INTERNETLAB

InternetLab's comment on Oversight Board case 2025-012-FB-UA

Catharina Vilela, lawyer and researcher at InternetLab
Clarice Tavares, anthropologist and head of research at InternetLab
Stephanie Lima, anthropologist and head of research at InternetLab

Violence against human rights defenders in Latin America is not a new phenomenon. Historically shaped by authoritarian regimes, colonial legacies, and democracies still in the process of consolidation, the continent has witnessed high rates of assaults, threats, and murders against those fighting for justice, equality, and fundamental rights throughout its history. These acts of violence reveal not only the vulnerabilities of Latin American democracies but also the persistence of social and political structures that promote and normalize such practices. The case under review by the Oversight Board involves an image shared on Facebook that incited physical violence against a Peruvian human rights defender. This incident is part of a broader pattern of political violence impacting the region.

In 2023, the Inter-American Commission on Human Rights (IACHR) recorded 126 murders of human rights defenders in Latin America. The attacks predominantly targeted individuals engaged in environmental and territorial defense, with a particular emphasis on Indigenous and Afro-descendant leaders, who face heightened risks in their activities.

According to the organization Global Witness, Latin America accounts for more than 70% of the world's murders of human rights defenders. Once again, the region reported the highest number of documented murders of land and environmental defenders in 2023, accounting for 85% of the cases, with 43% of the victims being Indigenous and 12% women.

Between 2020 and 2024, InternetLab, in partnership with Instituto AzMina, Núcleo Jornalismo, and the Digital Humanities Laboratory at UFBA, conducted MonitorA, an observatory that monitored online political violence during electoral periods in Brazil. In 2020, we found that female candidates were primarily attacked based on personal aspects—such as appearance, personal life, and social roles—while male candidates were mainly targeted for their political and professional actions. In 2024, it was observed that although women represented only 15% of the candidates in the second round of municipal elections in Brazil, they were the target of 68.2% of the offensive comments analyzed.

These findings highlight the need for an intersectional approach when analyzing political violence, emphasizing how individuals from historically marginalized groups face disproportionate risks and require specific political and legal responses that address overlapping forms of discrimination and vulnerability.

The evolution of digital technology and the rise of social media have introduced new dimensions to violence against human rights defenders, including journalists, politicians, and members of civil society. The digital space, which has the potential to serve as a platform for mobilization and the dissemination of information, often becomes an arena for threats,

disinformation campaigns, hate speech, and attacks. The lack of tools for society to monitor these dynamics hampers the identification of patterns and the development of preventive measures and policies, allowing online harassment and persecution—often trivialized—to persist and spill over into the physical world, resulting in serious consequences such as assaults, stalking, and even murder.

In light of this scenario of political violence against human rights defenders, the contribution presented, grounded in the experience and research conducted by InternetLab—a Latin American organization dedicated to studying human rights in the digital environment—seeks to provide a regional perspective on the challenges faced in content moderation and addressing online violence. The key points to be addressed, based on the issues highlighted by the Oversight Board, aim to assist in combating political violence—whether symbolic or through direct threats—with a focus on protecting human rights defenders:

1. ***Policy recommendations for the protection of human rights defenders that have already been made to social media platforms, as well as the outcomes of campaigns to implement those recommendations***

(a) Development of Specific Policies and Guidelines for Human Rights Defenders

Human rights defenders often hold highly visible positions, making them prime targets for defamation campaigns and coordinated attacks. To address this reality, platforms must go beyond a superficial analysis of posts, evaluating the scope and impact of these attacks, particularly in contexts marked by historical violence and political polarization. It is essential for platforms to develop policies and guidelines tailored to local political contexts, ensuring swift and effective responses during political crises and mass protests. These measures should include monitoring organized attacks, such as those aimed at discrediting defenders through visual and textual disinformation, as well as specific tools and channels for reporting incidents.

Furthermore, the formulation of these guidelines should take place in close collaboration with civil society and academia, which can provide insights into local specificities and various forms of discrimination, fostering more effective and context-sensitive approaches.

(b) Considering Markers of Difference in Policy Development

Online attacks often exploit vulnerabilities tied to markers such as gender, race, ethnicity, social class, or political position, amplifying the impact of hateful messages.

A study on attacks against journalists conducted by InternetLab, Instituto AzMina, Volt Data Lab, and INCT.DD on X revealed that women faced more than twice as many offenses compared to men. Among the tweets analyzed, 15.93% targeting female journalists were offensive, compared to 8.60% in the case of male journalists. Black and Indigenous women, in particular, were frequent targets, especially when addressing issues of racism.

In the Peruvian context, where groups like La Resistencia have a history of violence against human rights defenders and journalists, platforms must consider local dynamics and the additional risks faced by individuals from historically marginalized groups. Adopting an intersectional perspective would help platforms interpret threats within broader sociopolitical contexts, fostering fairer moderation decisions aligned with the protection of fundamental rights.

In this regard, we highlight the importance of the Oversight Board including data on gender identity, race, and other relevant social markers when describing cases for consultation. In cases of violence, such information is essential to broaden the understanding of how different forms of discrimination intersect, enabling the identification of vulnerability patterns that platforms could address more effectively.

2. ***The use of images, including digitally altered or AI-manipulated, to harass, intimidate and make threats of violence against activists and journalists.***

(c) Recognizing and Mitigating Risks of AI-Manipulated Images

The use of artificial intelligence technologies for image manipulation poses significant risks in contexts of digital political violence. The creation and dissemination of altered images illustrate the potential of such practices to amplify disinformation, incite violence, and delegitimize public figures.

In the case at hand, although detailed information about the victim is lacking, it is evident that the manipulation of images using artificial intelligence plays a strategic role in creating narratives of intimidation and delegitimization. The visual depiction of a bloodied human rights defender, combined with accusations of corruption and incitement to violence, operates as a symbolic message of violence, designed to reinforce negative perceptions and weaken the individual's position in a context of heightened political tension. This type of practice not only threatens the victim's safety and reputation but also generates systemic collateral effects, undermining the credibility of human rights organizations and the very legitimacy of social movements.

This phenomenon becomes even more complex when considering its disproportionate impact on women and other individuals from historically marginalized groups who engage in political activity or hold positions of influence. In these cases, image-based violence strategies are used to reinforce stereotypes that undermine the credibility and authority of the victims. These attacks have the potential to transcend the digital realm, escalating into threats, physical harm, and even withdrawal from public life.

During the 2024 municipal elections in Brazil, this reality was evidenced by cases of the creation and dissemination of false sexualized images targeting female candidates, exemplifying AI's use as a tool for gender-based violence. In São Paulo, the largest capital city in the country, both female candidates for mayor were subjected to this practice. These are just two instances among many others where women in politics and activism were targeted by such violence, aiming to undermine the exercise of their citizenship.

In this context, we emphasize the importance of developing advanced systems for detecting and removing AI-manipulated content, as well as creating specialized teams to conduct contextual analyses that identify threats to human rights and incitement to violence.

edifício itália
avenida são luís 50
1º andar conjunto 11b
1046-000 são paulo sp brasil
www.internetlab.org.br