

## **Submission to the United Nations Global Dialogue on AI Governance**

*Institute for Human Rights and Business (IHRB)*

May 2026

IHRB welcomes the opportunity to provide input to the first session of the United Nations Global Dialogue on AI Governance, to be held on 6 and 7 July 2026 in Geneva. This submission focuses on responses to consultation questions relating to the Global Dialogue's thematic clusters on safe, secure, and trustworthy AI systems, and on the importance of respect for and protection and promotion of human rights in the field of AI; and the transparency, accountability and robust human oversight of artificial intelligence systems in a manner that complies with international law.

### **Consultation question: In your opinion, what outcomes would make the first Global Dialogue on AI Governance a success?**

A successful first Global Dialogue on AI Governance would be both inclusive and operational - moving from aspirational statements to coordinated, measurable commitments grounded in shared responsibility. Critically, the Global Dialogue must establish that AI companies developing and deploying the most consequential systems bear responsibilities that voluntary self-regulation alone cannot discharge. Overall success requires progress on:

1. **Operational credibility** - A shared baseline of principles and standards is needed with explicit government and corporate endorsement. This means agreement on minimum safeguards for high-risk AI systems, including mandatory pre-deployment risk disclosure, independent auditing, and structured red-teaming, that would be binding on companies, not merely encouraged. The persistent gap between corporate AI ethics commitments and verifiable practice must be directly addressed.

2. **Meaningful convergence** - Across regulatory jurisdictions, greater alignment on which AI applications warrant heightened oversight, and what corporate obligations that oversight entails, would represent genuine steps towards coordinated governance. This includes convergence on incident reporting, liability frameworks, and the conditions under which AI deployment should be paused or prohibited.
3. **Substantive inclusion** - Voices from low- and middle-income countries, civil society, labour, and affected communities must shape outcomes, not merely observe them. Excluded voices carry knowledge about harms, including harms generated by specific corporate practices, and should be part of discussions on solutions. Permanent structural mechanisms that include this diversity of perspectives, not just one-off consultations, are required.
4. **Institutional architecture** - The Dialogue must define its relationship to existing multilateral processes, and establish follow up with clear deliverables. Crucially, it must address what leverage exists when major AI companies decline to participate or comply — a governance architecture silent on corporate non-participation will lack credibility. AI companies must be prepared to accept that the scale of their societal influence demands a commensurate level of public accountability.

**Consultation question: 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, and please briefly explain your selection.**

The four selected priorities listed below address the conditions under which AI systems pose systemic, cross-border, and potentially irreversible risks — to individuals, institutions, and societies. Both states and AI companies bear responsibility for meeting them.

**Safe, secure, and trustworthy AI is foundational.** Without robust safety practices, cybersecurity protections, and reliability standards, AI systems can produce large-scale harms through accidents, malicious use, or loss of control in high-stakes contexts. Trust must be built and verified by independent third parties, through mandatory incident disclosure, and against publicly accountable standards.

**Interoperability of governance approaches is essential** where AI systems, data flows, and supply chains cross borders. But regulatory fragmentation does not affect all countries equally: it systematically advantages jurisdictions where major AI developers operate and disadvantages those with weaker regulatory capacity. Aligning core concepts including risk classifications, audit standards, and incident reporting, is not only a technical imperative but an equity one as well.

**Protection and promotion of human rights must be central, not peripheral.** AI systems increasingly shape access to work, finance, healthcare, and information. Without explicit safeguards, they can entrench discrimination, enable surveillance abuses, and restrict fundamental freedoms. Critically, affected individuals must have access to effective remedy when AI causes harm — a right that current corporate and regulatory frameworks rarely guarantee. The populations most exposed to these risks are also those with the least governance capacity to address them.

**Transparency, accountability, and human oversight are distinct and non-negotiable.** Transparency is an information obligation, enabling scrutiny and informed decision-making. Accountability is a legal and institutional relationship, ensuring that states and companies answer for harms and compliance failures. Human oversight is an operational safeguard, providing a necessary check on automated systems, particularly in high-stakes and increasingly agentic AI contexts. These distinctions are critical and should not be conflated.

**Consultation question: In your opinion, are there any cross-cutting or emerging issues not captured by the listed themes above? If so, please explain.**

Addressing the issues below would strengthen the Global Dialogue by linking high-level principles to structural incentives, real-world impacts, and enforceable outcomes.

1. **Concentration of power across the AI value chain.** A small number of companies control critical infrastructure (compute, data, foundation models), shaping market access, innovation pathways, and governance norms. This raises competition, dependency, and sovereignty concerns, especially for low- and middle-income countries, and affects who can meaningfully comply with or influence governance regimes.
2. **Labour and economic impacts require explicit focus.** Beyond job displacement, AI is reshaping working conditions through algorithmic management, surveillance, and the hidden workforce behind data labeling and content moderation. It is also changing the way work is

done, and is influencing the trajectory of career prospects across a range of fields. Governance should address fair work, collective bargaining implications, and equitable distribution of productivity gains. Further research is needed to rethink, reorient, and restructure education so as not to widen or sharpen inequities.

3. **Environmental sustainability is an emerging but material issue.** The energy, water, land, and resource demands of large-scale AI systems have climate and local ecological impacts. Without standards for measurement and disclosure, sustainability risks remain opaque and ungoverned.
4. **Information integrity and democratic resilience cut across safety, rights, and accountability but deserve prioritization.** Generative AI amplifies risks of disinformation, fraud, and erosion of shared reality, particularly during elections and crises. Its dependence on accessible materials in dominant languages risks marginalizing languages spoken less widely.
5. **Liability and access to remedy remain underdeveloped.** Affected individuals often lack clear pathways to contest decisions or obtain redress, especially in cross-border contexts involving multiple actors in the AI supply chain.
6. **Measurement and assurance is a practical gap.** The AI field lacks widely accepted, independent methods to evaluate compliance with safety, fairness, and rights standards.