



Poem About Political Protest in Argentina

2025-043-IG-UA

Summary

The Oversight Board has raised concerns about repeated overenforcement arising from Meta’s practices that involve Instagram carousels. In this case of political speech critical of Argentina’s government and its policies’ social effects, part of a carousel containing multiple text images to form a political poem was incorrectly removed by Meta. Only one of the images, containing slurs, was assessed by moderators and therefore vital context missed.

So that full context can be considered, the Board recommends Meta ensure that moderators reviewing carousels are able to see all content within a carousel post before making a decision. The Board overturns Meta’s original decision to remove the content.

About the Case

In January 2025, days before protests against Argentina’s President Javier Milei, who had given a speech criticizing “radical feminism” and the “LGBT agenda,” an Instagram user posted a text-only image carousel. The words on the eight images form a poem that is broadly critical of Argentina’s government and of people’s apathy during a period when, according to the user, policy changes are impacting vulnerable groups. The poem appeals to readers to protest.

On the carousel’s second image, the text includes two slurs, “puto” and “trava,” used to refer to gay men and trans women, respectively. A day after the carousel was posted, Meta’s automated systems detected the second image and sent it to a human moderator, who decided it broke the Hateful Conduct rules. The moderator could not see the other seven images in the carousel. Only the second image was removed from



Instagram, and a strike was applied against the user's account. The user appealed to Meta. A second reviewer upheld the decision. The user then appealed to the Board.

When the Board selected the case, Meta reversed its original decision, restoring the second image to the carousel.

Key Findings

The Board found the content does not violate the Hateful Conduct policy because it qualifies for exceptions, allowing the use of slurs “to condemn or raise awareness” and in an “empowering” way. Meta’s eventual decision to restore the content was supported by the interpretation that the text condemns Milei’s government by employing terms often used in anti-LGBTQIA+ rhetoric. This interpretation is in line with how the Board has previously told Meta to assess content and exceptions to its policies. However, the Board highlights an additional crucial point: the post is condemning people’s indifference in the face of social, political and economic changes that the posting user thinks negatively impact certain vulnerable groups.

The slurs used in the post did not negatively target a particular individual or group but were invoked to criticize the Argentine government’s policies. They were used to advocate against social indifference to government measures impacting groups, including LGBTQIA+ people. Ahead of the October 2025 elections in Argentina, Meta must be mindful of ensuring political speech, including reclaimed speech, is not unnecessarily removed.

The Board notes it was virtually impossible in this case for the reviewer to determine the slurs were used in a permissible way, without having access to the full carousel of text images making up the poem. The Board is concerned about the possibility of repeated overenforcement resulting from reviewers not having access to full carousels when making enforcement decisions about specific carousel images, as happened in this case, and not being empowered to effectively assess intent. Users’ freedom of



expression can be impacted when select content is removed from carousels in which speech unfolds over multiple images.

The Oversight Board's Decision

The Board overturns Meta's original decision to take down the content.

The Board recommends that Meta:

- Ensure that, when reviewing content within carousels, moderators are able to see all content within the post before making a decision, even when only one image is sent for human review.
- Develop an integrated process for ensuring that, when a content type is introduced or significantly updated, the company's procedures and tooling allow for its moderation in line with the company's human rights responsibilities.

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

Full Case Decision

1. Case Description and Background

In January 2025, an Instagram user in Argentina posted a carousel of eight Spanish text-only images that form a poem. It was shared a few days after the country's President Javier Milei made a [speech](#) at the World Economic Forum's annual meeting, in which he defended his economic policies and criticized "radical feminism" and the "LGBT agenda."

The poem is broadly critical of Argentina's government, its policies and what it describes as the "violence" they have created. It also criticizes political apathy in the face of social and economic policy changes that, according to the user, impact



vulnerable groups. It calls for protest and was posted shortly before thousands of people took to the streets of Argentina’s capital, Buenos Aires, to [demonstrate](#) in what was known as the [Federal March of Anti-Fascist and Anti-Racist Pride](#).

The poem uses two slurs, “puto” and “trava” (terms commonly used to target gay men and trans women, respectively), in making its critique. In the post’s second image, the poem speculates that the reader may not feel impacted by the political context because they are not a “*puto, trava, woman, retiree or a student.*” The fourth image also uses “puto” to speak about people who claim to have gay friends but do not protest on their behalf. It makes a similar point about parents with daughters who support abortion rights. While “puto” and “trava” are considered slurs in Latin American countries, including Argentina, the terms have also been reappropriated by LGBTQIA+ people and are used in self-identifying and empowering contexts. The poem then appeals to the reader, saying the author will protest for the rights of those people choosing to stay at home when “they come looking for you.” The post was liked about 1,000 times and the second image was viewed around 6,000 times. No users reported it.

One day after the content was posted, Meta’s automated systems identified the second image with the two slurs as potentially violating and sent it for at-scale review by a human moderator. The fourth image with one slur was not detected by Meta’s automated systems, nor sent for at-scale review. Only the identified image with two slurs – rather than all the images in the carousel – was visible to the reviewer, who determined it violated Meta’s [Hateful Conduct](#) policy. As a result, Meta removed this image, leaving the rest of the carousel visible, and applied a standard strike against the user. On the same day, the user appealed Meta’s decision, and a second reviewer upheld the original decision.

The user who posted the content then appealed Meta’s decision to the Oversight Board. When the Board selected this case, Meta’s policy experts reviewed the post again, with access to all images. The company reversed its original decision, restored the image to the carousel and removed the strike on the user’s account.



The Board notes the following context in reaching its decision:

In 2024, President Milei enacted [economic reforms](#) in response to Argentina's prolonged economic crisis. Additionally, President Milei issued several [executive decrees](#). Civil society groups have [condemned](#) the decrees, saying they adversely impact the rights of LGBTQIA+ people. These groups have also [raised concerns](#) about anti-LGBTQIA+ rhetoric by government officials. The government's economic and social policies have sparked several [national demonstrations](#) by students, pensioners, civil society organizations, labor unions and opposition parties.

The government introduced new [guidelines](#) for controlling street demonstrations in the same month it assumed power, in December 2023, including measures that civil society groups [warn](#) may discourage or even criminalize protests. Since then, the Inter-American Commission on Human Rights (IACHR) has [expressed concern](#) over reports of alleged excessive use of force by state security forces during protests. The IACHR has [noted](#) Argentina's "strong tradition of civic engagement" and called on Argentina to "uphold the rights to freedom of expression and peaceful assembly." Several organizations, including [PEN International](#), [Amnesty International](#) and the [Argentine Journalism Forum](#), have tracked a decline of freedom of expression in Argentina since the Milei government took office.

2. User Submissions

In their appeal to the Board, the user who posted the content said the post "used artistic and thought-provoking language to engage audiences in a meaningful way, without promoting hate, discrimination or violence." They said that the post was "shared with the intention of fostering understanding and inspiring conversations around collective accountability and human rights." About the form of the post, they said, "This type of expression is rooted in cultural traditions of storytelling and critique, which are essential for building empathy and community."



3. Meta's Content Policies and Submissions

I. Meta's Content Policies

Under its [Hateful Conduct](#) policy, Meta removes content that “describes or negatively targets people with slurs.” Slurs are defined as “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.”

In the policy rationale, Meta recognizes that there are instances where slurs are used to “condemn the speech or report on it.” Meta also acknowledges that there are cases in which “speech, including slurs, that might otherwise violate our standards is used self-referentially or in an empowering way.” In its [January 7 update to the Hateful Conduct policy](#), Meta made clear that slurs qualify for these exceptions only when the user's intent is clear.

Internal guidelines for reviewers further define the exception for condemnation. Condemnation is described as “denouncing or challenging the use of the slur or hateful conduct,” which can include expressing disbelief, criticizing and exposing, and rejecting the use of the slur or hateful conduct.

II. Meta's Submissions

Meta has designated both “puto” and “trava” as slurs in its internal guidance. However, when the Board selected the case, Meta reversed its original decision to remove the carousel image, concluding that it did not violate the [Hateful Conduct](#) policy because the slurs are used in a “condemning context.” Meta said: “While the post uses slurs, they are not being used to target a person or group of people based on their protected characteristics in a hateful way. Instead, the slurs are being used artistically to condemn Milei's government by using terms often used in anti-LGBTQ+ rhetoric. In other words, the post condemns the same discrimination that the slur use also references.”



When the Board asked how Meta determined that the slurs were being used in a “condemning context,” Meta said the content “holistically condemn[ed] the perceived anti-LGBTQ+ rhetoric of Milei’s government ... but in a non-hateful and critical way.” Additionally, “while the slurs themselves are not the explicit objects of the author’s condemnation, the slurs are being used as an example of how the LGBTQ+ community can be attacked” and the “author is condemning the generalized use of these types of words and hateful conduct more broadly.”

Meta has previously told the Board that its policies generally do not grant at-scale reviewers discretion to determine user intent. According to the company, to ensure consistent and fair enforcement of its rules, it does not require at-scale reviewers to “infer intent or guess at what someone ‘is really saying’” because “divining intent for hate speech invites subjectivity, bias, and inequitable enforcement” (see [Violence Against Women](#)). When the Board asked Meta to clarify this position, considering the new focus on user intent in the slur exception section of the Hateful Conduct policy, the company responded that it does not see a discrepancy. According to Meta, the “policy strikes a balance where we do not ask reviewers to infer a speaker’s intent but may consider that intent when it is clear on the face of the content, to ensure we are allowing permissible speech.” In this case, “it is evident from the text itself [...] that the author’s intent is to discuss and challenge those who voted for the current government, which the author feels has committed violence.”

When asked how the company ensures that reviewers’ tooling is responsive to new content types, Meta said that all teams conduct a risk assessment to ensure that different content types (like Instagram carousels, stories, or reels) integrate into the company’s content moderation tools when launched, and that new tooling can be built or adapted to accommodate the specific requirements of these content types.

The Board asked questions on: exceptions to the Hateful Conduct policy’s prohibition on slurs; enforcement practices around Instagram carousel posts; and human rights due diligence undertaken to mitigate risks related to the launch of different content



types, such as images, videos and albums of images like carousels. Meta responded to all questions. However, Meta refused to let the Board publish information related to the company’s enforcement practices around Instagram carousel posts, specifically relating to the Board’s questions on when reviewers have access to the full carousel of images.

4. Public Comments

The Board received two public comments that met [the terms for submission](#). One of the comments was submitted from Latin America and the Caribbean and one from Asia Pacific and Oceania. To read public comments submitted with consent to publish click [here](#).

The submissions covered the following themes: the meaning and uses of the terms “puto” and “trava” in context; reclamation and reappropriation of slurs; the impact of Meta’s renamed Hateful Conduct policy on LGBTQIA+ people; and approaches to moderating artistic expression that includes political speech.

5. Oversight Board Analysis

The Board selected this case to address the recurring issue of overenforcement of Meta’s Community Standards to remove political speech. This is particularly important considering the legislative elections in Argentina scheduled for October 26, 2025. The Board also selected this case to examine how Meta moderates multi-part content types, like Instagram carousels, and the impact of such enforcement practices on freedom of expression. The case falls within the Board’s [strategic priorities](#) of Elections and Civic Space and Hate Speech Against Marginalized Groups.

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

I. Content Rules

The Board finds that, while two slurs are used, the post does not violate the Hateful Conduct policy. As the user noted in their statement to the Board, the slurs are used as “thought-provoking language” and do not negatively target a person or a group based on their protected characteristics. The Board finds that the slurs are used to typify people who consider themselves LGBTQIA+ allies — and thus feel comfortable using slurs in a reclaimed manner — but are unwilling to protest on behalf of LGBTQIA+ people. When read in its entirety, the post broadly condemns indifference to recent government measures impacting vulnerable groups, including LGBTQIA+ people and calls for action on their behalf in the form of protest. Reclaiming and reappropriating slurs for empowering political expression has a history among LGBTQIA+ activists in Argentina (see PC-31290), especially with respect to one of the terms (“trava”) at issue in this case. Therefore, the post qualifies for the exceptions to “condemn or raise awareness” and “empowering” use.

The Board notes that the post does not fit neatly into any one of these exceptions, instead, it touches on several of them (see [Reclaiming Arabic Words](#) for a similar approach). This is because the post does not specifically condemn the use of the slurs referenced in the post, nor does it explicitly condemn “hateful conduct,” as Meta’s internal guidelines on condemnation require. Rather, the post requires a broader understanding of the condemnation of the political context. The Board finds that Meta’s eventual conclusion that the post condemns “Milei’s government by using terms often used in anti-LGBTQ+ rhetoric” is a possible interpretation of the content. The Board also acknowledges that Meta’s revised approach to assessing the content aligns more closely with how the Board has previously suggested the company evaluate content (i.e., non-literally and assessing the content as a whole and in context). However, the Board highlights as a crucial point to the understanding of the post that it also



condemns people's indifference in the face of social, political and economic changes that the user thinks negatively impact certain vulnerable groups, and calls on people to protest.

However, the Board is unconvinced that Meta's approach to moderating carousel posts in this case would have allowed for at-scale reviewers to accurately interpret the post and reach the correct enforcement decision. In being shown only the potentially violating image from the carousel post, the reviewers were unable to evaluate the content in its entirety. Moreover, internal guidance prevents reviewers from considering context within and outside the post to determine intent and meaning. These are distinct issues that are discussed, respectively, in the "Enforcement Action" and "Legality" sections below. While each would have produced an incorrect enforcement outcome, the Board notes that in this case they compounded each other.

II. Enforcement Action

This case raises serious concerns about potential overenforcement stemming from Meta's enforcement of Instagram carousels, where the meaning of a post is meant to unfold over multiple images.

The Board has repeatedly pushed Meta for content to be reviewed as a whole at-scale, rather than making assessments based on isolated parts of the content (see, among others, [Wampum Belt](#)). In [Images of Partially Nude Indigenous Women](#), the Board found that, similar to this case, Meta's reviewers only had access to one image in a carousel post and should have considered additional context from other photos in the carousel to determine that content qualified for an allowance to remain on Instagram. The enforcement actions in this case, where only the identified image with two slurs was visible to the reviewer, meant that, in practice, the reviewer could not assess the carousel post as a whole. The Board finds that it was virtually impossible for a reviewer to determine that the slurs in this case were used in an allowable context without access to the full carousel of images, given that together they formed a poem, which could only be understood in its entirety.



The Board notes that Instagram carousel posts were [launched](#) over eight years ago. The Board is concerned about repeated overenforcement resulting from errors like the one seen in this case, where reviewers did not have access to the full carousel when making enforcement decisions about a specific carousel image. It is also concerned about potential impacts to users' freedom of expression resulting from removing select images from carousel posts. Previously, in [Reclaiming Arabic Words](#), the Board considered a carousel post that was largely informational and presented as a series of distinct statements. In that case, the Board found that removing an entire carousel would not be a proportionate response "even if the carousel had included one image with impermissible slurs not covered by an exception." On the other hand, the content in this case demonstrates how carousel posts can also allow for users' speech to unfold over multiple images in a narrative manner. While removing select images could result in a lesser restriction of expression (see [Reclaiming Arabic Words](#)), it could also change or distort a user's intended meaning, which is exacerbated if it is not clear to viewers that an image has been removed.

Meta's enforcement practices should be responsive to the different modes of expression enabled by carousel posts. Similar to what the Board has previously recommended (see [Pro-Navalny Protests in Russia](#), recommendation 6), one way Meta could do this is by notifying the user that part of their content has been removed (in this case, part of a carousel post). Meta could then give them the ability to amend or delete the entire post if their intended meaning has been changed. The Board has also previously recommended that Meta allow users to indicate in their appeal submissions that their content falls into one of the exceptions to the Hateful Conduct policy (see ["Two Buttons" Meme](#), recommendation 4).

In this case, the Board recommends that Meta develop a process for periodic review of its enforcement practices and content moderation tools to ensure that they integrate with different content types, such as images, carousels of images and videos. The first phase of this process could incorporate the risk assessments the company informed the Board it already undertakes prior to the launch of new content types, but it should



specifically address how Meta plans to meet its human rights responsibilities when moderating distinct content types that enable users to express themselves in different ways. The post-launch periodic review should assess how users are actually using content types and consider how Meta is meeting its human rights responsibilities in practice through continued live testing around moderation of new content types, problem identification and mitigation.

The Board also expects that in parallel with the release of new content types, Meta will develop the necessary automated detection tools for the enforcement of its policies, integrating human rights standards in their design and implementation. The Board notes that Meta has publicly [acknowledged](#) leveraging large-language models for some enforcement-related tasks, like removing potentially non-violating content from review queues in certain circumstances. It is important that these tools have the capacity to prevent overenforcement by reviewing signals from posts as a whole, rather than looking for isolated potential violations in parts of the post that are not violative when viewed holistically.

Finally, to reduce the likelihood of errors like the one in this case, the Board also recommends that Meta allow reviewers to see full carousels when only one image is sent for review. The Board recognizes that there are several trade-offs involved in making a product decision of this sort. For example, expanding reviewer access to carousel posts could increase the amount of time reviewers spend reviewing posts. However, as the Board discussed in the [Cambodian Prime Minister](#) decision, Meta should implement product features and operational guidelines that allow for more accurate review of long-form content. In this case, Meta should ensure that human reviewers are able to see all components of a post, including all carousel images, when they deem it necessary.



5.2 Compliance With Meta’s Human Rights Responsibilities

The Board finds that keeping the content up on the platform, as required by a proper interpretation of Meta’s content policies, is also 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, as well as cultural and artistic expression ([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). This extends to expression that may be considered “deeply offensive” (General Comment No. 34, (2011), para. 11; see also para. 17 of the 2019 report of the United Nations (UN) Special Rapporteur on freedom of expression, [A/74/486](#) and [Posts Displaying South Africa’s Apartheid-Era Flag](#)). This post also relates to discrimination against LGBTQIA+ people. The UN High Commissioner for Human Rights has noted concerns over discriminatory limitations on advocacy for rights of LGBTQIA+ persons leading to restrictions on the freedom of expression ([A/HRC/19/41](#), para. 65).

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 [UN Guiding Principles on Business and Human Rights](#), which Meta 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). Here, the Board finds that Meta’s initial decision to remove the content under its policies did not meet these requirements.

I. Legality (Clarity and Accessibility of the Rules)

The principle of legality requires rules limiting expression to be accessible and clear. Rules should be formulated with sufficient precision to enable individuals to regulate their conduct accordingly (General Comment No. 34, para. 25). Additionally, these rules 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.

The Board finds that Meta’s prohibition on slurs and public-facing exceptions on allowable uses of slurs (condemning, reporting, self-referential and empowering) are sufficiently clear as applied in this case.

As the Board has noted in previous decisions, Meta could improve its at-scale enforcement of slur exceptions by providing clearer guidance for reviewers to evaluate content in its entirety and to consider local context when determining intent and meaning (see, for example, [Wampum Belt](#)). Meta policy experts relied on a series of contextual cues about the user’s intent and contemporary Argentinian politics to determine that the slurs in the second image were non-violating, after the Board identified the content. As the Board has discussed in multiple cases (see, for example, [Violence Against Women](#)), current reviewer guidance limits the possibility of this kind



of contextual analysis significantly, even when there are clear cues that the content may engage an exception.

While Meta told the Board that it does not ask reviewers to infer speaker intent, it said that reviewers may consider intent “when it is clear on the face of the content.” In the example provided to the Board, it said reviewers are empowered to conclude the speaker explicitly intended to condemn the use of a slur if the content states “I denounce the use of [slur].” While reviewers may be empowered to determine intent in cases like that, statements of intent can take far less formulaic forms (see [Nazi Quote](#)). Based on these factors, the Board is not convinced that reviewers are adequately authorized to consider intent in practice. In the poem in this case, for example, user intent is clear but is articulated implicitly with artistic language and references to contemporary Argentinian politics. Moreover, in the [Wampum Belt](#) and [“Two Buttons” Meme](#) decisions, the Board said it is not necessary for a user to explicitly state their intent for it to meet the requirements of an exception to the Hate Speech policy. In those cases, the Board said it is enough for a user to make intent clear in the context of the whole post. Allowing users to communicate intent in this way is in line with Meta’s human rights responsibilities regarding expression, as well as its own values. When Meta employs human reviewers, the company should rely on their agency and capacity for interpretation to make inferences from content to assess user intent.

As the Board has noted in previous decisions, it is indispensable that reviewers possess adequate linguistic and local contextual knowledge so they can accurately assess meaning and intent. Meta has [previously](#) told the Board that reviewers are “designated to their market based on their linguistic aptitude and cultural and market knowledge.” These factors are particularly important as, following the January 7 Hateful Conduct policy changes, Meta now relies on clear user intent to determine when content qualifies for a policy exception on slurs. The Board has previously recommended that Meta update its internal guidance to show what indicators it provides to reviewers to grant exceptions when considering content that may otherwise be removed under the Hate Speech policy (see [Violence Against Women](#), recommendation 2). Further empowering reviewers to truly “consider intent” for such exceptions requires Meta to



recognize that condemnation can be context dependent. Meta should implement internal guidance that allows reviewers to exercise judgment about such context when assessing content.

II. Legitimate Aim

Any state restriction on freedom of expression should pursue one or more of the legitimate aims listed in the [ICCPR](#), which include protecting the “rights of others.” The Board has previously recognized that the Hate Speech, and now the Hateful Conduct Community Standard, pursues the legitimate aim of protecting the rights of others. In the [Reclaiming Arabic Words](#) decision, for example, the Board found the Hate Speech policy pursued the legitimate aim of protecting the rights of others to equality, and protection against violence and discrimination based on sexual orientation and gender identity.

III. Necessity and Proportionality

Under ICCPR Article 19(3), necessity and proportionality requires that state 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).

The Board has previously recognized the potential for harms to the rights of LGBTQIA+ people from allowing homophobic slurs and hate speech to remain on Meta’s platforms, and the role the Hateful Conduct policy plays in respecting those rights (see, for example, [Post in Polish Targeting Trans People](#), [Colombia Protests](#), [Reclaiming Arabic Words](#)). The Board sought expert linguistic input and public comments that confirmed that the terms used in this case are considered slurs in Latin American countries, including Argentina. However, they can be used in reclaimed ways as markers of identity and “expression[s] of resistance” (PC-31290).



In this post, the slurs did not negatively target a particular individual or group but instead were invoked to criticize the Argentine government’s economic and social policies. The post pushes readers to consider their own position in Argentina’s political context and engage in protest. Argentines regularly [use social media to mobilize protests](#) on political and social issues. Here, the slurs were used to advocate against social indifference to recent measures impacting vulnerable groups, including LGBTQIA+ people. Based on these factors, the Board finds that removal of the content was not necessary to protect LGBTQIA+ people from discrimination.

As part of its preparations for the October 2025 legislative elections in Argentina, Meta should ensure that political speech that uses the slurs considered in this case and others like them in allowable contexts are not unnecessarily removed. This requires Meta to rely on the broader understanding of condemnation of the political context discussed above. The public comment by digital rights organization Derechos Digitales and civil society organization Conectando Derechos (PC-31290) notes that these slurs are “pillars of political identity” for LGBTQIA+ people and activists, and their use in the poem aims to “amplify and make visible the voices of groups directly affected by the Argentine government’s measures.”

Public comments received in this case advocate for considering the case content in light of Meta’s January 7, 2025, Hateful Conduct policy changes (see PC-31290 and PC-31289). In announcing those changes, Meta stated that it aimed to make fewer mistakes and remove less non-violating content in error. The Board has previously discussed the potential for disproportionate errors in the moderation of reclaimed or reappropriated speech by queer communities and noted the [adverse impact](#) of mistaken removals (see [Reclaiming Arabic Words](#) and [Reclaimed Term in Drag Performance](#)). In these cases, the Board [affirmed](#) that overenforcement on this type of speech is a “serious threat to their freedom of expression.”

In this case, the Board reiterates its [call](#) for Meta to be “particularly sensitive to the possibility of wrongful removal” of reclaimed speech, “given the importance of [such speech] for LGBTQIA+ people in countering discrimination.” Sexual orientation and



gender identity remain protected characteristics in the Hateful Conduct policy, and the exceptions to the slur policy offer protection for expression that reclaims and reappropriates slurs to advocate for the rights of LGBTQIA+ persons. Meta should ensure that its enforcement practices make that protection effective.

6. The Oversight Board's Decision

The Board overturns Meta's original decision to take down the content.

7. Recommendations

Enforcement

1. So that the full context of a post can be considered during review, Meta should ensure that, when reviewing content within carousels or multiple-image content types, moderators are able to see all content within the post before making a decision, even when only one image is sent for human review.

The Board will consider this recommendation implemented when Meta shares internal documentation with the Board detailing these changes in the moderator interface.

2. Meta should develop an integrated process for ensuring that, when a content type is introduced or significantly updated, the company's procedures and tooling allow for moderation in line with the company's human rights responsibilities. This process should include:
 - a. A pre-launch period where enforcement policies, operational guidelines and reviewer product decisions are set up, tested and red-teamed (i.e., proactively seeking vulnerabilities) by cross-functional teams following a pre-determined methodology.



- b. A time-bound post-launch period involving periodic live testing, problem identification and mitigation that specifically addresses the different modes of expression enabled by the content type.

The Board will consider this recommendation implemented when Meta shares internal documentation with the Board detailing this process and alerts the Board each time it is activated with an asynchronous update.

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. 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.