

January 28, 2025

Members of the Oversight Board META

Ref: Submitting public comments Content Targeting Human Rights Defender in Peru

In order to respond to the request for comments in the case “Content Targeting Human Rights Defenders in Peru”, the [Center for Justice and International Law \(CEJIL\)](#), based on our experience of more than 30 years in the protection of human rights defenders in the region, submit comments on the following aspects: A) The socio-political context in Peru, specifically regarding the risks faced by human rights defenders and civil society organizations, B) International standards of International Human Rights Law on veiled or implied threats, C) Non-compliance with the community standards of “Violence and Incitement” and “Bullying and Harassment” derived from the publication under analysis, and D) Suggestions to the Oversight Board on treatment of this publication and possible recommendations on modifications to Community Standards.

I. Socio-political context in Peru, specifically regarding risks to human rights defenders and civil society organizations.

Peru is currently facing a [profound political and social crisis](#) that has significantly weakened the checks and balances system, and has seriously affected the rule of law and the exercise of human rights. In this context, [the work of human rights defenders \(HRDs\) is crucial](#) for the protection of fundamental freedoms and constitutional order. However, as a direct consequence of their work, HRDs face increasingly serious attacks, which generates a climate of harassment and intimidation. Both the [United Nations High Commissioner](#) (Recommendation 7), the [Inter-American Commission on Human Rights](#) (IACHR, Recommendation 28) and the [Office of the Rapporteur for Freedom of the Inter-American Commission](#) (para. 1270) have called attention to this situation. Nonetheless, [there is no comprehensive policy](#) for the protection of HRDs in Peru to date.

In this context, extremist groups such as [“La Resistencia”](#), [“Los Insurgentes”](#) and [“La Sociedad Patriotas”](#) have emerged and play a central role in perpetuating aggressions against HRDs. These organizations have carried out [at least fifty aggressions](#) documented between 2023, 2024 and the first month of 2025. These aggressions are directed against [journalists](#), [protesters](#), members of [human rights organizations](#), [artists](#), [justice operators](#), [victims](#) and even [children](#). The aggressions include [death threats](#), [physical attacks](#), [harassment in the HRDs’](#)

[homes and workplaces](#), [doxing](#), [terruqueo](#)¹ and [smear and defamation campaigns on social networks](#). Despite the seriousness and intensity of these extremist groups' aggressions against HRDs, their acts [remain mostly unpunished](#).

The smear campaigns have legitimized the promotion of [legislative initiatives](#) that seek to unduly control and restrict the operation of civil society organizations in Peru, which various [rapporteurs of the United Nations](#) and [several countries](#) have already condemned. Despite such international condemnation, the Peruvian Congress has promoted bills that would grant the government broad powers to arbitrarily sanction organizations that receive international funding, placing both the defenders and the legitimate activities of these organizations at risk. It is worth noting that these types of initiatives [are being promoted throughout the region](#) in countries such as [Paraguay](#) and [Venezuela](#), seriously weakening civic space and defenders' ability to carry out their work.

Additionally, we highlight that the user who made the publication under review by the Oversight Board is the leader of one of these extremist groups, known for acts of intimidation and threats against human rights defenders HRDs in Peru. According to [publicly available information](#), he has been sentenced to prison for the crime of defamation against civil society organizations. Furthermore there is an [ongoing investigation](#) by the Criminal Prosecutor's Office Specializing in Human Rights against members of this group for crimes of harassment and coercion against a group of HRDs.

Moreover, the HRD targeted by the post in question holds a leadership position within a coalition of human rights civil society organizations in Peru, with more than 40 years of experience. These threats against a widely recognized activist who serves as a spokesperson and interlocutor are particularly concerning, as they may have a chilling effect on other HRDs.

II. Applicable international standards on veiled or implied threats

Various human rights protection mechanisms –both at the [regional](#) and [international](#) levels– emphatically support the important work of human rights defenders and the consequent obligation to protect them so that they can continue their work in a democratic society. Concerning threats to human rights defenders, the [Esperanza Protocolo](#) establishes specific guidelines based on international human rights law provisions to respond to threats to human rights defenders.

The *Esperanza Protocol* defines threats against HRDs as “intentional conduct that indicates a future harm or that intimidates an HRD, their family or community”, including “individual, collective, direct and indirect, explicit and symbolic threats, whether they take place in offline or online spaces” (Section II). Due to the intimidating nature of threats and their deterrent effect,

¹ Terruqueo" is a stigmatizing narrative that pejoratively associates a person or group with terrorism, without evidence to support such accusations, and is frequently used by extremist groups to delegitimize people who oppose their political ideals.

the *Esperanza Protocol* recognizes that threats not only reflect an intention to cause harm but can themselves violate human rights. In this regard, the Inter-American Court has recently summarized and developed its standards on HRDs in the [judgment of the CAJAR case](#), emphasizing that defending human rights can only be carried out freely when those who engage in it are not subjected to threats (paras. 468–478).

Regarding the responsibility of companies in the face of threats against HRDs, the *Esperanza Protocol* establishes that technology companies –including social media companies– must adopt measures to prevent their platforms from being used to target, harass, and intimidate HRDs (Section IV B 2). It also refers to social media platforms’ responsibility to engage in context-specific moderation globally, taking into account different languages, usage, and contexts. Finally, the *Esperanza Protocol* recognizes that symbolic threats are pervasive in some contexts. It specifies that such threats can occur in different spheres –including online– and be expressed in different forms (Section VIII F I iii).

Therefore, the international human rights standards systematized by the *Esperanza Protocol* highlight the importance of collaboration between States, companies and civil society to protect human rights defenders from threats. In this context, technology companies have a crucial responsibility: to ensure that their platforms are not used as tools to intimidate or harass HRDs. To comport with this obligation, technology companies may have to adopt specific content moderation measures, sensitive to different global contexts and realities, to safeguard the indispensable space that HRDs occupy in the promotion of democracy and human rights.

III. Non-compliance of publication with Meta’s Community Standards

In the following, we refer to the reasons why the publication of a death threat against a female human rights defender does not comply with Meta’s principles to “promote the best of what people can do together by keeping people safe and preventing harm.” Additionally, we argue it breaches Meta’s standards on “[Violence and Incitement](#)” and “[Bullying and Harassment](#).”

First, the publication fails to comply with the “[Violence and Incitement](#)” standard because it constitutes a credible threat. This policy states that content to be suppressed includes “threats of violence against various targets.” It also states that the policy requires information or context to suppress threats when they constitute “[c]oded statements where the method of violence is not clearly articulated, but the threat is veiled or implicit, as shown by the combination of both a threat signal and contextual signal. . . .”

In the present case, the artificial intelligence modification of the defender’s face to cover it with blood constitutes an implicit death threat given the publication’s context. In particular, the parameters set out in the [Rabat Plan of Action](#), which has been used by the Oversight Board in several cases (see, for example, [Iranian women confronted on the Streets](#) and [Gender based cases](#)) to assess the threshold for defining restrictions on freedom of expression, are met:

- i. In relation to the context, as documented *above*, Peru is facing a profound political and social crisis characterized by systematic harassment of HRDs. This harassment stems both from government actions aimed at weakening civil society organizations and from acts of extremist groups that resort to intimidation, threats and, in documented cases, physical attacks against HRDs.
- ii. The author of the publication, is a leader of an extremist group known in Peru for leading acts of intimidation and threats against HRDs (see *above*).
- iii. The speech had a clear intimidating intent: to threaten a prominent human rights defender with an image of her face covered in blood. In the Peruvian cultural context, this representation constitutes a death threat. Unlike expressions that may be metaphorical ([Metaphorical statement against the President of Peru](#)) or aspirational ([Anti-colectivos content in post-election Venezuela](#)), this act was explicitly aimed to produce a deterrent effect against human rights defense work through a direct threat. Additionally, leaving the post on social networks amplified the threat, generating a chilling effect to other HRDs ([Cambodian Prime Minister](#)). The seriousness of this threat was recognized by the [UN Rapporteur, who condemned the action](#). Finally, we highlight that the intent to replicate this content is evident: the author of the post took a screenshot of this threat and reproduced its content again in a Facebook post on January 16, 2025.
- iv. The content of the post includes an image of the human rights defender covered in blood, accompanied by a caption accusing civil society organizations of being funded by foreign sources, promoting violent protests, blaming law enforcement, manipulating victim figures, bringing cases to international courts, and profiting from compensations. In the Peruvian context, the image of the HRD is understood as a death threat, while the caption seeks to legitimize this threat by alleging the HRD's involvement in corrupt activities.
- v. The threat becomes more serious considering that the speaker has 6.5 K followers on Facebook, most of whom are members or sympathizers of the extremist group. These followers [have demonstrated, on several occasions](#), their willingness and ability to execute [acts of physical aggression](#) against the targets indicated by their leader.
- vi. There is a high probability that the threat will materialize, as there is a documented history of members of this group [perpetrating violent acts motu proprio](#), targeting individuals previously threatened by the group. In addition, the [impunity with which these harassments have been handled](#) and the [absence of a protection policy for HRDs](#) fosters an environment in which these acts can be carried out without legal consequences, further aggravating the situation.

Moreover, the publication also violates Meta's "[Bullying and Harassment](#)" Community Standard. According to the standard, the person to whom the threat was directed is a public figure of limited scope, as she is widely known in the country for her activist work. The Standard states that directed mass harassment postings must be removed if they are directed at "individuals at heightened risk of offline harm," which includes "human rights defenders." In this regard, we highlight that the user who posted this threat constantly posts messages on his social networks aimed at attacking HRDs. As has been assessed by the Oversight Board in previous cases ([Image of gender-based violence](#)), the cumulative effect of these threats

contributes to a pattern of discrediting and delegitimizing HRDs' work, which affects both the person to whom the message is directed and other people who carry out similar activities in the defense and protection of human rights.

In addition, we note that this publication also constitutes gender-based violence, since the aggression was specifically directed at a woman human rights defender, so that its publication generates a differentiated intimidating effect on other defenders. Thus, as stated [by the Special Rapporteur on violence against women](#), violence against women online can lead women to refrain from using the Internet and withdraw from public life (para. 26). In this regard, we recall that in other cases the Oversight Board has recognized how certain content can be discriminatory because of its cumulative effect, without the need to demonstrate how each specific publication causes direct and imminent harm ([Depiction of Zarte Piet , Knin cartoon](#) and [South African Slurs](#)).

Therefore, we consider that the publication described above seriously violates Meta's policies on "[Violence and Incitement](#)" and "[Bullying and Harassment](#)." This type of content not only represents a direct threat to the safety of a woman human rights defender, but also generates a chilling effect that inhibits other defenders from exercising their fundamental rights. In addition, it perpetuates patterns of gender-based violence and reinforces dynamics of exclusion, delegitimization and harassment that affect those who work in the defense of human rights in highly vulnerable contexts.

IV. Suggestions to the Oversight Board on treatment of this publication and possible recommendations on policy modifications

For the foregoing reasons, the only appropriate measure to ensure that Meta complies with its responsibility to prevent online violence and harassment is the immediate removal of the post and any subsequent posts that replicate the initial content, because the posts are in clear violation of Meta's Community Standards. Furthermore, we point out that a simple warning that the image is violent or may be modified by artificial intelligence ([Derogatory Image of Candidates for US Elections](#)) would not be sufficient, as the harm and chilling effect to other HRDs is caused by the display of the threat itself. Finally, we recommend that the "Bullying and Harassment" and "Violence and Incitement" policies include specific provisions to protect HRDs from threats, including those of a symbolic nature.