Artificial Intelligence and Freedom of Expression: Contribution to the Joint Declaration Consultation

Introduction

This document presents the official contribution of Derechos Digitales to the public consultation organized by the international Freedom of Expression Mandate Holders: the United Nations (UN) Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression, and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information in Africa.

This consultation, that received inputs until July 28th of 2025[1], will inform a Joint Declaration on Artificial Intelligence and Freedom of Expression.

This contribution is grounded in our regional experience in Latin America and focuses on the impacts of AI and generative AI (GenAI) systems on freedom of expression, media pluralism, and democratic participation.

[1] The consultation was addressed in writing through a SurveyMonkey form, with a limit on contributions of up to 500 words or fewer. The consultation had eleven question. Those questions are included as a reference in this submission.

Risks and challenges to freedom of expression across the AI lifecycle: Perspectives from Latin America[2]

I. In the data collection and processing layer of AI/GenAI
a. For the press, a major challenge is the unauthorized, unpaid, and unacknowledged use of content they produce for the training of AI and Generative AI (GenAI) systems, particularly those developed for commercial or profit-driven purposes. This exploitation exacerbates the already precarious situation of media outlets whose business models have been undermined by Big Tech, some of which are also behind the most widely used generative AI models on the market.

b. From a freedom of expression standpoint, the intensive exploitation of internet users' data, often classified as "public data", has been used to unlawfully expand the training datasets of AI/GenAI systems. This practice poses serious risks, including state or corporate surveillance, the erosion of online privacy (a fundamental enabler of free speech in digital spaces), and a broader reduction in the expectation of privacy in the public sphere.

II. In the content generation layer (Text, Images, Audio) of AI/GenAI
a. For the press, a significant risk lies in the replacement of human input in newsrooms, a process already underway in several outlets where GenAI systems are used to assist or even substitute human writers in creating news reports, analyses, and other content. This trend deepens the labor crisis faced by media professionals.

b. A second risk for the press concerns the weakening and degradation of journalistic content due to the uncritical and growing adoption of GenAI systems. The outputs of these tools often jeopardize key journalistic standards such as veracity, accuracy, context, timeliness, and content quality.

c. A third risk relates to the use of GenAI systems as fact-checking tools, despite the dangers they pose such as hallucinations, reinforcement of biases, and decreased accuracy with more complex prompts. A broader risk tied to the widespread use of GenAI is its increasing

[2] Input to question 1: "What do you consider to be the biggest risks and challenges for freedom of expression, including media freedom, independence and pluralism, when it comes to the design, development and deployment of AI and in particular GenAI? Please consider the entire lifecycle of the AI (and GenAI) and specify the layer where the risk or challenge is to be identified. If possible, please also provide examples on what you, in your field of work, expertise or region, would consider to be the most important issue(s)."

substitution for traditional news consumption, which undermines public perception of the press as essential to democracy threatening media pluralism.

III. In the predictive and pattern recognition layer of AI/GenAI

a. At the level of output generation, risks to online freedom of expression emerge from the use of AI and GenAI by governments to classify, monitor, predict, and profile public discourse and internet users, particularly those perceived as critical or misaligned with current administrations. This phenomenon, increasingly observed in parts of Latin America, is often accompanied by coercive or policing measures. These can result in consequences such as censorship, self-censorship, the loss of freedom for certain forms of expression, and the stigmatization of specific groups, collectives, or political movements.

Key recommendations for addressing AI-related threats to freedom of expression and their intended stakeholders[3]

To address the risks outlined above, the Mandate Holders should develop targeted recommendations for the following stakeholders: AI developers and tech companies, governments and media organizations.

For AI developers:

- Mandate full transparency around the datasets used for training, including the inclusion of public and private data and the terms under which they were obtained.
- Require the disclosure of model performance metrics, methodologies for bias and accuracy calibration. As well as typical error types and use cases.
- Publish environmental, social, and democratic impact assessments, in line with instruments like the Escazú Agreement.

For governments:

- Require and implement mandatory human rights impact assessments throughout their development, deployment and use.

[3] Inputs to question 2: "What key recommendations should the Freedom of Expression Mandate Holders issue to respond to the risks and challenges you identified? For which stakeholders should the recommendations be developed?"

- Apply a moratorium or a ban on AI Systems that don't comply with basic human rights criteria.
- Prohibit the surveillance-based use of AI to classify or suppress online expression, especially for political purposes.
- Ensure that any AI systems procured by public institutions follow transparent procurement processes and uphold human rights standards.
- Guarantee the protection of journalistic sources, editorial independence, and media workers' rights in the face of GenAI adoption.

For media organizations:

- Strengthen editorial standards and human oversight in the use of GenAI tools.
- Invest in training journalists and editors to critically assess GenAI in an ethical, rights-based perspective.

Key recommendations on transparency in the design, development and use of AI and genAI[4]

For developers of generative AI (GenAI) models:

- Publish clear, open, and accessible information about the types of data used to train and fine-tune their systems. This should include explicit details on which public and private data are included in their datasets.
- Provide researchers with access to data regarding the performance of their AI models, including:
- Methodologies used to calibrate key factors such as veracity, accuracy, and bias;
- Data on the frequency and nature of errors generated by the model;
- Insights into predominant use cases of the model according to user profiles, among other metrics.

[4] Inputs to question 3: "What key recommendations should the Freedom of Expression Mandate Holders issue to tackle the role of transparency with regards to the design, development and use of AI and GenAI? Please consider providing inputs on what you believe to constitute the minimum required information that should be public, on how to guarantee its accessibility and intelligibility and/or on any other standard on transparency that you consider relevant."

- Produce and publicly disclose information on the impact on these tools in facilitated gender violence and disinformation specially in electoral processes, among others.
- Ensure the public availability and intelligibility of environmental impact data -in line with the standards set by the Escazú Agreement- regarding how AI models contribute to climate change and affect environmental sustainability. This should include:
- Information on the location and natural resource consumption of data centers;
- Details on the local deployment of infrastructure and compliance with requirements such as Prior Consent (or Consulta Previa), where applicable.
- Disclose comprehensive information about licenses acquired by governments, including the purpose of each license, its cost, duration, and the procurement method (e.g., direct contracting, public bidding, etc.).
- Develop and publish transparency and accountability mechanisms to assess and report the social, ethical, and human rights impacts of their models on different populations, especially vulnerable or marginalized communities.

AI and freedom of expression: opportunities worth protecting[5]

Despite the significant risks associated with the development and deployment of AI and GenAI, these technologies also offer meaningful opportunities to enhance freedom of expression, media freedom, independence, and pluralism provided that their design and use are grounded in human rights principles, transparency, and accountability.

One of the clearest opportunities lies in improving access to information. GenAI tools can facilitate the discovery, translation, and summarization of complex or inaccessible content, helping users engage with diverse sources of information across language and literacy barriers. For example, automated translation systems can make global news content accessible to minority language speakers, while text summarization tools can help users better understand lengthy policy documents or legal texts. When properly designed and

[5] Input to questión 4: "What do you consider to be the major opportunities for freedom of expression, including media freedom, independence and pluralism, when it comes to the design, development and deployment of AI and in particular GenAI? Please specify and, if possible, provide examples on what you, in your field of work, expertise or region, would consider to be the most important issue(s). Where relevant, please specify also what measures and guarantees must be put in place to preserve these advantages."

deployed, these tools can promote informed public discourse and reduce structural barriers to participation.

In the newsroom, GenAI can support -not replace- journalistic work. AI systems can assist with time-consuming tasks such as transcription, data visualization, or content tagging. This can free up reporters to focus on investigative journalism and in-depth analysis. In regions where media outlets face financial constraints or operate under pressure, such tools -when used ethically and under human oversight- could strengthen journalistic capacity and resilience.

In our research, we in fact show how a good practice in journalism is linked to the use of AI systems to enhance and expand the capacity to monitor cases and news related to femicide. When used under supervision and with methodological transparency, these tools can strengthen the capabilities of journalists and newsrooms.

However, for these opportunities to be realized sustainably and without compromising rights, certain measures and safeguards must be in place:

- Transparency and explainability: Developers must disclose the datasets, logic, and assumptions behind AI systems to enable public scrutiny, especially when deployed in public-interest contexts.
- Human oversight: GenAI should never replace editorial judgment or fact-check reporting. Media organizations should ensure that any use of GenAI supports journalistic ethics and includes human verification.
- Inclusive design and deployment: AI tools should be developed in consultation with diverse stakeholders -including media workers, civil society, and marginalized communities- to avoid reinforcing existing inequalities and biases.
- Public accountability and impact assessments: Governments and private companies must conduct and publish human rights impact assessments related to freedom of expression, media pluralism, and democratic participation.

Safeguarding inclusion: ensuring AI benefits marginalized voices[6]

[6] Input to question 5: "What guarantees should be put in place to ensure that the opportunities that AI offers for advancing freedom of expression benefit everyone, including women, minorities, vulnerable, underrepresented, and other marginalised groups and communities across the globe?"

First, it is urgent to advance regulations grounded in human rights. While there have been advances in legislation, these are not always aligned with human right standards, and in some cases may even undermine them. It is essential that regulatory frameworks include specific protections for press workers, and adopt a with a gender-sensitive perspective and intersectional lens. Moreover, a range of binding international standards remain applicable and must be upheld even in the absence of specific legislation.

Second, it is essential to promote transparency and accountability in the design and deployment of AI systems. This includes ensuring that datasets used to train AI are diverse, inclusive, and free from bias that may reinforce existing structural inequalities. Marginalised voices must be actively included in the development processes of AI technologies -not only as subjects of data, but also as decision-makers, developers, and evaluators-.

In fact, our research (https://ia.derechosdigitales.org/en/casos/) on the use of AI systems by Latin American states also reveals a strong appetite by governments for citizens' data, while automation is outsourced to private-sector actors whose operations and technological tools lack transparency and often do not align with local legal frameworks on data protection and transparency.

Third, digital literacy must be expanded globally. Marginalised communities are often left behind in terms of digital infrastructure and training, which limits their ability to participate fully in the digital public sphere. Ensuring access to safe, open, and inclusive digital spaces is critical for advancing freedom of expresssion.

Fourth, independent monitoring bodies, including civil society organizations, must be supported to conduct regular audits of AI systems. These bodies play a critical role in holding both governments and private actors accountable for discriminatory or harmful uses of AI, as well as its instrumentalization as a surveillance tool that enables broader practices infringing on the right to privacy and other fundamental freedoms. Whistleblower protections and accessible mechanisms for redress must also be established to address both abuses and unintended consequences.

Finally, special attention should be paid to linguistic diversity and cultural contexts to avoid replicating the dominance of certain languages and worldviews in AI-generated content. Freedom of expression must be protected not just in form, but in meaningful access to participation for all communities.

Safeguards for state deployment of AI in media and public discourse[7]

Media outlets face multiple challenges, including threats to their economic sustainability of and the reduced availability of independent press funding. Another concern involves the unauthorized and non-consensual exploitation of journalistic content by generative AI systems. Additionally, many journalists face precarious working conditions and high levels of violence, particularly in regions and countries where they are targeted by armed non-state actors or even by state agents.

In this context, the design and deployment of AI and generative AI systems by public and private actors may worsen or amplify these pre-existing challenges.

On the other hand, when states develop or deploy AI systems -especially in the context of mass surveillance- they risk violating fundamental rights, including the right to privacy and freedom of expression. Journalists and media workers, in particular, are disproportionately affected by such surveillance practices, which can lead to self-censorship, hinder investigative reporting, and endanger sources.

Furthermore, public-private partnerships involving AI must be subjected to robust transparency, oversight, and regulatory frameworks. Without clear safeguards, these collaborations may create opaque mechanisms for censorship, content moderation, and data collection, further threatening media independence and pluralism.

To mitigate these risks, it is crucial to establish human rights–based frameworks for the design and implementation of AI systems, including mechanisms for accountability, public consultation, and meaningful participation of journalists, civil society, and marginalized communities.

Assessment of international, regional and national AI governance instruments: Strengths, weaknesses and existing gaps in protecting freedom of expression and media pluralism[8]

[7] Input to question 6: "What are the specific challenges and needs for freedom of expression and freedom of the media that arise when AI systems are designed, developed and deployed by the State, either alone or within the framework of public-private partnerships?"

[8] Input to question 7: "A variety of international, regional and national instruments (among others: the UN Global Digital Compact, the Council of Europe Framework Convention on AI and Human Rights, the Africa Declaration on Artificial Intelligence, the OECD AI principles, the UNESCO Recommendation on the Ethics of Artificial Intelligence, the EU AI Act, the Ministers and high-level authorities of science and technology of the Americas' Declaration and Plan of Action: Towards the Safe, Secure and Trustworthy Development and Deployment of Artificial Intelligence in the Americas) have already been

Gaps in relation to Latin America:

- The UN Global Digital Compact includes only a brief and insufficient reference to the protection of workers displaced, replaced, or subjected to more precarious conditions by the mass adoption of AI in their workplaces. It merely emphasizes the need for upskilling and reskilling. Its section on Information Integrity highlights the importance of quality information and media literacy but lacks concrete strategies or targets to protect journalism as a public good or ensure the survival of news media beyond vague or general statements. The GDC also enables the military use of state AI systems, which poses an additional threat to the protection of freedom of expression, especially in socially turbulent contexts.
- The OECD AI Principles (2024) include general references to combating misinformation and safeguarding information integrity but fail to outline specific measures to protect media workers, support public funding for journalism, ensure physical safety for journalists, or explore fair compensation mechanisms from Big Tech for press content.
- The OAS Declaration and Plan of Action on AI lacks any reference to protecting the media and its workers. It omits how to address state deployment of AI systems for surveillance, censorship, or monitoring of journalists and the general public.

Overall, many of these frameworks emphasize principles without clear objectives, metrics, or compliance mechanisms. As voluntary frameworks -such as those from OECD and UNESCO- they create a fragmented and uneven AI governance landscape across Latin America.

Is worth mentioning that the EU AI Act is influencing regulatory debates in the region, but it does not offer concrete protections for journalistic work or workers, nor robust safeguards for online freedom of expression in contexts where AI is used by states for surveillance or censorship. Also, there is a need to contextualize regulatory processes to local settings, as the export and adoption of legal models -particularly from Europe- can be problematic.

Critically, these frameworks often fail to strengthen local institutional capacity or address issues like press content IP protection from generative AI developers especially in Latin American countries with limited negotiation power.

---

put in place to deal with the impact of AI on freedom of expression and, to a certain extent, on the media. What are the main strengths and weaknesses of these instruments and what are the gaps still to be filled?"

Concentration of power in generative AI: risks to freedom of expression, media independence and pluralism[9]

Vertical integration: Big Tech holds concentrated control over the digital sphere -from data collection, to model training, to content distribution via search engines, app stores, and social platforms- leaving no space for independent oversight or competition.

Structural dependency: newsrooms are increasingly reliant on AI tools and cloud infrastructure owned by the same companies extracting their content and advertising revenue.

Search engines and abusive practices: absence of traffic redirection to news sites by search engines that use GenAI to summarize media outlets content, which contributes to the growing perception that news media are no longer necessary.

Marginalization of local and independent media: media outlets in the Global South, particularly in Latin America, face asymmetrical power dynamics with tech giants who use their content to train GenAI systems without compensation or consent, while weakening their ability to monetize their own work.

Gatekeeping of information ecosystems: a handful of powerful companies (e.g., OpenAI/Microsoft, Google, Meta, Amazon) dominate the full AI stack, from infrastructure (cloud computing) to foundation models and distribution platforms. This gives them disproportionate control over what content is generated, amplified, or suppressed.

Amplification of biases and linguistic inequality: Dominant GenAI models tend to prioritize content in English and reflect Western worldviews. This undermines pluralism by making local languages, cultures, and alternative perspectives less visible online.

Predatory licensing or non-Compensatory Use of Content: current licensing practices often exclude local or small publishers, creating economic dependency and reinforcing monopolies.

Selective compliance with court rulings by Big Tech: as well as the use of intimidating diplomatic and legal practices by authorities in the Latin American countries where they operate.

[9] Input to question 8: "What is the key impact of the current concentration of power across the AI, and in particular the GenAI, stack on freedom of expression and media freedom, independence and pluralism? What market and business dynamics (including market concentration, vertical integration, structural or economic dependencies) need to be addressed to solve the issue(s) and what are the specific instruments to be used? Please specify and, if possible, provide examples on what you, in your field of work, expertise or region, would consider to be the most important issue(s)."

Ensuring human agency in the design and deployment of AI and generative systems[10]

To ensure that GenAI strengthens rather than undermines human agency, it is essential to embed human rights and democratic values into every stage of AI system development and deployment. This is particularly urgent in regions like Latin America, where historical inequalities and weak regulatory oversight increase the risk of abuse and exclusion.

First, meaningful transparency must be guaranteed. GenAI systems should clearly disclose their nature (for example, when a user is interacting with a machine), the sources of their training data, and how they make decisions. Users must have the right to opt out of interacting with AI-generated content, especially in sensitive contexts such as public information, journalism, or education.

Second, democratic oversight and public participation are essential. The development of AI governance frameworks must include civil society, academia, and marginalized communities not just industry and government actors. Without inclusive participation, AI risks reinforcing dominant power structures and suppressing alternative or dissenting voices.

Third, algorithmic explainability and contestability are crucial. Users must be able to understand, challenge, and appeal decisions made or influenced by GenAI, particularly in public services (the use of GenAI by the judiciary is spreading in Latin American countries at a fast pace (https://www.derechosdigitales.org/wp-content/uploads/Derechos-Digitales_AI-and-justice_final.pdf), employment, and content moderation)

Fourth, media literacy and digital education should be scaled up. People need tools to critically evaluate AI-generated content and understand how these systems affect their choices, opinions, and freedoms.

[10] Input to question 9: "What are, in your opinion, the key safeguards and measures to be put in place to guarantee that the design, development and deployment of AI, and in particular GenAI, guarantee and enhance human agency, rather than limiting or suppressing it? Please specify and, if possible, provide examples on what you, in your field of work, expertise or region, would consider to be the most important issue(s)."

Finally, cultural and linguistic diversity must be actively protected. GenAI should support local languages and knowledge systems, rather than centralizing cultural narratives in dominant global languages and norms.

The role of AI in the media sector: legitimate uses and critical boundaries[11]

AI -especially Gen AI- should serve as a supportive tool in the media sector, not a substitute for human judgment, creativity, or ethical responsibility. In particular, certain red lines must be respected.

AI should never be used to manage journalist–source interactions. Recent statements by OpenAI CEO Sam Altman have confirmed that ChatGPT does not offer confidentiality guarantees for interactions between journalists and their sources (or between lawyers and clients). This represents a serious threat to investigative journalism and source protection, particularly in high-risk environments like Latin America. Journalists must be made aware of this risk when using GenAI tools.

AI should also never be used to fully write original news content or opinion pieces. Doing so undermines editorial independence, journalistic responsibility, and the authenticity of public discourse. Likewise, delegating the creation of images or illustrations to GenAI can endanger the livelihoods of artists, photojournalists, and designers, and may perpetuate bias or misinformation if used without ethical oversight.

However, AI can be encouraged for certain supportive and assistive functions, such as:

- Improving accessibility through automated captions, image descriptions, and translation;
- Summarizing long reports or documents to aid newsroom efficiency;
- Detecting content manipulation or misinformation;
- Supporting language diversity through localized content generation.

In Latin America, where many newsrooms operate under severe resource constraints, GenAI can help optimize workflows --but this must never come at the expense of journalistic ethics, authorship, or employment rights--.

[11] Input to question 10: “What are the main purposes that AI, and in particular GenAI, should serve with regards to the media sector? Are there any tasks, processes and functions in the media sector that should never be delegated to or performed by AI systems? Please specify and, if possible, provide examples on what you, in your field of work, expertise or region, would consider to be the most important goals and/or red lines.”

Clear ethical guidelines, transparency in tool design, and active collaboration with journalists and media workers are essential to ensure GenAI supports -not replaces- human editorial agency and media pluralism.

Examples of good practices for AI and GenAI use in journalism and media[12]

The project **"Datos Contra el Feminicidio"** (Data Against Feminicide) offers a tool for understanding, supporting, and promoting the critical use of AI. The developed a machine learning platform to detect cases of feminicide in press reports, in ongoing collaboration with other activists.

The team is making progress towards the use of large-scale language models to expand analysis of press coverage around feminicide, given that most independent projects on these subjects are based on press monitoring and require tools that are more situated to the local context.

They use guides and manuals for best practices in the coverage of feminicide and gender-based violence, they are designing a tool that not only points out areas where the media could be reproducing biases or stereotypes, but also suggests recommendations for producing a report that is committed to human rights and a gender perspective. See: https://www.derechosdigitales.org/wp-content/uploads/Glimpse_2024_ENG.pdf

La Silla Vacía, a Colombian media outlet (https://www.lasillavacia.com/curso-inteligencia-artificial-una-solucion-para-aumentar-su-productividad/) developed an AI course for journalists that critically explores its capabilities and examines the benefits it can offer to complement and support the work of journalists and newsrooms.

Aos Fatos, a Brazilian initiative (https://www.aosfatos.org/) uses AI to assist human fact-checking efforts. They developed a fact-checking bot trained with their own data to generate and distribute journalist-verified information from Aos Fatos on platforms such as Telegram and WhatsApp - along with Escriba, an automatic transcription service.

[12] Input to question 11: "Can you identify best practices – whether from the private sector, governments or civil society organisations – related to the use of AI (and GenAI) in the media ecosystem? If so, please indicate them. Links to these materials are appreciated."