United Nations

XRSI INTERVENTION

UN GLOBAL DIALOGUE ON AI GOVERNANCE(2026)

From Principles to Proof in AI Governance
Intervention Submission | Public Brief for Global Stakeholders, Partners, and Policymakers

Prepared BY : Kavya Pearlman

XRSI™
Human Intelligence In The Loop

CONTEXT

The world does not have an AI capability gap. It has a trust vacuum.

The United Nations General Assembly established the Global Dialogue on Artificial Intelligence Governance on 26 August 2025 as a platform for governments and stakeholders to advance international cooperation on AI governance. The Dialogue is designed to enable the exchange of best practices, lessons learned, and coordinated approaches to governing AI systems through open, transparent, and inclusive engagement. The first convening will take place in Geneva on 6 and 7 July 2026, followed by a second session in New York in 2027. Stakeholders were invited to submit inputs and recommendations to inform this process.

This document represents XRSI's intervention submission to the UN Global Dialogue on AI Governance.

XRSI™
Human Intelligence In The Loop

WHY THIS MATTERS NOW

AI is no longer just a tool. It is becoming an actor.

Artificial intelligence is transitioning from tools to agentic systems capable of planning, acting, and coordinating across environments. With the emergence of world models, these systems can simulate and act within complex realities. At the same time, AI is shaping geopolitical dynamics, influencing cyber operations, information ecosystems, and decision making in high stakes environments. These systems are also entering homes, interacting with children, and influencing cognition, behavior, and identity. **Capability is accelerating. Governance is not.** This convergence has created a widening trust vacuum where capability is advancing faster than governance.

Governance MUST move from principles to proof.

XRSI™

Human Intelligence In The Loop

OUTCOMES FOR A SUCCESSFUL GLOBAL DIALOGUE

The outcome must be operational, not aspirational.

A successful first Global Dialogue must establish minimum global guardrails for agentic AI and world models, with enforceable requirements for human intervention, decision traceability, and accountability by design. It must enable verifiable transparency rather than self attestation, recognize high risk contexts including conflict environments and systems interacting with children, and define shared accountability across governments, industry, and civil society.

Success is not alignment on principles. It is alignment on implementation.

XRSI™
Human Intelligence In The Loop

PRIORITY AREAS FOR ACTION

Agentic systems without enforceable control create systemic risk.

Transparency, Accountability, and Traceability

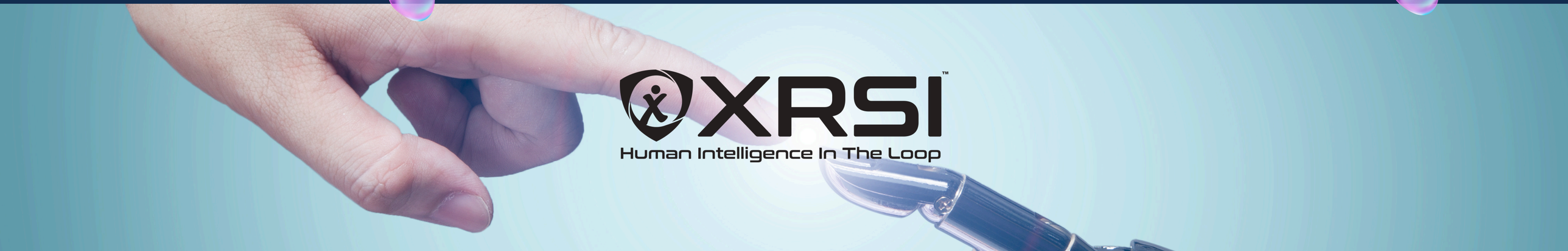
to make data usage and decision pathways visible, auditable, and verifiable

Safe and Trustworthy AI with Human Intelligence in the Loop

to ensure systems are reliable, decisions are controllable, and meaningful intervention remains enforceable through real human authority

Interoperable Governance anchored in Human Rights

to enable consistent safeguards across jurisdictions while protecting dignity, agency, and vulnerable populations



CROSS CUTTING AND EMERGING ISSUES

1 **Agentic AI and autonomy governance**

where systems plan, act, and coordinate across environments, requiring defined limits of autonomy and enforceable human intervention

2 **World models and simulation driven decision making**

that rely on internal representations of reality, creating risks of misalignment, emergent behavior, and decisions that are difficult to audit

3 **Data exposure and inferred data risks**

where systems generate sensitive insights beyond collected data, including behavioral profiling, biometric inference, and synthetic outputs

4 **Child and developmental safety in AI mediated environments**

where systems influence cognition, emotional development, and identity, requiring the highest standards of governance and protection

A persistent gap remains between stated principles and implemented controls. Governance is often declared, but not demonstrated. This requires mechanisms for continuous monitoring, auditability, and verifiable compliance embedded directly into system design and deployment.

IMPACT OF GOVERNANCE GAPS

The cost of inaction is already visible.

AI deployment is outpacing governance maturity, leaving limited visibility into data flows, decision logic, and system behavior, which drives operational risk and erodes trust. Fragmented regulations and standards create inconsistent safeguards and make accountability difficult across jurisdictions. Agentic systems add unpredictability and weaken effective human intervention. At the same time, demand is rising for standardized, operational governance that translates principles into verifiable controls. Embedding governance into system design enables proactive risk management, especially for systems that directly interact with individuals, including children.

The opportunity is clear : Move from fragmented approaches to coordinated, verifiable governance.

XRSI™

Human Intelligence In The Loop

THE ECOSYSTEM EXISTS. IT IS FRAGMENTED.

AI for Good

unesco

1

Abundance without integration:

AI-powered platforms can connect gig workers with opportunities, offering greater flexibility and control over their work schedules.

2

Principles without proof:

AI-driven platforms create new opportunities for gig workers in areas like data annotation, content moderation, and AI-assisted tasks.

3

Opportunity for connective tissue:

The rise of AI in the gig economy raises questions about job security, fair wages, and worker rights.

The goal is not to replace existing efforts, but to connect, operationalize, and validate them.

ROLE OF THE AI DIALOGUE IN INTERNATIONAL COOPERATION

From principles to observable system behavior	By enabling a global signal system where stakeholders share insights on performance, failures, and edge cases, creating collective visibility into risk
From isolated claims to comparable governance maturity	Through structured disclosures of how organizations implement oversight, Human Intelligence in the Loop, and accountability
From fragmented testing to shared evaluation environments	Where agentic systems and high risk use cases, including those involving children, are assessed collaboratively before large scale deployment
From static commitments to continuous verification	By establishing mechanisms for ongoing monitoring, auditability, and evidence based reporting

This approach shifts global cooperation from alignment in principle to coordination in practice. Trust is not assumed. It is built, observed, and maintained collectively.

XRSI™
Human Intelligence In The Loop

STAKEHOLDER CONTRIBUTION AND INCLUSION

Governance must be distributed, coordinated, and inclusive.

Governments define policy direction and risk thresholds; the private sector implements systems and controls; academia develops evaluation methods and research; civil society ensures alignment with societal impact; and standards bodies translate these inputs into measurable controls. At the same time, those most affected remain underrepresented, including children and youth, caregivers and families, engineers and auditors, communities in high risk environments, and interdisciplinary experts in human development. Inclusion must be structured and tied to real world use cases rather than symbolic participation.

Governance fails when those most affected are not meaningfully included in shaping it.

XRSI™
Human Intelligence In The Loop

INNOVATIVE ENGAGEMENT FORMATS

Immersive sandbox environments

Using virtual worlds to simulate real conditions and test safeguards

Live governance simulations

To model failure scenarios and escalation risks across stakeholders

Multi stakeholder design sprints

To co create and rapidly test practical governance solutions

Structured evidence sharing

To capture real world system behavior and emerging risks

GOOD PRACTICES AND POLICY APPROACHES

Risk based Governance

that classifies AI systems by impact and applies proportionate controls, with stricter requirements for high risk use cases

Sandbox and Multi Stakeholder Implementation

using regulatory sandboxes and collaborative models to test safeguards, evaluate edge cases, and ensure enforceable Human Intelligence in the Loop

Data Lifecycle Lovernance

embedding controls from collection to disposal, including inventory, classification, lineage, and accountability with continuous assess, implement, and monitor cycles

Auditability and Traceability

through logging of data usage, decision pathways, and system behavior to enable accountability for automated and agentic systems

No system should operate beyond our ability to supervise, interrupt, and audit it.

THE CORE SHIFT : TRUST AND ACCOUNTABILITY

AI governance must evolve to match AI capability. This is not a technology question. It is a question of trust, accountability, and human agency.

The shift is clear: from principles to protocols, from claims to verification, and from fragmented approaches to interoperable systems. Trust must be measurable, enforceable, and continuous. In a world shaped by agentic systems, geopolitical tension, and AI embedded in daily life, the defining question remains:

Do we retain meaningful human agency over the systems we create?

This intervention advances a path to ensure the answer is yes.

United Nations

XRSI INTERVENTION

https://www.un.org/global-dialogue-ai-governance/en

UN GLOBAL DIALOGUE ON AI GOVERNANCE (2026)

XRSI™

Human Intelligence In The Loop

www.xrsi.org | info@xrsi.org