40+ Organizations from the U.S., Latin America, and Europe Call on Meta to Reinstate Protections for Users Facing Heightened Levels of Hateful Content Since January Rollbacks

We, the undersigned organizations from across the world, call on Meta leadership to urgently reinstate protections and widespread content moderation for users, and to hire an independent third party to formally assess the impact of its January 2025 policy changes on harmful content centering user experiences.

Since the early 2000s, social media platforms have been experimenting with how to best protect users from potentially harmful content online, including hate speech and disinformation. The balance has never been perfect, but most platforms—with Meta often in the lead—have generally worked toward improving their trust and safety protocols. On January 7, 2025, to appease the Trump administration, CEO Mark Zuckerberg made the [alarming decision](#) to gut those protections, in turn reversing over a decade of progress made in creating a safer online environment for all users.

Marginalized communities—including women, people of color, LGBTQ people, and more— are already experiencing the real-world consequences: threats of violence, calls for self-harm, dehumanizing abuse and harassment, and diminished freedom of expression online. While this decision was made by Meta leadership, **every Meta employee who enables or defends it shares responsibility for this harm.**

As you are aware, these core changes included: The dissolution of Meta's fact-checking program in the U.S.; the removal of Meta policy protections related to "immigration, gender identity, and gender"; and the cessation of "proactive" enforcement of some policies on harmful content. Meta also added terms ("transgenderism" and "homosexuality") that are well-known to convey animus against LGBTQ people, especially transgender and gender nonconforming people.

The policy rollbacks, and subsequent revelations of leaked revised internal content moderation guidelines, have drawn [widespread concern](#) from [human rights organizations](#), as well as from [Meta's own Oversight Board](#). These warnings have remained unaddressed and even written off by Meta leadership. Just weeks ago, Meta announced it would [effectively ignore](#) the recommendation of the Oversight Board to remove the term "transgenderism" from its hateful conduct policy. Meta also declined to commit to the Board's other three recommendations. In effect, Meta's actions suggest that the Oversight Board functions more as a public relations tool than as an independent body designed to provide meaningful checks and balances over the world's largest social media company, which serves billions of users globally.

It has been exactly six months since the rollbacks were implemented, and already we are witnessing significant adverse impacts, particularly on the marginalized communities we work with as civil society organizations. According to Meta's own recent [quarterly report](#), published May 29, 2025, the company reported a rise in violent content, bullying and harassment alongside an overall decrease in false flags since the changes in content moderation policies.

This report, however, does not appear to have been independently conducted and likely underrepresents the true scale of harm, given Meta's recent narrowing of what qualifies as harmful content. In addition, the report did not investigate the nature of violent content escalating or represent users' perspectives. Given these limitations and gaps in data, a coalition of nonprofits—UltraViolet, All Out, and GLAAD—surveyed more than 7,000 active users across Facebook, Instagram, and Threads—primarily women and LGBTQ people. **Key findings include:**

- 1 in 6 respondents report being the victim of some form of gender-based or sexual violence on Meta platforms.
- 92% are concerned about harmful content increasing since the rollbacks.
- 72% see more hate targeting protected groups.
- 92% feel less protected from being exposed to or targeted by harmful content.
- Over 25% say they have been targeted directly with hate or harassment.
- 66% have witnessed harmful content in their feeds.
- 77% feel less safe expressing themselves freely on Meta platforms.

Select qualitative testimonies–among thousands–are as follows:
- "As a non-binary trans person, violence against me has skyrocketed since January. I live in daily fear."
- "I have seen greater incidents of LGBTQ+ individuals and the community as a whole being called 'sick,' 'deranged,' and 'mentally ill.'
- "Homophobic slurs like 'f*ggot' and demands that I seek therapy are common."
- "Reading comments by men glorifying femicide is deeply painful and triggering."
- "I have been called the N word and have reported racist, misogynistic and misogynoir comments, and Facebook has not removed the offensive comments. A number of men have even stalked and trolled my Facebook profile."
- "I have seen on a number of LGBTQ+ posts a number of 'memes' that encourage violence or suicide which META takes no action in when reported despite giving you the option to report such content. What is the point in a report function that never actually takes action unless it is simply for show?"
- "There are times when I am afraid to comment on a post because of the violence expressed by others in their comments."
- "The impact of this new policy affects not only those that have already been targeted as minorities but creates a sense of unsafe space for freedom of speech."
- "I have come across TONS of death and sexual assault threats. I have seen far too many heartbreaking videos where these women are scared for their lives."

The report's findings offer stark evidence that harmful content has increased since Meta announced these sweeping changes, with real offline consequences. The findings also clearly indicate that freedom of expression is *decreasing* for vulnerable groups.

As Meta is the largest social media company, with billions of users worldwide, the implications of these policy rollbacks for humanity cannot be overstated. Users deserve online spaces where they can feel safe and thrive. They cannot do that when continually targeted with hate and harassment on the basis of who they are. As organizations committed to safety and freedom of expression for all, we demand that Meta leadership urgently reinstate protections and widespread content moderation for users.

Sincerely,

Accountable Tech
ADT - Associação da Diversidade Tabuleirense
Agência Ambiental Pick-upau
All Out
Articulação Brasileira de Gays - ARTGAY
Asociación por las Infancias Transgénero
Asociación Red Nacional de Consejeros, Consejeras y Consejeres de Paz LGBTIQ+
Bits of Freedom
Casa Marielle Franco Brasil
Centro de Defesa dos Direitos da Criança e do Adolescente Padre Marcos Passerini-CDMP
Coletivo LGBTQIAPN+ Somar de Porto Velho/RO
Coletivo Viveração
Corporación Caribe Afirmativo
Danes je nov dan
Democracy for America Fund
EDRi - European Digital Rights
Eletronic Frontier Foundation (EFF)
Forbidden Colours
Friends of the Earth
Fundación Polari
GLAAD
Global Project Against Hate and Extremism
Inequality Media Civic Action
ITTC - Instituto Terra, Trabalho e Cidadania
Kairos Fellowship
Kurytiba Metropole
La Junta Marica
Mariposas Negras Colectiva
National Campaign for Justice
National Hispanic Media Coalition
National Women's Law Center
Núcleo Ypykuéra
ProgressNow New Mexico

Reproaction
Sexual Violence Prevention Association
SUPERRR Lab
TakeItBack.org
The Juggernaut Project
UltraViolet
Voz da Diversidade
Women's March