Toward interoperability and inclusive participation in AI Governance

White paper v1.1
June 2026

AI GOVERNANCE FOR HUMANITY LAB

Disclaimer

This document presents findings, analysis, and recommendations synthesised by the Lab based on contributions from diverse stakeholders, and does not reflect the official positions, consensus, or individual endorsement of any single contributor, participant, or their respective affiliated organisations. The views and opinions expressed in this document do not represent the views of the United Nations. The designations employed, including geographical names, and the presentation of the materials in the present publication, including the citations, maps and bibliography, do not imply the expression of any opinion whatsoever on the part of the United Nations concerning the names and legal status of any country, territory, city or area or of its authorities or concerning the delimitation of its frontiers or boundaries and do not imply official endorsement or acceptance by the United Nations. Information contained in the present publication emanating from actions and decisions taken by States does not imply official endorsement, acceptance or recognition by the United Nations of such actions and decisions, and such information is included without prejudice to the position of any State Member of the United Nations. Information contained in this report concerning specific companies or of certain manufacturers' products does not imply official endorsement or recommendation by the United Nations (including by virtue of any omission of references to other products of a similar nature). The present publication does not affect the position of any State Member of the United Nations with regard to any international multilateral agreement. The inclusion of references to data, analysis, or commentary from participating organisations is for analytical purposes only and does not constitute endorsement by, nor should it be interpreted as representing the views, strategies, or official positions of, the companies and commercial entities engaged in this process.

How to cite this document: AI Governance for Humanity Lab. (2026). *Toward Interoperability and Inclusive Participation in AI Governance.* White paper.

Note on this edition: *This version of the white paper includes minor revisions to correct technical errors and editorial inaccuracies from the original release. The core findings and conclusions remain unchanged.*

CONTENTS

Executive summary

Over the past few years, countries have evolved their regulatory frameworks and AI strategies to respond to the unprecedented pace of AI innovation. Efforts around the world to govern AI reflect diverse realities, interests and priorities and pursue distinct, and often divergent, approaches to incentivise innovation, boost adoption to harness the economic benefits of the technology, and counter its potential adverse impacts. Against this backdrop, the notion of interoperability in AI governance has emerged as a strategic option to enable divergent and often incompatible AI governance philosophies to effectively address the cross-border challenges of the technology. Supporting the Global Dialogue on AI Governance, this white paper offers a timely assessment of the state of play of AI governance, a structure to navigate the multifaceted opportunities that interoperability offers, and a suite of actionable pathways toward interoperability and inclusivity in AI governance.

Part I:

Mapping The State of Play of AI Governance

The global arena is predominantly shaped by three dominant regulatory paradigms: the European Union, the United States, and China. These paradigms are accompanied by the adaptive principles of middle powers and diverse pragmatic, developmental, capacity-building frameworks emerging globally.

Emerging Trends

There is an emerging global drive towards "AI sovereignty" as nations seek to reduce structural dependencies on concentrated foreign technology infrastructures. Concurrently, shifting public attitudes reflect a mix of public optimism and mounting civic resistance against transnational labour concerns and the massive environmental footprints of data centres.

Assessing Governance Fragmentation

Governance fragmentation introduces severe structural challenges, including regulatory arbitrage, cross-border accountability gaps, and heavy compliance friction that disproportionately distributes costs globally. At the same time, strategic fragmentation can serve as a vital protective mechanism, offering lower-capacity states the necessary policy space to experiment and align rules with local socio-economic realities.

The 'What', 'Why' and 'How' of Interoperability

Governance Interoperability provides a pragmatic mechanism for sovereign systems to communicate and cooperate. It requires coordinated action across six distinct entry points: semantic, ethical, legal, organisational, technical, and institutional layers. While it offers significant opportunities to address structural risks, it also has clear limits worth recognising.

Part II:

Paths Forward

This white paper makes the case for the international community to establish global baseline conditions built strictly as protective "floors" rather than restrictive "ceilings." These minimums should guarantee non-negotiable human rights and safety standards while preserving the sovereign flexibility of nations to implement more stringent domestic protections.

14 Actionable Pathways

1 **Globally Agreed Glossaries and Taxonomies** Advances international alignment and other interoperability measures by building shared definitions for foundational terms (like "AI systems" and "high-risk").

2 **Targeted AI governance literacy** to embed practical governance knowledge within public administrations, judiciaries, private sector, diplomats and civil society.

3 **Global Network for AI Governance Knowledge and Practice** Deploys a decentralised polycentric network to weave diverse multistakeholder viewpoints into global rules.

4 **Cross-Border AI Regulatory Sandboxes** Establishes an internationally connected network of controlled testing environments governed by shared evaluation protocols, helping states access rigorous validation infrastructure without prohibitive domestic costs.

5 **Crosswalks and Tools for Semantic Interoperability** Advances crosswalks and automated frameworks to systematically map how risk metrics correspond across borders, providing clear incentives for alignment without the friction of enforced legal harmonization.

6 **Multistakeholder open toolbox** Delivers a shared open toolbox featuring model laws, multilingual procurement templates, and Human Rights Impact Assessments.

7 **Mutual Recognition Agreements (MRAs)** Enables targeted adequacy agreements to accept foreign conformity evaluations, leveraging process-oriented technical standards (like ISO/IEC 42001) as neutral bridging mechanisms to cut duplicative costs.

8 **Equitable Participation in Global Governance and Standard-Setting** Addresses structural exclusion and high-income bias in technical standard bodies by implementing targeted support for equitable participation, regional convenings, and sustained funding for inclusive participation in standard setting.

9 **Interoperable Digital Public Infrastructure** Coordinates multilateral funding to deploy open-source models, data trusts, and public compute facilities, establishing a democratically accountable alternative to mitigate technological dependence.

10 **Community auditing** as a first-class governance mechanism drawing on existing participatory harm auditing workbenches and methodologies.

11 **Shared Boundaries and Global "Red Lines"** Establishes a non-negotiable protective floor by universally restricting specific AI applications that severely violate fundamental human rights or pose uncontainable, catastrophic transboundary security risks.

12 **Shared Incident Reporting and Crisis Protocols** Implements a global registry using common machine-readable schemas to log serious AI failures and capability surprises, backed by emergency scenario stress-tests among regional policymakers.

13 **Transnational Labour Governance** Connects AI governance directly with formal international labour governance to advance mechanisms against opaque algorithmic management while granting platform and data-labelling workforces inclusive standing.

14 **Bridge AI Governance and Climate Frameworks** Structures an environmental impact reporting framework that aligns rapid computational growth with global emissions planning and oversight.

Acknowledgements

This document is the result of a collaborative and deliberative process convened by the AI Governance for Humanity Lab. It has been informed by the substantial written contributions, oral interventions, and strategic input of experts from academia, civil society, public sector, industry, and international organisations.

The findings, analyses, and conclusions presented herein have been synthesised by the Lab through an iterative process, combining comprehensive written submissions, expert interventions and commentary, and stakeholder consultation inputs gathered during the Valencia Dialogue on May 28–29, 2026.

Head of the Lab & co-editor: *Ana García Robles* (AI Governance for Humanity Lab)

Lead researcher and main editor: *Andrés Domínguez Hernández* (AI Governance for Humanity Lab)

Co-drafting and integration of empirical data: *Fola Adeleke, Leah Junck, David Leslie.*

Working group (substantive written contributions):

Allison Wylde, Ana Brandusescu, Aryamehr Fattahi, Blair Attard-Frost, Chinasa T. Okolo, Cüneyt Yılmaz, Dwayne Woods, Elena Simperl, Emmie Hine, Fola Adeleke, Francisco Javier Torres Gella, Huw Roberts, Idoia Salazar, Isabella Wilkinson, Jake Okechukwu Effoduh, Johannes Geith, Jonas Tallberg, Julia Pomares, Luca Belli, Luciano Charlita de Freitas, Magnus Lundgren, Markus Furendal, Michal Natorski, Mubarak Raji, Naim Ahmad, Natalie Bertels, Nathalie A. Smuha, Olajide Olugbade, Shyam Krishna, Vicente Silva, Virginia Dignum, Yik Chan Chin.

Valencia Dialogue consultation and workshop participants

Alessandro Mantelero, Shreeppriya Gopalakrishnan, David Leslie, Rahaf Harfoush, Linda Bonyo, Jake Okechukwu Effoduh, Mario Hernandez Ramos, Yik Chin, Johannes Geith, Aryamehr Fattahi, Fola Adeleke, Huw Roberts, Francisco Javier Torres Gella, Florian Ostmann, Idoia Salazar, Natalie Bertels, Asunción Gómez-Pérez, Jon Ander Gómez Adrián, and delegations from UNESCO, UNICEF, OECD, and UNIDO.

High-level expert review

Alessandro Mantelero, Rahaf Harfoush, Eugenio V. Garcia, Rachel Adams, David Leslie.

Special thanks to the *UN Under-Secretary-General and Special Envoy for Digital and Emerging Technologies, Amandeep Gill and to Quintin Chou-Lambert* for strategic guidance and review; to *Francesco Stabilito, Mehdi Snene* for technical insight and guidance; and to *Brian Shung Seun Lau and Karoline Hassfurther* for the logo, the main Lab design elements, and guidance. Graphic design by *A2Colores.*

The Lab has been established with the support of the Kingdom of Spain.

Declaration of generative AI use:

This white paper was conceptualised, researched, and entirely drafted by human authors, who retain full editorial and ethical responsibility for its content. Generative artificial intelligence (Google Gemini 3.5 and Microsoft 365 Copilot) was utilised for copy-editing, syntax polishing, and assisting designers to render graphic visualisations. All computational processing was restricted to a UN secure enterprise environment to protect data privacy, and every output was rigorously reviewed by the authors to eliminate algorithmic bias and ensure factual integrity.

Introduction

The rapid proliferation and integration of artificial intelligence (AI) around the world and across virtually every aspect of society have brought the international community to a critical governance juncture. Over the past few years, countries have evolved their regulatory frameworks and AI strategies to come to terms with the unprecedented scale and speed of AI innovation. Efforts around the world to govern AI reflect diverse socio-economic realities, constitutional traditions, social contracts, geopolitical interests, and cultural values. In their various manifestations, these efforts pursue distinct, and often divergent, approaches to incentivise innovation, boost adoption to harness the economic benefits of the technology, counter its potential adverse impacts, and ensure that benefits are distributed equitably in society. But AI models and systems know no borders; they are distributed globally, and are also underpinned by transnational supply chains of data, labour, knowledge, and infrastructure. As a result, AI raises several cross-border challenges that demand international cooperation and alignment of strategies in a context of geopolitical rivalry and divergent regulatory philosophies.

Against this backdrop, the Global Digital Compact (GDC) and the Secretary-General's High-Level Advisory Body on AI's Final Report converge on the need to establish interoperable governance approaches anchored in global norms and principles in the interests of all countries. The notion of interoperability in AI governance offers a strategic option to enable divergent and often incompatible AI governance regimes and philosophies to effectively address the cross-border challenges of the technology absent a unified global approach (which might not even be necessary at this stage), and an agreed set of norms.

Crucially, the international community does not need to build a global interoperable framework from scratch. Member States and stakeholders can leverage and reinforce existing interoperability infrastructures at the international, regional, and local levels. This includes anchoring multilateral commitments within the valuable, ongoing technical and normative work already being pioneered by international organisations, standard-setting bodies, and national regulators and supervisory agencies. As the current assessment reveals, the public and private sectors, academia, and civil society are already testing such opportunities for alignment and cooperation across borders.

The **UN Global Dialogue on Artificial Intelligence Governance**— with a dedicated cluster on Safe, secure and trustworthy AI: interoperability and compatibility of approaches—serves as a unique inclusive forum to advance these critical conversations and translate global deliberations on governance interoperability into coordinated action. To support this global effort, this white paper offers **a timely assessment of the state of play of AI governance**, mapping the consolidating and emerging trends in a world of geopolitical multipolarity, and providing a structure to navigate the multifaceted opportunities that interoperability presents. This resource also puts forward a suite of **actionable pathways** aimed at addressing the most pressing challenges posed by regulatory and governance fragmentation. Rather than a linear roadmap or prescriptive formulas, these pathways represent open-ended guideposts for multistakeholder cooperation that are anchored in existing infrastructures and ongoing processes, and that require continuous exploration, co-design, and shared responsibility across national, regional, and international levels, societal actors, sectors and institutions.

Part I:
The state of play of AI governance: a landscape in flux

The goal of Part I is to help inform the conversations at the UN Global Dialogue on AI Governance by offering a timely assessment of the current state of play of AI governance globally, identifying emerging trends related to how different regulatory philosophies and governance approaches around the world coexist and interact, while acknowledging that this is a highly dynamic and rapidly changing landscape. Rather than treating fragmentation as a governance failure or understanding it as the opposite of "order," it is more accurate —and productive— to understand it as a natural expression of a plural world in which different regulatory philosophies and capacities coexist. With this in mind, this section looks at how interoperability in AI governance can play a key role to tackle some of the most pressing challenges stemming from governance fragmentation, identifying the potential upsides but also acknowledging the limitations of interoperability.

Navigating the global AI governance landscape

To navigate the complex AI governance landscape, it is worth looking at the different manifestations of AI governance beyond formal rules. More broadly, AI governance can be defined as *the establishment of regulations, standards, policies, practices, norms, institutions, and incentives to steer the production and future trajectory of AI systems toward a desired goal.*[1] It is helpful to recognise that different goals may require different configurations across these dimensions and levers of governance.

Across national, regional and international levels

Efforts to regulate AI are taking place simultaneously at the local, national, regional, and international levels. Which level is most predominant varies across jurisdictions. For example, the European Union relies heavily on regional regulation (the EU AI Act), the US sees significant impact from state and local laws, and China relies on a top-down state-led approach. At the international level, instruments such as the OECD AI Principles, Council of Europe's Framework Convention on AI, UNESCO's Recommendation on the Ethics of AI, the Hiroshima AI Process, and BRICS Leaders' Statement on the Global Governance of AI have emerged as efforts to coordinate international action to govern the technology.

[1] Domínguez Hernández, A., Perini, A.M., Hadjiloizou, S. et al. (2025) Towards a sociotechnical ecology of artificial intelligence: power, accountability, and governance in a global context.

A range of hard and soft law instruments

Governance interventions sit on a continuum between hard and soft law instruments, varying across the degree of legal obligation, the precision of the rules, and the extent of delegation to parties to interpret, monitor, or enforce those rules. Various jurisdictions are moving toward high delegation, granting authority to third-party entities like standard-setting bodies to implement, monitor, and operationalise rules. For example, China has been relying on standards bodies such as TC260 to develop guidance on how regulations should be operationalised. The EU AI Act delegates the task of operationalisation to European standard-setting organisations, most notably CEN and CENELEC, tasked with formulating harmonised standards.

Horizontal vs. vertical approaches

Many advanced AI technologies are general-purpose, meaning they can be applied across contexts. To address the cross-cutting risks, jurisdictions are adopting either horizontal or vertical governance approaches. Horizontal approaches promote cross-cutting regulations that govern all AI technologies across all contexts, whereas vertical approaches focus on specific sectors or AI applications. Horizontal approaches (like the EU AI Act)[2] promote regulations that govern all AI technologies across all contexts, providing regulatory clarity. They increasingly target the "foundation model" layer where systemic risks and infrastructural dependencies tend to concentrate. Conversely, vertical approaches (like some elements of those in the UK or some elements of the approach in China) tend to delegate governance to sectoral regulators or target specific applications and use cases, enabling greater flexibility in response to technological change. For example, in China, matters relating to professional education and reskilling in AI are governed by a General Office of the Ministry of Human Resources and Social Security guideline, while a separate Special Action Plan addresses the environmental impact of data centres, and the AI Safety Governance Frameworks (2.0) governs safety-related issues such as AI-facilitated misinformation and violence, bias, and safety and security.

From project lifecycle to ecosystem level governance

At the project lifecycle level AI systems are governed through risk assessments, auditing processes, and procurement standards. Conversely, ecosystem-level governance manages the broader social, legal, and economic structures—including markets and institutions—that influence AI adoption. There has been a significant turn toward lifecycle governance approaches in recent years. For instance, the Council of Europe's HUDERIA, Egypt National Guidelines, the ASEAN Guide on AI Governance, and the EU AI Act have established lifecycle perspectives. Integrating both perspectives is useful for interventions to be both targeted and systemic.

[2] The horizontal approach of the EU AI Act however does not capture military AI use.

A patchwork of global responses to AI

The term "patchwork" has often been used to describe the current international landscape of approaches to govern AI. Not only is this landscape characterised by a diversity of competing regulatory philosophies and national, regional, and continental strategies, but by persistent digital divides and global asymmetries in AI adoption and institutional capacity. Around the world and over the last few years, the landscape has transitioned from an initial concentration of governance among AI-leading nations toward a dense, multi-layered patchwork of binding legislation, soft-law instruments, technical standards, and voluntary codes of practice.

The scale of the expansion of the AI governance space is captured in the 2nd Edition of the Global Index on Responsible AI,[3] which provides one of the clearest empirical mappings of this evolving governance terrain. Assessing 135 countries across 17 responsible AI indicators spanning five dimensions, the Index identifies 200 new responsible AI-relevant frameworks since 2024 and 306 new cases of indicators addressed by at least one governance instrument, 203 of which are from Global South countries. In just two years, 29 countries have added frameworks on gender equality and AI, 30 now address AI safety and security, and 15 have introduced labour protections relevant to AI. Over the same period, policy coverage among Global South countries increased from an average of 2.5 to 4.7 key areas, while Global North countries rose from 8.2 to 11.1, reflecting their comparatively higher regulatory capacity. Yet, significant unevenness persists, with binding protections remaining weaker in many Global South countries, reinforcing the fragmented, multi-speed nature of the global governance patchwork.

The three dominant paradigms

The global AI governance landscape has gradually moved away from a sparse field of voluntary ethical principles toward a dense, unevenly fragmented environment defined by distinct, differing regulatory philosophies. The global architecture of AI governance is largely characterised by the consolidation of three dominant regulatory paradigms in the US, the EU and China that reflect fundamentally different assumptions regarding the relationship between innovation, risk containment, state authority, and fundamental rights.

European Union

Rights-Based & Risk-Tiered: Horizontal regulation (EU AI Act) prioritising fundamental rights through ex-ante obligations and a risk-based classification system with extraterritorial reach. The EU framework is designed with deliberate extraterritorial scope, requiring any organisation serving EU customers to comply regardless of its geographic headquarters, thereby projecting its regulatory influence globally through the "Brussels Effect".

United States

Innovation-First, Global Leadership and Security-Oriented: Sectoral, deregulatory approach relying on voluntary codes, agency guidance, and non-binding frameworks (NIST) to maintain commercial competitiveness in addition to measures designed to secure US leadership globally. Individual states such as California, Colorado, and New York have enacted their own binding laws focused on transparency, anti-discrimination, and individual data control.

China

State-directed with a focus on Security, Innovation and Development: Multifaceted governance approach that seeks to balance technological utility, security, ethics, mutual respect, and development, through product-specific rules and mandatory government assessments supplemented by pilot programmes and regulatory sandboxes to protect core socialist values.

[3] Adams, R., Grossman, N., Adeleke, F., et al. (2026). Global Index on Responsible AI (2nd ed.). Global Center on AI Governance.

Counter-balancing paradigms

Beyond these three dominant paradigms, other evolving and emerging paradigms reflect distinct national and regional priorities, traditions and philosophies, which stand as counterbalances (to varying degrees) to the dominant governance narratives at the global level. Emerging frameworks and philosophies across the global community add to an environment of geopolitical complexity, multipolarity, and constant fluidity, even if they do not yet accumulate into a full counterweight within the current political economy of AI at scale. ***Governance divergences are still experienced unevenly across institutional and infrastructural realities. The GIRAI data stresses this structural imbalance: while 73 countries have adopted AI policies, only 55% of cases globally show evidence of implementation, falling to 45% in the Global South.***

"Middle powers" such as the United Kingdom, Canada, Japan, India, Singapore, South Korea, and Brazil occupy a vital strategic position in the current geopolitical landscape. Consulted experts highlight that middle-sized countries are uniquely positioned to form coalitions of the willing on operational components of AI governance, such as model evaluations, incident reporting, and data governance. The United Kingdom, for example, illustrates a distinct institution-first and principles-based approach; it avoids an overarching horizontal statute by delegating AI oversight to existing sectoral regulators, while simultaneously asserting international leadership in the technical governance of frontier risks through initiatives like its AI Security Institute. Similarly, Canada demonstrates an adaptive governance trajectory that, despite shifting toward deregulation and voluntary instruments, remains a critical source of procedural interoperability by providing highly transferable regulatory templates for algorithmic impact assessments, such as its Directive on Automated Decision-Making. Countries like Singapore and Japan are pursuing a distinct approach that blends voluntary frameworks with selective binding elements to balance governance with economic growth, without imposing heavy compliance burdens. For example, Singapore governs AI through a combination of strategies and principles-based instruments, including the National AI Strategy, the Model AI Governance Framework (2nd edition), Singapore's Green Data Centre Roadmap, and a new Model AI Governance Framework for Agentic AI, while selectively deploying binding laws for labour protections (the Platform Workers Act 2024 No.30 of 2024) and the Elections (Integrity of Online Advertising) (Amendment) Bill 2024 prohibiting AI-facilitated misinformation. Japan, similarly, through the Hiroshima AI Process, has emphasised voluntary international coordination with over 60 countries now participating in its Friends Group, and has relied almost entirely on a mix of soft laws across the primary indicators measured in the GIRAI 2026— such as AI literacy, safety and security, transparency and accountability, human oversight, environmental impact, among others, but has a binding data protection framework which governs data privacy and data sharing and access. South Korea on the other hand has recently enacted a new risk-based law for artificial intelligence with mandatory rules similar to the EU AI Act.

The pragmatic and balanced approach within the Association of Southeast Asian Nations (ASEAN) is characterised by a deliberately voluntary, non-binding approach that seeks to pragmatically balance digital innovation with safety. Rather than pursuing rigid regulatory convergence, the region actively encourages industry growth and the deployment of regulatory sandboxes. This flexible approach avoids forced uniformity, strategically accommodating the diverse institutional and financial capacities of its member states while preserving contextual governance needs. A defining feature of the ASEAN approach is the operationalisation of high-level principles through practical, machine-readable governance tools for cross-border interoperability. The ASEAN Guide on AI Governance and Ethics provides Member States with actionable risk-impact templates, detailed case studies, and a tailored three-tier risk system (minimal, medium, and high). This regional taxonomy effectively addresses locally significant harms while remaining structurally compatible with global, comprehensive models such as the European Union's risk framework. This is the case with Vietnam's new Law on Artificial Intelligence.

The Investment-Driven Hub Strategy of Gulf States: Nations such as the United Arab Emirates and Saudi Arabia are pursuing active AI deployment and infrastructural investment with comparatively less developed rights architectures, positioning themselves as economic hubs and mediators. Gulf states are also making strides to develop local models that address the cultural and linguistic deficits of large language models produced in the AI-leading nations. These countries play an important role, not merely as norm-takers, but as evaluators, infrastructure hosts, model producers, and power brokers capable of leveraging their massive compute facilities in the current geopolitical environment. The UAE's performance in GIRAI 2026 serves as an example. Despite an overall low score in responsible AI governance, the country performs among the highest of any Global South country in the areas of capacity building such as reskilling and upskilling initiatives, and public sector skills development, driven by investment partnerships and government funded programmes. For instance, the UAE hosts the world's first AI dedicated university; the Mohamed bin Zayed University of Artificial Intelligence (MBZUAI).

Developmental and capacity building paradigms in the Global Majority: The Global Majority (often used interchangeably with "the Global South") is not a monolith, nor a passive recipient of external rules and governance frameworks authored elsewhere. Countries across Africa, Latin America and the Caribbean, and Southeast Asia are increasingly operating as active originators of AI models and applications, developing context-specific governance instruments, and building regulatory capacity that reflects their distinct democratic traditions and public service delivery. This distinction can be seen with Brazil, for example, combining several AI policies (the AI Bill, the AI Plan, and the Brazil AI Strategy) with multiple binding instruments, including dedicated legislation on children's digital rights and AI-enabled violence against women, to govern its AI landscape. Peru became the first country in the LAC region to enact horizontal AI legislation, establishing obligations for public authorities while simultaneously adapting sectoral laws and penal codes to regulate private entities. India has positioned itself as a convening power for Global South perspectives by hosting the first AI summit in the Global South and deploying digital public infrastructure to support domestic AI capacity like compute, water and energy. Meanwhile, Rwanda primarily uses its National AI Policy to address capacity-focused indicators such as AI literacy, public procurement, and public sector skills development, while relying on non-binding ethical guidelines for safety and ethics-based issues such as safety, security, and gender equality.

In contrast to the risk-based, innovation-first, or state-driven security models of the dominant governance paradigms, many Global Majority jurisdictions are adopting regulatory philosophies grounded in developmental and capacity building approaches. This is evidenced by the GIRAI 2026 finding that only 22% of responsible AI frameworks adopted in the Global South are binding compared to 58% in the Global North and may be interpreted as an unwillingness to introduce stringent environments that could stifle foreign AI investments, which, in turn appears to fit a paradigm that frames AI primarily as a mechanism for socioeconomic catch-up, and leapfrogging in these contexts. Yet, the limited use of binding laws sits alongside a quick and vast expansion of responsible AI activity, including 828 government-led initiatives and 52 countries investing in local language and culturally inclusive AI. This suggests an understanding of leapfrogging and catching up that does not always merely reflect a linear race towards the models of more developed nations, but that also involves deliberate attention to building capacity, infrastructure and context-appropriate technologies that respond to local realities.

Consequently, governance priorities in these regions often differ from those in advanced economies. For example, Indonesia explicitly prioritises domestic capacity development over the immediate adoption of international frameworks, while Rwanda has indicated that governance approaches must account for foundational infrastructural realities, such as low internet penetration. Furthermore, cultural philosophies actively shape these paradigms; Kenya's AI Strategy, for instance, advocates for the respect of African culture through the concept of Ubuntu ("I am because we are"), promoting community interdependence and mutual responsibility as an alternative to highly individualised Global North rights frameworks.

This distinct capacity-focused approach in the Global South should not be mistaken for an absence of responsible AI governance. In fact, GIRAI 2026 shows several Global South countries outperforming the two leading AI nations (US and China) across key indicators: Brazil, Chile, Colombia, Ethiopia, Romania, Peru, and Jordan all score higher than both the United States and China on AI labour protections, while Colombia, Uruguay, Pakistan, and Costa Rica outscore the UK and China on public disclosure of government algorithms. These examples illustrate that Global Majority governance is not "behind," but is, in some cases, advancing context-specific, capacity-oriented models that prioritise public-service realities and social needs while also addressing extreme and systemic risks. At the same time, many Global South countries exhibit a significant gap between de jure protections and de facto ones. For example, while some countries in Latin America and the Caribbean outperform advanced economies in AI-related labour protections, much of its workforce (47.6%) remains in informal employment, evidencing a gap in the practical enforceability of such protections.[4]

International instruments and networks

Complementing the national regulatory paradigms of major technological powers and the diverse strategies of regional alternatives, the international AI governance architecture encompasses a dense, layered ecosystem of multilateral treaties, soft-law principles, minilateral agreements, and operational networks.

Governance Category	Instrument or Framework	Description and Strategic Focus
Binding Multilateral Treaties	**Council of Europe Framework Convention on Artificial Intelligence (2024)**	This represents the first legally binding international AI treaty, signed by 20 states including the United States and the United Kingdom. It adopts a horizontal approach to ensure AI systems respect human rights, democracy, and the rule of law, functioning as an interoperable framework despite relying on soft-law enforcement mechanisms.

[4] International Labour Organization (ILO). (2024). 2024 Labour Overview of Latin America and the Caribbean. https://www.ilo.org/publications/2024-labour-overview-latin-america-and-caribbean

Governance Category	Instrument or Framework	Description and Strategic Focus
Universal and Multilateral Normative Frameworks	**UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)**	Adopted by 193 UNESCO Member States, this is the most universally ratified instrument and provides a comprehensive ethical foundation. It is accompanied by practical implementation tools, such as the Readiness Assessment Methodology (RAM), which assist nations—particularly in the Global Majority—in evaluating their institutional capacities against global benchmarks.
	OECD AI Principles (2019)	Adopted by over 40 countries, these principles established a foundational vocabulary for trustworthy AI, focusing on transparency, robustness, and accountability. They have also provided a definitional baseline for AI systems that subsequently informed binding regulations, thereby facilitating semantic interoperability.
Minilateral and Voluntary Coordination Mechanisms	**G7 Hiroshima AI Process (HAIP) (2023)**	This process focuses on voluntary international coordination, delivering a Comprehensive Policy Framework and an International Code of Conduct. It includes a reporting framework that fosters transparency by providing a structured mechanism for organisations to disclose their AI risk management practices.
	BRICS AI Governance Principles (2024)	Articulating an alternative multilateral vision, these principles emphasise state sovereignty, equitable participation, and narrowing the digital divide. They advocate for equal rights in global AI governance and actively support capacity building for developing countries.
Operational and Technical Networks	**International Network for Advanced AI Measurement, Evaluation and Science**	This network (formerly the International Network of AI Safety Institutes) represents an operational interoperability mechanism coordinating frontier model evaluation methodologies and red-teaming findings across institutes from the UK, US, Canada, Japan, Kenya, South Korea, Australia, Singapore, and the EU. It bridges divergent regulatory philosophies through scientific consensus and common testing protocols.[5]
Voluntary technical standards	**Technical Standardisation Bodies (ISO/IEC, IEEE, ITU)**	These organisations translate normative aspirations into operational artifacts. For example, the ISO/IEC 42001 standard provides a technology-neutral AI Management System framework that organisations can implement across different domestic jurisdictions to support technical interoperability. The IEEE 7000 series on ethical AI provides complementary technical guidance.

[5] UK AI Security Institute. (2025). International consensus and open questions in AI evaluations. https://www.aisi.gov.uk/blog/international-ai-network-consensus-and-open-questions

Emerging trends

The drive toward "AI sovereignty" and "digital sovereignty"

In a landscape of geopolitical fragmentation, and high concentrations of AI infrastructure in the hands of a few large tech firms, narratives of "AI sovereignty" and "digital sovereignty" have marked a dominant and defining trend in the global AI governance landscape. AI governance is no longer solely about operationalising ethical principles, alignment of values, safety, or risk management; it is increasingly inseparable from questions of technological sovereignty, industrial policy, and global geopolitical and geoeconomic power. Digital sovereignty can be understood as the capacity to understand, develop, and regulate digital technologies according to locally legitimate democratic, developmental, and constitutional priorities.[6]

This trend manifests in different ways across countries as a deliberate shift toward treating AI as a strategic national resource and critical infrastructure. Some states and blocs like the EU are actively seeking to exert control over domestic computational resources, data infrastructures, and foundation models to ensure strategic independence. This drive is fundamentally a reaction to the structural dependencies created by the current global AI ecosystem, where advanced capabilities, cloud infrastructures, and standard-setting authority are hyper-concentrated in a handful of transnational corporations and dominant technology-producing nations.[7] These dependencies manifest unevenly: countries that have weaker regulatory capacities and lack sovereign digital infrastructures risk becoming dependent on foreign technological ecosystems for critical public and private functions, including public administration, healthcare, education, security, and communications.

A primary concern is the risk of regulatory capture and the consolidation of corporate power through the emergence of "private sovereignty" offerings. Critical perspectives highlight that state-led AI sovereignty initiatives frequently frame infrastructure solely as a matter of national security and technocratic control, rather than public interest and democratic oversight. Consequently, these initiatives often fail to provide genuine democratic assurances; instead, they risk encouraging the concentration of power within large domestic tech monopolies while pushing civil society and affected communities to the margins of the governance process.

Emerging economic and political blocs, such as BRICS, are increasingly framing technological dependence not merely as an economic concern, but as a critical vulnerability to their strategic resilience. Consequently, states across Africa and Latin America are deploying data localisation requirements, developing sovereign cloud policies, and investing heavily in domestic AI talent pipelines. These measures are strategic attempts to retain the economic and strategic value of data and AI within national borders, build local AI ecosystems, and reduce long-term structural dependence on foreign technological infrastructures. Whether AI sovereignty can be achieved in practice depends on structural factors of capacity and the layer of the AI stack (e.g., compute, model application, data exchange, etc.) For many countries, developing sovereign compute capabilities is particularly challenging due to the investment and technical expertise required.

---

[6] Jiang, M., & Belli, L. (Eds.). (2024). Digital Sovereignty from the BRICS Countries: How the Global South and Emerging Power Alliances Are Reshaping Digital Governance. Cambridge University Press.
[7] Belli, L., & Gaspar, W. B. (Eds.). (2025). The Quest for AI Sovereignty, Transparency and Accountability: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition.

The distinct role of indigenous sovereignty

It is important to add nuance to this trend, distinguishing state-level digital sovereignty from indigenous perspectives on sovereignty and AI governance. Anchored in frameworks such as the CARE Principles,[8] the First Nations Principles of OCAP®,[9] and Te Mana Raraunga,[10] Indigenous sovereignty demands non-extractive, culturally grounded approaches to AI and data governance based on collective consent, benefit-sharing, and the right to refusal.

Initiatives that draw on different legal and ethical traditions and diverge from the notion of AI as a predominantly strategic resource also raise broader questions around what (AI-driven) technology as a public good may mean and what legal approaches should consist of. Consulted experts stress that state-level AI sovereignty initiatives must recognise and work in parallel with these indigenous legal and ethical traditions, ensuring they are not erased, instrumentalised, or merely subsumed under generic national data protection frameworks.[11]

Evolving public attitudes on AI

Global public sentiment toward artificial intelligence is characterised by a coexistence of optimism and anxiety that is highly stratified by geography. According to the 2026 Stanford HAI AI Index Report[12] which synthesises large-scale surveys across more than 30 countries, 59% of respondents globally believe that AI products and services offer more benefits than drawbacks, yet 52% report feeling nervous about the technology. Southeast Asian nations and China remain the most enthusiastic, with over 80% of citizens anticipating profound life changes from AI in the near term and demonstrating high levels of trust. Conversely, Western nations display deep caution, and Japan stands out for its persistent public pessimism. Furthermore, India recorded the sharpest global spike in public nervousness in a single year, even as workplace AI adoption in India as well as China, Nigeria, and Saudi Arabia exceeded 80%, vastly outstripping advanced Western economies.

Public unease is deeply tied to public perceptions regarding AI's long-term impact on the future of work. These labour anxieties are shifting from passive worry to active, coordinated resistance across global supply chains. Case studies compiled by the International Labour Organisation (ILO) demonstrate how worker representatives across Europe, North America, Asia, South America, and Africa are utilising collective bargaining and social dialogue to contest algorithmic management and push for "labour-complementing" rather than "labour-replacing" technologies.[13]

---

[8] https://www.gida-global.org/careprinciples
[9] https://fnigc.ca/ocap-training/
[10] https://www.temanararaunga.maori.nz/
[11] Nyamnjoh A (2025) Beyond a buzzword: Can Ubuntu reframe AI Ethics? https://health.uct.ac.za/ethics-lab/articles/2025-09-09-beyond-buzzword-can-ubuntu-reframe-ai-ethics-0
[12] Stanford University, 2026 AI Index Report (Chapter 9: Public Opinion), 2026, hai.stanford.edu/ai-index/2026-ai-index-report/public-opinion.
[13] Doellgast, V., Appalla, S., Ginzburg, D., Kim, J., Thian, WL. 2025. Global case studies of social dialogue on AI and algorithmic management, ILO Working Paper 144 (Geneva, ILO). https://doi.org/10.54394/VOQE4924

Simultaneously, the physical footprint of AI is triggering local resistance, turning data centre expansion into a contested environmental issue. Within the United States, which hosts the highest density of these facilities, heightened pressures on grid integration, escalating environmental costs, and resource concerns are sparking public backlash. A May 2026 Gallup survey revealed that 71 percent of the American population would explicitly object to the establishment of a data centre within their local community.[14] In Latin America, severe droughts have driven water-justice protests against the construction of data centres. A recent report from the United Nations University Institute for Water, Environment and Health (UNU-INWEH) reveals that data centres globally consumed 4.5 trillion litres of water in 2025—enough to meet the basic needs of over 600 million people in Sub-Saharan Africa—whilst generating 189 million tonnes of carbon dioxide emissions.[15]

Ultimately, evolving public awareness and optimism toward AI, along with pushback against AI's environmental costs and transnational labour issues, will significantly dictate the political legitimacy and priorities of global AI governance. International frameworks will be required to bridge the gap between industry-led innovation agendas and the need to address public demands for structural safety, labour rights, and climate accountability.

A nuanced view of global fragmentation and its implications

While there might be a tendency to frame global regulatory fragmentation as a problem to be addressed through a diversity of governance measures, it is worth considering the two sides of the issue. Rather than inherently problematic, fragmentation is a reflection of a pluralistic world in which diverging governance priorities, regulatory philosophies, cultural values, social contracts, and developmental priorities coexist. Fragmentation therefore raises many structural challenges but it also offers opportunities for productive adaptation and reorganisation of the governance landscape.

The structural challenges of fragmentation

- **Regulatory Arbitrage and the Race to the Bottom:** A primary structural challenge stemming from global fragmentation is regulatory arbitrage. In the absence of coordinated international baselines, developers and multinational vendors can shift their research, model training, data processing, and deployment to jurisdictions with the least restrictive oversight. This dynamic actively weakens safety and rights protections across all jurisdictions, as systems prohibited or tightly regulated in one market can operate freely elsewhere and still be accessed cross-border. Consequently, fragmentation sets the conditions for a race to the bottom, where countries competing for AI investment may deliberately lower regulatory standards, prioritising speed of innovation over safety and public accountability.

- **Accountability Gaps and the Attribution of Cross-Border Harms:** AI systems are developed and deployed through complex, globalised value chains: training data may originate from one region, models developed and fine-tuned in another, and applications deployed worldwide via cloud APIs. This structural reality means that when cross-border harms occur—such as algorithmic discrimination, unfair extraction of intellectual property for AI training, labour exploitation, or privacy breaches—determining the applicable law

[14] Gallup. (May 2026). Public Perceptions of Artificial Intelligence Infrastructure and Localized Environmental Impact. Gallup Poll Report. https://news.gallup.com/poll/709772/americans-oppose-data-centers-area.aspx
[15] Doellgast, V., Appalla, S., Ginzburg, D., Kim, J., Thian, WL. 2025. Global case studies of social dialogue on AI and algorithmic management, ILO Working Paper 144 (Geneva, ILO). https://doi.org/10.54394/VOQE4924

and assigning responsibility becomes extraordinarily difficult. National regulators often lack the jurisdiction, mandate, or mutual legal assistance tools required to compel compliance from, or enforce remedies against, foreign developers. As a result, the individuals and marginalised communities most impacted by AI harms are frequently left without accessible avenues to request explanations, contest automated decisions, or seek meaningful redress.

- **Compliance Burdens, Market Stratification, and Monopolisation:** From an economic perspective, divergent regulatory frameworks impose compounding and sometimes structurally irreconcilable compliance obligations on AI developers and deployers. Navigating this complex patchwork of rules requires vast legal and administrative resources. Consequently, fragmentation disproportionately burdens Small and Medium-sized Enterprises (SMEs), academic researchers, and developers from lower-income countries, effectively freezing them out of global markets. This compliance friction actively drives market consolidation, cementing the oligopolistic dominance of massive multinational technology corporations that possess the resources and legal infrastructure to maintain parallel compliance architectures. Over time, this dynamic stifles innovation diversity and reinforces the structural power of a few dominant actors.

- **Amplification of Systemic and Global Security Risks:** Fragmentation fundamentally undermines the global community's ability to govern severe, catastrophic, and systemic risks. Threats such as AI-enabled cyberattacks, the proliferation of deepfakes and synthetic media for electoral interference, and the potential uplift of chemical, biological, radiological, and nuclear (CBRN) capabilities do not respect national borders. Cross-border risks cannot be effectively managed through fragmented mitigations; a secure regulatory environment in one jurisdiction provides little protection if a dangerous capability—such as a highly capable open-weight model—is released without safeguards in a jurisdiction with lower security oversight. Furthermore, the lack of centralised information-sharing protocols for near-misses, capability surprises, and threat intelligence exacerbates these security blind spots, preventing timely and coordinated international threat assessment.

- **Definitional Friction and Incompatible Terminology:** The current landscape is hindered by severe semantic and procedural fragmentation. Jurisdictions define critical regulatory concepts like "high-risk," "high-impact," and "transparency" differently, creating structural incompatibilities that prevent governance systems and frameworks from communicating effectively. When enforcement agencies and oversight bodies do not share unified taxonomies, audit methodologies, or incident reporting frameworks, the ability to learn from AI-related harms from other places, and ensure cross-border compliance is severely limited. This lack of a shared evidence base hinders comparative policy evaluation, limits cross-border supervisory cooperation, and prevents effective transnational oversight.

- **Uneven Governance Capacity and Structural Dependency:** Finally, global fragmentation is profoundly asymmetrical, reflecting deep power imbalances between a few technology-producing nations and the rest of the world. Technical standards, evaluation benchmarks, and regulatory norms are disproportionately shaped by a small set of advanced economies and well-resourced industry actors, structurally excluding the Global Majority from agenda-setting. Global Majority countries are frequently forced into an alignment dilemma: either bear the high compliance costs of aligning to dominant frameworks (EU AI Act) they had no role in designing, or face exclusion from global data flows, digital trade, and investment. This dilemma is compounded by asymmetries in international standard-setting. Standardisation processes in bodies like ISO or IEEE are dominated by high-income countries and industry experts where participation in drafting is structurally inaccessible to most Global Majority states, civil society organisations, and affected communities, owing to fees, travel costs, English-language work, technical specialism, and dominance by industry experts.[16]

[16] Roberts, H., & Ziosi, M. (2025). Can we standardise the frontier of AI? Oxford AI Governance Institute. https://aigi.ox.ac.uk/publications/can-we-standardise-the-frontier-of-ai/

The structural asymmetries of the AI ecosystem also facilitate data colonialism, where resources and data are extracted globally to train proprietary models that yield little local benefit, all while intensifying uneven environmental costs and reinforcing dependencies on foreign cloud and computational infrastructures.[17]

Strategic fragmentation as a protective mechanism

Global regulatory fragmentation however is not exclusively a coordination failure; in many instances, it can be a strategic necessity. Premature harmonisation around a single regulatory paradigm, especially one designed for high-income economies with mature institutional enforcement capacities, can chill the space that low- and middle-income countries require to develop their own AI governance agendas.

Fragmentation can provide smaller and lower-capacity states with the critical breathing room to experiment against a backdrop of persistent normative divergence between governance models, to refuse ill-fitting but dominant external frameworks, and to align their AI governance strategies with their own constitutional, developmental, and socioeconomic commitments.

Furthermore, governance fragmentation places a corresponding demand on policymakers, AI developers, and tech firms to build more inclusive and equitable approaches to global AI standard-making, regulation, and governance. **In a fragmented international governance landscape, inclusive participation is a necessary condition for global AI governance to achieve democratic legitimacy rather than merely declare it.**

What is interoperability in AI governance, and why is it important now?

Interoperability is widely recognised as a strategic option to address some of the challenges arising in the current landscape of global fragmentation and plurality of regulatory philosophies. When applied to the context of AI governance, the notion of interoperability is multilayered and continuously evolving. There is no single, universally agreed-upon definition, largely because Member States and stakeholders approach the concept with differing assumptions regarding the ultimate goal of international cooperation, and because of the growing repertoire of potential avenues for enacting interoperability.

Broadly, interoperability in AI governance refers to **the capacity of diverse regulatory philosophies, technical frameworks, value systems, and governance institutions to communicate and cooperate effectively across borders without necessitating complete legal harmonisation or institutional uniformity.**

In this way, interoperability measures can act as the “connecting tissue” between distinct, sovereign systems rather than a single or harmonised global regime.

The development of shared vocabularies, taxonomies, reference points, or agreed-upon minimums creates the conditions and basic guarantees that allow distinct systems to interact without collapsing their differences.

[17] Regilme, S. S. F. (2024). Artificial intelligence colonialism: Environmental damage, labor exploitation, and human rights crises in the Global South. SAIS Review of International Affairs. https://muse.jhu.edu/article/950958

The multiple entry points of interoperability

As shown earlier, AI governance is a multifaceted, layered concept, and therefore governance interoperability correspondingly presents multiple entry points and levers of intervention. To effectively operationalise international cooperation, stakeholders must approach interoperability as a series of distinct layers, recognising that progress or gaps within one layer are not automatically compensated by advancements in another. The contributions of the consulted experts, alongside existing models such as the European Interoperability Framework (EIF)[18] and the United Nations University (UNU) Policy Report on Interoperability in AI Safety Governance,[19] provide a good starting point to understanding the different entry points interoperability:

Semantic interoperability

Semantic interoperability focuses on the establishment of a "shared language" between different regulatory and technical systems. It involves developing common vocabularies, shared data and metadata standards, and comparable risk taxonomies. Without semantic alignment on core regulatory concepts (such as the legal definitions of "high-risk AI," "foundation models," or "transparency") conformity assessments or audits conducted under one regime cannot translate meaningfully to another. This layer ensures that data and regulatory categories can be correctly interpreted and reused across organisational and jurisdictional boundaries without loss of meaning or accuracy. Furthermore, semantic approaches where regulatory requirements are written in machine-readable ontologies, offer opportunities to enable distinct regimes to meet directly at the semantic layer, allowing compliance checks to be queried and partially automated across borders.

Ethical interoperability

Ethical interoperability addresses the ability of diverse institutions, systems, and actors to collaborate across different moral frameworks, rights regimes, and value systems. While absolute ethical uniformity is unfeasible given differing political, cultural, and developmental contexts, this layer seeks to align frameworks around shared foundational principles—such as the protection of human rights, accountability, transparency, safety, and fairness. Such normative alignment provides a common reference point for facilitating cross-border regulatory coordination, and for promoting mutual trust and interoperability in AI among countries. Countries have adopted multi-tiered approaches to establish compatible ethical governance systems at both domestic and international levels. Shared principles such as UNESCO's Recommendation on the Ethics of AI and the OECD AI Principles provide a common reference for assessing AI practices, facilitating cross-border regulatory coordination, and promoting mutual trust and interoperability in AI among countries.

[18] The European Interoperability Framework https://interoperable-europe.ec.europa.eu/collection/iopeu-monitoring/european-interoperability-framework-detail

[19] Chin, Y. C., Raho, D. A., Kim, H.-M., Bi, C., Ong, J., Huang, J., & Stinckwich, S. (2025). Interoperability in AI safety governance: Ethics, regulations, and standards [Policy report]. United Nations University. https://collections.unu.edu/eserv/UNU:10363/Interoperability_in_AI_Safety_Governance.pdf

Legal and regulatory interoperability

Legal interoperability concerns the capacity of differing legal frameworks and constitutional traditions to coexist, coordinate, and facilitate cross-border cooperation without demanding identical substantive rules or complete harmonisation. It functions as a "third way" between unmanaged fragmentation and full harmonisation, allowing Member States to protect their specific domestic policy priorities and accommodate different levels of regulatory and institutional capacity. Member States can enact this layer of interoperability through mechanisms such as mutual recognition agreements, equivalence and adequacy assessments, model contractual clauses, harmonised digital trade agreements, and common risk and impact assessment practices. This dimension is essential for overcoming barriers posed by mixed legal systems (such as the varying traditions of common and civil law) and differing jurisdictional mandates.

Organisational interoperability

This dimension involves the alignment of workflows, business processes, formal relationships, and responsibilities between distinct organisations. This layer of interoperability determines whether day-to-day governance practices can interface effectively between organisations and enable transparency and compliance across jurisdictions. Key mechanisms to operationalise this layer include documentation, shared incident reporting frameworks, coordinated audit schedules, aligned conformity assessment procedures, and synchronised timelines for compliance deadlines.

Technical interoperability

Technical interoperability addresses the compatibility of technological systems, applications, and infrastructures to communicate and exchange data securely across different environments. This layer relies on shared technical standards, Application Programming Interfaces (APIs), common data exchange mechanisms, model interfaces, and safety testing protocols. It dictates whether an AI system developed in one jurisdiction can technically function, integrate, and be evaluated within the digital infrastructure of another. By relying on open standards and shared protocols, technical interoperability can facilitate innovation and reduce a state's structural dependency on proprietary, closed ecosystems controlled by dominant technology monopolies.

Institutional interoperability

Institutional interoperability refers to the formal and informal capacity of governance bodies, national regulatory authorities, standards organisations, and international organisations to coordinate effectively. This is operationalised through joint procedures, cooperative oversight arrangements, cross-border enforcement cooperation, and structured information-sharing mechanisms. Given the complex ecosystem of actors involved in global AI development and the asymmetries in capacity and standard-setting power, it is crucial that cooperative mechanisms ensure interoperability does not entrench institutional dominance by a small number of states or corporate actors.

Opportunities

Establishing and strengthening interoperability in AI governance presents a **transformative opportunity to foster a resilient, equitable, and functional global architecture**. Pursuing interoperability through international cooperation and coordination offers a pragmatic and pluralistic approach that acknowledges diversity while actively bridging deep capacity asymmetries among Member States. Rather than demanding politically complex and developmentally undesirable total harmonisation, negotiating interoperable frameworks allow jurisdictions to maintain institutional pluralism and tailor regulations to their distinct realities and priorities while helping to address some of the structural challenges of fragmentation described earlier. This approach empowers states to experiment with context-specific governance—such as regulatory sandboxes or capacity-building AI strategies—while remaining connected to global norms and participating in international digital trade. The pursuit of interoperability is relevant now because as regulatory regimes mature and consolidate, retroactive alignment becomes progressively harder, and the cost of non-interoperability will compound.

Cross-border interoperability is an essential prerequisite for **closing transnational accountability gaps and mitigating systemic security risks.** Coordinated mechanisms, such as joint incident reporting protocols and cross-border supervisory cooperation, ensure that clear channels for attribution and redress exist when AI systems cause transboundary harms or violate fundamental rights.This cross-border connectivity secures a shared "floor" of rights-based protections, extending baseline safeguards to individuals and vulnerable communities regardless of where an AI system is developed or deployed. In the face of catastrophic threats, such as AI-enabled cyberattacks or the proliferation of dangerous capabilities, (institutional, technical and organisational) interoperability enables the international community to **pool threat intelligence, standardise safety evaluations**, and mount collective, preventative responses rather than relying on isolated and inadequate national defences.

Finally, advancing interoperability in AI governance presents a clear opportunity to **foster a more inclusive AI innovation ecosystem** by removing artificial frictions that disproportionately burden smaller actors and nations. Currently, the fragmentation of governance frameworks and the proliferation of divergent conformity assessment requirements impose compounding, and sometimes irreconcilable, compliance burdens. By enacting interoperability and capacity building mechanisms—such as the mutual recognition of conformity assessments, interoperable digital public infrastructures, and more equitable representation in standard setting—Member States can substantially lower these prohibitive compliance and access barriers and reduce duplicative administrative costs. Advancing governance interoperability can function as an enabler of digital trade and cross-border safety in AI-enabled services. By lowering compliance costs, easing market access for new firms and developers operating internationally, and reducing switching costs to reduce vendor lock-in, interoperability can mitigate the risk of monopolisation and consolidation in the AI industry. It can enable smaller market entrants—particularly those in the Global Majority who are building innovative, context-specific services on top of existing foundation models—to compete on a level playing field. This transition can shift the global landscape from one of exclusionary market stratification toward a more competitive, resilient, and inclusive global AI innovation ecosystem.

Interoperability is relevant now because as regulatory regimes mature and consolidate, retroactive alignment becomes progressively harder, and the cost of non-interoperability will compound.

The limits of AI governance interoperability

For Member States, advancing governance interoperability presents significant strategic advantages as discussed so far. Pragmatically, it reduces compliance friction for cross-border innovation, mitigates the risks of regulatory arbitrage, allows diverse, context-specific regulatory philosophies to coexist, and enables opportunities for participatory engagement, adaptation and mutual learning.

However, interoperability must be approached with a clear recognition of its structural limits and potential risks. Interoperability cannot resolve deep normative divergence over ethical values or fundamental rights, nor can it eliminate underlying geopolitical power asymmetries. Without strict, rights-based guardrails, interoperability mechanisms risk serving as a vehicle for dominant global actors to export their regulatory preferences under the guise of technical neutrality.

An important caution highlighted by consulted experts is that *interoperability is not adequacy*. **Connecting two inadequate regulatory regimes does not inherently make them protective**. If interoperability prioritises corporate compliance convenience over robust public-interest protections, or reduced administrative burden over protection of rights, it risks driving a regulatory "race to the bottom" where interoperability agreements allow the least-regulated jurisdiction to set the effective global ceiling. Furthermore, procedural interoperability without substantive normative alignment may facilitate smooth cross-border coordination without necessarily delivering actual protection for vulnerable communities.

Interoperability mechanisms also face a profound timing challenge: institutional negotiations, treaty frameworks, and standardisation processes operate on timelines measured in years, whereas AI capabilities evolve in months. Consequently, interoperable frameworks risk addressing outdated technological paradigms and struggling to govern sudden developments, such as the rise of agentic AI. Finally, operationalising some interoperability measures (e.g., equivalence determinations and mutual recognition) demands significant institutional and technical capacity. For Member States lacking robust regulatory infrastructure, participating in interoperability arrangements on equal terms remains highly challenging without dedicated multilateral capacity-building support.

Procedural interoperability without substantive normative alignment may facilitate smooth cross-border coordination without necessarily delivering actual protection for vulnerable communities.

Part II:
Paths forward

Part II aims to offer all States and stakeholders involved in discussions on international AI governance with a preliminary suite of actionable pathways aimed at addressing the challenges stemming from global fragmentation. Important governance interoperability efforts are already underway and offer a space of opportunity. The international community does not need to start from scratch. It can leverage robust existing infrastructures of international cooperation and draw highly transferable lessons from other complex transnational policy domains.

Rather than prescribing a definite roadmap, this white paper brings attention to a series of policy, financial, methodological, technological, knowledge- and participation-related measures that respond to AI's most pressing challenges and help develop the building blocks and tools for a global AI governance architecture. Consideration has been given to what transferable lessons can be drawn from successes and governance efforts in other domains, as well as how these successes and efforts have leveraged existing initiatives and architectures of multilateral cooperation. The actionable pathways presented here are not recipes for success nor do they capture all ongoing efforts exhaustively; instead, they represent open-ended guideposts that will necessarily evolve through multi-stakeholder dialogue and through continuous identification of synergies and exploration in different fora and geographical contexts.

The identification of pathways begins with a call for global alignment on minimum rights-based and safety protections that can serve as an overarching guiding principle for all measures. By defining a threshold below which no jurisdiction should fall, minimums leave ample space for states to pursue more ambitious domestic regimes, experiment with policy, or adapt regulations to local priorities.

The case for minimums in AI governance: Toward floors not ceilings

To address the structural challenges stemming from a fragmented global landscape, it is important for the international community to **establish baseline conditions for governance interoperability that can sustain cooperation across diverse jurisdictions.** Identifying, agreeing, and defining these minimums is not merely a technical exercise; it is a profoundly social and political process that determines how costs, benefits, and regulatory authority are distributed across the international community and at each the interoperability entry points described in Part I.

A defensible floor establishes the essential conditions under which a plurality of legitimate regulatory responses can co-exist without discouraging the existence of stronger national measures. A ceiling, by contrast, freezes global ambition at the level of the least demanding consenting state and chills the developmental policy space that smaller jurisdictions and rights-protective majorities need. International AI governance floors should not pre-empt higher national standards and policies; the sovereign right of Member States to govern more ambitiously must be treated as non-negotiable.

A defining consensus among the consulted experts is that global baselines must be strictly conceptualised as a protective "floor," not a restrictive "ceiling."

While establishing minimums offers a pragmatic route to interoperability, it carries severe structural risks that Member States must actively manage. First, there is the risk of **standard-setting asymmetry**, where minimum standards gradually evolve into de facto global norms without inclusive participation and deliberation, effectively embedding the institutional preferences and economic interests of a few dominant technology-producing nations. Second, there is a risk of creating a mere **façade of norms** if the basic layer of obligations is set too low to provide meaningful protection against some of the most pressing AI risks. International baseline agreements have been historically invoked by regulated corporate actors as a ceiling against stronger national measures, and lessons from environmental and data protection governance demonstrate the pattern clearly: once a minimum is established, industry resistance to higher requirements intensifies, and the minimum is progressively treated as the effective global norm.[20] Finally, without differentiated capacity building efforts, the implementation of these minimums is likely to generate **highly uneven distributional effects**. While countries with advanced regulatory and technical infrastructures can easily absorb new compliance requirements, lower-capacity jurisdictions may face overwhelming administrative and financial burdens. A rigid, one-size-fits-all approach is unworkable across countries with vastly different legal traditions and institutional, technical, and financial capacities.

To ensure that minimums do not function as exclusionary barriers for the Global Majority, robust flexibility must be built into their design without compromising the protection of non-negotiable human rights and safety redlines. Adopting principles analogous to the UNFCCC "common but differentiated responsibilities," could allow for tiered implementation timelines and graduated obligations scaled to the specific risk profile of the local AI ecosystem. Crucially, any governance framework founded upon shared minimums must be accompanied by a viable capacity-building framework, directly linking governance commitments to dedicated technical and financial assistance, open tools, shared templates, and model impact assessments. This structural flexibility ensures that global minimums serve as an inclusive foundation for cooperation rather than a mechanism that reinforces existing global inequalities.

[20] Kuzio, J., Ahmadi, M., Kim, K.-C., Migaud, M. R., Wang, Y.-F., & Bullock, J. (2022). Building better global data governance. Data & Policy, 4, e25. https://doi.org/10.1017/dap.2022.17

Actionable pathways for interoperability and inclusive participation in AI governance

The following fourteen (14) actionable pathways comprise a combination of policy, financial, methodological, technological, knowledge-sharing and participatory measures that have emerged from substantive contributions and deliberations with consulted experts. The actionable pathways presented here have been considered with a preliminary assessment of their feasibility, timeline of implementation, and identification of how actions extend and complement one another. As such, the pathways are not rigid recommendations but calls for multistakeholder cooperation and identification of productive new synergies.

The pathways are organised according to how they relate to four guiding pillars: 1) Establishing shared understanding and an infrastructure of meaning; 2) Building the "connecting tissue" between AI governance frameworks; 3) Addressing structural asymmetries in the global AI ecosystem; and 4) Safeguarding fundamental rights, trust and accountability.

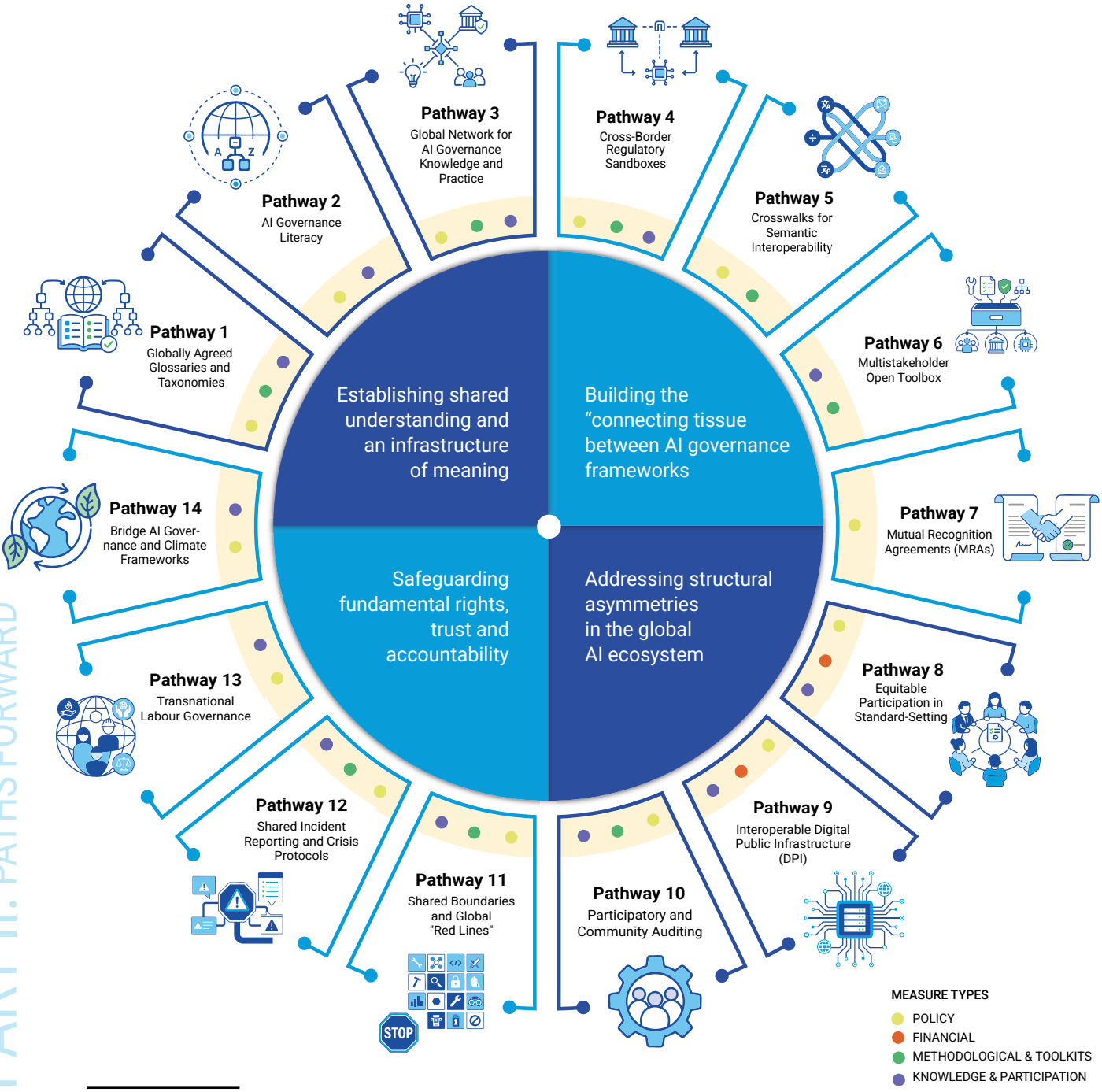
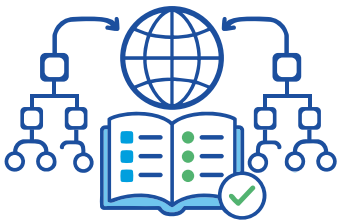


Figure: The 14 actionable pathways at a glance

Pathway 1

Advance and consolidate **globally agreed glossaries and taxonomies**



Policy · Methodological & Toolkits · Knowledge & Participation

Description: Different jurisdictions frequently define critical regulatory concepts in ways that may appear similar on the surface but diverge substantively in practice. For instance, the "high-risk" classification under the European Union's approach differs fundamentally from the "high-impact" designation in South Korea or the capability-based definitions emerging elsewhere. This lack of agreement on terminology and conceptual clarity acts as a significant point of confusion, hindering meaningful international cooperation. This pathway focuses on reducing semantic friction by establishing a shared language between different governance frameworks. It involves agreeing on how to define core concepts, such as what constitutes an "AI system," "high-risk," or "foundation model," and developing shared taxonomies for classifying AI risks and capabilities.

Potential impact: Establishing a common definitional baseline provides a highly achievable mechanism that can significantly reduce interpretive fragmentation and facilitate smoother communication across jurisdictions and sectors. Consolidating shared glossaries establishes mutual intelligibility, reducing compliance friction for transnational innovation and preventing costly regulatory overlap; shared vocabularies thereby lower the friction associated with cross-border dialogue and governance coordination. Furthermore, common taxonomies enable transparency tools, such as incident reporting mechanisms, to generate genuinely comparable data globally, providing a shared empirical reference point for policymakers, regulators, and developers. Ultimately, this foundational baseline acts as an essential precondition for more advanced layers of interoperability, supporting the development of a cohesive international AI ecosystem.

How: The UN Global Dialogue on AI Governance can facilitate an inclusive, multistakeholder process to agree on risk definitions—including core cross-border risks like CBRN, cyber, autonomy, mass manipulation in the non-military domain—leveraging existing frameworks such as the OECD AI definitions or ISO/IEC standards to support comparable evaluation. This pathway must be approached carefully to navigate instances where terminology inevitably embeds normative choices about acceptable risk thresholds.

Pathway 2

Expand and sustain **targeted AI governance literacy** across all stakeholder groups

Policy · Knowledge & Participation

Description: Meaningful AI governance requires going beyond top-down regulations by building specialised institutional, diplomatic, and organisational knowledge required to effectively coordinate and implement interoperable frameworks. This pathway aims to embed AI literacy and governance expertise deeply into public administrations, the private sector, diplomats, judiciaries, and civil society ensuring all stakeholders keep abreast of the state of the art and of practical knowledge on rights risks, procurement, assurance, public-sector AI, risk and impact assessment, and redress. This pathway is an enabler of other measures and has direct synergies with pathways 3, 6 and 10.

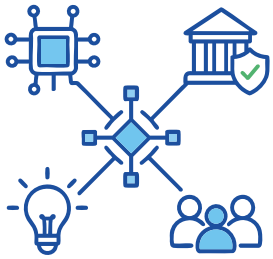
Potential impact: Strengthening literacy closes the severe capacity asymmetries; it empowers local regulators to enforce rules effectively, enables judges to adjudicate complex AI-related harms, public sector and private sector practitioners to conduct audits and assessment throughout the life cycle, and citizens to critically engage with AI systems and demand accountability. Targeted and differentiated AI governance literacy across jurisdictions and stakeholder groups is a key building block for ensuring interoperability measures are legitimate and capacity sensitive.

How: UN system and multilateral partners can establish dedicated capacity-building funds and training programs explicitly targeted at policymakers, public officials, and civil society in underrepresented regions. This includes utilising existing UN infrastructure, such as the United Nations University (UNU), to provide specialised technical and policy training to Member States and their diplomats; cooperation exchanges for public officials, expanding successful technical and governance capacity-building models (like the EU TAIEX project,[21] UNESCO's Judges Initiative,[22] or UNDP human rights impact toolkits[23]) to a global scale, and launching UN on-site training, fellowships, and multilingual digital campaigns tailored to diverse socio-cultural realities to raise public awareness. The AI capacity building fellowship for government officials and research programmes, called for in A/RES/80/173, and the UN-supported Network of Centers for Exchange and Cooperation on AI Capacity Building, could also contribute to this effort.

[21] https://international-partnerships.ec.europa.eu/funding-and-technical-assistance/technical-assistance/taiex-technical-assistance-and-information-exchange_en
[22] https://www.unesco.org/en/artificial-intelligence/rule-law
[23] https://www.undp.org/eurasia/publications/human-rights-impact-ai-assessment-toolkit

Pathway 3

Establish a **global network for AI governance knowledge and practice**



Policy · Methodological & Toolkits · Knowledge & Participation

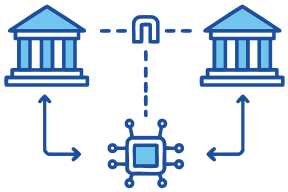
Description: This pathway proposes a polycentric, decentralised network connecting regional capacity hubs, academic institutions, industry practitioners, civil society, artists, and policymakers. This network functions as an infrastructure of meaning, connecting diverse knowledge systems across different cultural, linguistic, legal, and developmental contexts. This also involves incorporating indigenous sovereignty perspectives and African philosophies of communal responsibility into AI governance.

Potential impact: This network shifts global AI governance from an exclusionary, "one-size-fits-all" approach into a pluralistic, culturally grounded framework. It guarantees that the perspectives of structurally marginalised communities, platform and data workers, and the Global Majority serve as authoritative evidence in shaping global rules. It also creates a mechanism for continuous adaptive governance and peer learning on rights risks, procurement, assurance, public-sector AI and redress.

How: The UN can develop a globally connected ecosystem of AI governance knowledge and practice by linking existing regional and national hubs or knowledge/practice nodes and fostering new ones where gaps remain. Rather than creating new centralised structures, this approach would build a **network of networks** that leverages established centres and initiatives across regions. Such a network could facilitate continuous knowledge exchange, co-learning, and the emergence of Communities of Practice across countries, regions and globally. It would enable the sharing of experiences, tools, and lessons learned, while supporting more coordinated responses to emerging challenges. To support its sustainability and inclusivity, this effort must be underpinned by dedicated financing, including through a multilateral and multi-stakeholder funding mechanism, to strengthen existing hubs and catalyse new ones where needed. Over time, this interconnected architecture could also contribute to shaping a shared understanding of AI governance, serving as a **global infrastructure for knowledge, dialogue, and meaning-making** that enriches the UN decision-making processes. This should involve regional economic commissions, and relevant UN system partners while leveraging the success of established regional bodies. The AI Governance for Humanity Lab, under the UN ODET's mandate to facilitate inclusive, multi-stakeholder policy dialogue on AI governance, is very well positioned to establish and facilitate this network coordinating and synergising efforts with the UN system and Member States.

Pathway 4

Support networked **cross-border regulatory sandboxes**



Policy · Methodological & Toolkits · Knowledge & Participation

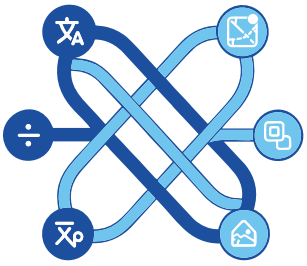
Description: Regulatory sandboxes have been adopted in several jurisdictions, including the EU, the UK, China, South Korea, Chile, Colombia, and Brazil, serving as controlled environments where developers, private sector actors, and public authorities can jointly test AI applications under regulatory supervision before widespread deployment. To prevent regulatory learning from remaining confined within isolated national jurisdictions, these mechanisms must be linked internationally. This pathway entails establishing networked regulatory sandboxes supported by shared protocols for cross-border pilots, model evaluation, and evidence sharing.

Potential impact: Establishing globally connected regulatory sandboxes allows policymakers to test, adapt, and iterate regulatory tools in cooperative environments. It enables regulatory learning to accumulate globally, providing Member States with empirical evidence regarding which governance interventions are effective in practice. For developers, aligned testing methodologies and shared sandbox protocols across borders lower the friction and compliance costs of operating internationally. Crucially, for lower-capacity jurisdictions, participating in cross-border sandboxes provides vital access to rigorous testing infrastructure and evaluation data that would be highly resource-intensive to develop independently, which can foster a more inclusive innovation ecosystem.

How: The establishment of an internationally connected platform of AI regulatory sandboxes, supported by shared protocols for cross-border pilots and evidence sharing, could be facilitated by suitable UN system efforts coordinated through the AI Governance for Humanity Lab. Furthermore, international capacity-building initiatives, drawing on successful models such as the EU's Public Administration Cooperation Exchange (PACE) programme, could be expanded globally to assist lower-capacity states in building the institutional expertise required to meaningfully participate in and benefit from these cross-border sandboxing initiatives.

Pathway 5

Support development and adoption of **crosswalks and novel approaches** for semantic interoperability



Policy · Methodological & Toolkits · Knowledge & Participation

Description: To manage the scale and speed of AI deployment, governance can benefit from novel automated approaches for semantic interoperability involving the translation of policy requirements, risk categories, and documentation duties into machine-readable formats. Concurrently, crosswalks operate as non-binding mapping tools, systematically illustrating how the rules, risk classifications, and definitions of one jurisdiction correspond to those of another or to international standards (like ISO/IEC 42001). Automation tools should integrate a socio-technical perspective that recognises that developing semantic approaches is not a purely technical exercise; any machine-readable representation inherently hard-codes choices about what constitutes a "risk," a "harm," or a relevant data category.

Potential impact: By making governance gaps visible and structured, this pathway can generate powerful incentives for cross-border alignment without the political resistance of harmonisation. Furthermore, semantic alignment and crosswalks ensure that differing terminologies, such as a "high-risk" designation under the EU AI Act versus a "high-impact" designation elsewhere, become legible and comparable across borders. Crucially, adopting a socio-technical approach ensures that interoperability does not become a vehicle for dominant technology-producing nations and corporate actors to impose their institutional assumptions and risk tolerances on the rest of the world.

How: To transition from abstract commitments to functional infrastructure, the UN Global Dialogue on AI Governance can support the adoption of novel approaches for semantic interoperability. The AI Governance for Humanity Lab can host and coordinate the advancement of this pathway, leveraging its network mobilisation function to accelerate global distribution, scale, and uptake in direct collaboration with universities, independent institutes, specialised developers, and civil society organisations. The UN Global Dialogue on AI Governance can support the development and adoption of open-access tools, crosswalks, and reference schemas such as the AI Risk Ontology (AIRO) for encoding regulatory requirements,[24] and the Croissant metadata standard for machine-readable dataset documentation.[25]

[24] https://oecd.ai/en/catalogue/tools/airo-ai-risk-ontology
[25] https://docs.mlcommons.org/croissant/docs/croissant-spec-1.0.html

Pathway 6

Develop a **multistakeholder open toolbox to operationalise governance approaches**

Methodological & Toolkits Knowledge & Participation

Description: This pathway entails creating a comprehensive suite of open-access operational tools tailored for a diverse array of stakeholders, including policymakers, developers, researchers, and civil society. Crucially, this involves translating normative commitments into procedural guidance and strengthening mechanisms that link AI governance to existing international human rights obligations through standardising tools such as Human Rights Impact Assessments (HRIAs)[26] and Ethical Impact Assessments (EIAs)[27] which provide structured methodologies for evaluating the potential harms of AI systems prior to deployment. Furthermore, this pathway includes the provision of open-access and multilingual model AI legislation and standardised public-sector procurement templates to assist capacity-constrained jurisdictions in enacting robust governance without needing to build administrative frameworks entirely from scratch.

Potential impact: Providing shared, multilingual, open-source operational tools drastically lowers the barrier to entry for practitioners and public-sector authorities, particularly in lower-capacity jurisdictions and the Global Majority. By offering shared templates and model laws, Member States can reduce duplicative compliance burdens and provide immediate regulatory clarity, preserving sovereign policy space while facilitating cross-border coordination. Implementing standardised HRIAs and EIAs ensures that ethical and human rights commitments are actively translated into concrete engineering practices and verifiable evidence requirements, thereby protecting vulnerable populations from algorithmic discrimination and safeguarding democratic processes. Additionally, integrating AI governance directly into public procurement through standardised clauses leverages the significant purchasing power of states to enforce baseline safety and rights requirements across global AI supply chains.

How: The UN can provide the foundations for a shared, open-access toolbox containing actionable, multilingual templates for HRIAs, EIAs, public-sector procurement clauses, risk classification schemas, and audit checklists that can be easily localised by domestic authorities and developers. This initiative should aggregate and expand upon existing resources, such as the UNDP HRIA toolkit, UNESCO's Ethical Impact Assessment and Readiness Assessment Methodology, and the World Economic Forum's procurement models, creating a comprehensive UN-wide implementation package.

- **Model Legislative Provisions:** Building on successful precedents like UNCITRAL's work on international trade law,[28] the UN system can advance non-binding model legislative provisions for AI governance. These model laws would provide lower-capacity nations with adaptable, high-quality templates for addressing AI transparency, incident reporting, and accountability, mitigating

[26] https://www.undp.org/eurasia/publications/human-rights-impact-ai-assessment-toolkit
[27] https://www.unesco.org/ethics-ai/en/eia
[28] UNCITRAL Model Legislative Provisions on Public-Private Partnerships (2019) https://uncitral.un.org/en/mlpppp

the technical burden of original legislative drafting. Inclusive drafting processes are essential to ensure these models do not embed assumptions inadequate for the Global Majority.

- **Mandating and Supporting Impact Assessments**: Member States can mandate HRIAs and EIAs for high-risk AI deployments, particularly in sensitive public sector domains such as welfare, policing, and migration. To support compliance, public bodies could develop and provide free, open-access agentic AI tools designed specifically to assist entities in fulfilling these rigorous impact assessments and risk management protocols.
- **Standardising Procurement Clauses:** Specialised agencies in the UN system can coordinate the creation of standardised procurement clauses that public administrations globally can adopt. These templates will ensure that any acquired AI systems are contractually bound to align with international human rights baselines, algorithmic transparency requirements, and post-deployment monitoring standards.

Pathway 7

Enable

Mutual Recognition Agreements (MRAs)

Policy

Description: Rather than harmonising substantive regulatory requirements, mutual recognition agreements (MRAs) for AI conformity assessments could reduce duplication of compliance costs without requiring full regulatory convergence. Under an MRA, one jurisdiction formally agrees to accept another jurisdiction's regulatory determinations, conformity assessments, or audit results as satisfying its own equivalent requirements. This pathway shifts the focus from mandating identical substantive rules to agreeing on what constitutes a credible, rigorous evaluation or audit process. This approach can draw on successful precedents in pharmaceutical regulation, telecoms, and the Common Criteria Recognition Arrangement (CCRA) in cybersecurity.

Potential impact: MRAs drastically reduce duplicative compliance burdens and testing costs, which currently stratify the market and disadvantage smaller actors who lack the resources to maintain parallel compliance architectures across multiple jurisdictions. For smaller economies and lower-capacity states, participating in MRAs allows them to leverage the rigorous assessments conducted by better-resourced regulatory authorities abroad, ensuring high safety standards for their citizens without needing to duplicate the complex, costly evaluation infrastructure domestically.

How: Member States can move toward negotiating MRAs incrementally, starting with procedural alignment between coalitions of jurisdictions or bilateral arrangements that other states can subsequently accede to. To facilitate these agreements, jurisdictions could use internationally recognised, process-oriented technical standards (such as ISO/IEC 42001) as neutral bridging mechanisms; a jurisdiction might accept certification under this standard as partial evidence of compliance, establishing the foundational trust required for mutual recognition.

Pathway 8

Support **global participation and equitable representation in standard-setting**

Policy · financial · Knowledge & Participation

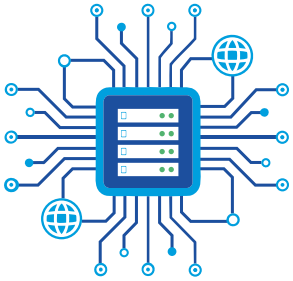
Description: International technical standard-setting processes currently operate with severe participation asymmetries; organisations such as ISO, IEC, and IEEE are predominantly shaped by industry experts and delegations from high-income, technology-producing nations. Representation challenges are not unique to AI, but they are especially consequential because AI standards encode value-laden decisions on topics like fairness, safety, and accountability. True participatory governance requires moving beyond nominal consultation to genuine co-design, redistributing agenda-setting power. This pathway entails targeted support for improving the inclusion of Global Majority states, civil society, labour unions, and structurally marginalised groups in global AI governance fora and standard setting.

Potential impact: Embedding equitable representation can safeguard global AI governance frameworks and standards against regulatory capture by dominant corporate actors and prevent the exportation of politically contested normative choices under the guise of technical neutrality. Resourced participation guarantees that global rules are informed by the lived realities, cultural contexts, and vulnerabilities of the communities most adversely affected by AI systems, thereby enhancing the democratic legitimacy, equity, and operational effectiveness of international AI governance frameworks and standards.

How: Influential standards-making initiatives are being undertaken by non-governmental organisations (e.g., ISO/IEC) and private technical consortia (e.g., MLCommons). Because of this, stakeholders excluded from these organisations have little direct oversight of, and influence over, the technical standards being developed for AI. Many standards-making institutions recognise the need to improve equitable participation, but these efforts have not gone far enough. To address this problem, the UN Global Dialogue on AI Governance could explore mechanisms to establish more equitable AI standards-making; for example by advocating for targeted financial support for more equitable participation and aggregation of interests in standard-making, defaulting to multilingual access, and running regional convenings before global standard-making negotiations through the ITU. Given the time and resources needed to meaningfully participate in these processes, funding would need to be sustained and include the provision of stipends and travel support, translation services, and multi-year accreditation.

Pathway 9

Support interoperable Digital Public Infrastructure (DPI) and public-interest AI



Policy · financial · Knowledge & Participation

Description: The global concentration of computational infrastructure, foundation models, and data ecosystems in a few private corporate actors and jurisdictions generates profound structural dependencies between AI-producing and AI-adopting countries. Countries also face unequal institutional, technical, and regulatory capacities, which can affect their ability to participate meaningfully in interoperable AI governance arrangements. Supporting interoperable Digital Public Infrastructure (DPI) and public-interest AI entails coordinated multilateral investment in sovereign or pooled compute provisioning, shared data trusts, and open-source models, functioning as a vital, democratically accountable counterweight to private tech monopolies.

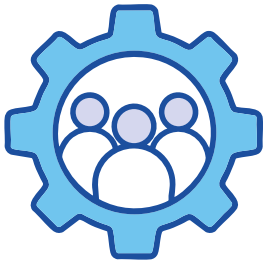
Potential impact: Developing robust public-interest AI and DPI can provide states with the technical sovereignty necessary to exert regulatory autonomy, significantly reducing reliance on foreign commercial providers for critical functions in healthcare, education, and public administration. Interoperable DPI creates genuine competitive entry points for local developers and small-to-medium enterprises (SMEs) by removing artificial frictions and vendor lock-in. Ultimately, it balances capabilities more equitably across the globe, allowing lower-capacity nations to proactively shape AI trajectories and integrate into the global AI value chain, rather than merely consuming externally designed and governed models.

How: The international community could advance the UN Global Digital Public Infrastructure (DPI) efforts to finance and deploy secure, inclusive, and interoperable DPIs. Member States could foster public-interest AI ecosystems by establishing procurement-linked open architectures, standardising portability rules, and supporting open-access, contextually representative training datasets and evaluation benchmarks. To operationalise this, capacity-sensitive mechanisms for AI capacity development—such as the Global Fund for AI proposed by the Secretary-General—would be needed to fund the expansion of localised, open-source sovereign cloud infrastructures. Concurrently, international frameworks should facilitate South-South cooperation to pool resources for regional supercomputing hubs and culturally adapted foundation models (such as Singapore's SEA-LION, African language initiatives or LATAM-GPT).

Pathway 10

Implement

Participatory and community Auditing as legitimate governance mechanisms



Policy · Methodological & Toolkits · Knowledge & Participation

Description: Most international AI governance debates treat governance as an activity performed by states, upon companies, on behalf of citizens, leaving the publics in whose name AI is regulated largely absent from the rooms where rules are written and operationalised. Implementing participatory and community auditing corrects this by treating community oversight and the infrastructure that supports it—such as tools, methodologies, training, and certification—as first-class governance mechanisms rather than optional consultation exercises. This applies to AI systems, foundation models, and the data that feeds them. This pathway involves formally integrating social dialogue and community-led evidence-gathering into the auditing process, ensuring that the lived experiences of those directly affected by AI harms, including workers, structurally marginalised groups, and indigenous communities, serve as authoritative input.

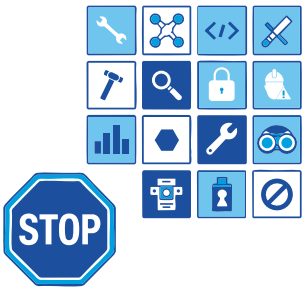
Potential impact: Participatory auditing ensures that global rules are informed by the lived realities of those most affected by AI systems, counteracting the blind spots of traditional expert panels and developer-defined tests. Comparative evidence on social dialogue around AI and algorithmic management suggests that governance works significantly better when workers and their organisations have a voice in how systems are introduced, monitored, and challenged. This approach effectively connects ethical principles with the institutions and communities capable of enforcing them. It transforms interoperability from a tool of corporate compliance convenience into a pluralistic, democratic infrastructure that actively mitigates the structural power imbalances inherent in global AI development.

How: Member States and the international community could co-fund and institutionalise standardised participatory harm auditing methodologies. The AI Governance for Humanity Lab could host and coordinate the advancement of this pathway. Acting as a central convener, the Lab can broker collaboration among universities, research institutes, and civil society organisations to facilitate the development, scaling, and deployment of collaborative models building on existing resources such as the Participatory Harm Auditing Workbenches and Methodologies (PHAWM) project[29] and the OECD citizen participation guidelines.[30] These resources provide the tools and training necessary to empower citizens without technical AI backgrounds to formally audit systems for societal harms. Furthermore, the UN Global Dialogues on AI Governance could involve participatory evidence-gathering in high-impact domains through community hearings, multilingual consultations, worker testimony, and case documentation from affected groups, treating different ways of knowing and experiencing AI as essential governance evidence.

[29] The Participatory Harm Auditing Workbenches and Methodologies (PHAWM) Project https://phawm.org/
[30] OECD Guidelines for Citizen Participation https://www.oecd.org/en/publications/oecd-guidelines-for-citizen-participation-processes_f765caf6-en/full-report/component-5.html

Pathway 11

Define
shared boundaries and global "red lines"



Policy · Methodological & Toolkits · Knowledge & Participation

Description: This pathway involves agreeing globally on a specific set of prohibited AI use-cases that either severely violate fundamental human rights or pose uncontainable, catastrophic cross-border risks. This includes the exploration of "fundamental utility assessment" which interrogates the use of AI technologies, allowing governments or victims to demand the removal of AI systems making unreasonable, harmful decisions.

Potential impact: Establishing this non-negotiable global protective floor prevents regulatory arbitrage situations where dangerous systems banned in strict jurisdictions are exported to less stringent environments.

How: **Anchor in Existing Human Rights Law:** global prohibitions can be based firmly in universally ratified international human rights law to establish a universally recognised normative floor, rather than developing new frameworks from scratch.

- **Implement Pre-Release Thresholds**: boundaries for frontier models can be operationalised by mandating rigorous pre-release evaluation with a public summary of findings for models above a defined capability or compute threshold; halting a model's release entirely if it crosses a defined red line threshold (e.g., demonstrating dangerous autonomous replication).
- **Address open-weight release:** The international community can recognise that pre-release thresholds and territorial controls become largely ineffective once a model's weights are released openly. This makes coordinated, pre-release redlines for frontier open-weight models a structural rather than a domestic question, applied at the point of release, before weights enter open circulation.
- **Leverage Procurement and Investment:** red lines can be enforced through market mechanisms, such as advocating for governments to make public technology procurement and investments conditional upon companies proving they do not develop, use, or trade in prohibited systems.

Pathway 12

Establish **shared incident reporting and crisis protocols**

Policy · Methodological & Toolkits · Knowledge & Participation

Description: This pathway entails establishing a global system for reporting serious AI incidents, capability surprises, and near-misses, drawing on highly transferable models like the aviation sector's voluntary incident-reporting culture and pharmacovigilance networks. Alongside incident reporting, this pathway entails the creation of pre-emptive global crisis preparation protocols for transboundary threats (e.g., AI-enabled cyberattacks or biological threats), and can be supported by the provision of open-source operational tools—such as model impact assessments, audit checklists as defined in Pathway 6.

Potential Impact: A shared incident reporting framework closes severe jurisdictional blind spots, ensuring that when an AI system fails or behaves dangerously in one country, the rest of the world can anticipate and respond to the threat. Establishing crisis protocols—including technical threat diagnosis and policy backchannels—maximises the international community's ability to weather complex, transnational AI emergencies.

How: The UN could endorse and administer a common, machine-readable schema for AI incident reporting that Member States can adopt by reference. To operationalise crisis protocols, national safety and security institutes can host scenario exercises on the sidelines of global dialogues to stress-test emergency backchannels and response mechanisms among regional policymakers.

Pathway 13

Advance
transnational labour related AI governance

Policy Knowledge & Participation

Description: AI governance is profoundly linked to power imbalances in the workplace and platform work. Automated systems, particularly in the platform economy, increasingly shape hiring, task allocation, compensation, performance management, and disciplinary actions, often with deep opacity and limited accountability. Furthermore, the development of AI relies fundamentally on a vast, often invisible global supply chain of data workers—including data labellers and content moderators—who are frequently situated in the Global South. This hidden economy is highly susceptible to exploitation, as technology companies can easily relocate these operations across jurisdictions to exploit weaker labour protections and evade regulatory friction. Promoting transnational labour governance entails moving beyond broad, human-centred ethical language to establish concrete protections that connect AI principles directly with existing labour institutions, ensuring that algorithmic management is governed fairly across borders.

Potential Impact: A dedicated focus at the Global Dialogue on AI Governance on transnational labour governance prevents AI from deepening socioeconomic insecurity, work intensification, and workplace inequality. By ensuring that AI deployment is shaped by public oversight, worker voice, and social dialogue, it protects vulnerable populations from opaque, cross-border automated decisions. Treating labour rights as core AI governance issues empowers workers and unions to formally challenge harmful algorithmic management practices, reducing the risks of privacy violations and unjust automated assessments. Crucially, formalising the protection of data workers in the AI supply chain would ensure that the digital economy respects fundamental human rights and decent work standards globally, preventing the externalisation of social risks to the most vulnerable populations.

How: This pathway requires shifting from voluntary guidelines to binding international instruments and concrete enforcement mechanisms. The International Labour Organization (ILO) can serve as the central node for this effort, building upon its current push to establish a new Convention and Recommendation on decent work in the platform economy. To protect workers in the global AI supply chain, governance frameworks must extend labour protections and standards directly to data labellers and content moderators. Additionally, the Global Dialogue on AI Governance should formally integrate platform-worker collectives, gig-worker unions, and civil society organisations into social dialogue and collective bargaining mechanisms, granting them standing to co-design governance of AI systems at the workplace.

Pathway 14

Bridge

AI governance and climate frameworks

Policy · Knowledge & Participation

Description: The intersection of AI development and environmental sustainability represents a critical, yet frequently overlooked, governance gap. The rapid scaling of AI models and the proliferation of globally distributed data centres consume vast amounts of energy and water, compounding the environmental footprint of AI development. Currently, climate governance instruments, such as those under the Paris Agreement or the UNFCCC processes, do not systematically account for AI-driven energy demand, while AI governance frameworks frequently fail to address environmental externalities. Bridging these domains requires structural alignment between technology regulation and global climate commitments to treat sustainability as a core component of responsible AI.

Potential Impact: Integrating these frameworks will ensure that the exponential growth of AI technologies does not derail global climate targets or severely complicate emissions planning. It enables the international community to accurately measure and mitigate the environmental costs of the AI industry, preventing jurisdictions with less restrictive environmental regulations from becoming havens for highly energy-intensive operations that externalise their ecological costs onto the rest of the world. It would also stimulate the development of more efficient models that do not assume infinite access to data and computational power.

How: The UN can facilitate formal coordination between AI governance bodies and environmental mechanisms (such as the UNFCCC) to systematically incorporate AI-driven resource demands into global emissions planning and climate commitments. Furthermore, following UN Secretary-General's recent announcement of the AI Environmental Transparency Initiative,[31] Member States can mandate environmental impact reporting as a core component of AI system documentation and pre-deployment assessments, requiring developers to rigorously assess and publicly disclose the energy consumption, computational resource utilisation, and broader ecological footprint associated with training and deploying large-scale foundation models.

[31] Asadullah M. UN Secretary-General launches AI environmental transparency initiative. United Nations University. https://unu.edu/inweh/news/un-secretary-general-launches-ai-environmental-transparency-initiative-calling-ai

Conclusion

The actionable pathways delineated within this white paper represent **strategic avenues** for fostering interoperability and ensuring inclusive participation in the global AI governance architecture. Each pathway raises multifaceted operational, social, financial, and political complexities that will inevitably necessitate continuous refinement and iterative advancement. Rather than barriers, these challenges should be viewed as **dynamic friction points** requiring sustained, multistakeholder cooperation and shared responsibility. While certain interventions offer prospects for immediate, short-term impact, others demand profound diplomatic coordination, significant capital commitment, and long-term institutional alignment across jurisdictions.

Against this evolving backdrop, the AI Governance for Humanity Lab is uniquely positioned to harness the convening power of the United Nations while embracing the adaptive spirit of the Laboratory as a vital space for experimentation and normative innovation. The Lab is primed to utilise its network mobilisation function to accelerate key knowledge-sharing and participatory measures (Pathways 2, 3, 4, and 8). Furthermore, it can play a pivotal role in catalysing innovative responses to the rapid pace of technological change by co-developing novel interoperability approaches and operational toolkits (Pathways 5, 6, and 10). Ultimately, looking forward, it is clear that AI governance interoperability is not a static destination but an ongoing process of institutional and societal adaptation. This white paper thus offers key steps toward building a more resilient and inclusive AI governance landscape in the years to come.

The AI Governance for Humanity Lab will continue to connect communities and explore opportunities for innovation in AI governance. If you are interested in collaborating with the Lab contact ai-governance-lab@un.org

About the Lab

The AI Governance for Humanity Lab is a UN initiative anchored in the United Nations Office for Digital and Emerging Technologies (UN ODET) and based in Valencia, Spain. Established with the support of the Kingdom of Spain, the AI Governance for Humanity Lab advances the recommendations of Governing AI for Humanity, the report of the United Nations Secretary-General's High-level Advisory Body on Artificial Intelligence, and delivers on ODET's mandate to facilitate inclusive, multistakeholder policy dialogue on AI governance. In particular, it contributes practice-based insights, operational learning, and collaborative networks that can inform and support the work of UN-led processes—including the Global Dialogue on AI Governance—while remaining distinct from scientific assessment and intergovernmental deliberation functions.

More information about the lab can be found on the website https://www.un.org/digital-emerging-technologies/content/ai-governance-humanity-lab.

References

Aczel, M. and others. (2026). *Environmental cost of AI's energy use: Carbon, water and land footprints. United Nations University Institute for Water, Environment and Health.* https://doi.org/10.53328/INR26RMA002

Adams, R. and others. (2026). Global Index on Responsible AI (2nd ed.). Global Center on AI Governance. https://www.global-index.ai/download/report/2026

Asadullah M. UN Secretary-General launches AI environmental transparency initiative. United Nations University. https://unu.edu/inweh/news/un-secretary-general-launches-ai-environmental-transparency-initiative-calling-ai

Belli, L., & Gaspar, W. B. (Eds.). (2025). The Quest for AI Sovereignty, Transparency and Accountability: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition. https://cyberbrics.info/wp-content/uploads/2025/02/The-Quest-for-AI-Sovereignty-Transparency-and-Accountability-Pre-Print.pdf

Chin, Y. C. and others (2025). *Interoperability in AI safety governance: Ethics, regulations, and standards* [Policy report]. United Nations University. https://collections.unu.edu/eserv/UNU:10363/Interoperability_in_AI_Safety_Governance.pdf

Doellgast, V. and others. (2025). Global case studies of social dialogue on AI and algorithmic management, ILO Working Paper 144 (Geneva, ILO). https://doi.org/10.54394/VOQE4924

Domínguez Hernández, A., and others (2026). Towards a sociotechnical ecology of artificial intelligence: Power, accountability, and governance in a global context. Springer AI and Ethics. https://doi.org/10.1007/s43681-025-00902-6

European Commission. *The European Interoperability Framework.* https://interoperable-europe.ec.europa.eu/collection/iopeu-monitoring/european-interoperability-framework-detail

European Commission. TAIEX https://international-partnerships.ec.europa.eu/funding-and-technical-assistance/technical-assistance/taiex-technical-assistance-and-information-exchange_en

The First Nations Principles of OCAP®. https://fnigc.ca/ocap-training/

Gallup. (2026). *Public Perceptions of Artificial Intelligence Infrastructure and Localized Environmental Impact.* Gallup Poll Report. https://news.gallup.com/poll/709772/americans-oppose-data-centers-area.aspx

Global Indigenous Data Alliance. *CARE principles for Indigenous data governance.* https://www.gida-global.org/careprinciples

MLCommons. Croissant Commons Specification http://mlcommons.org/croissant/1.0

International Labour Organization (ILO). (2024). 2024 *Labour Overview of Latin America and the Caribbean.* https://www.ilo.org/publications/2024-labour-overview-latin-america-and-caribbean

Jiang, M., & Belli, L. (Eds.). (2024). Digital Sovereignty from the BRICS Countries: How the Global South and Emerging Power Alliances Are Reshaping Digital Governance. Cambridge University Press. https://doi.org/10.1017/9781009531085

Kuzio, J. and others (2022). Building better global data governance. Data & Policy, 4, e25. https://doi.org/10.1017/dap.2022.17

Nyamnjoh, A. (2025). B*eyond a buzzword: Can Ubuntu reframe AI Ethics?* https://health.uct.ac.za/ethics-lab/articles/2025-09-09-beyond-buzzword-can-ubuntu-reframe-ai-ethics-0

OECD Guidelines for Citizen Participation. https://www.oecd.org/en/publications/oecd-guidelines-for-citizen-participation-processes_f765caf6-en/full-report/component-5.html

OECD AI Risk Ontology https://oecd.ai/en/catalogue/tools/airo-ai-risk-ontology

Regilme, S. S. F. (2024). Artificial intelligence colonialism: Environmental damage, labor exploitation, and human rights crises in the Global South. *SAIS Review of International Affairs.* https://muse.jhu.edu/article/950958

Roberts, H., & Ziosi, M. (2025). *Can we standardise the frontier of AI?* Oxford AI Governance Institute. https://aigi.ox.ac.uk/publications/can-we-standardise-the-frontier-of-ai/

Sajadieh, S. and others (2026). The AI Index 2026 annual report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University. https://hai.stanford.edu/ai-index/2026-ai-index-report/public-opinion

Te Mana Raraunga. *Māori Data Sovereignty Network.* https://www.temanararaunga.maori.nz/

The Participatory Harm Auditing Workbenches and Methodologies (PHAWM) Project https://phawm.org/

UK AI Security Institute. (2025). *International consensus and open questions in AI evaluations.* https://www.aisi.gov.uk/blog/international-ai-network-consensus-and-open-questions

UNCITRAL. (2019). Model Legislative *Provisions on Public-Private Partnerships.* https://uncitral.un.org/en/mlpppp

UNDP Human Rights Impact Assessment. https://www.undp.org/eurasia/publications/human-rights-impact-ai-assessment-toolkit

UNESCO. Ethical Impact Assessment. https://www.unesco.org/ethics-ai/en/eia

UNESCO. Artificial Intelligence and the rule of law https://www.unesco.org/en/artificial-intelligence/rule-law

Toward interoperability and inclusive participation in AI Governance

White paper
June 2026

AI GOVERNANCE FOR HUMANITY LAB