MAP AI

Multistakeholder Approach to Participation in AI Governance

AI GOVERNANCE, SAFETY AND INFRASTRUCTURE

A high-level mapping of contemporary debates, developments, and challenges

GLOBAL NETWORK INITIATIVE

NLU DELHI

CENTRE FOR COMMUNICATION GOVERNANCE

This briefing note was commissioned by the Centre for Communication Governance at National Law University Delhi (CCG-NLUD) and the Global Network Initiative (GNI), and written by Amrita Sengupta.

Date of publication: March 2026

It is intended to provide context for and help inform engagement in the activities convened alongside the AI Impact Summit as part of the Multistakeholder Approaches to Participation in AI governance (MAP-AI) project. The analysis and presentation of facts herein should not be understood to represent the views of CCG-NLUD, GNI, or any of those who participate in those activities.

Table of Contents

About MAP-AI

The Multistakeholder Approaches to Participation in AI Governance (MAP-AI) initiative's objective is to foster meaningful and effective multistakeholder engagement across a range of critical AI governance-focused convenings, processes, and initiatives, with a particular focus on elevating underrepresented voices and perspectives.

MAP-AI was launched by the Global Network Initiative (GNI) and the Centre for Communication Governance (CCG) at National Law University Delhi in 2025, with the immediate goal of enhancing multistakeholder engagement and academic and civil society influence at the India AI Impact Summit, hosted by the Government of India in New Delhi in February 2026. Learn more about our Impact Summit activities.

Building on our activities at the Impact Summit, MAP-AI will synthesize lessons learned, continue to facilitate multistakeholder engagement, and publish insights to inform future international AI governance efforts.

For more information, please reach out to:
info@globalnetworkinitiative.org or ccginfo@nludelhi.ac.in

MAP AI
Multistakeholder Approach to Participation in AI Governance

GLOBAL NETWORK INITIATIVE

NLU DELHI

CENTRE FOR COMMUNICATION GOVERNANCE

SECTION 1

Introduction

Artificial intelligence encompasses computational systems that process data to generate predictions, classifications, decisions, or content. The Organisation for Economic Co-operation and Development (OECD) defines an AI system as "a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment."[1] General-purpose AI, are usually capable of solving a wide array of tasks without being specifically designed for them.[2] Particularly foundation models (which are also considered general purpose models given the generality of capabilities of the model) like GPT-4 process vast datasets to generate text, images, code, and other outputs based on learned patterns.[3][4]

While much contemporary attention focuses on generative AI[5] due to its rapid advancement and deployments, and especially the massive uptake from end-users, traditional machine learning has delivered significant societal benefits that risk being overshadowed. Applications including disease diagnosis, climate modeling, agricultural optimization, and accessibility tools for persons with disabilities demonstrate machine learning's positive potential.[6][7]

As AI systems continue to advance and integrate across societies, important questions emerge. This mapping document provides a high-level overview of key debates, specifically around general-purpose AI systems, across three broad themes: AI governance, safe and trusted AI, and AI infrastructure. Drawing on various policy documents, academic and scholarly research, and civil society research and advocacy materials, this mapping seeks to inform ongoing conversations about approaches to AI governance and development.

SECTION 2

Global AI Governance

As a result of the increased update of AI systems, the conversation around AI governance has accelerated. From the G20 AI Principles to the UN and Global Partnership on AI (GPAI) Resolutions on AI, the African Declaration on AI, and most recently, the Hamburg Declaration on Responsible AI, these efforts reflect unprecedented international recognition that AI's impact transcends borders and requires coordinated global action.[8 9 10]

AI safety and AI governance represent overlapping but distinct domains. AI safety refers to the technical, institutional, and regulatory efforts aimed at ensuring that artificial intelligence systems function reliably, securely, and in ways that prevent harm to individuals, communities, democratic institutions, and societies at large[11 12]. AI governance encompasses the broader frameworks, norms, institutions, and processes through which societies collectively shape the development, deployment, and impact of AI technologies[13 14]. Standards play a critical mediating role, translating abstract principles into concrete technical specifications and operational practices that can be implemented, audited, and enforced[15]. This section will examine the existing governance processes related to AI, clarify AI safety concepts and initiatives and also delve into the role of standard setting bodies in AI governance, while also highlighting critical challenges in governing AI systems.

Multilateral Frameworks

Figure 1: Major AI governance process at a glance

Year	Governance Process	Mandate
2019	OECD AI Principles	Established five principles for trustworthy AI: inclusive growth, human rights, transparency, robustness, and accountability. Endorsed by the G20 to guide national AI governance[16 17].
2020	Global Partnership on AI (GPAI)	Multi-stakeholder initiative to bridge AI research and policy, advance responsible AI, and support applied projects aligned with public policy priorities. Integrated with OECD AI work in 2024[18 19].
2021	UNESCO Recommendation on the Ethics of AI	Adopted by 193 member states. Provides a global normative framework centered on human rights, dignity, transparency, fairness, and oversight, with policy action guidance across sectors[20].

2023	G7 Hiroshima AI Process	Produced International Guiding Principles and Code of Conduct for advanced AI systems, emphasizing transparency, risk reporting, lifecycle governance, and international coordination[21].
2023	BRICS AI Study Group	Established to monitor AI developments and promote cooperation and innovation across BRICS member states[22].
2024	Council of Europe Framework Convention on AI	First legally binding international treaty on AI. Mandates protection of human rights, democracy, and rule of law in AI deployment[23].
2024	UN High Level Advisory Body on AI Report	Proposed an integrated global AI governance architecture addressing political, economic, social, technical, and environmental dimensions[24].
2025	UN General Assembly Resolution A/RES/79/325	Established an Independent International Scientific Panel on AI and a Global Dialogue on AI Governance to provide evidence-based guidance and support policy interoperability[25].
2025	Hamburg Declaration on Responsible AI for the SDGs	Multistakeholder declaration focused on rights based, inclusive, and sustainable AI in development cooperation, with emphasis on supporting developing countries[26].

Source: multiple, cited in the table.

States and international institutions have been long involved in developing ethical guidelines, principles, and frameworks for the governance of AI systems. Research from Global Partners Digital, as of October 2023, identified more than 50 ongoing international initiatives focused on AI governance, driven by a wide range of forums, institutions, and actors. The majority of these are embedded within existing international cooperation frameworks, with nearly one quarter situated in the UN system, including the UN Secretary-General's High-level Advisory Body on AI, the Global Digital Compact process, and AI-related work under UNESCO, the Human Rights Council, the ITU, and the Internet Governance Forum (IGF). Of these efforts, only the High-level Advisory Body on AI constitutes a new UN entity[27].

Key initiatives include:

The OECD's 2019 adoption of AI Principles established five core tenets including inclusive growth, human rights respect, transparency, robustness, and accountability, which was subsequently endorsed by G20 leaders[28 29].

The Global Partnership on AI, established in 2020 under Canadian and French leadership, brought together 15 member economies in an international, multi-stakeholder initiative that convened leading experts from science, industry, civil society, international organizations, and governments, united by a shared commitment to the responsible advancement of AI[30]. Its core objective is to close the gap between AI research and real-world deployment by supporting frontier research and applied projects aligned with key AI policy priorities[31 32]. In July 2024, the GPAI initiative was integrated with the OECD's AI-related work, bringing both efforts together under the GPAI name to form a unified partnership. Under this renewed structure, all GPAI member countries engage on an equal basis, while the framework also enables additional countries to join in the future, contingent on consensus approval by the full membership[33].

In 2021, 193 member states adopted UNESCO's Recommendation on the Ethics of Artificial Intelligence[34], and it is applicable to all 194 member states of UNESCO[35]. The Recommendation centers protecting human rights and human dignity, grounded in principles such as transparency, fairness, and meaningful human oversight of AI systems. Further its Policy Action Areas provide governments with frameworks to operationalize these principles across domains including data governance, environmental sustainability, gender equality, education and research, and health and social wellbeing[36].

The G7's Hiroshima AI Process was initiated in May 2023 under Japan's presidency for enhanced cooperation in AI governance[37]. It produced the Hiroshima AI Process Comprehensive Policy Framework, the first global effort of its kind, which includes *International Guiding Principles for all AI actors and an International Code of Conduct* for organizations developing advanced AI systems. These elements set out expectations for transparency, risk reporting, digital literacy, and responsible behaviour across the AI life cycle. The initiative aims to broaden support among partner governments and coordinate with multilateral forums such as the OECD and UN while also enabling Japan to strengthen cooperation and research on AI safety[38]. Specifically, the International Code of Conduct sets out a reporting framework for organizations[39] encouraging them to release public reports describing their AI governance and risk management practices, in alignment with the Code of Conduct[40].

In 2023, during the BRICS summit in Johannesburg, BRICS countries agreed to form an AI study group. The initiative's central aim is to closely monitor and evaluate developments in AI technologies. A key objective of the BRICS AI study group is also to promote and support innovation in artificial intelligence across member countries of the BRICS bloc[41]. In 2025, during the 17th BRICS Summit, countries signed the Leaders' Declaration on Global Governance of Artificial Intelligence, which put together a set of guidelines aimed at promoting responsible development, deployment, and use of AI technologies to support sustainable development and inclusive growth. Key themes in the declaration included openness, public-interest AI, and sovereignty[42].

The Council of Europe's Framework Convention on AI and Human Rights, Democracy and the Rule of Law, signed by 46 member states and 36 others in 2024, became the first global *"legally binding instrument designed*

to ensure that AI upholds common standards in human rights, democracy and the rule of law, and to minimise the risk of those rights and principles being undermined as a result of the use of AI.[43]"

The Hamburg Declaration on Responsible AI for the Sustainable Development Goals, endorsed in June 2025 at the Hamburg Sustainability Conference, represents the first global multi-stakeholder declaration with explicit focus on AI in international development cooperation[44]. The Declaration commits to rights-based, inclusive, and sustainable AI, with a focus on supporting developing countries through education, local innovation, resource efficiency, and protections for marginalized communities[45].

In September 2024, the United Nations High-level Advisory Body on Artificial Intelligence published its final report, *Governing AI for Humanity* which proposes a global approach to AI governance, arguing that the technology's risks and benefits span political, economic, social, ethical, human rights, technical, and environmental dimensions and therefore require an integrated international response[46].

The UN General Assembly adopted Resolution A/RES/79/325 in 2025, and established two new governance mechanisms: an Independent International Scientific Panel on AI and a Global Dialogue on AI Governance to convene relevant experts across states, civil society, industry, and scientific experts to share best practices and foster policy interoperability[47]. These forums are designed as principal venues for collective global AI governance, offering evidence-based guidance and annual reporting to inform national policies and international cooperation. Additionally, AI governance was embedded within the Global Digital Compact, which commits all member states to a multistakeholder, human-rights-anchored approach to AI governance and capacity building, particularly for developing countries[48]. As recent as on February 5, 2026, the UN Secretary-General submitted a list of 40 experts from every region to the UN General Assembly (A/80/619) to form the new Independent International Scientific Panel on Artificial Intelligence, as mandated by the Pact for the Future and the Global Digital Compact[49].

Regional AI Governance Efforts

Apart from a few key AI governance processes, there are certain regional processes that have been articulated in the past years. The table below documents a few major regional AI governance initiatives, declarations, and frameworks adopted across different regions.

Figure 2: A summary of select regional AI governance efforts

Region	Process	Scope and Mandate
Latin America and Caribbean	Santiago Declaration to Promote Ethical AI (2023)	First regional ministerial declaration on AI ethics in Latin America and the Caribbean. Establishes fundamental principles including proportionality, security, fairness, non-discrimination, gender equality, accessibility, sustainability, privacy and data protection. Based on UNESCO's Recommendation on the Ethics of AI. Led to formation of an Intergovernmental Council on AI for the region[50].

Latin America and Caribbean	Montevideo Declaration on AI and Regional Roadmap (2024)	Adopted at the 2nd UNESCO Ministerial Summit on AI Ethics. Confirms regional commitment to AI that protects human rights, fundamental freedoms, and democracy. Includes a 2024-2025 roadmap with five priorities: governance and regulation, talent development, protection of vulnerable groups, environment and sustainability, and capacity building. Annual summits established to review implementation[51].
Latin America and Caribbean	Cartagena de Indias Declaration for AI Governance, Ecosystems and Education (2024)	Focuses on strengthening regional coordination and cooperation in the digital field, particularly for the UN Global Digital Compact. Addresses governance, construction of AI ecosystems, and promotion of AI education. Emphasizes technological development with respect for human rights, dignity, sustainability, and gender equality[52].
African Union	Continental Artificial Intelligence Strategy (2024)	The first AI-focused agreement within the continent. It focuses on applying AI to improve livelihoods and public services, managing risks through rights-based and culturally grounded governance, strengthening infrastructure and skills, promoting regional and global cooperation, and mobilizing public and private investment to build sustainable AI capacity across the continent[53].
ASEAN	ASEAN Guide on AI Governance and Ethics (2024)	A practical guidance tool for organizations designing, developing, and deploying traditional AI technologies in commercial and non-military applications. Seven guiding principles: transparency & explainability, fairness & equity, security & safety, human-centricity, privacy, accountability, and reliability. Focuses on encouraging alignment within ASEAN and fostering interoperability of AI frameworks[54].

Source: multiple, cited in the table.

Global AI Summits

The global AI summit series, inaugurated at Bletchley Park in November 2023, catalyzed international dialogue on AI safety and also has been intertwined with the establishment of national, regional and global AI Safety Institutes[55]. The table below provides an overview of major AI summits held globally from 2023 to 2026.

Figure 3: Overview of major AI Summits

Country	Focus	Key Outcomes
United Kingdom (Bletchley Park) **AI Safety Summit 2023**	AI Safety – First global summit on AI safety focused on "frontier AI" and existential risks	The Bletchley Declaration signed by 29 countries and the European Union, articulated shared concern about frontier AI risks and the need for collaborative action on frontier AI safety[56].
South Korea (Seoul) **AI Seoul Summit 2024**	AI Safety, Innovation, and Inclusivity – Expanded beyond pure safety to include economic and social benefits[57]	The summit resulted in 3 outcome documents: • Seoul Declaration for safe, innovative and inclusive AI: AI Seoul Summit 2024 • Seoul Statement of Intent toward International Cooperation on AI Safety Science, AI Seoul Summit 2024 (Annex) • Seoul Ministerial Statement for advancing AI safety, innovation and inclusivity: AI Seoul Summit 2024[58].
France (Paris) **AI Action Summit 2025**	Aimed at "strengthening international action towards artificial intelligence serving the general interest.[59]"	• Publication of the International AI Safety Report that brings together expert perspectives on AI capabilities and risks • The AI Action Summit Declaration - Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet[60]
India (New Delhi) **AI Impact Summit 2026**	Focused on People, Planet, and Progress, it "envisions a future where AI advances humanity, fosters inclusive growth, and safeguards our shared planet.[61]"	• AI Impact Summit New Delhi Declaration signed by 92 countries and international organisations, endorsing "a shared global vision for collaborative, trusted, resilient and efficient AI.[62]"

Source: multiple, cited in the table.

AI Safety Institutes

AI Safety Institutes (AISIs) usually combine state backing with technical expertise to evaluate AI systems, conduct safety research, and facilitate knowledge sharing among governments, industry, and academia[63]. The United Kingdom established the world's first publicly funded AI Safety Institute in November 2023, shortly after the Bletchley AI Safety Summit, with a mandate centered on evaluating frontier AI models and reducing systemic risks[64]. However, in February 2025, it was rebranded as the AI Security Institute, signaling a shift toward narrower security priorities and away from broader societal impacts[65].

The United States established the Center for AI Standards and Innovation (CAISI) within NIST in November 2023, focusing on developing technical standards, voluntary safety benchmarks, and measurement science for AI systems[66]. Singapore's AI Safety Institute, located within the Digital Trust Centre, has prioritized multilingual safety testing, content assurance, and public sector deployment safeguards, reflecting Southeast Asia's linguistic and cultural diversity[67]. India announced the IndiaAI Safety Institute on January 30, 2025, explicitly designed to promote research grounded in Indian social, economic, cultural, and linguistic diversity using Indian datasets[68]. In 2025, the Australian Government noted that they will establish an AISI to strengthen AI safety across the nation. The role of the AISI will be to provide expertise towards monitoring, evaluating, and disseminating information on emerging AI technologies, associated risks, and potential harms[69]. Through the AI6 framework, the Australian AISI sets up 6 key tenets for implementing responsible AI governance in Australia[70].

The International Network of AI Safety Institutes, launched at the AI Seoul Summit in May 2024, brings together Australia, Canada, the European Commission, France, Japan, Kenya, Singapore, South Korea, the United Kingdom, and the United States. Its mandate is to advance AI safety science, coordinate technical evaluations, share knowledge, and support evidence-based policymaking[71]. In December 2025, it was announced that the International Network of AI Safety Institutes would now be called the Network for Advanced AI Measurement, Evaluation and Science, re-emphasising the shared focus on the science of AI measurement and evaluations, with the UK taking on the role of Network Coordinator[72].

Kenya's inclusion represents significant symbolic progress as one of the first Global South countries to participate in such a technical governance network, though capacity asymmetries persist and concerns remain that shared benchmarks and testing protocols may not adequately reflect linguistic, cultural, and socio-economic diversity[73]. The Centre for Communication Governance's report provides a detailed mapping of existing AISI's including the objectives of the AISI and their governance structure. An excerpt of that mapping and some key AISIs are shared below[74]. Global South civil society organizations have also undertaken work to put forward recommendations on how country AI Safety Institutes could be structured[75].

Figure 4: Overview of AI safety Institutes

Country / Name of Institute	Year Established	Mandate
United Kingdom AI Security Institute (formerly AI Safety Institute)	November 2023	Conducting research and building infrastructure to understand the capabilities and impacts of advanced AI and to develop and test risk mitigations. Also working with diverse stakeholders to affect how AI is developed and to shape global policymaking on this issue[76].
United States Center for AI Standards and Innovation (formerly US AI Safety Institute)	November 2023	Advancing a risk-based approach to maximize AI benefits and minimize its potential negative consequences. Focused on fundamental research to improve AI measurement science, standards, and related tools, including benchmarks and evaluations[77].
Japan AI Safety Institute	Feb 2024	Examine and promote evaluation methods and standards for AI safety, with the aim of realizing safe, secure, and trustworthy AI[78].
European Union EU AI Office	May 2024	Established within the European Commission, unique among AISIs in possessing regulatory enforcement powers under EU AI Act. Balances AI safety evaluation with broader mandate including driving innovation and AI adoption[79].
Singapore AI Safety Institute[80] (Digital Trust Centre)	May 2024	Focused on multilingual safety testing, content assurance, public sector deployment safeguards, red-teaming exercises. Its mandate has been given to a national centre housed within the Nanyang Technological University (NTU) - the Digital Trust Centre (re- purposed and designated as Singapore's AI Safety Institute) in partnership with Infocomm Media Development Authority of Singapore[81].
Canadian Artificial Intelligence Safety Institute	November 2024	Advances AI safety science through understanding risks of advanced AI systems including synthetic content, disinformation, cybersecurity. Mandated to ensure the safe40 and responsible development and deployment of artificial intelligence[82].

South Korea AI Safety Institute	November 2024	The AI Safety Institute serves as a dedicated organization designed to systematically and professionally address diverse AI risks, including those caused by technological limitations, human misuse, and loss of control over AI systems[83].
India AI Safety Institute	January 2025	Ensures ethical and safe application of AI models. Promotes domestic R&D grounded in India's social, economic, cultural, and linguistic diversity based on Indian datasets. Part of Safe and Trusted Pillar of IndiaAI Mission. Emphasizes context-appropriate AI safety[84].
France - National Institute for the Evaluation and Security of Artificial Intelligence	January 2025	National institute dedicated to assessing the security of artificial intelligence, examining its impacts from a scientific perspective as well as from a security standpoint[85 86].
Australia AI Safety Institute	November 2025	The role of the AISI will be to provide expertise towards monitoring, evaluating, and disseminating information on emerging AI technologies, associated risks, and potential harms[87].

Source: Multiple (cited in the table) including borrowing from a resource developed by CCG-NLU Delhi.
https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/exploring-aisis-for-the-global-south-805.pdf

While recognized as important institutions for bridging the technical with the policy dimensions of AI and ensuring AI governance keeps pace with the evolving nature of AI, open questions remain around how identified AI risks are mitigated and guardrails may be operationalised in practice. Further, without hard regulatory responses, AI safety may face similar challenges as other AI governance mechanisms which rely on voluntary disclosure, principle based approaches, without creating tangible requirements or mandates for those developing or deploying AI systems[88 89].

AI Standard Setting

Standards represent technical specifications and operational guidelines that translate abstract principles into concrete, auditable practices governing the development, deployment, and oversight of AI systems[90] [91]. As AI becomes increasingly embedded in critical infrastructure, public services, and everyday applications, standards have emerged as essential instruments for ensuring safety, interoperability, trustworthiness, and regulatory compliance across jurisdictions. For example, the Global Digital Compact, adopted under the UN Pact for the Future, and mentioned above, sets out a framework to strengthen international AI governance in the public interest. It encourages standards development organizations to work together to advance interoperable AI standards that support safety, reliability, sustainability, and the protection of human rights[92].

International Telecommunication Union's World Telecommunication Standardization Assembly (WTSA-24), held in New Delhi in October 2024, featured the first session of the new International AI Standards Summit, convened jointly with ISO and IEC, during a special session focused on the importance of international standards for promoting the safe, responsible, and inclusive use of AI. The subsequent summit took place in Seoul, Republic of Korea, on 2–3 December 2025[93]. AI standards were also positioned as a central focus of the ITU led AI for Good Global Summit held from 8 to 11 July in Geneva, Switzerland in 2025, which concluded with a dedicated AI standards day. Further, in early 2025, the Internet Engineering Task Force (IETF) set up the AI Preferences (AIPREF) Working Group, which is mandated to work on "standardizing building blocks that allow for the expression of preferences about how content is collected and processed for Artificial Intelligence (AI) model development, deployment, and use.[94]"

The December 2025 International AI Standards Summit in Seoul marked a significant moment of convergence as ISO, IEC, and ITU jointly issued the Seoul Statement, committing to actively incorporate socio-technical dimensions in standards development, deepen understanding of the interplay between international standards and human rights, strengthen an inclusive and dynamic multistakeholder community, and enhance public-private collaboration on AI capacity building[95]. This represented recognition that AI standards must address not only technical interoperability but also broader societal implications.

Regional standards organizations play critical mediating roles between global frameworks and national contexts and are increasingly becoming involved in AI standards, though the extent and scope of this engagement varies between bodies and across regions. As examples, the European Telecommunications Standards Institute (ETSI) develops technical standards supporting European Union regulatory frameworks including the AI Act, with the European Commission authorized to request development of harmonized standards by recognized European standards organizations that can override member-state level standards[96]. These may adopt or adapt ISO/IEC standards with modifications reflecting regional priorities. The Korean Agency for Technology and Standards (KATS) hosted the 2025 International AI Standards Summit, reflecting increasing engagement from Asian regional bodies[97]. The Bureau of Indian Standards (BIS) has developed a committee focused on AI[98], and is recognized in the 2026 AI Governance Guidelines as a core institution responsible for developing common standards and engaging in global standards setting bodies[99].

Key Dynamics in AI Governance

Multiple structural aspects influence global AI governance. Geopolitical competition[100] including national security concerns, frontier dominance, sovereignty concerns, export controls[101]; regulatory fragmentation[102 103] and extraterritorial application of some laws; cost of governance, overlapping institutions[104], and lack of adequate governance capacities[105] are all factors influencing the governance of AI. Recognizing the above, in this subsection, we specifically discuss multistakeholderism and Global South participation in Global AI governance.

The Role of Multistakeholderism

Multistakeholder governance models bring together governments, private companies, researchers, and civil society to share responsibility for shaping how AI is developed and deployed. The Bletchley Declaration explicitly affirmed that all actors have a role to play in ensuring the safety of AI, stating that nations, international fora and other initiatives, companies, civil society, and academia will need to work together[106]. However, experts have noted that meaningful multistakeholder participation requires centering AI governance in multistakeholderism, addressing power asymmetries that systematically exclude marginalized voices through deliberate processes, and ensuring accountability in multistakeholder AI governance processes[107]. Similarly, principles such as the São Paulo Multistakeholder Guidelines provide important guidance on enabling multistakeholder governance processes[108]. Civil society organizations have emphasized that AI policy processes at the national, regional, and international level are often opaque and dominated by state actors, and lack clear rules for civil society participation thereby allowing for limited meaningful multistakeholder participation[109 110 111 112]. They suggest that multistakeholder governance must move beyond consultation toward co-design and co-governance, with civil society, affected communities, and marginalized groups possessing genuine decision-making authority[113].

Even in standard setting bodies for instance, while ISO/IEC JTC 1/SC 42 reports that half of delegations attending plenary meetings are led by women, representing progress on gender diversity, the geographic and institutional diversity of participants remains skewed toward high-income countries and large technology corporations.[114]

Global South Participation

Global South countries face structural barriers to participating meaningfully in safety science, standard setting, and institutional governance. High-income countries dominate international AI governance, hosting major forums and producing most global AI policies and principles, with activity heavily concentrated in Europe, the United States, and China. Research shows that nearly two-thirds of over 470 AI governance documents originated in these regions, while contributions from Latin America, the Caribbean, and Africa remain minimal[115]. Without deliberate corrective measures, AI governance risks entrenching existing inequalities within the global governance architecture.

SECTION 3

Safe and Trusted AI

The AI Impact Summit being held in India in February 2026 is organised around three foundational pillars ('Sutras'): people, planet, and progress. Building upon these pillars the discussions at Summit are slated to be organized around seven areas of multilateral cooperation ('Chakras'), one of which is Safe and Trusted AI, towards which a working group has been set up, with Japan and Brazil as country Co-Chairs alongside India[116]. This working group acknowledges the risks and harms that arise out of wide scale deployment of AI, "from algorithmic bias and misinformation to autonomous decision errors and unanticipated model behavior," and how they impact public trust. It aims to address these through: promoting adoption of interoperable AI safety tools, promoting international cooperation to operationalise AI safety and trust and finally to build inclusive capacity to strengthen AI safety research and governance[117]. We will explore some of these issues in greater detail with a view to understand the harms and risks already seen with the use of AI systems, technical responses to building safe AI, and what it really means to build trusted AI that works for people and societies.

Trust in AI and Trustworthy AI

It is important to clarify the conceptual distinction between trust in AI systems vs trustworthy AI. Trust in AI predominantly refers to users' perceptions, beliefs and willingness to rely on AI systems, contingent on experience, context and so on[118]. As early as 2004, in writing about trust in automated systems, the authors examine trust through multiple lenses: psychological, sociological, organizational, and neurological. They argue that trust strongly influences whether users over-rely on or under-use automation[119]. OpenAI's report on how people use AI showed that 73% of users were using ChatGPT for non-work related activities such as practical guidance, seeking information, self expression and so on[120]. However, other reports show that even though 66% of people surveyed in a study are already intentionally using AI with some regularity, less than half of global respondents are willing to trust it (46%)[121]. The high usage coupled with hesitance in trusting AI systems is creating a unique paradox.

Academic work suggests that trustworthy AI, in contrast, is a comprehensive framework that is built on three pillars, legal compliance, ethical soundness, and technical and social robustness; supported by seven core requirements: human agency and oversight, robustness and safety, privacy and data governance, transparency, diversity and fairness, societal and environmental wellbeing, and accountability[122]. Other works take a lifecycle approach to discuss how AI trustworthiness can be enhanced at each stage; "from data to AI models and from system deployment to its operation.[123]"

AI Safety

There are several frames for understanding AI safety that can at times be conflicting. These include: existential risk, near-term sociotechnical harms, cybersecurity and misuse, safety vs innovation[124]. One influential framing situates AI safety around low-probability but high-impact scenarios associated with highly capable or general-purpose AI systems, emphasizing existential or catastrophic risks arising from loss of human control, misalignment of advanced systems, or unintended systemic effects at scale[125] [126]. An extension of this framing has been to ask for AI moratoriums[127] and for temporary pause on giant AI experiments[128], led by technical industry actors, research communities and philanthropic actors.

In contrast, another framing, championed by many civil society organizations, Global South governments, and digital rights advocates, defines AI safety in relation to immediate and observable harms including algorithmic discrimination, gender-based harms, labor precarity, misinformation, privacy violations, and the exclusion of marginalized communities from access to remedies[129] [130]. Rather than focusing on speculative future scenarios, this approach prioritizes governance mechanisms addressing real-world deployments of AI systems across different sectors, public services and digital platforms.

Risks and Harms of General-Purpose AI Systems

The 2026 AI Safety Report highlights various risks that the quick advancements in general purpose AI technology is creating. Malicious use for conducting frauds and scams, influencing and manipulating people's beliefs as well as increased vulnerability to AI-enabled cyberattacks are suggested as dominated risks. Further, systemic risks such as large scale labor market impacts and loss of human autonomy are some longer term impacts that the advancements in AI without sufficient guardrails may pose[131].

Further, over the last many years, researchers, civil society organisations and in some instances, States, have documented and presented evidence of the extent of harm that AI systems can produce, when not built ethically, transparently and with the right guardrails in place. From non-consensual deepfakes[132] [133] [134] [135] to caste[136] [137], gender[138], racial biases[139] in LLMs, from exploitative labor practices[140] [141] to environmental impacts of AI[142] [143], general purpose AI systems can exacerbate existing societal biases and harmful impacts, especially for those at different axes of vulnerabilities. These are a few illustrative sets of concerns that AI systems have brought to fore, and merit deliberate interventions to develop AI that is safe, trustworthy, and accountable.

Socio-Technical Approaches to Building Safe and Trusted AI

AI systems rely on safeguards at multiple stages, including development measures to improve robustness, deployment controls such as filtering and human oversight, and post-deployment tools like provenance tracking and content detection. However, these mechanisms are imperfect, as users can circumvent controls and tools like watermarking can be removed, making single safeguards insufficient and necessitating layered defence-in-depth approaches[144]. The International AI Safety Report (2026) documents various technical safeguards.

These range from model development safeguards such as data curation, alignment training, adversarial testing, unlearning, and interpretability tools, to deployment controls like content filtering, human oversight, sandboxing and monitoring mechanisms, as well as ecosystem-level measures including watermarking, model identification, and AI-generated content detection[145]. Other tools such as regulatory sandboxes[146 147], algorithmic impact assessments and AI audits provide structured environments for testing AI systems before deployment, though adoption remains limited particularly in resource-constrained jurisdictions[148]. The Safety Report also noted that most evaluation frameworks remain voluntary without enforceable verification mechanisms, creating challenges for policymakers seeking visibility into how risks are identified and managed in practice[149].

Several research studies advocate for socio-technical evaluations as an approach for building trusted and accountable AI systems. These evaluations are meant to extend beyond technical capability evaluations to include human interaction and systemic impacts, recognizing that context determines whether and how capabilities cause harm. The authors argue that sociotechnical approaches combining technical assessments with human factors and societal context provide tractable pathways toward robust and comprehensive AI safety evaluation[150 151 152].

Ultimately, to build long term trust, public accountability and engagement is essential. Public accountability requires moving beyond technical audits toward comprehensive governance frameworks with enforceable obligations. Some mandatory pre-deployment impact assessments, as proposed in various jurisdictions, include language around genuine engagement with affected communities rather than perfunctory consultation[153]. Public registries documenting AI system deployments in high-stakes contexts, including performance data disaggregated by demographic groups, may also enable informed scrutiny and challenge[154].

SECTION 4

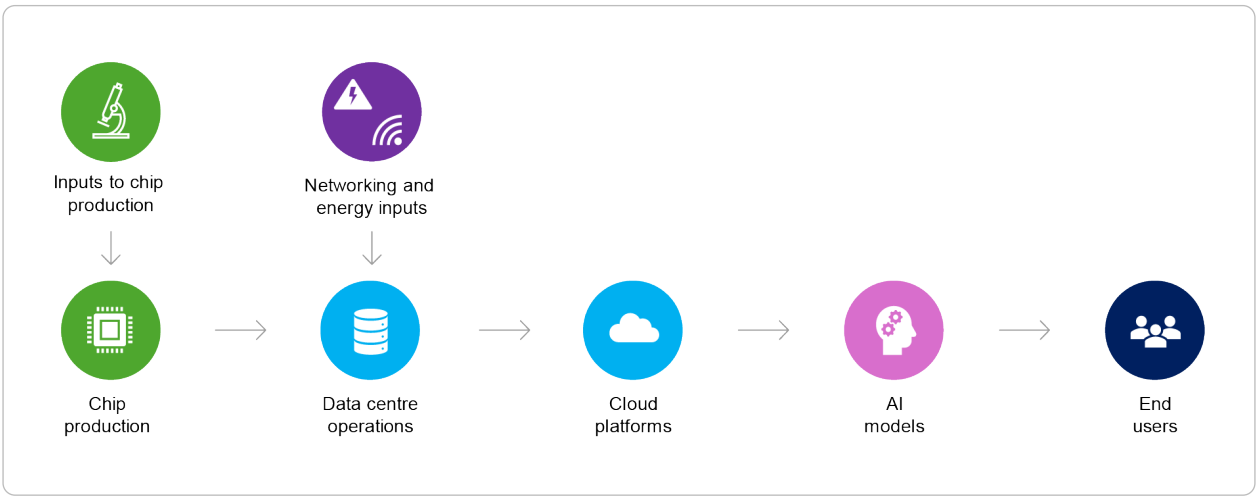
Context-driven AI Infrastructure

AI infrastructure helps determine who can build AI systems, whose languages and contexts are represented, and which communities benefit from AI applications. This section explores the foundations of AI infrastructure, its effects on and the effects of market factors, and its implications for Digital Public Infrastructure (DPI). Finally, it will briefly cover multilingual AI and community AI initiatives.

The Foundations on which AI is Built

While AI infrastructure is highly complex and operates across interconnected layers, it broadly consists of physical and digital layers. The physical layer includes data centers for storage and processing, specialized processing units like GPUs and TPUs for model training, and high-performance computing clusters that determine computational capacity, as well as resources that AI infrastructure is built on including natural resources, as well as public and private infrastructure[155] [156]. OECD has elaborated that the physical AI supply chain layer is multi-layered, beginning with chip designers and manufacturers who produce hardware for AI data centers, supported by specialized suppliers of materials and equipment. These chips are integrated into server racks and deployed by data center operators, who rely on energy and networking infrastructure to power and interconnect servers. Cloud providers then sell access to this computing power, enabling AI developers to train and deploy AI systems[157].

Figure 4: Overview of AI safety Institutes



Source: OECD Analysis https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/11/competition-in-artificial-intelligence-infrastructure_69319aee/623d1874-en.pdf adapted from Pilz (2023) An Assessment of Data Center Infrastructure's Role in AI Governance, https://www.konstantinpilz.com/datacenters/assessment

The digital layer typically constitutes the data layer, the model development and operations layer, and the application layer. The data layer encompasses training datasets, evaluation benchmarks, and processes to understand data provenance[158] [159]. The model layer includes foundation models, specialized algorithms, and the platforms hosting these systems. The AI model layer determines how effectively applications can understand, predict, and respond to real-world needs[160]. The application layer comprises services and tools enabling end-user interaction with AI capabilities[161] [162]. Having access to, and being in the position to develop various capacities across these different layers have an outsized role to play in determining how AI systems are shaped, who benefits from it societally and economically, and the exclusions it may create.

The physical and digital layers of AI infrastructure rest on and require a significant amount of natural resources. The manufacturing of specialised processors, chips and hardware that AI systems run on depend on several critical minerals; for instance silicon for microchips, cobalt for memory and logic components, and rare earth elements for high efficiency magnets and semiconductor technologies, to name a few[163] [164]. There is also a huge demand for electricity primarily to continuously power the datacenters. In 2024, electricity consumption from data centres was estimated to amount to around 415 terawatt hours (TWh), or about 1.5% of global electricity consumption[165], which is projected to rise to around 945 TWh by 2030[166]. Some research studies have found that multi-purpose, generative architectures are orders of magnitude more energy-intensive than task-specific systems[167]. Data centres also consume large quantities of water, used primarily to cool the chips and servers that would otherwise overheat under sustained computational load[168] [169] [170].

Market Considerations, Geopolitics and Sovereignty

Certain large companies have become vertically dominant across several of these layers, which affords them certain network efficiencies, economic benefits, and political influence. Their access to and control over data, compute infrastructure, talent, and market channels can reinforce existing market dominance, reducing opportunities for broad competition or public benefit[171] [172]. Competition concerns such as vertical integration and reduced contestability are evident in this space and are leading to some concerted efforts aimed at preserving competitive markets and mitigating the broader societal implications of concentrated AI power[173]. Pointing to geographic asymmetries, reports suggest that Africa accounts for less than 1 percent of global data center capacity despite being home to 18 percent of the world's population. India would need to nearly double capacity by 2026 just to meet domestic demand. Most of South Asia, Southeast Asia, and Latin America depend on external infrastructure[174].

Access to high-performance computing infrastructure is shaped by several interrelated factors, including the supply of advanced semiconductors and AI-optimised chips, cutting-edge GPUs that are increasingly critical for generative AI workloads, and specialised servers designed to handle AI-specific processing demands within data centres[175]. Training state-of-the-art foundation models requires thousands of specialized GPUs and costs, leading to uneven progress across research groups and raising concerns about equity, environmental cost, and concentration of AI capabilities[176] [177].

Studies show how compute divides drive modern AI research, with large industrial and well-resourced labs dominating compute-intensive machine learning research topics, reducing the representation of academic-only

teams in these areas[178] [179]. Similarly UNCTAD noted that in 2022, just 100 companies, primarily in the US and China, made up for 40% of global AI research and development (R&D). Combined, the two countries held 60% of all AI patents and produced a third of global AI publications[180].

This concentration, together with increased geopolitical volatility[181] [182] [183] is also creating momentum for various initiatives and investments, including across and within the Global South, that aim to expand AI capacities at various layers of the AI stack and increase digital and AI sovereignty. In 2024, the Brazilian government launched the Brazilian Artificial Intelligence Plan (PBIA) 2024–2028, with an estimated investment of R$ 23 billion dedicated to AI research, development, and infrastructure[184]. India has committed over ₹10,300 crore investment over five years for IndiaAI Mission with 38,000 GPUs deployed[185]. The African Union's Continental AI Strategy emphasizes building local capacity while acknowledging infrastructure gaps requiring international cooperation and technology transfer[186]. How these efforts and investments materialise, whether they promote competition and reduce dependencies on specific providers, and how they ultimately further public value and diversified economic growth, remains to be seen.

Digital Public Infrastructure and AI Integration

State-led efforts to provide access to and facilitate the use of AI for public benefit can create efficiencies, improve service delivery, and foster innovation. The DPI agenda significantly advanced during India's G20 presidency in 2023, with the G20 New Delhi Declaration framing DPI as "an evolving concept" that refers to *"a set of shared digital systems, built and leveraged by both the public and private sectors, based on secure and resilient infrastructure, [which] can be built on open standards and specifications, as well as open-source software [that] can enable the delivery of services at societal scale.[187]"* In principle, DPI seeks to operate as open, interoperable platforms facilitating identity verification, payments, data exchange, and service delivery without dependence on proprietary systems controlled by private corporations[188] [189]. In February 2026, India signed an MoU with 23 countries for cooperation on digital public infrastructure[190].

Some academic research theorises a two-way relationship between artificial intelligence and digital public infrastructure, AI for DPI and DPI for AI. AI for DPI describes the use of artificial intelligence to support, enhance, or expand the operation of digital public infrastructure. The use and promotion of chatbots in public service delivery is a key example of AI for DPI. DPI for AI comprises ways in which DPI is leveraged to advance a country's AI-related interests. Examples range from the creation, collection, and collation of large datasets for AI training to the Open Cloud Compute (OCC) system proposed by India's People+ai[191]. In February 2026, Latam-GPT, was presented as the first open large language model (LLM) developed from and for Latin America and the Caribbean. As the first regional GPT model in the world, it is also being positioned as a technological public good, a foundation layer that will enable others to build on top[192]. The Centre for Digital Public Infrastructure's DPI-AI Framework presents a model for integrating artificial intelligence into public digital systems by building on existing digital public infrastructure rather than treating AI as a separate layer, positioning AI capabilities as interoperable functions that connect to foundational DPI components such as digital identity, data exchange, and payments. The framework is intended to help policymakers and practitioners align AI with DPI principles and architectures to create adaptive, coherent, and publicly governed digital infrastructure[193]. How these frameworks are implemented and their impacts remains to be seen.

Governments are increasingly integrating AI capabilities into digital public infrastructure and/or public service delivery. India's BHASHINI platform deploys AI-powered translation services across 23 Indian languages, enabling citizens to access government services in their mother tongues[194]. Singapore's Singpass digital identity system uses machine learning to detect fraudulent authentication attempts and streamline service delivery[195]. Denmark's Muni AI chatbot assists citizens across 37 municipalities with queries about public services[196]. These integrations have the potential to create and improve public service delivery, when in compliance with existing and appropriate safeguards[197] [198] [199]. To this end, several organizations have developed principles and practices for developing effective and rights respecting DPI[200].

Multilingual AI and Community Approaches

While over 7,000 languages are globally spoken, a very small percentage receive meaningful representation in AI training data and digital resources[201]. Multilingual machine translation tools have the potential to reduce the digital language divide by extending support to many languages, as seen with Google Translate supporting 249 languages and initiatives to expand to 1,000, ChatGPT reporting support for over 90 languages, but these efforts still exclude many under-represented languages[202]. For instance, of the top 34 languages used on the internet globally as of 2024, none were African[203]. Further systematic issues persist in democratising language models because training datasets often contain mislabeled or poor-quality data[204], notwithstanding efforts to address these gaps[205]. This exclusion perpetuates marginalization by disadvantaging those who don't speak dominant languages[206] [207].

Certain community-led/public initiatives have demonstrated promising approaches to multilingual AI development. Masakhane, founded at Deep Learning Indaba 2019, is a distributed open-source research initiative focused on advancing machine translation and NLP for African languages through continent-wide collaboration. It aims to strengthen African NLP research capacity, develop datasets and tools for underrepresented languages, and establish replicable models for community-driven research in emerging regions[208].

The AI4Bharat-IndicNLP Corpus initiative introduced a large-scale general-domain text corpus of 2.7 billion words covering ten Indian languages and associated resources, and demonstrated that these resources outperform existing public embeddings on multiple NLP evaluation tasks, thereby aiming to accelerate research for Indic language processing[209] [210].

South Africa's Esethu Framework establishes protocols for ethically collecting speech datasets from native speakers while redistributing revenue back to communities whose linguistic data enables commercial AI development[211]. Further, AI is also being used to support the revitalisation and everyday use of te reo Māori through speech recognition, transcription, and language learning tools trained on extensive community-curated audio archives. Initiatives led by organisations such as Te Hiku combine AI development with indigenous data sovereignty principles, in an attempt to ensure Māori communities retain control over the collection, governance, and application of their linguistic and cultural data[212].

SECTION 5

Concluding Remarks

This mapping document has examined the evolving landscape of AI governance, safety, and infrastructure, identifying notable variations in participation, resources, and governance capacity across regions. Evidence indicates that while various governance approaches are being institutionalised, governance capacity varies considerably, particularly affecting resource-constrained contexts, while geopolitical considerations increasingly influence technology development. As the second section of this report reveals, building safe, trusted and accountable AI may not merit a one-size-fits-all approach, but requires a more contextual understanding of real-world environments in which AI is deployed and its consequences. The section on infrastructure reveals concentrated computational resources in certain regions, raising questions about dependencies and access, while also pointing to alternative approaches being taken to democratise AI access and use. More inclusive governance approaches may benefit from broader stakeholder participation, including Global South leadership, affected communities, civil society organizations, and diverse institutional actors. This mapping provides a foundation on governance models and approaches to AI development that consider multiple contexts and perspectives with the hope of sparking conversations and further inquiry into some of these areas.

Endnotes

1. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf
2. https://www.sciencedirect.com/science/article/abs/pii/S1566253523004517?utm_source=blog&utm_medium=taming-saas-sprawl
3. https://arxiv.org/abs/2005.14165
4. https://link.springer.com/article/10.1007/s10506-024-09412-y
5. The terms foundation models, generative AI, general purpose AI and even large language models are sometimes used interchangeably. There have been attempts to define these terms, but they tend to have overlaps in articulation. See more here - https://cset.georgetown.edu/article/what-are-generative-ai-large-language-models-and-foundation-models/
6. https://doi.org/10.1038/s41591-018-0300-7
7. https://arxiv.org/abs/1906.05433
8. https://www.caidp.org/resources/g20/
9. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy
10. https://www.undp.org/news/hamburg-declaration-responsible-ai
11. https://arxiv.org/abs/1606.06565
12. https://www.nist.gov/itl/ai-risk-management-framework
13. https://academic.oup.com/ia/article/100/3/1275/7641064
14. https://www.fhi.ox.ac.uk
15. https://standards.ieee.org/ieee/7000/6781/
16. https://www.oecd.org/going-digital/ai/principles
17. https://www.g20.org
18. https://indiaai.gov.in/globalindiaaisummit/about
19. https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html
20. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence#:~:text=UNESCO's%20first%2Dever%20global%20standard%20on,194%20member%20states%20of%20UNESCO
21. https://www.oecd.org/en/publications/2023/09/g7-hiroshima-process-on-generative-artificial-intelligence-ai_8d19e746.html
22. https://indiaai.gov.in/news/brics-creates-an-ai-research-group
23. https://www.coe.int/en/web/artificial-intelligence
24. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
25. https://docs.un.org/en/A/RES/79/325
26. https://www.undp.org/news/hamburg-declaration-responsible-ai-global-leaders-commit-responsible-ai-sustainable-development
27. https://www.gp-digital.org/navigating-the-global-ai-governance-landscape/ , https://docs.google.com/spreadsheets/d/1gOZbs7dC9VsV2h0zJn60L7cUVwu9mccqz_B265tNyQY/edit?gid=0#gid=0
28. https://www.oecd.org/going-digital/ai/principles
29. https://www.g20.org
30. https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html
31. https://indiaai.gov.in/globalindiaaisummit/about

32. https://www.oecd.org/en/about/programmes/global-partnership-on-artificial-intelligence.html
33. https://oecd.ai/en/about/about-gpai
34. https://www.ohchr.org/sites/default/files/2022-03/UNESCO.pdf
35. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence#:~:text=UNESCO's%20first%2Dever%20global%20standard%20on,194%20member%20states%20of%20UNESCO.
36. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
37. https://www.oecd.org/en/publications/2023/09/g7-hiroshima-process-on-generative-artificial-intelligence-ai_8d19e746.html
38. https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html
39. https://transparency.oecd.ai/
40. https://www.brookings.edu/articles/haip-reporting-framework-ai-governance/
41. https://indiaai.gov.in/news/brics-creates-an-ai-research-group
42. https://brics.br/en/news/artificial-intelligence-must-not-be-a-privilege-for-the-few-nor-a-tool-of-manipulation-in-the-hands-of-millionaires-declared-lula-at-the-brics-summit
43. https://www.coe.int/en/web/artificial-intelligence
44. https://www.bmz-digital.global/en/hsc/
45. https://www.undp.org/news/hamburg-declaration-responsible-ai-global-leaders-commit-responsible-ai-sustainable-development
46. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf
47. https://docs.un.org/en/A/RES/79/325
48. https://www.un.org/pact-for-the-future/en/annex-i-global-digital-compact
49. https://docs.un.org/en/a/80/619
50. https://blog.investchile.gob.cl/chile-ethical-artificial-intelligence
51. https://www.telefonica.com/en/communication-room/blog/2nd-unesco-ministerial-summit-ethics-ai-latin-america-caribbean/
52. https://thecostaricanews.com/latin-american-countries-adopt-cartagena-declaration-on-ai/
53. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy
54. https://asean.org/book/asean-guide-on-ai-governance-and-ethics/
55. https://www.gov.uk/government/organisations/ai-security-institute
56. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration
57. https://www.techuk.org/resource/key-outcomes-of-the-ai-seoul-summit.html
58. https://www.industry.gov.au/publications/seoul-declaration-countries-attending-ai-seoul-summit-21-22-may-2024#seoul-declaration-1
59. https://www.elysee.fr/en/sommet-pour-l-action-sur-l-ia
60. https://www.techuk.org/resource/what-were-the-outcomes-of-the-paris-ai-action-summit.html
61. https://impact.indiaai.gov.in/
62. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2231208&v=4&reg=3&lang=2
63. https://alltechishuman.org/all-tech-is-human-blog/the-global-landscape-of-ai-safety-institutes
64. https://www.gov.uk/government/publications/ai-safety-institute
65. https://www.gov.uk/government/organisations/ai-security-institute
66. https://www.nist.gov/caisi
67. https://www.imda.gov.sg

68. https://indiaai.gov.in/article/india-takes-the-lead-establishing-the-indiaai-safety-institute-for-responsible-ai-innovation
69. https://www.aidataanalytics.network/responsible-ai/news-trends/australia-forms-new-institute-to-strengthen-ai-safety
70. https://safeaiaus.org/safety-standards/guidance-for-ai-adoption-ai6/
71. https://www.nist.gov/news-events/news/2024/11/fact-sheet-international-network-ai-safety-institutes
72. https://www.gov.uk/government/news/efforts-to-share-best-practices-on-ai-measurement-and-evaluations-driven-forward-through-the-international-network-for-advanced-ai-measurement-evalua
73. https://www.diplomacyandlaw.com/post/global-governance-of-artificial-intelligence
74. https://ccgdelhi.s3.ap-south-1.amazonaws.com/uploads/exploring-aisis-for-the-global-south-805.pdf
75. https://itsrio.org/en/publicacoes/an-ai-safety-and-opportunities-institute-for-brazil-a-design-of-features-functions-and-potential-challenges/
76. https://www.aisi.gov.uk/
77. https://www.nist.gov/artificial-intelligence
78. https://www.meti.go.jp/english/
79. https://digital-strategy.ec.europa.eu/en/policies/ai-office
80. https://sgaisi.sg/
81. https://www.imda.gov.sg/
82. https://ised-isde.canada.ca/site/ised/en/canadian-artificial-intelligence-safety-institute
83. https://www.msit.go.kr/eng/bbs/view.do?nttSeqNo=1058&bbsSeqNo=42&mId=4
84. https://indiaai.gov.in/article/india-takes-the-lead-establishing-the-indiaai-safety-institute-for-responsible-ai-innovation
85. https://www.campusfrance.org/en/actu/creation-d-un-institut-national-pour-l-evaluation-et-la-securite-de-l-ia
86. https://www.lne.fr/en/news/2025/lne-role-ai-evaluation-and-security-within-inesia
87. https://www.aidataanalytics.network/responsible-ai/news-trends/australia-forms-new-institute-to-strengthen-ai-safety
88. https://dig.watch/updates/are-ai-safety-institutes-shaping-the-future-of-trustworthy-ai
89. https://www.techpolicy.press/ai-safety-institutes-crop-up-around-the-world-helpful-or-hype/
90. https://aistandardshub.org/resource/main-training-page-example/1-what-are-standards/
91. https://www.responsible.ai/news/ai-standards-deep-dive-decoding-different-ai-standards-and-the-eus-approach/
92. https://www.itu.int/hub/2025/04/standards-help-unlock-trustworthy-ai-opportunities-for-all/
93. https://www.iec.ch/blog/iec-iso-and-itu-launch-2025-international-ai-standards-summit
94. https://www.ietf.org/blog/aipref-wg/
95. https://www.iso.org/news/2025/12/ai-standards-summit
96. https://www.brookings.edu/articles/small-yards-big-tents-how-to-build-cooperation-on-critical-international-standards/
97. https://www.aistandardssummit.org/event/354f4a77-ee25-47e3-8e84-291a55519c0c/summary
98. https://bis.gov.in/wp-content/uploads/2018/11/agenda-compo-litd-30.pdf
99. https://indiaai.s3.ap-south-1.amazonaws.com/docs/guidelines-governance.pdf
100. https://doi.org/10.2307/2010357. In https://academic.oup.com/ia/article/100/3/1275/7641064
101. https://oxgs.org/2024/04/08/oxgs-report-navigating-geopolitics-in-ai-governance/
102. https://bisi.org.uk/reports/global-fragmentation-of-ai-governance
103. https://itif.org/publications/2026/01/26/how-brussels-effect-hinders-innovation-in-global-south/

104. https://academic.oup.com/ia/article/100/3/1275/7641064
105. https://arxiv.org/abs/2601.17191
106. https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration
107. https://www.cigionline.org/publications/advancing-multi-stakeholderism-for-global-governance-of-the-internet-and-ai/
108. https://globalnetworkinitiative.org/shaping-the-road-to-the-ai-impact-summit-in-india/
109. https://cipesa.org/2025/11/now-more-than-ever-africa-needs-participatory-ai-regulatory-sandboxes/
110. https://www.ohchr.org/Documents/Issues/DigitalAge/Submissions/CSOs/Derechos-Digitales.pdf
111. https://www.alsur.lat/sites/default/files/2024-10/ALSUR%20-%20IA%20en%20Latam%20%5BENG%5D-1.pdf
112. https://www.eff.org/files/2024/10/11/human_rights_implications_and_operational_framework.pdf
113. https://medium.com/opendatacharter/ai-governance-for-and-from-the-global-south-07326645b053
114. https://www.unesco.org/ethics-ai/en/articles/how-iso-and-iec-are-developing-international-standards-responsible-adoption-ai
115. https://www.t20brasil.org/media/documentos/arquivos/TF05_ST_05_Bridging_the_AI_gov66cdcbf06f991.pdf
116. https://impact.indiaai.gov.in/working-groups/safe-trusted-ai
117. https://impact.indiaai.gov.in/working-groups/safe-trusted-ai
118. https://www.sciencedirect.com/science/article/pii/S1389041725000373
119. https://journals.sagepub.com/doi/10.1518/hfes.46.1.50_30392
120. https://www.nber.org/system/files/working_papers/w34255/w34255.pdf
121. https://mbs.edu/news/global-study-reveals-trust-of-ai-remains-a-critical-challenge
122. https://www.sciencedirect.com/science/article/pii/S1566253523002129
123. https://dl.acm.org/doi/10.1145/3555803
124. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
125. Bostrom, N. (2014). Superintelligence: Paths, dangers, strategies. Oxford University Press.
126. https://internationalaisafetyreport.org
127. https://moratorium.ai/
128. https://futureoflife.org/open-letter/pause-giant-ai-experiments/
129. http://proceedings.mlr.press/v81/buolamwini18a.html
130. https://www.accessnow.org
131. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
132. https://doi.org/10.1177/15248380251384271
133. https://www.unwomen.org/en/digital-library/publications/2025/01/ai-powered-online-abuse
134. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf
135. https://www.derechosdigitales.org/en/recursos/deepfakes-when-technologies-promote-digital-gender-based-violence/
136. https://www.nature.com/articles/d41586-025-04041-0
137. https://arxiv.org/abs/2505.14971
138. https://arxiv.org/abs/1607.06520
139. http://proceedings.mlr.press/v81/buolamwini18a.html
140. Hao, K. (2025). Empire of AI: Dreams and Nightmares in Sam Altman's OpenAI. Penguin Press.
141. https://time.com/6247678/openai-chatgpt-kenya-workers/

142. https://huggingface.co/blog/sasha/ai-environment-primer
143. https://wedocs.unep.org/items/5f3afe87-5419-439b-a6ee-14240b2605e9
144. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
145. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
146. https://www.oecd.org/en/publications/regulatory-sandbox-toolkit_de36fa62-en.html
147. https://www.adalovelaceinstitute.org/blog/regulatory-sandboxes/
148. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4486259
149. https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026
150. https://arxiv.org/abs/2310.11986
151. https://datasociety.net/library/ai-governance-needs-sociotechnical-expertise/
152. https://journals.sagepub.com/doi/10.1177/21695067231192298
153. For example, Article 40 of the EU AI Act references consultation with stakeholders. https://artificialintelligenceact.eu/article/40/, as do the 2026 India AI Governance Guidelines. https://www.pib.gov.in/PressNoteDetails.aspx?id=157370&NoteId=157370&ModuleId=3&reg=3&lang=2
154. https://inventory.algorithmwatch.org/about
155. https://psa.gov.in/CMS/web/sites/default/files/publication/WP_Democratising%20Access_V3.0_29122025A.pdf
156. https://www.nvidia.com/en-in/glossary/ai-infrastructure/
157. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/11/competition-in-artificial-intelligence-infrastructure_69319aee/623d1874-en.pdf
158. https://doi.org/10.1145/3458723
159. https://www.itic.org/documents/artificial-intelligence/ITI_AITechnologyStack.pdf
160. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=157204&ModuleId=3&reg=3&lang=1
161. https://cacm.acm.org/research/the-ai-tech-stack-model/
162. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=157204&ModuleId=3&reg=3&lang=1
163. https://www.sfa-oxford.com/knowledge-and-insights/critical-minerals-in-low-carbon-and-future-technologies/critical-minerals-in-artificial-intelligence/
164. https://www.science.org/doi/10.1126/science.aef6678
165. https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai
166. https://iea.blob.core.windows.net/assets/de9dea13-b07d-42c5-a398-d1b3ae17d866/EnergyandAI.pdf
167. https://dl.acm.org/doi/10.1145/3630106.3658542
168. https://www.nature.com/articles/s41545-021-00101-w
169. https://www.goethe.de/ins/in/en/kul/fmd/afu/26335486.html
170. https://arxiv.org/abs/2304.03271
171. https://ainowinstitute.org/publications/research/executive-summary-artificial-power
172. https://journals.sagepub.com/doi/10.1177/20539517241232630
173. https://www.ineteconomics.org/uploads/papers/WP_228-Korinek-and-Vipra.pdf
174. https://www.csis.org/analysis/divide-delivery-how-ai-can-serve-global-south
175. https://cyberbrics.info/wp-content/uploads/2023/08/AI-sovereignty-updated-CLEAN.pdf
176. https://arxiv.org/abs/2004.08900
177. https://arxiv.org/abs/2203.15556
178. https://arxiv.org/abs/2401.02452
179. https://arxiv.org/pdf/2010.15581

180. https://unctad.org/news/ai-market-projected-hit-48-trillion-2033-emerging-dominant-frontier-technology
181. https://www.weforum.org/stories/2025/07/ai-geopolitics-data-centres-technological-rivalry/
182. https://www.foreignaffairs.com/united-states/geopolitics-age-artificial-intelligence
183. https://indianexpress.com/article/explained/explained-sci-tech/us-ai-export-rule-india-9780103/
184. https://community.sap.com/t5/joinville-blog-posts/brazil-as-one-of-the-leading-hubs-for-artificial-intelligence-2024-2025/ba-p/14293868
185. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=156786&ModuleId=3&reg=3&lang=2
186. https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy
187. https://www.mea.gov.in/Images/CPV/G20-New-Delhi-Leaders-Declaration.pdf
188. https://www.chathamhouse.org/sites/default/files/2025-10/2025-10-27-digital-public-infrastructure-wilkinson-et-al.pdf
189. https://www.weforum.org/stories/2024/12/can-digital-public-infrastructure-help-guide-the-transformation/
190. https://www.pib.gov.in/PressReleasePage.aspx?PRID=2224505&reg=3&lang=1
191. https://ajic.wits.ac.za/article/view/20800/24560
192. https://accesspartnership.com/opinion/launch-of-the-first-open-large-language-model-for-latin-america-the-caribbean-latam-gpt/
193. https://digitalpublicinfrastructure.ai/wp-content/uploads/2026/01/Vision_Paper_-DPI-AI-Framework_2026_CDPI.pdf
194. https://bhashini.gov.in/
195. https://www.singpass.gov.sg/main/security/
196. https://eurocities.eu/latest/accessing-danish-digital-public-services-with-muni/
197. https://cipesa.org/2025/07/digital-public-infrastructure-in-africa-a-looming-crisis-of-equitable-access-digital-rights-and-sovereign-control/
198. https://www.dataprivacybr.org/en/report-digital-identity-and-digital-public-infrastructure-recommendations-for-a-fair-information-architecture/
199. https://carnegieendowment.org/research/2025/02/digital-public-infrastructure-a-practical-approach-for-africa?lang=en
200. https://www.dpi-safeguards.org/
201. https://aclanthology.org/2020.acl-main.560/
202. https://www.brookings.edu/articles/closing-the-gap-a-call-for-more-inclusive-language-technologies/
203. https://www.weforum.org/stories/2024/09/ai-linguistic-diversity-gap-missed-opportunity/
204. https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00447/109285/Quality-at-a-Glance-An-Audit-of-Web-Crawled
205. https://pml4dc.github.io/iclr2023/pdf/PML4DC_ICLR2023_39.pdf
206. https://aclanthology.org/2022.acl-long.482/
207. https://news.stanford.edu/stories/2025/05/digital-divide-ai-llms-exclusion-non-english-speakers-research
208. https://arxiv.org/abs/2003.11529
209. https://arxiv.org/abs/2005.00085
210. https://github.com/AI4Bharat
211. https://aclanthology.org/2025.acl-long.1487.pdf
212. https://www.itu.int/hub/2022/08/ai-indigenous-languages-maori-te-reo/

MAP AI

Multistakeholder Approach to Participation in AI Governance