C.A.S.C.A

COMMENTS TO META OVERSIGHT BOARD

How Meta Must Respect Legitimate Speech in Argentina

July, 2025

Comments to Meta Oversight Board

How Meta Must Respect Legitimate Speech in Civic Sphere

Authors: Prakhar Singh, Chirkankshit Bulani, R Dayasakhti, Raima Singh

Research Consultant: Dr. Ivneet Walia, Associate Professor of Law and Officiating Registrar, RGNUL

CENTRE FOR ADVANCED STUDIES IN CYBER LAW AND ARTIFICIAL INTELLIGENCE [CASCA] is a research-driven centre at RGNUL dedicated to advancing scholarly research and discourse in the field of Technology Law and Regulation. As a research centre of a leading institution in India, we are committed to promoting interdisciplinary research, fostering collaboration, and driving innovation in the fields of cyber law, artificial intelligence, and other allied areas.

For more information

Visit **cascargnul.com**

Disclaimer

OVERSIGHT BOARD'S CALL FOR PUBLIC COMMENTS

(Released June 17, 2025)

In January 2025, an Instagram user in Argentina posted a Spanish-language poem in a text-image carousel criticizing government oppression and urging protest. Timed alongside demonstrations against President Javier Milei's speech at the World Economic Forum—where he denounced "radical feminism" and the "LGBT agenda"—the poem addressed marginalized groups, using reclaimed slurs like "puto" and "trava" (referring to gay men and trans women). The post received limited engagement, with around 1,000 likes and 6,000 views on the second image, and no user reports. Meta's automated systems flagged the second image for violating its Hateful Conduct policy, and a human reviewer, seeing only that image, removed it and applied a strike. After the user appealed, the decision was upheld. In their defense, the user argued the post was artistic and intended to promote human rights and solidarity. Upon further review—after the Oversight Board took up the case—Meta reversed its decision, reinstated the content, and removed the strike, acknowledging the slurs were used not to attack but to highlight discrimination. The Board selected the case to underscore the ongoing challenge of moderating political expression and to reinforce protections for civic discourse.

Public comments were solicited to provide insights on:

(i) The appropriate framework for Meta's content moderation policies in relation to slurs used within political discourse or artistic expression, particularly where such usage may convey solidarity or critique rather than incite hatred.

(ii) The semantic and socio-cultural significance of the terms "puto" and "trava" in the present context and in broader Latin American discourse, along with the potential social and psychological effects stemming from their usage.

(iii) The risks posed by hate speech directed at LGBTQIA+ individuals on digital platforms, and the essential role these platforms play in fostering visibility, community engagement, and advocacy for LGBTQIA+ rights in Argentina.

The comments contributed by CASCA strive to help Meta strengthen its approach embracing linguistic sensitivity, uphold user trust, and ensure that its platform does not become a weapon for suppression of legitimate speech. The original Call for Public Comments can be accessed here.

Introduction

In the complex and continually shifting landscape of social media—where algorithmic filters intersect with human adjudication—a single Instagram publication from Argentina has come to epitomize both the potential and the pitfalls of large-scale content moderation. Meta now dominates the global social media landscape with almost 3.43 billion active users,[1] It is imperative that the content moderation is done properly for the world. While Meta's "Free Expression" regime focusses on lower content removal,[2] thus leading to almost one-third reduction since January 2025 in the content post the new regime.[3]

In January 2025, an anonymous individual posted a sequence of text-only images presenting a concise, Spanish-language poem that scrutinized the government's treatment of marginalized communities and urged civic engagement.[4] The poem's deliberate use of potent terminology sought to illuminate the injustices endured by gay men, transgender women, students, retirees, and other socially excluded groups. Our submission will emphasize that Meta's moderation systems, encompassing both automated processes and human review, must develop a more sophisticated understanding of linguistic, cultural, and political context. A rigid, decontextualized application of policies risks suppressing legitimate and vital forms of expression essential for robust public discourse and social commentary.

Contextual Nuances in Regional Linguistic Expression

The yardstick of political and artistic speech, particularly when such speech is exercised in reappropriated terms must be studied in its socio-linguistic and cultural contexts. A prime example of this is the poem in the case at hand, where the terms 'Puto' and "trava', which are historically slurs against gay men and trans individuals[5]. The poem concerned was removed by Meta, but was reinstated after a human review. This case highlights the concerns with moderation with contextualization, and establishes the importance of speaker intent, community affiliation and understanding by the audience.

Reappropriation is defined as a cultural process by which a group takes back words or artifacts that were previously used in a disparaging manner against them, thereby initiating a specific form of semantic change.[6] In

---

[1] Dixon SJ, 'Meta Global Family Dau 2025' (Statista, 15 May 2025) <https://www.statista.com/statistics/1092227/facebook-product-dau/> accessed 1 July 2025.

[2] Dave P, 'Meta's "free Expression" Push Results in Far Fewer Content Takedowns' (Wired, 29 May 2025) <https://www.wired.com/story/meta-content-moderation-changes-decrease-removals/> accessed 1 July 2025.

[3] 'Integrity Reports, First Quarter 2025' (Transparency Center) <https://transparency.meta.com/hi-in/integrity-reports-q1-2025/> accessed 1 July 2025.

[4] 'Oversight Board to Assess Political Protest in Argentina' (The Oversight Board, 17 June 2025) <https://www.oversightboard.com/news/board-to-examine-removals-of-legitimate-speech-in-argentinian-political-poem-case/?_hsmi=367024897> accessed 1 July 2025

[5] G Olivera and M Ortega-Breña, 'Reframing Identities in Argentine Documentary Cinema: The Emergence of LGBT People as Political Subjects in Rosa Patria (Loza, 2008–2009) and Putos Peronistas (Cesatti, 2011)' (2021) 48(2) Latin American Perspectives 155.

[6] R Brontsema, 'A Queer Revolution: Reconceptualizing the Debate Over Linguistic Reclamation' (2004) 17(1) Colorado Research in Linguistics https://doi.org/10.25810/dky3-zq57 accessed 24 June 2025.

context of Argentina, it stands as a deeply political act, where the terms such as *"puto"* and *"trava"* stand as visibility and cultural assertion. The use of these terms in protest poetry, art and cinema is to demand political inclusion and create a change in interpretation of the term from a negative to a positive one.[7]

Reappropriation functions through a multitude of ways, including value reversal of the term, neutralization, or stigma exploitation (retention of a term to acknowledge injustice).[8] The crucial distinction in reappropriation is not often use of the term, but who uses it. An in-group usage vs usage by outsiders can make the term empowering or offensive. In the present case, use of the terms in the poem were done in the context of speaking in favour of marginalised voices.

This particular phenomenon is also widespread. In the US, the term *queer"* has followed the trajectory from slur to identity marker and academic term.[9] In India, the term *"Dalit"* has been reclaimed by the oppressed classes to fight against centuries of discrimination.[10] In another case being considered by Meta, the use of the term 'tugeges' for Kikuyu people in Kenya is also another example of how context can affect speech.[11] Such 'reappropriation' of terms has been prominent in Argentina, along with the rest of Latin America as well. The term '*trava*', which Meta has identified to be a slur in the present instance, has also been a part of similar linguistic reappropriations. In the past, the term '*trava'* was a derogatory way of referring to the trans-women community. Now, there has been a significant shift in the usage of the term as the trans-women community of Argentina choose to identify as 'trava' as a way of reclaiming language and fighting back the oppression they face.[12]

Meta's reversal marks progress in acknowledging marginalized voices. However, treating all speech uniformly ignores the cultural and political significance of terms like "puto" and "trava" in Argentina, which serve as expressions of resistance.

Protecting Marginalized Voices in Digital Spaces

Argentina had once spearheaded action to promote the rights of the LGBTQIA+ community by becoming the first Latin American country to legalise same-sex marriages in 2010. It subsequently strengthened the community's rights by passing the Gender Identity Law in 2012, which permitted people to change their gender

---

[7] G Olivera and M Ortega-Breña, 'Reframing Identities in Argentine Documentary Cinema: The Emergence of LGBT People as Political Subjects in Rosa Patria (Loza, 2008–2009) and Putos Peronistas (Cesatti, 2011)' (2021) 48(2) Latin American Perspectives 155.

[8] C Groom and others, 'The reappropriation of stigmatizing labels: implications for social identity' in *Identity Issues in Groups, Research on Managing Groups and Teams* vol 5 (Emerald Group Publishing Limited 2003) 221 https://doi.org/10.1016/s1534-0856(02)05009-0.

[9] 'Reclaiming Slurs in Popular Culture' (Number Analytics, 27 May 2025) https://www.numberanalytics.com/blog/reclaiming-slurs-in-popular-culture accessed 24 June 2025.

[10] 'Dalits in India' (Minority Rights Group, undated) https://minorityrights.org/communities/dalits/ accessed 24 June 2025.

[11] Oversight Board, 'Board to Consider How Meta Should Respect Political Expression in Kenya' (Oversight Board, undated) https://www.oversightboard.com/news/board-to-consider-how-meta-should-respect-political-expression-in-kenya/ accessed 24 June 2025

[12] Joaquim Renato Alves de Souza, 'Learning About Trans Technologies and Work in Brazil and Argentina' (Digilabour, 12 March 2025) <https://digilabour.com.br/learning-about-trans-technologies-and-work-in-brazil-and-argentina/amp/> accessed 1 July 2025.

on official documents based on self-determination without undergoing surgeries, therapy, or other invasive or bureaucratic requirements and ensured that transgender people received comprehensive medical care.[13] Such changes for progress were challenged in 2023 when Mr. Javier Milei was elected to be President of Argentina. His presidency raised and realised the concerns of the community, beginning with his very first speech in which he hinted at reducing "woke" action.[14] Subsequently, his government repealed the Femicide Laws[15] and also dismantled the Ministry for Women, Gender and Diversity.[16] These, along with several other actions and policy changes by the government has led to a rise in violence, hate speech and hate crimes against women as well as the LGBTQIA+ community. The rise in animosity was reflected when the National Observatory of LGBT Hate Crimes registered 133 hate crimes in Argentina in 2023, immediately post the election of President Javier Milei.[17] Adding on to this, last year, 4 lesbian women were brutally set on fire in Buenos Aires, where the government explicitly and arbitrarily dismissed the notion that it was a hate crime.[18]

With the government reflecting homophobic sentiments, similar feelings have been aggravated in sections of society which do not belong to the community. Discriminatory and harmful practices make the community members feel unsafe and unable to openly participate in the societies they inhabit, preventing them from engaging in civil and political life.[19] Such a status quo necessitates digital activism for two important reasons: (i) visibility and representation; (ii) safe space and anonymity. The online platform enables digital activists to disseminate their message while also protecting their identity. In the past, Argentina has seen successful digital activism through the #NiUnaMenos.[20] This was a social media campaign which raised awareness about violence against women, mobilised protests against femicide and subsequently spread to other countries in Latin America. The region has seen similar success through #MeuPrimeiroAssedio, focused on street harassment, #Restencia and thus, establishing the importance and effectiveness of digital activism in the region.[21]

---

[13] Alberto de Belaunde, 'Argentina at the Ballot Box: The Uncertain Future of LGBTQ Equality' (Outright International, 13 November 2023) <https://outrightinternational.org/insights/argentina-ballot-box-uncertain-future-lgbtq-equality> accessed 26 June 2025.

[14] Kate Fitz-Gibbon, 'Argentina's president is vowing to repeal 'woke' femicide law.' (The Conversation, 29 January 2025) <https://theconversation.com/argentinas-president-is-vowing-to-repeal-woke-femicide-law-it-could-have-ripple-effects-across-latin-america-248435> accessed 26 June 2025.

[15] Noël James, 'Argentina femicide women's rights law' (The Guardian, 29 January 2025) <https://www.theguardian.com/world/2025/jan/29/argentina-femicide-womens-rights-law> accessed 26 June 2025.

[16] Noël James, 'Women This Week: Milei Administration Dissolves Argentina's Ministry of Women' (Council on Foreign Relations, 14 June 2024) <https://www.cfr.org/blog/women-week-milei-administration-dissolves-argentinas-ministry-women> accessed 26 June 2025.

[17] Amnesty International, 'Argentina: Ongoing Criminalisation Against LGBT Activist'<https://www.amnesty.org.uk/urgent-actions/argentina-ongoing-criminalisation-against-lgbt-activist> accessed 26 June 2025.

[18] Erin Kilbride, 'Lesbian Women Set on Fire in Argentina' (Human Rights Watch, 14 May 2024) <https://www.hrw.org/news/2024/05/14/lesbian-women-set-fire-argentina> accessed 26 June 2025.

[19] United Nations High Commissioner for Human Rights, *Youth and Human Rights* (A/HRC/39/33, 2018) Para 33.

[20] Vaishnavi Pallapothu, 'Ni Una Menos Argentina' (The Gender Security Project, 10 March 2021) <https://www.gendersecurityproject.com/subversion-diaries/ni-una-menos-argentina> accessed 26 June 2025.

[21] Sarah Lee, 'Digital Activism in Latin America: Empowering Social Change through Technology and Online Movements' (Number Analytics, 18 June 2025) <https://www.numberanalytics.com/blog/digital-activism-latin-america#google_vignette> accessed 26 June 2025.

Addressing Procedural Weaknesses and Guiding Future Feature Development

The present case sheds light upon the issues associated with AI based content moderation, which must be adequately supplemented by human moderators. This underlines the need for Meta to push for a system of contextualized assessment, where content which is not outrightly offensive must be subject to a systematic and tiered review, considering the contextual nuance of each post.[22] However, the company's new policy on hate speech has raised concerns regarding the protection of marginalized groups on the platform. The recent update on Meta's content review policy has sparked an international concern among digital rights advocates, with stakeholders describing it as a 'deliberate regression' from an already flawed system.

This new content review policy recently adopted by Meta, namely the 'Hateful Conduct Community Standards', has been adjudged as a liberalized moderation policy, which has multiplied the flaws in the original review process. This further removes the protections laid out for the marginalized groups, while discussing 'political' issues, a categorization which remains undefined, yet.[23] Although this policy was introduced with the premise of ensuring expanded freedom of speech and expression, it misaligns with the human rights standards which must be guaranteed and upheld by online platforms.

In reference to the Oversight Board's charter,[24] the body has the inherent power to issue opinions and recommendations on Meta's content moderation policies. The functions of the Board correspond to human rights guidelines established under instruments such as Universal Declaration of Human Rights (UDHR).[25] However, with limited enforceability, the recommendations of the Board are not being effectively incorporated into the system. It has been rightly stated that content moderation is an ever-evolving process.[26] Rather than introducing guidelines which are in a stark contradiction to established principles of human rights, it is essential that an integrated system of content review is incorporated. Involving and investing in various stakeholders, such as human rights groups or the very communities which are being targeted through such posts, can lead to consistent application of standards.

Conclusion and Reccomendations

The case of the reinstated Argentine protest poem illustrates a systemic mismatch between Meta's content filters and the real-world use of language in political art. An automated classifier flagged the poem's second image for

---

[22] Castro D, "Content Moderation in Multi-User Immersive Experiences: AR/VR and the Future of Online Speech" (*ITIF*, June 3, 2022) https://itif.org/publications/2022/02/28/content-moderation-multi-user-immersive-experiences-arvr-and-future-online/

[23] "Meta's New Policies: How They Endanger LGBTQ+ Communities and Our..." (*HRC*, January 27, 2025) https://www.hrc.org/news/metas-new-policies-how-they-endanger-lgbtq-communities-and-our-tips-for-staying-safe-online

[24] "Oversight Board Charter" (2019) report https://about.fb.com/wp content/uploads/2019/09/oversight_board_charter.pdf

[25] Enarsson T, "Navigating Hate Speech and Content Moderation under the DSA: Insights from ECtHR Case Law" [2024] Information & Communications Technology Law 1 https://doi.org/10.1080/13600834.2024.2395579.

[26] "Insider Q&A: Trust and Safety Exec Talks about AI and Content Moderation | AP News" (*AP News*, April 22, 2024) https://apnews.com/article/alex-popken-webpurify-twitter-ai-content-moderation-28e540b1021d6bb7ecd5fe7584db2976.

containing Spanish slurs ("puto," "trava")[27] even though the full carousel of images – a feminist/queer critique of the government – showed these terms were being reclaimed for solidarity. Meta's *Hateful Conduct* policy explicitly forbids slurs and dehumanizing speech, yet it also notes that slurs used "to condemn" or "in an empowering way" may be allowed when intent is clear.[28]

Automated moderation flagged the single slide without understanding its broader context, exemplifying a common flaw in off-the-shelf hate-detection systems that over-remove counter-speech and reclaimed language. As a result, the poem's deliberate reappropriation of slurs, an empowering political strategy was misread as hateful. Without integrating cultural and linguistic nuance, such moderation risks repeatedly silencing creative resistance. Therefore, a linguistically sensitive approach is required.

LGBTQ+ activists rely on online spaces for solidarity and protest, but feel increasingly under threat. Meta's recent policy shifts, which explicitly allow anti-LGBTQ+ slurs (e.g. describing queer people as "mentally ill")[29] and downgrade moderation of non-illegal hate, are already driving users to self-censor or withdraw. In Argentina, this risk is amplified by the current government's rollback of queer rights. In just one year, Javier Milei's administration dismantled the Ministry of Gender and Diversity and allowed employers to fire people for their sexual orientation with impunity[30].

When governments stoke homophobic/transphobic sentiment, social media become crucial venues for counter-speech. Such over-broad censorship contravenes international human rights norms: under Article 19 of the UDHR, "everyone has the right to freedom of opinion and expression," and Article 27 protects participation in culture and the arts[31]. Meta's own Charter and Oversight Board recognize free expression as fundamental.[32] Therefore, Meta must strengthen enforcement against hate speech on its platform. Furthermore, there must be transparency and an enhanced appeals process with stricter resolution timelines. Along with this, user-satisfaction surveys will help them to address the issues better.

---

[27] Oversight Board: Improving How Meta Treats People and Communities around the World' (The Oversight Board, 2 June 2025) https://www.oversightboard.com accessed 1 July 2025..

[28] 'Transparency Center' (Transparency Center) <https://transparency.meta.com/>

[29] 'Human Rights Campaign' (*HRC*) <https://www.hrc.org/> accessed 1 July 2025.

[30] 'Washington Blade Newspaper' (Washington Blade: LGBTQ News, Politics, LGBTQ Rights, Gay News, 1 December 2023) <https://www.washingtonblade.com> accessed 1 July 2025.

[31] 'Slavery and Torture in Cambodian Scamming Compounds' (Amnesty International, 27 June 2025) <https://www.amnesty.org/en/> accessed 1 July 2025.

[32] 'Oversight Board: Improving How Meta Treats People and Communities around the World' (The Oversight Board, 2 June 2025) <https://www.oversightboard.com/> accessed 1 July 2025.