Contribution to the Office of the High Commissioner for Human Rights' call: Protection of human rights defenders in the digital age

March, 2026

Derechos Digitales[1] is an independent non-profit Latin American organization founded in 2005, whose mission is the defense, promotion, and development of fundamental rights in digital environments in Latin America.

1. Legislative and regulatory measures

A. What impacts have recent trends in legislative and regulatory efforts at local, regional and global levels – including, for example, on information integrity, online safety and cybercrime – had on the work and safety of HRDs offline and online?

- Derechos Digitales has documented the impact of cybercrime regulatory measures globally and of "anti-NGO" legislative efforts regionally on the work and safety of HRDs[2] online and offline. In the first case, cybercrime laws promoted under the narrative of protection and fighting cybercrime, have become a pretext for criminalization. By 2020, over 180 countries had adopted substantive and procedural legislation on cybercrime and electronic evidence. Through our analysis of eleven cases, from Cuba, Egypt, Jordan, Libya, Nicaragua, Russia, Saudi Arabia, Uganda, and Venezuela, we found that, while cybercrime regulations multiply globally, they have not only been in some cases ineffective at safeguarding the expression of HRDs, particularly women and LGBTQIA+ people, but also have exposed them to greater risk. This is especially pronounced in countries where cultural and legal restrictions narrow the civic space, exacerbated by the rise of anti-NGO laws.
- The cybercrime laws analyzed apply legal concepts that criminalize online speech in an overly broad manner, violating existing freedom of expression standards and undermining the work of HRDs. This trend impacts HRDs' safety, as legislation relies on generic terms that lack precise definitions for the criminal offenses invoked, such as disinformation, terrorism, or hate crimes, making them open to abusive interpretation by authorities. The risks are even greater in contexts of fragile democratic institutions in the face of democratic backsliding.
- Cybercrime laws are often used in conjunction with other laws, like cybersecurity laws, criminal codes, and laws governing ICTs, telecommunications and counterterrorism, among others.
- As noted in a IACHR recent report, a growing trend is observed in Latin America towards the adoption of laws and regulations that restrict the legitimate work of civil society, the so called "anti-NGO laws", which include: Peru (Law N° 32301), Paraguay (Law N° 7363), Guatemala (reform of the Law on Non-Governmental Organizations for Development and the Civil Code), Nicaragua (National Assembly passed the Foreign Agents Act), Ecuador (Integrity Strategy for NGOs), Honduras (reforms to the Money Laundering Law were approved to include Politically Exposed Persons), Venezuela (Law for the control, regularization, performance and financing of non-governmental and related organizations), El Salvador (Foreign Agents Law). These laws are based on broad and ambiguous

[1] More information, at: https://www.derechosdigitales.org/en/home/. This contribution was prepared by Laura Mantilla-León and Marina Meira, and reviewed by Paloma Lara-Castro. For more information, please contact: paloma.lara.castro@derechosdigitales.org

[2] Under this contribution, we define civil society within the UN framework as encompassing all individuals, institutions, and collectives that voluntarily engage in forms of public and political participation and action aimed at promoting, protecting, and advancing human rights, as well as justice, equality, and respect for human dignity. The term lists a wide range of actors, including human rights defenders, NGOs, victim support groups, coalitions for specific rights (e.g. LGBTQIA+, the environment), indigenous and rural community groups, trade unions, social movements.

definitions; impose restrictions; spread narratives that explicitly or effectively silence and close off the space of civil society; and limit free participation in public affairs online and offline.

B. What legal or regulatory instruments and institutional procedures are commonly used to restrict the rights to freedom of expression, association and privacy of HRDs online?

- Together with the legislation previously detailed, within our research we have found institutional procedures such as cyberpatrolling and online surveillance used to restrict HRDs rights to freedom of expression, association and privacy.
- To mention a recent case, in Argentina, the current government approved the new Argentina Federal Police (PFA) statute by decree, granting the police powers that threaten fundamental rights and freedoms. Without judicial authorization, the police can now conduct "crime prevention tasks in digital public spaces," including on social media and websites. This provision reinforces a porous mechanism with little oversight and no clear limits in Latin America: cyberpatrolling. This form of intelligence imposes an extraordinary restriction on rights such as privacy and freedom of expression, and is especially for critical voices such as HRDs. In contexts of social tension and growing authoritarianism as the country is currently facing, it can be used as a tool that threatens democratic debate and social protest, enabling mass surveillance and persecution.
- In Brazil and Colombia, we documented the existence of social media monitoring contracts deployed by public sector entities to measure the approval of the government and its authorities, and to design strategies to ensure their positioning on social media. The targets of these cyberpatrolling strategies are often political figures, journalists, media outlets, HRDs, among others.
- In Bolivia, cyberpatrolling has been deployed to identify misinformation, a practice that has been instrumentalized for the persecution of political opponents -often encompassing HRDs-, resulting in individuals convicted for allegedly participating in "destabilizing movements".
- In Paraguay, Law N° 7280 grants the National Police broad surveillance powers, including cyberpatrols and communication interceptions, without clearly defining legal limits or safeguards. These measures threaten the fundamental rights and freedoms and leave HRDs in specific vulnerability.

C. How have legislative and regulatory efforts in one country or region impacted similar legal and regulatory measures in other countries or regions?

- Legislative and regulatory efforts in one country/region are increasingly shaping similar measures elsewhere through processes of norm diffusion and policy transfer, particularly in the area of restrictions on civic space. In Latin America, attacks HRDs have acquired a transboundary dimension, especially in the digital sphere. Practices such as online harassment, coordinated disinformation campaigns, and cross-border surveillance often operate beyond national jurisdictions and are frequently enabled by opaque or transnational funding structures.
- In parallel, the role of big-tech companies has become increasingly central, as their platforms are used both to amplify harmful campaigns against HRDs and, in some cases, to facilitate state-aligned surveillance practices. As we have highlighted insufficient transparency and accountability in the governance of these platforms can contribute to the spread and normalization of authoritarian practices and legislation.

2. Digital communications

A. Which risks do internet shutdowns, network interferences, geo-blocking or other forms of restrictions of connectivity and communications pose to HRDs' work and safety?

- Meaningful access to the internet is a key tool for HRDs as it allows them to share their work, obtain and disseminate information at key political moments, report human rights violations, organize their actions, and exercise their right to defend rights. We have documented localized Internet disruptions in several Latin American countries, particularly in contexts of social protest, impacting HRDs' work and safety:
- In Brazil, in 2019, the website Women on Waves focused on abortion rights was reportedly blocked by major Internet providers. The site has faced censorship around the world since 2017, negatively affecting the exercise of sexual and reproductive rights.
- In Ecuador, during the 2019 social protests, civil society organizations documented temporary shutdowns and disruptions of certain social media platforms, mobile communications, websites, and internet connections. This impacted their activities related to organizing, coordination, planning, and assembly.
- We have followed closely the case of environmental HRDs. In Paraguay, we have documented violations related to connectivity restrictions in areas of socio-environmental conflict and possible deliberate Internet shutdowns during security operations, which directly affect the ability of HRDs to act.
- In parallel, in the Amazon region, structural connectivity gaps severely limit HRDs' work. Research by Derechos Digitales shows that many communities depend on expensive, unstable and low-speed mobile connections, with extremely low broadband penetration in rural areas. This restricts HRDs' ability to access, produce and share information, weakening their capacity to document violations, coordinate actions, and seek protection.

B. What forms of technology-facilitated attacks do HRDs face on social media platforms and digital communications services? How do these online attacks intersect with offline events?

- As technologies are constantly evolving, so too are the forms of violence exercised through them, with new tools and platforms enabling increasingly sophisticated, scalable and targeted attacks against HRDs.
- Among technology-facilitated attacks HRDs face online are: hacking and takeover of personal, collective, and media accounts; removal or sabotage of content for censorship or retaliation; identity theft and malicious use of public profiles; non-consensual sharing of intimate images; the use of artificial intelligence to manipulate images, voices, or videos for discrediting, fraud, or disinformation purposes; infiltration and covert surveillance in HRDs WhatsApp groups; hate speech; smear campaigns.
- Technology-facilitated attacks transcend online spaces and compromise integral safety and well being. They also disproportionately affect women and LGBTQIA+ people. This is because online and offline spaces are a continuum, where rights exercised digitally and consequences experienced physically are inextricably linked.

C. What specific risks to HRDs emerge via online platforms and communications services in situations of armed conflict, instability and/or elections?

- In contexts of armed conflict, HRDs especially women and LGBTQIA+ individuals face automated harassment; the creation of fake intimate images using generative models; disinformation campaigns

with gender bias; and increasingly invasive forms of digital surveillance assisted with AI. Regarding this last risk, AI systems process enormous volumes of data obtained from years of mass surveillance of the population, which is used to generate lists of individuals and structures considered targets. These tools, often presented as neutral or necessary for security or humanitarian management, can particularly expose HRDs and other communities in vulnerable situations.

- In electoral contexts, HRDs face disproportionate disinformation campaigns that operate in a coordinated and cross-platform way. These campaigns do not affect everyone equally. Women public figures and LGBTQIA+ individuals are systematically subjected to attacks that combine disinformation, harassment and stigmatization, thereby facing digital political violence.

D. What specific risks do women HRDs and HRDs from groups affected by marginalisation and discrimination face on online platforms and communications services?

- To start, for groups such as women and LGBTQIA+ people, who have been historically excluded from the political sphere and from the full enjoyment of fundamental rights, online platforms have been a key instrument to expand the exercise of freedom of expression, association and assembly. However, the online context not only replicates the structural misogyny and violence to which they have been subject for centuries offline, but can even exacerbate it.
- As stated by the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, women and gender-diverse people are more commonly targeted technology-facilitated violence facing death and rape threats, “sextortion”, doxxing, gender-based trolling, bullying and harassment, stalking, sexual harassment and the non-consensual sharing of intimate images, spread of AI manipulated images or videos for discrediting or disinformation purposes, smear campaigns.
- These forms of violence create a chilling effect, discouraging participation and often forcing women and gender-diverse HRDs to withdraw from online spaces, ultimately limiting their presence, voices, and influence in the public sphere.

E. How do companies’ policies and practices relating to content moderation and engagement with law enforcement and government authorities affect HRDs’ work and safety?

- Content moderation companies’ policies can either guarantee HRDs safety and rights or restrict it, especially in contexts of contraction of civic space. Dominant platform business models, driven by engagement and attention metrics, tend to amplify sensational, polarizing and violent content, which can increase the visibility of attacks against HRDs, exacerbate harassment dynamics, and further endanger their work and safety.
- A concrete example is the case of Peruvian HRDs who have faced several online attacks including harassment and stigmatization campaigns by an organization called "La Resistencia". HRDs in that country have witnessed unfounded accusations of corruption and systematic digital violence that spill over into physical environments. One case related to this group's actions involved a Facebook post targeting the executive secretary of the Coordinación Nacional de Derechos Humanos, which had been manipulated using AI to show her face covered in blood, alongside a message accusing her and her organization of alleged acts of corruption and violence.
- Meta initially dismissed the image as unproblematic, but after human rights organizations submitted comments to the Oversight Board, it was recognized as an "implicit or veiled" threat of violence.

Given elements of the Peruvian context, it could be understood that an image such as the one disseminated would inevitably have a harmful effect on the targeted HRD.

- Companies' content moderation policies should recognize that, when content involves HRDs, social context is a key factor in determining whether a veiled threat exists. Companies should also report on the frequency of false positives and false negatives in the application of these policies regarding human rights defenders and the circulation of political speech.

F. How do advances in AI technologies exacerbate risks to HRDs' operations and presence on online platforms and communications services?

- AI assisted and facilitated attacks on HRDs risk their presence online by discrediting their work through the spread of manipulated images, voices or videos for disinformation and criminalization proposes. Deepfakes represent a direct threat to democracy, freedom of expression, and HRDs privacy. As we have argued, these attacks often have several consequences leading to self-censorship, withdrawal from online spaces, damage to reputation, and physical safety risks.

3. Digital restrictions to privacy

A. What risks have emerged for HRDs with the increasing procurement, use and abuse of digital surveillance tools, including spyware and interception technologies, by State and non-State actors?

- In a recent AlSur Consortium report, we documented several cases of digital surveillance and interception of communications to HRDs' movements across Latin America, highlighting how the adverse effects of privacy violations can also lead to the violation of other rights. These include equality before the law; the right to life, liberty, and personal integrity; the right to a fair trial and due process; freedom of expression; the right to protest and freedom of association; and freedom of movement.

B and D. What risks have emerged for HRDs with the expansion of biometric surveillance infrastructure and increased monitoring of public and digital spaces? How do advances in AI technologies exacerbate risks to the privacy and safety of HRDs?

- As we have argued in other consultation processes, the expansion of biometric surveillance infrastructure and the increased monitoring of public and digital spaces represent a threat that has the potential to produce an intimidating effect, or chilling effect, on the population, especially on HRDs.
- There is a growing trend in Latin America toward the acquisition of such technologies, as documented in AlSur Consortium's recent reports. In 2025, we identified facial recognition technologies (FRTs) currently active in Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay. These technologies are generally deployed for citizen security; license plate recognition and other vehicle security-related purposes; and immigration and border control.
- These systems are frequently implemented without prior consultation, impact assessments, or effective transparency mechanisms, creating serious risks for HRDs in a context of shrinking civic space. In particular, the lack of transparency regarding how biometric data is collected, stored, shared, and used raises significant concerns about misuse, function creep, and the potential targeting of HRDs.

- Additionally, numerous studies have shown that biometric and AI-driven systems, especially FRTs, are affected by structural biases (particularly racial, gender, and ethnic biases), which can lead to disproportionate surveillance, misidentification, and criminalization of already marginalized groups, including indigenous and black communities.

C. How have technological and regulatory developments relating to encryption eased or exacerbated risks to HRDs?

- Since online platforms are crucial for HRDs' work, and given the increase in attacks on their communications, end-to-end encryption takes on special relevance. It helps reduce risks in accessing and transferring information, prevents unauthorized intermediaries from intercepting communications, while also preventing service providers from accessing the content of messages.
- However, the role of legislation in areas affecting encryption implementation cannot be overlooked. For instance, under a treaty between the US and the UK, US based platforms such as Facebook and WhatsApp were compelled to share users' encrypted messages with the British police, establishing a global precedent. This highlights the need for institutional frameworks that guarantee effective privacy and freedom of expression in digital environments through encryption, especially for HRDs and historically vulnerable groups.

4. Corporate responses

A and B. How are companies meeting their responsibilities to identify, assess, mitigate and respond to risks posed to HRDs on their platforms and services?; Are existing corporate models and approaches to risk assessment, due diligence, remedial mechanisms and engagement with HRDs on protection concerns and reports of violations sufficient and/or effective?

- Evidence shows that platforms often fail to respond adequately to reports of violence and harassment, particularly in cases involving gender-based violence and attacks against HRDs. Several researches document persistent gaps in response, including delayed or absent content removal and lack of effective user support.
- Corporate content moderation and safety policies are frequently designed and implemented without sufficient local context or consideration of local languages, leading to both over-censorship and under-enforcement in cases of coordinated harassment or political violence.
- Existing due diligence and risk assessment frameworks are insufficient and poorly implemented, particularly in relation to intersectional harms and systemic risks, as companies rarely conduct meaningful human rights impact assessments prior to deployment or ensure ongoing evaluation with participation from affected communities.
- In Latin America, the absence of robust and enforceable platform regulation has resulted in a structural lack of accountability, allowing companies to operate with limited oversight and to systematically deprioritize risks affecting HRDs in the region.

C. What challenges do civil society and companies face in ensuring corporate policies, processes and initiatives – including in relation to internal mechanisms and external engagement – adequately and effectively address the range and extent of risks faced by HRDs in the digital age?

- A key challenge is the structural asymmetry between global platforms and HRDs. While they often rely on digital platforms to access and disseminate information, platforms lack transparency, including opaque moderation criteria and algorithmic systems, which restricts meaningful oversight and limits civil society's ability to assess risks and impacts.

- There is also a lack of regulatory frameworks in Latin America guaranteeing access to platform data for independent scrutiny, which severely limits the ability of HRDs affected to assess systemic risks, document harms, and hold companies accountable. This data asymmetry further entrenches power imbalances and undermines evidence-based responses to threats faced by HRDs.
- In Latin America, there are cases in which companies have failed to comply with local laws or judicial decisions in repeated disputes involving platform accountability and content moderation obligations, revealing tensions between corporate governance models and domestic legal frameworks.

D. What steps should companies take to improve identification, assessment and prevention of risks posed to HRDs' work and safety on their platforms and services?

- Companies should align their policies and practices with the UN Guiding Principles on Business and Human Rights, which must serve as the primary framework guiding the identification, assessment, and prevention of risks to HRDs. In addition, standards such as the Protocolo de la Esperanza provide important guidance, including the expectation that companies refrain from designing, developing, producing, or selling technologies that may be used to repress fundamental rights and freedoms, including the exercise of human rights defense.
- In this sense, companies should implement ongoing human rights due diligence, including the identification and assessment of actual and potential impacts on HRDs, with particular attention to heightened and context-specific risks, integrating practices for local languages and context-specific moderation; conduct regular human rights impact assessments prior to and throughout the lifecycle of their services, incorporating gender and intersectional perspectives and ensuring meaningful consultation with affected stakeholders, including HRDs; ensure access to effective remedy, through accessible, transparent, and timely grievance mechanisms that are responsive to the specific risks faced by HRDs; and strengthen transparency and accountability, including clear communication about how risks are identified and addressed, and how their systems may contribute to or mitigate harm.