachagua, who is from the Kikuyu ethnic group, is only choosing to endorse Odinga, who is from the Luo ethnic group, to increase his popularity among Luos.

A second user responded in a comment, mocking the original posting user’s content, dismissing it as meant for “tugeges” (“retarded Kikuyu”) – a direct reference to Gachagua’s supporters.

At the time, Meta designated the term “tugeges” a slur under the Hateful Conduct Community Standard. The second user’s comment was removed by automation and a strike applied to their account. However, Meta removed the term from its slur list in June 2025 when the Board asked questions about its designation. The company also restored the comment and revoked the strike the following month.



## **Key Findings**

The Board encourages Meta to be judicious in its designation process before banning words. The term “tugeges” does not meet Meta’s definition of a slur. At the time of Kenya’s 2022 elections, the way in which the word was being used meant it could have qualified. However, its use has quickly evolved, raising timing questions about the January 2024 designation. When this content was posted, “tugeges” should not have qualified as a slur.

While the term can be derogatory in terms of ethnicity, many use it to criticize blind political loyalty. Other Kenyan dialects have equivalents, typically to express political criticism. Kenya’s National Coalition on Freedom of Expression and Content Moderation has not found the term to be hate speech. Experts and public comments point out that the term’s discriminatory ethnic overtones are not fixed and the word has various colloquial meanings. This should have been considered by Meta and underscores the importance of the company engaging national stakeholders in both designating and auditing slur lists.

Deliberations for this case also included discussions on how Meta can continue to offer users an opportunity to act when their content may violate the Hateful Conduct policy. The Board’s recommendation suggests implementing a product feature for potential Hateful Conduct and Bullying and Harassment violations, as well as markets with an active crisis.

## **The Oversight Board’s Decision**

The Board overturns Meta’s original decision to remove the content.

The Board recommends that Meta:



- Provide users with an opportunity for self-remediation comparable to the post time friction intervention that was created as a result of the Pro-Navalny Protest in Russia, recommendation no. 6. If this intervention is no longer in effect, Meta should provide a comparable product intervention.

The Board also reiterates a recommendation from an earlier decision, which Meta has fully committed to and is implementing:

- When auditing its slur lists, Meta should ensure it carries out broad external engagement with relevant stakeholders.

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

## **Full Case Decision**

### **1. Case Description and Background**

In the lead-up to the African Union chairperson election in February 2025, a Facebook user posted an image with text overlay of Kenya’s former Deputy President Rigathi Gachagua supporting the nomination of Raila Odinga as African Union chairperson. In a caption, the user states that Gachagua, who is from the Kikuyu ethnic group, was only endorsing Odinga, who is from the Luo ethnic group, to increase his popularity among the Luos. A second user, who was tagged in the post, responded in a comment, mocking the original poster’s reaction to Gachagua’s statement and dismissing the statement as meant for “tugeges,” referring to Gachagua’s supporters. The commenting user clarified that Gachagua’s endorsement is for an external audience, not for Kenyans.

Meta’s proactive detection tool identified and removed the user’s comment for containing the term “tugeges” and violating the Hateful Conduct Community Standard. Meta designated the word as a slur in January 2024. At that time, Meta defined the term to mean “retarded Kikuyu.” Meta removed the comment and applied a standard strike against the user. The user appealed Meta’s decision to remove the comment, but a



moderator who reviewed the appeal upheld the decision. The posting user then appealed to the Oversight Board. In June 2025, following questions from the Board concerning the designation of the term, Meta decided to remove it from the slur list. As a result, the content was restored in July 2025 and the strike against the user was revoked. The user was notified of the content restoration and informed that this was the result of the Board reviewing their case.

The Board notes the following context in reaching its decision.

In Kenya's multi-party democracy, politicians and political parties often leverage ethnicity and inter-group allegiances to seek votes during elections. Of the individuals discussed in the post, Odinga garners significant support from the [Luo community](#). Current President [William Ruto](#) hails from the [Kalenjin community](#), Kenya's third largest ethnic group, while [Gachagua](#) is from the [Kikuyu community](#), the largest ethnic group. President Ruto and Gachagua teamed up and won the 2022 general elections, with Gachagua serving as deputy president from 2022 until his [impeachment](#) in 2024.

How politics and ethnicity intersect is often a subject of public discussion in Kenya and can be contentious. Advocacy of ethnic hatred inciting violence among communities, particularly during elections, has been a concern. The 2008 post-election violence, which [resulted](#) in 1,200 deaths and the displacement of hundreds of thousands of people, had multiple and complex drivers, but was in part [fueled](#) by hate speech and incitement on the basis of ethnicity. Similar incidents [occurred](#) after the 2017 general elections. The National Cohesion and Integration Commission (NCIC), Kenya's social cohesion agency, has [warned](#) against the use of hate speech on social media platforms, including from politicians such as Gachagua [hinting at](#) potential violence in the upcoming 2027 general elections. In September 2023, responding to speculation, the NCIC said it had [no plans](#) to ban the word "tugeges" as hate speech, but in June 2024 it [warned](#) against using the term to refer to people supporting a different political ideology, as the word is "hurtful, demeans and dehumanizes people." However, the term is not included in the NCIC's [hate speech lexicon](#).



In June 2024, youth-led protests [erupted](#) in Kenya after the government had proposed tax hikes. The protests eventually evolved into broader calls for a more accountable government, [transcending ethnic allegiances](#). In June 2025, youth-led protests broke out again, calling for President Ruto’s resignation and marking [continued dissatisfaction](#) with the Ruto government. In response, law enforcement authorities have [cracked down](#) on protesters, including opening fire on them, leading to [deaths](#), injuries and arrests. In July 2025, President Ruto [accused](#) the opposition that now includes Gachagua of planning to depose him, which Gachagua denied.

## **2. User Submissions**

In their statement to the Board, the user who posted the comment stated they did not insult, threaten or use abusive language. They said they were making a simple political comment “in a civil way” and that it was unfair for Meta to remove the post.

## **3. Meta’s Content Policies and Submissions**

### *1. Meta’s Content Policies*

#### Hateful Conduct Community Standard

Under the [Hateful Conduct policy](#), Meta removes content that “describes or negatively targets people with slurs.” Meta defines hateful conduct as “direct attacks against people – rather than concepts or institutions” – based on protected characteristics such as race, ethnicity or disability, among others. Slurs are “words that inherently create an atmosphere of exclusion and intimidation against people on the basis of a protected characteristic, often because these words are tied to historical discrimination, oppression and violence.”

As stated in the Hateful Conduct policy rationale, Meta believes that people “use their voice and connect more freely when they don’t feel attacked on the basis of who they are.” The policy rationale also specifies instances in which the use of slurs is allowed: to



condemn the slur, report on it, use it self-referentially or use it in an empowering way. Meta says it allows such uses where the speaker's intention is clear.

## *II. Meta's Submissions*

Meta initially removed the content for violating the Hateful Conduct Community Standard rule against slurs. The term “tugege” was designated as a slur at the time it was posted, and Meta found its use was not permitted under any of the exceptions.

Meta added the term “tugege” to its slur list for Swahili markets in January 2024 after the word gained traction during the 2022 Kenyan general elections. The designation of the word followed Meta's [designation process](#). According to qualitative and quantitative analysis the company conducted at the time, the term was understood to mean a “retarded Kikuyu” and is the plural form of “kagege,” a Kikuyu term to describe “a person who is extremely confused to the point of gaping vacantly at the world.” The term derives from the word “gega” which means “to stare in puzzlement.” Meta's regional and policy teams determined then that the word was used to attack Kikuyus based on their ethnicity.

When announcing its January 7, 2025 policy changes, Meta stated that Hateful Conduct policy violations would be treated as “less severe policy violations” and that the company would be reducing its reliance on automation for the enforcement of these violations. The company will instead rely on “user reports and other external signals” to complement classifiers in detecting potential violations. It may also still use classifiers to proactively remove violating content once detected without human review. In countries experiencing crises (whether or not designated under the company's Crisis Policy Protocol), Meta will continue proactively detecting and removing Hateful Conduct violations without relying on user reports or other external signals.

At the time the content was posted, Meta had deployed proactive detection and removal of potential Hateful Conduct policy violations in Kenya as part of the



company's broader integrity measures in anticipation of the 2027 national elections, as political campaigning had already begun. Meta pointed to Kenya's long history of ethnic tension around elections, as well as politics being heavily shaped by tribal affiliations, as reasons for this, though the company had not designated the country a crisis CPP for the African Union Chairperson election. In the [Australia Electoral Commission Voting Rules](#) decision, Meta informed the Board that its automated system can be configured to proactively detect content and either enqueue it for human review or automatically remove it.

In the present case, Meta used its automated system to identify and remove the case content for violating the Hateful Conduct policy. The automated system detects the use of Meta-designated slurs in Swahili markets as part of a targeted search initiative responding to heightened political tension and informed by Meta's internal assessment of off-platform risks. This automated system for detecting and removing violating content will continue to operate until the 2027 general elections, after which Meta will assess whether to stop or retain it.

In June 2025, while deliberating this case, the Board questioned Meta on the designation of the term "tugeges" as a slur, given expert and stakeholder input to the Board highlighting various colloquial meanings of the term in Kenyan political discourse indicating it is not always considered an attack based on ethnicity. This prompted further investigation from Meta, which led them to reevaluate the designation in consultation with its internal public policy teams and remove it from the slur list in July 2025. Meta found that the term is now also used "to target politicians based on their roles as political figures," not their protected characteristic. Meta confirmed that the case content was restored and the strike against the user removed on July 23, 2025, five months after the content was posted.

Meta conducts global audits of its slur list as part of standard practice to ensure that the meaning of designated terms remains current. The frequency of such audits depends on Meta's allocation of its resources including regional expertise. Outside of the audit, Meta's regional and content policy teams also review ad hoc requests to add or remove



terms as necessary. This year’s audit, which includes Swahili markets, started in July – after the term “tugeges” had been removed from the slur list as a result of this case – and is expected to be completed by September.

The Board asked questions about the designation and subsequent removal of the term “tugeges” from the slur list, measures to address potentially violating content, how the company audits its slur list and about proactive detection of potential hate speech as part of Meta’s electoral integrity efforts following enforcement changes announced in January. Meta responded to all questions.

#### **4. Public Comments**

The Oversight Board received four public comments that met [the terms for submission](#). Two of the comments were submitted from Sub-Saharan Africa and two from Central and South Asia. To read public comments submitted with consent to publish, click [here](#).

The submissions covered the following themes: the colloquial uses of the term “tugeges”; how automated enforcement does not consider nuances in language and context; and how Meta’s automated moderation of content for hate speech can lead to both underenforcement and overenforcement.

#### **5. Oversight Board Analysis**

The Board selected this case to help Meta reconcile its commitment to voice with the enforcement of the Hateful Conduct policy against slurs in countries with a history of ethnic tension. The Board analyzed Meta’s decision in this case against Meta’s content policies, values and human rights responsibilities. The Board also assessed the implications of this case for Meta’s broader approach to content governance.

##### **5.1 Compliance With Meta’s Content Policies**





The Board finds that the content in the case does not violate the Hateful Conduct policy, as the use of the term “tugeges” does not meet Meta’s definition of a “slur.”

The Board recognizes that when Meta removed the content the term was designated as a slur, and the company's automation and human reviewers removed the content in line with internal guidance. Given Kenya's history of ethnic tension, and the potential for further violence around elections, the company's attentiveness to removing hate speech was in line with its procedures and slur designations at the time. However, use of the designated term at issue quickly evolved, bringing into question the timeliness of Meta’s January 2024 designation. By the time this content was posted, and arguably in early 2024 when it was first designated, it should not have qualified as a slur.

In this case, the term “tugeges” was used to discuss the politics surrounding the African Union chairperson nomination and express political criticism on the matter. While the comment is negatively describing Gachagua’s supporters, the term does not inherently create an atmosphere of discriminatory exclusion and intimidation, and therefore is not a slur. Public comment submissions noted the importance of content moderation to be sensitive to the context and nuances of political discourse in Kenya to prevent over-enforcement (see PC-31266, PC-31269, PC-31270, PC-31272).

Experts the Board consulted confirmed that during the 2022 general elections, the term “tugeges” was used widely to describe Kikuyu voters perceived by some as voting without fully considering the consequences of their vote. While the term can be derogatory in relation to ethnicity in certain circumstances, the term has in recent years also developed a broader connotation to refer to individuals, regardless of ethnicity, who vote without a deep interrogation of their politics and political allegiances. Many use the term to criticize blind political loyalty or lack of discernment without meaning it as hate speech, even in situations where political loyalty and ethnicity are related. Experts pointed out that other Kenyan dialects have equivalents of the term, showing that the use of such words is typical to discuss politics or express political criticism in Kenya. The Board notes that, in June 2024, the NCIC warned against using the term in political discourse because it is “hurtful, demeans and dehumanizes people,” but



stopped short of treating the word as hate speech or suggesting its prohibition. According to Kenya’s National Coalition on Freedom of Expression and Content Moderation, the use of Kikuyu words is not atypical in political discourse and Meta’s failure to distinguish hate speech and protected political commentary can lead to freedom of expression being jeopardized (PC-31272). The Kenya-based non-profit, Global Initiative on Tech and Human Rights, stated that the use of the word in the case content is satirical “to critique political posturing and perceived gullibility among a specific political following,” instead of denigrating others based on ethnicity (PC-31266).

The Board notes that Meta now has a different approach to automated (what Meta calls “proactive”) removal of Hateful Conduct violations following Meta’s announcement in January 2025, except in countries experiencing a crisis where automated removal is still used. The Board recommended in previous decisions for Meta to develop a protocol to help the company address harms particular to a crisis or conflict (see e.g., Tigray Communication Affairs Bureau, Sudan Graphic Video, Former President Trump’s Suspension), as well as to improve integrity measures for periods even outside of elections (Brazilian General’s Speech). The Board has also recommended in the Pro-Navalny Protests in Russia decision that Meta explore measures promptly notifying users that a word or phrase in their post may violate Meta’s policy on negative character claims.

Given the context in Kenya, the Board understands why Meta chose to deploy automated detection and removal of hate speech far in advance of the national elections in 2027. It is important that Meta does not wait for a crisis to erupt (e.g., into violence) to take precautionary actions to ensure appropriate enforcement of Hateful Conduct policy violations. The Board also acknowledges Meta’s rationale in defaulting to proactive enforcement of the Hateful Conduct policy in countries where a crisis is ongoing. Situations of protracted armed conflict may merit a long-term crisis designation to deploy proactive removal of violating content in addition to proactive detection. However, the Board encourages Meta to be cautious about such approaches as these may lead to more opportunities for overly broad content takedowns. Caution



is also needed for automated enforcement more generally, as this can be overbroad and may lead to erroneous removal of speech, particularly in relation to elections, as what happened in the present case.

## **5.2 Compliance With Meta’s Human Rights Responsibilities**

The Board finds that removing the content from the platform was not consistent with Meta’s human rights responsibilities.

### *Freedom of Expression (Article 19 ICCPR)*

Article 19 of the International Covenant on Civil and Political Rights (ICCPR) provides for broad protection of expression, including views about politics, public affairs and human rights ([General Comment No. 34](#), paras. 11-12). It gives “particularly high” protection for “public debate concerning public figures in the political domain and public institutions” as an essential component for the conduct of public affairs ([General Comment No. 34](#), para. 38, 20; see also [General Comment No. 25](#), para. 12 and 25) and protects speech that may be considered “deeply offensive” (General Comment No. 34, (2011), para. 11).

When restrictions on expression are imposed by a state, they must meet the requirements of legality, legitimate aim, and necessity and proportionality (Article 19, para. 3, ICCPR). These requirements are often referred to as the “three-part test.” The Board uses this framework to interpret Meta’s human rights responsibilities in line with the United Nations (UN) Guiding Principles on Business and Human Rights, which Meta itself has committed to in its Corporate Human Rights Policy. The Board does this both in relation to the individual content decision under review and what this says about Meta’s broader approach to content governance. As the UN Special Rapporteur on freedom of expression has stated, although “companies do not have the obligations of governments, their impact is of a sort that requires them to assess the same kind of questions about protecting their users’ right to freedom of expression” ([A/74/486](#), para. 41).



### *I. Legality (Clarity and Accessibility of the Rules)*

The principle of legality requires rules limiting expression to be accessible and clear, formulated with sufficient precision to enable an individual to regulate their conduct accordingly (General Comment No. 34, para. 25). Additionally, these rules “may not confer unfettered discretion for the restriction of freedom of expression on those charged with [their] execution” and must “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” (ibid). The UN Special Rapporteur on freedom of expression has stated that when applied to private actors’ governance of online speech, rules should be clear and specific (A/HRC/38/35, para. 46). People using Meta’s platforms should be able to access and understand the rules and content reviewers should have clear guidance regarding their enforcement.

As applied to the content in this case, the Hateful Conduct Community Standard rule on slurs is sufficiently clear. The Board notes Meta’s [increased transparency](#) on its slur designation process as a direct result of the company implementing recommendations from the Board (see e.g., [Reclaiming Arabic Words](#), [Political Dispute ahead of Turkish Elections](#), [Criticism of EU Migration Policies and Immigrants](#)).

### *II. Legitimate Aim*

Any restriction on freedom of expression should pursue one or more of the legitimate aims listed in the ICCPR, which includes protecting the rights of others (Article 19, para. 3, ICCPR). The Board has previously recognized that the Hateful Conduct Community Standard pursues the legitimate aim of protecting the rights of others ([Posts Displaying South Africa’s Apartheid-Era Flag](#); Article 2, para. 1, ICCPR; Article 2 and 5 [International Convention on the Elimination of All Forms of Racial Discrimination](#)).

### *III. Necessity and Proportionality*



Under ICCPR Article 19(3), necessity and proportionality require 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 No. 34, para. 34). As part of their human rights responsibilities, social media companies should consider a range of possible responses to problematic content beyond deletion to ensure restrictions are narrowly tailored ([A/74/486](#), para. 51; see [Claimed COVID Cure](#)).

The Board finds that the removal of the comment from Facebook was not necessary as the term did not attack anyone on the basis of ethnicity but was squarely a political discussion. Removing the post did not serve the aim of protecting the rights of others, but rather led to an unnecessary restriction on political speech (see [Political Dispute Ahead of Turkish Elections](#)). Several stakeholders observed how ethnicity has been instrumentalized by politicians in the past to foment division and make political gains (see PC-31272 National Coalition on Freedom of Expression and Content Moderation, Kenya), at times leading to electoral violence. While Meta is right to be attentive to these potential harms, the term’s general use and its use in this case was not to intimidate, exclude or otherwise incite violence or discrimination.

While the use of terms such as “tugeges” may raise concerns when deployed as an attack based on ethnicity, especially if they are deployed by an influential politician, harms cannot be said to result from all its uses to justify its inclusion on a banned word list. This case demonstrates that Meta’s ban of the term was overbroad and led to suppressing political speech and public debate. Meta should be judicious with its designation process before adding terms to what is essentially a banned word list. The NCIC did not find the word to constitute hate speech, though it urged the public against its problematic uses. Expert research and public comments confirmed that the discriminatory ethnic undertones of the term are less fixed and subject to contestation. This should have been considered by Meta when it carried out its qualitative and quantitative analysis before designating the term in January 2024. This underscores the importance of Meta engaging stakeholders at the national level on which terms are



appropriate to designate as slurs and remaining vigilant that terms that previously were thought to meet its definition can change meaning over relatively short periods of time.

In line with the Board’s recommendation in the [Pro-Navalny Protests in Russia](#) decision, Meta should develop product interventions less intrusive than content removal to address the harm posed by hate speech and discriminatory content. However, when asked by the Board, the information provided by Meta on the efficacy of such interventions was still tentative. The Board encourages Meta to carefully continue exploring such measures, and ensure they are effective and would not lead to adverse human rights impacts.

## **6. The Oversight Board’s Decision**

The Board overturns Meta’s original decision to remove the content.

## **7. Recommendations**

### Enforcement

1. To ensure Meta’s current product interventions help users avoid Hateful Conduct policy violations, Meta should provide users with an opportunity for self-remediation comparable to the post time friction intervention that was created as a result of the Pro-Navalny Protests in Russia recommendation no. 6. If this intervention is no longer in effect, Meta should provide a comparable product intervention.

The Board will consider this recommendation implemented when Meta provides enforcement data that demonstrates the efficacy of these product interventions.



The Board also reiterates the following recommendation no. 3 of the [Criticism of EU Migration Policies and Immigration decision](#), which Meta fully committed to and is in the process of implementing:

When Meta audits its slur lists, it should ensure it carries out broad external engagement with relevant stakeholders. This should include consulting with impacted groups and civil society. The Board will consider this recommendation implemented when Meta amends its explanation of how it audits and updates its market-specific slur lists on its Transparency Center.

**\*Procedural Note:**

- The Oversight Board's decisions are made by panels of five Members and approved by a majority vote of the full Board. Board decisions do not necessarily represent the views of all Members.
- Under its [Charter](#), the Oversight Board may review appeals from users whose content Meta removed, appeals from users who reported content that Meta left up, and decisions that Meta refers to it (Charter Article 2, Section 1). The Board has binding authority to uphold or overturn Meta's content decisions (Charter Article 3, Section 5; Charter Article 4). The Board may issue non-binding recommendations that Meta is required to respond to (Charter Article 3, Section 4; Article 4). Where Meta commits to act on recommendations, the Board monitors their implementation.
- For this case decision, independent research was commissioned on behalf of the Board. The Board was assisted by Duco Advisors, an advisory firm focusing on the intersection of geopolitics, trust and safety, and technology. Linguistic expertise was provided by Lionbridge Technologies, LLC, whose specialists are fluent in more than 350 languages and work from 5,000 cities across the world.