WITNESS

WITNESS submission to the Global Dialogue on AI Governance

7 May 2026

WITNESS submits this present document to the Global Dialogue on AI Governance consultation.

It advocates for a human rights-based approach to AI governance that prioritizes equitable participation, transparency, and accountability. Our organization also identifies key thematic priorities for urgent action, including ensuring safe, secure, and trustworthy AI to combat media manipulation like deepfakes, promoting the protection and promotion of human rights, and establishing interoperable governance approaches. The submission emphasizes the critical need to explicitly address emerging issues such as AI-enabled non-consensual intimate imagery (NCII) and recommends strengthening inclusive participation by elevating the voices of civil society, frontline defenders, and the Global Majority in the Dialogue.

Response to the submission

Priorities

8. In your opinion, what outcomes would make the first Global Dialogue on AI Governance a success? (Max. 300 words) *

A successful first Global Dialogue on AI Governance would advance equitable and inclusive participation while delivering clear, actionable outcomes.

A key priority is ensuring that co-chairs and panelists reflect geographical balance and diversity, including strong representation from civil society in the Global South. This strengthens both legitimacy and the quality of discussions by grounding them in lived realities. Beyond representation, the Dialogue should include dedicated opportunities for civil society interventions during sessions, rather than limiting participation to pre-inscribed formats.

Accessibility is also essential. Hybrid participation, timely communication, and inclusive formats can reduce barriers for under-resourced stakeholders, particularly in developing contexts.

Regarding speakers, success would require a transparent and inclusive nomination process, allowing Member States and stakeholders to propose candidates. Special

attention should be given to underrepresented regions and communities, especially those most affected by AI-related risks. Ensuring balanced thematic coverage, including capacity-building and bridging AI divides, will further enhance relevance.

Ultimately, success would mean that these inclusive processes lead to sustained collaboration and trust-building, ensuring the Dialogue reflects diverse realities and contributes to more effective international cooperation on AI governance.

9. From your perspective, which of the following thematic areas identified by the General Assembly Resolution 79/325 for the AI Dialogue reflect your priorities for urgent action and active engagement by your entity? Please select up to 4 priorities.*

• Safe, secure and trustworthy AI
• Interoperability of governance approaches
• Protection and promotion of human rights
• Transparency, accountability, and human oversight

10. Please briefly explain your selection. (Max. 300 words)*

The selected themes align closely with the practical, rights-based approach advanced by WITNESS, particularly our focus on audiovisual evidence, media integrity, and the human rights impacts of generative AI.

We trust that safe, secure, and trustworthy AI reflects the urgent need to ensure that synthetic media tools are not weaponized to mislead, harass, or undermine the evidentiary value of frontline documentation. WITNESS has documented how manipulated or AI-generated content can erode trust in authentic footage—the "liar's dividend"—making technical safeguards, content provenance, and responsible deployment essential for protecting the truth.

Added to that, interoperability of governance approaches is critical because media ecosystems and human rights documentation are inherently cross-border. Our engagement with initiatives like the Coalition for Content Provenance and Authenticity (C2PA) highlights the importance of shared, open technical standards that work across platforms and jurisdictions to verify content authenticity and maintain the chain of custody for human rights evidence.

Another foundational issue is the protection and promotion of human rights. WITNESS grounds its work in ensuring that responses to AI risks—such as content moderation or automated detection—do not inadvertently undermine freedom of expression, access to information, or the safety of those documenting abuses. We emphasize that AI-enabled harms, including misinformation and non-consensual intimate imagery, disproportionately affect marginalized communities, activists, and journalists, and must be addressed through local cultural and linguistic contexts.

Lastly, transparency, accountability, and human oversight are essential to operationalize these principles. We advocate for clear disclosures around AI-generated content, meaningful accountability for platforms and developers, and human-centered review processes that prevent automated systems from introducing bias or silencing vulnerable voices.

Taken together, these themes form a coherent framework that reflects both the risks to our information ecosystem and the transformative potential of AI when governed through a human rights lens.

10. In your opinion, are there any cross-cutting or emerging issues not captured by the listed themes above? If so, please explain. (Max. 300 words)

One important cross-cutting issue that could be made more explicit is the rise of AI-enabled gender-based harms, particularly AI-generated non-consensual intimate imagery (NCII). While other submissions reference technology-facilitated gender-based violence (TFGBV), the category is broad and risks obscuring the speed and scale at which specific harms are evolving. AI-generated NCII—often referred to as "deepfake pornography"—is one of the most acute examples. It disproportionately targets women and marginalized groups, is increasingly easy to produce with off-the-shelf tools, and can spread rapidly across platforms, creating severe and often irreversible personal, social, and economic consequences.

What makes this issue especially significant is how it cuts across multiple governance domains. It raises questions about platform accountability (content moderation and takedown mechanisms), legal frameworks (definitions of consent, liability, and jurisdiction), and technical safeguards (watermarking, provenance, and detection tools). It also exposes gaps in existing remedies: victims often face fragmented reporting systems, slow response times, and limited legal recourse, particularly in cross-border contexts.

In this regard, explicitly naming AI-generated NCII within the broader TFGBV category would strengthen the framework and allow for discussions in real life and rapidly scaling harm. It would also help align policy responses with emerging technical approaches, such as content provenance standards (e.g., efforts like the Coalition for Content Provenance and Authenticity (C2PA), and WITNESS efforts on deepfakes detection), which can support authenticity verification, even if they are not a complete solution.

More broadly, this example highlights a recurring governance gap: the lag between general policy categories and highly specific, fast-evolving AI harms. Addressing this gap may require more adaptive governance mechanisms that can quickly identify, name, and respond to new risk patterns as they emerge, rather than relying solely on high-level classifications.

WITNESS

Impact of AI governance

12. How are the governance gaps and related developments/advances in the thematic areas you selected above affecting your country, region, or sector? Please highlight the most significant challenges and opportunities. (Max. 300 words)

Deepfakes & Remedying Harms: Generative AI has increased the scale and sophistication of media manipulation, eroding trust. There is an urgent need for aligned policies on deepfake detection, particularly for cases involving Non-Consensual Intimate Imagery (NCII).

On access to detection tools, we must expand equitable access to knowledge, ensuring the right to reliable information and creating shared spaces for discussing technical challenges and growing misuse complexity. In order to facilitate this, some necessary changes include: (a) Incorporate sociotechnical considerations into regulations to ensure detection evaluations reflect real-life applications and design accountability mechanisms that safeguard fairness across the system lifecycle; and (b) Prioritize collaborative, global multistakeholder engagement to ensure tools respond to diverse needs and protect against adversarial attacks.

Transparency and Provenance: WITNESS would welcome if the Global dialogue helps us advance watermarking and labeling approaches that protect rights without enabling surveillance. We need more alignment across industry policies and regulation establishing concrete guidelines for downstream transparency and content provenance.

On this note, we would like to highlight that AI transparency should be enabled on the basis of respecting the following principles: (a) Privacy protection: markers must not embed personally identifiable information (PII) by default; (b) "How, not who": focus on the chain of custody and tools used to alter content rather than a traceability model identifying individual users.

International cooperation on AI governance

13. What role can the AI Dialogue play in advancing international cooperation on AI governance? (Max. 300 words)

In a world with fragmented policy and regulatory-related efforts, The Global Dialogue on AI Governance can play a pivotal role as a bridge between initiatives, helping align technical, policy, and human rights–based approaches across regions and stakeholders.

In this sense, we trust that the Dialogue can advance shared norms and common language around AI governance grounded in international human rights standards and

guiding principles such as transparency. By bringing together governments, civil society, and industry, it can help reduce policy fragmentation and support more coherent approaches to issues such as accountability, and risk management.

Lastly, we see the dialogue with a good potential to strengthen inclusive participation and trust-building, particularly by ensuring that perspectives from the Global Majority, frontline defenders, and affected communities are meaningfully integrated. This is essential for legitimacy and for identifying governance gaps that may not be visible from dominant policy centers.

14. What are some of the existing initiatives, partnerships, or mechanisms that the AI Dialogue should build upon or connect with, and what added value could the AI Dialogue bring? (Max. 300 words)

The following initiatives are good examples of spaces and networks that should be connected and considered in shaping the agenda of the dialogue. These spaces have long term experience in shaping standards and frameworks towards AI leveraging on multi-stakeholder discussions on topics of expertise.

- The Partnership on AI is a multi-stakeholder initiative that brings together companies, civil society organizations, academic institutions, and media actors to advance responsible AI. It focuses on developing best practices, conducting research, and fostering dialogue on issues such as transparency, fairness, safety, and the societal impacts of AI systems.
- The Coalition for Content Provenance and Authenticity (C2PA) is a cross-industry effort to establish technical standards for content provenance. Its work enables digital media to carry verifiable information about origin and edits, helping users and platforms assess authenticity and address challenges such as misinformation and AI-generated content.
- The Global Network Initiative is a multi-stakeholder initiative that brings together technology companies, civil society, investors, and academics to protect and advance freedom of expression and privacy in the digital age. It develops principles, implementation guidelines, and accountability mechanisms to help companies navigate government demands and human rights risks.

Inclusive participation

15. How can different stakeholders contribute to the AI Dialogue? Please share recommendations for the format and structure of the AI Dialogue. (Max. 300 words)

As highlighted in the MAP-AI submission to this same consultation: The Global Dialogue should facilitate robust, meaningful and continuous engagement by civil society and other stakeholders.

In this sense, we support and echo the following recommendations:
1. Implement formats that effectively protect and promote meaningful multistakeholder throughout the Global Dialogue: Participation mechanisms should be well-structured, facilitating early and ongoing consultation throughout the design, deliberation and implementation phases of the Global Dialogue. Formats should allow non-governmental and governmental stakeholders to respond to each other in real time, rather than purely ad hoc and non-dialogue formats. Feedback loops should be established to demonstrate how contributions influence decisions: this requires cyclical reporting on the contributions received, how they were or were not incorporated into the Dialogue's outputs, and signalling areas of ongoing consultation.
Concretely, enhanced engagement may be achieved through focusing Thematic Discussion sessions on specific pre-identified topics, identified through consolidation of consultation input. Thematic Discussions could be linked with intersessional Working Groups, designed to ensure follow-up and transmit the Dialogue's outcomes to other UN fora, such as the IGF.

2. Ensure all stakeholders, and specifically traditionally excluded groups, have the necessary information, resources, skills, and equitable access to meaningfully participate in the Global Dialogue. This involves centralising all information, providing timely notice and sharing of calls for inputs or consultation events, clearly communicating timelines and milestones for input, and making use of satellite and other intersessional mechanisms to convene diverse stakeholders. It requires elevating the voices of traditionally excluded or underrepresented groups, ensuring hybrid and accessible formats, offering financial and logistical support to aid participation, and taking concrete steps to ensure accessibility. Member States should host national consultations to shape their national positions.

16. Which voices, communities, or perspectives are currently underrepresented in global discussions on AI governance? How could they be included? (Max. 300 words)

Global discussions on AI governance must ensure meaningful participation of civil society, women human rights defenders, LGBTQIA+ rights advocates, labour organisations, frontline defenders, journalists, and communities in the Global Majority across the entire AI policy cycle—from agenda setting to design and deployment, as well as evaluation and assessment frameworks.

A concerted effort is needed to engage affected communities and provide the accommodations required for full participation in the Dialogue. We reiterate the procedural recommendations outlined by Derechos Digitales and co-signed by over thirty organisations, particularly those aimed at elevating traditionally excluded voices.

To ensure meaningful inclusion, several elements are critical. First, governments and

non-government actors should provide financial and logistical support based on need, enabling participation from under-resourced stakeholders. Second, the Dialogue should adopt hybrid and accessible modalities, including options for low-bandwidth participation, and scheduling that accommodates different time zones and working contexts.

Third, it is essential to promote gender balance and intersectional representation, recognizing that risks and barriers are unevenly distributed. Fourth, safety and security measures must be in place to prevent, monitor, and respond to reprisals or intimidation linked to participation, particularly for defenders and journalists who may face heightened risks.

Finally, careful consideration should be given to the selection of host countries and venues, ensuring conditions that allow safe, non-discriminatory, and barrier-free participation, including for those facing visa, mobility, or immigration challenges. We believe these steps would help ensure that the Global Dialogue reflects diverse lived realities and supports inclusive, effective, and rights-respecting AI governance.

17. What innovative engagement formats could most effectively foster meaningful and dynamic engagement during the AI Dialogue? (Max. 300 words)

On this point, we also support the following recommendations made by the MAP-AI project submission:

1. Convene thematic, cross-sectoral and sector-specific consultations with the Co-Facilitators, Secretariat and the non-governmental stakeholder community early in the process and on an ongoing basis to provide feedback to the development of the Co-Chair summary Outcome and other outputs.
2. Appoint a civil society representative liaison to support the Co-Chairs to facilitate meaningful multistakeholder engagement ahead of, during, and following the Global Dialogue. Consider leveraging the support of this liaison when drafting the Co-Chair Summary Outcome of the Global Dialogue (see e.g. the Multistakeholder Advisory Group of the IGF and the WSIS+20 Informal Multistakeholder Sounding Board).
3. Structure session formats that allow civil society and other non-governmental and governmental stakeholders to respond to each other in real time. There should be a minimum of 50% rotating time allocated for non-governmental stakeholders to engage in dialogue with each other and with governments.
4. Focus Thematic Discussion sessions on specific pre-identified topics using the consultation input consolidation. Consider linking Thematic Discussions with intersessional Working Groups to ensure follow-up and transmit the Dialogue's outcomes to other UN fora, such as the IGF.

Good practices and policy approaches

WITNESS

18. Please share examples of policies, practices, platforms, or approaches that promote effective AI governance or offer concrete solutions to addressing its challenges. (Max. 300 words)

Effective AI governance is emerging through an ecosystem of standards, institutional practices, and technical mechanisms that prioritize human rights and accountability. WITNESS actively shapes this landscape by bridging technical development with frontline advocacy to protect vulnerable voices from AI-induced harms.

Institutional frameworks, such as those advanced by the Council of Europe (CoE), emphasize a human rights–based approach that WITNESS operationalizes through participation in global policy debates. This includes leading advocacy for transparency and rights-respecting disclosures within the EU AI Act and contributing to the Digital Content Provenance Act. To move governance from abstract principles to practice, we trust that methodologies like HUDERIA assessments and the TRIED Benchmark can help ensure that AI detection tools are effective, equitable, and resilient to real-world challenges like low-resolution media.

Technical standards like the Coalition for Content Provenance and Authenticity (C2PA) are critical for verifying digital origin. As a co-chair within the C2PA, WITNESS works to ensure these standards remain open, accessible, and designed to protect privacy rather than enable surveillance. This is complemented by the Deepfakes Rapid Response Force, which provides direct support to journalists and human rights defenders by analyzing suspected AI-generated content in high-stakes contexts like elections and conflicts.

Ultimately, effective AI governance requires an integrated ecosystem: technical standards like C2PA for provenance, institutional safeguards like CoE frameworks, and operational tools like HUDERIA assessments. The most promising solutions combine verifiability, rights-based oversight, and practical implementation to protect truth and maintain the credibility of frontline voices in an increasingly automated information environment.