

The First Global Dialogue on AI Governance: Who Gets to Govern AI?

DOI: 10.62692/gppd.2026.01

Published: September 15, 2026

CrossMark

ABSTRACT

The inaugural session of the United Nations Global Dialogue on AI Governance, convened in Geneva on 6–7 July 2026, was the first annual session of a universal UN platform for discussion of international cooperation on AI governance. This paper uses the Geneva Dialogue not as an event to be reported but as a diagnostic case through which to examine a deeper political question: who actually possesses the power to govern artificial intelligence? Drawing on official United Nations documents, the preliminary report of the Independent International Scientific Panel on Artificial Intelligence, and contemporary institutional and statistical evidence available before the publication deadline, the analysis distinguishes seven analytically separate dimensions of governing power—participation, representation, influence, agenda-setting, decision-making, implementation capacity and technological power—and argues that international AI-governance practice and commentary can conflate procedural participation with substantive governing agency. The paper develops an original Advocacy Unified Network (AUN) framework of “AI Governance Agency,” understood as a bundle of epistemic, normative, standard-setting, infrastructural, regulatory, implementation and accountability capabilities. It then maps the global distribution of these capabilities across compute, semiconductors, data, models, cloud, talent, standards, financing and regulatory capacity, and tests the proposition that universal participation does not automatically produce universal governing agency. The findings indicate that the Dialogue’s universalism is procedurally substantial but does not by itself redistribute the capabilities that make a seat meaningful. The paper concludes with concrete recommendations at global, regional, national, technical and civil-society levels, and proposes an institutional test for success that the Dialogue must meet if it is to become more than another layer in an already fragmented governance landscape.

TABLE OF CONTENTS

1. Introduction and Research Question

Research question and analytical propositions

Author disclosure and positionality

2. The Emergence of Global AI Governance

3. The Geneva Dialogue: Mandate, Architecture and Participation

3.1 Mandate and institutional basis

3.2 Architecture: co-chairs, secretariat and the role of ITU

3.3 Participation: who was in the room

3.4 Agenda-setting

4. Who Was Represented—and How?

- 4.1 Governmental representation
- 4.2 Non-governmental participation
- 4.3 Global South representation: a quantitative and qualitative reading
- 4.4 Affected communities and the missing voices

5. Participation versus Governing Agency

- 5.1 The seven dimensions of governing power
- 5.2 The AUN concept of AI Governance Agency
- 5.3 What the Geneva Dialogue can and cannot decide

6. The Global Distribution of AI Capabilities and Governance Power

- 6.1 Compute and cloud infrastructure
- 6.2 Semiconductors: the deepest choke point
- 6.3 Models, data and talent
- 6.4 Financing and investment
- 6.5 Standards and regulatory capacity

7. Global South Representation and Institutional Agency

- 7.1 Global South leadership: where agency is already being exercised
- 7.2 Differences within the Global South: against homogenisation
- 7.3 Structural vulnerability and the special case of SIDS and LDCs
- 7.4 Alternative governance approaches from the Global South

8. The Role of Science, Standards and Technical Capacity

- 8.1 The Independent International Scientific Panel on AI
- 8.2 Technical standard-setting: the hidden layer of governance
- 8.3 Technical capacity and the right to govern

9. Fragmentation of the International AI Governance Landscape

- 9.1 The crowded field
- 9.2 Does the Dialogue reduce fragmentation or add a layer?
- 9.3 The US question: governance under superpower dissent

10. The Geneva Dialogue's Institutional Strengths and Limitations

- 10.1 Strengths
- 10.2 Limitations
- 10.3 The institutional test: nine criteria applied
- 10.4 Study limitations and inference boundaries

11. An AUN Framework for Genuinely Inclusive AI Governance

- 11.1 From universal participation to universal agency: the design inversion
- 11.2 The four pillars of a capability-oriented architecture
- 11.3 The capability gradient as design constraint

12. Policy Architecture and Recommendations

- 12.1 Global-level recommendations
- 12.2 Regional-level recommendations
- 12.3 National-level recommendations

1. Introduction and Research Question

On 6 and 7 July 2026, in the Palexpo convention centre in Geneva, the United Nations convened the inaugural session of what it called the Global Dialogue on AI Governance. The Secretary-General of the United Nations, António Guterres, used the occasion to appeal for “far-reaching, worldwide controls” on artificial intelligence, warning that AI chips designed for civilian use were already shifting to the battlefield and insisting that any future agreement must be “worthy of global trust” and put safety first [1][2]. The President of the General Assembly, Annalena Baerbock, framed the moment in still more expansive terms, describing the Dialogue as an opportunity to define “a shared vision in which technological progress goes hand in hand with human dignity, equity, and sustainable development” and to demonstrate that the United Nations was “able and willing to move on the most pressing and emerging issues of our time” [3]. The optics were deliberate: a universal membership organisation, an explicitly inclusive platform, co-chaired by a small developing country of the Global South (El Salvador) and a small technologically advanced state of the Global North (Estonia), convening alongside the World Summit on the Information Society (WSIS) Forum and the International Telecommunication Union’s AI for Good Global Summit.

The premise of this paper is that the Geneva Dialogue is too important to be read as an event, and too interesting to be read as a press release. It is better understood as a diagnostic window onto the deeper political question that gives this study its title: who gets to govern artificial intelligence? That question cannot be answered by listing who attended. It requires examining who possesses meaningful power to shape AI rules, standards, infrastructure, evidence, implementation and accountability—and, just as importantly, who does not. The core analytical move of this paper is to disaggregate “governance” into seven analytically distinct dimensions: participation, representation, influence, agenda-setting, decision-making power, implementation capacity, and technological power. These dimensions are related but not cumulative or necessarily sequential: an actor can possess substantial technological power without formal participation in a diplomatic forum, or substantial regulatory power without frontier technological capacity. The core argument advanced and tested here is that conflating procedural presence with substantive governing agency is a structural feature of contemporary global AI governance, and that the Geneva Dialogue, for all its procedural ambition, has not yet escaped it.

The motivation for this inquiry is intellectual and institutional rather than ideological. The Advocacy Unified Network (AUN), the author’s institutional affiliation, has argued across its work on trade, digital and development policy that the governance of transformative technologies is shaped not only in the rooms where treaties are discussed but also in upstream spaces where evidence is produced, standards are written, infrastructure is built, talent is concentrated and financing is allocated. If that proposition holds for AI, then important loci of AI governance lie in the foundries, data centres, model laboratories, standards committees and regulatory agencies in which a relatively small number of actors make decisions whose effects are felt globally. The Geneva Dialogue matters because it is the United Nations’ first universal platform specifically dedicated to AI governance; it also provides a useful window onto how much of the underlying capability distribution is established before multilateral deliberation begins.

This study proceeds in five analytical moves. First, it situates the Geneva Dialogue within the longer emergence of global AI governance, from the 2023–2024 wave of national strategies and summitry through the 2024 Global Digital Compact to the August 2025 General Assembly resolution that formally established both the Dialogue and its sister mechanism, the Independent International Scientific Panel on Artificial Intelligence [4][5]. Second, it examines the Dialogue’s mandate, architecture and participation in granular detail, asking whether the institutional design matches the universalist claims. Third, it develops an original AUN framework of “AI Governance Agency”, which treats the capacity to govern AI as a bundle of seven capabilities—epistemic, normative, standard-setting, infrastructural, regulatory, implementation and accountability—and uses this framework to read the Dialogue’s outputs. Fourth, it maps the global distribution of the structural capabilities underlying AI governance (compute, semiconductors, data, models, cloud, talent, standards, financing and regulatory capacity) and asks whether the Dialogue’s universalism is substantive or merely procedural. Fifth, it situates the Dialogue within the fragmented international AI governance landscape and asks the genuinely difficult question: does the Geneva Dialogue reduce fragmentation, or does it add another layer to it?

The paper is deliberately evidence-driven and uses sources available before a clear evidence cutoff of 14 September 2026. It privileges official United Nations and General Assembly documents, the preliminary report of the Scientific Panel, official government and regional-organisation statements, and authoritative institutional research. It does not invent participation statistics, demographic distributions, or claims about Global South representation; where reliable data are not available, it says so. It is neither anti-technology nor anti-Western in orientation. The judgement it offers is institutional: that the problem of global AI governance is not principally that some countries are underrepresented in international discussions, but that the capabilities required to shape AI’s rules, evidence, infrastructure, standards and implementation remain highly unevenly distributed, so that universal participation does not automatically produce universal governing agency. The paper’s task is to test, refine and qualify that proposition—not to assume it.

Before proceeding, a note on method and limits is warranted. This is a qualitative single-case institutional analysis informed by quantitative indicators of capability concentration. It draws on the Stanford Institute for Human-Centered Artificial Intelligence (HAI) AI Index 2026 for investment, talent and compute metrics; on the United Nations, the International Telecommunication Union (ITU), the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the Organisation for Economic Co-operation and Development (OECD) for governance instruments; and on think-tank, civil-society and industry analyses for critical interpretation. Primary institutional sources are preferred for claims about mandates, participation, institutional design and legal status; secondary sources are used mainly for interpretation or for indicators not published by the relevant public institution. The Geneva Dialogue is treated as a diagnostic case, not as a representative sample of global AI governance. The paper does not claim causal identification, population-level generalisation or empirical validation of the AGA framework. Its purpose is to establish a transparent analytical model, test it against the inaugural Dialogue, and identify institutional conditions that can be examined longitudinally.

Research question and analytical propositions

Research question. To what extent does the Geneva Dialogue convert universal participation into substantive governing agency, and what institutional and capability conditions mediate that conversion?

Proposition 1. Universal participation is necessary for legitimate global deliberation but is insufficient, by itself, to produce substantive governing agency.

Proposition 2. Governing agency is multidimensional: meaningful influence depends on epistemic, normative, standard-setting, infrastructural, regulatory, implementation and accountability capabilities rather than on

participation alone.

Proposition 3. The Dialogue is more likely to reduce fragmentation if it develops durable interfaces with the scientific, technical, regulatory and capacity-building layers of the wider AI governance regime complex.

Proposition 4. Global South inclusion becomes substantively meaningful when participation is coupled with capability-building, technical capacity and pathways into agenda-setting, standards and implementation.

Author disclosure and positionality

The author is Chairman of the Advocacy Unified Network (AUN), which is also the institutional home of the AUN AI Governance Agency (AGA) framework developed in this paper. That institutional position is a relevant potential source of perspective and is disclosed deliberately. The paper is not an official United Nations document, does not speak for the United Nations or any UN entity, and should not be read as an institutional endorsement by the UN. The descriptive and empirical claims are grounded in the cited public sources; the AGA framework and the policy recommendations are the author's analytical contributions and are presented as proposals for scrutiny rather than as established institutional facts. Funding and competing interests: no external funder or financial conflict is identified in the manuscript; the author's leadership role at AUN is the relevant institutional disclosure.

2. The Emergence of Global AI Governance

Global AI governance did not begin in Geneva. It accreted, in layers, over roughly a decade, through a sequence of national strategies, voluntary principles, soft-law instruments, summit declarations and, eventually, a treaty intended to be legally binding upon entry into force. Understanding the Geneva Dialogue requires understanding this accumulation, because the Dialogue was not created in a vacuum but was positioned atop an already dense and partly competitive governance landscape. The Decision to establish it—committed in the Global Digital Compact of September 2024 and operationalised by General Assembly resolution A/RES/79/325 of 26 August 2025—was, in important respects, a decision about how to relate to what already existed [4][5].

The first layer is normative and largely declaratory. In November 2021, UNESCO's 193 Member States adopted the Recommendation on the Ethics of Artificial Intelligence, the first global standard-setting instrument on AI ethics, which is applicable to all 194 member states of the organisation and which established a set of values and principles around human rights, dignity, transparency and accountability [6] [7]. Two years earlier, in May 2019, the OECD had adopted its AI Principles—the first intergovernmental guidelines on AI—which the G20 subsequently endorsed, and which were updated in 2024 to reflect generative AI; the OECD AI Policy Observatory now tracks more than 1,000 national AI policies and initiatives [8]. These instruments are influential but non-binding. They shape how officials speak about AI; they do not, on their own, constrain what firms do or what states build.

The second layer is regulatory and increasingly binding in domestic law. The European Union's AI Act, which entered into force in August 2024, is the first comprehensive horizontal AI regulation adopted at EU level and, through its extraterritorial reach, has acquired a significant international standard-setting effect often discussed in terms of the "Brussels effect" [9]. At the international level, the Council of Europe's Framework Convention on Artificial Intelligence, opened for signature on 5 September 2024, is the first international legally binding treaty specifically devoted to AI governance. Ten signatories joined at the opening ceremony, including the United States and Israel as non-member states of the Council of Europe; the European Union signed separately in its own institutional capacity [9][10]. As of 12 September 2026, the Convention had not yet entered into force: the Treaty Office requires five ratifications, including at least three Council of Europe

member States [10]. This is the layer at which “governance” begins to acquire teeth—but it remains regionally concentrated, and its binding force depends on ratification that, at the time of writing, is still incomplete.

The third layer is diplomatic and summit-driven. The Bletchley Declaration of November 2023, the Seoul Summit of May 2024, and the AI Action Summit held in Paris in February 2025 constituted a rapid succession of leaders’-level engagements that produced declarations, voluntary commitments, and, in Paris, a statement on inclusive and sustainable AI signed by 63 countries and the African Union Commission and European Union, but not by the United States or United Kingdom [11][12]. Alongside the summitry, the G7’s Hiroshima AI Process produced an International Code of Conduct for organizations developing advanced AI systems in 2023, and the G20, under South Africa’s 2025 presidency, established a dedicated AI Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development [11][12]. Each summit added rhetorical weight; none produced a binding global AI treaty. Their cumulative effect, however, was to normalise the idea that AI was a matter for heads of state, not merely for regulators and engineers.

The fourth layer is technical and standard-setting. The International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC), through its subcommittee SC 42, have produced a growing family of AI standards, including ISO/IEC 42001 on AI management systems. The IEEE has its own portfolio. National standards bodies—the US National Institute of Standards and Technology (NIST) with its AI Risk Management Framework, the UK and others—have generated de facto reference standards that travel internationally through procurement and certification. This is the layer at which “governance” becomes engineering: a management-system standard is not a law, but a firm that cannot certify against it may find itself excluded from public procurement and from supply chains. Technical standard-setting is, in this sense, one of the most consequential and least visible sites of AI governance.

The fifth layer is the scientific. The United Nations Secretary-General’s High-level Advisory Body on Artificial Intelligence, convened in October 2023, delivered its final report, *Governing AI for Humanity*, in September 2024, after extensive global consultations involving more than 2,000 participants [13][14]. The report proposed seven recommendations: the establishment of an International Scientific Panel on AI; the launch of a Policy Dialogue on AI Governance; the creation of an AI Standards Exchange; an AI Capacity Development Network; a Global Fund for AI “to put a floor under the AI divide”; a Global AI Data Framework; and an AI Office within the UN Secretariat [13][15]. Two of these—the Scientific Panel and the Policy Dialogue—were carried forward into the Global Digital Compact and then into the General Assembly resolution of August 2025. The others, at the time of writing, remain proposals. This selective uptake is analytically significant, but the paper does not infer political motive from the pattern. Two recommendations—the Scientific Panel and the Policy Dialogue—were institutionalised, while several resource- or authority-redistributing proposals, including a Global Fund, a Global AI Data Framework and an AI Office, were not carried into the August 2025 resolution.

It is against this layered inheritance that the Geneva Dialogue must be read. The Dialogue inherits the universalism of UNESCO and the General Assembly; the technocratic ambition of the OECD and the standards bodies; the binding-treaty precedent of the Council of Europe; the summit momentum of Bletchly, Seoul and Paris; and the architectural blueprint of the High-level Advisory Body. It does not inherit, however, the regulatory teeth of the EU AI Act, the technical authority of ISO/IEC, or the resources that a Global Fund would have provided. The Dialogue is, in this sense, the universalist layer on top of an architecture that is, in its binding and operational core, regional, technical and concentrated. The question is whether a universalist layer can discipline what lies beneath it, or whether it will instead be disciplined by it.

3. The Geneva Dialogue: Mandate, Architecture and Participation

To assess the Geneva Dialogue as an institutional object rather than as a ceremonial occasion, it is necessary to specify its mandate, its institutional basis, its architecture, the composition of its participation, and the process by which its agenda was set. Each of these elements reveals something about the kind of governance the Dialogue is capable of exercising—and, more revealingly, the kind it is not.

3.1 Mandate and institutional basis

The legal basis of the Dialogue is unusually clear for an emerging multilateral mechanism. The commitment to establish both an AI scientific panel and a global dialogue on AI governance was made in the Global Digital Compact, adopted as part of the Pact for the Future at the Summit of the Future in September 2024 [16]. That political commitment was then operationalised in General Assembly resolution A/RES/79/325, adopted by consensus on 26 August 2025, which established the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance and set out their terms of reference and modalities [4][5]. The pairing is doctrinally important: the Panel produces a shared scientific evidence base, and the Dialogue is the intergovernmental platform that consumes it. This is an explicit borrowing from the architecture of the Intergovernmental Panel on Climate Change (IPCC) and the climate regime, in which science and diplomacy are linked but institutionally separated.

The mandate, as the United Nations itself describes it, is unambiguous in its modesty. The Dialogue is “the United Nations platform where all governments, private sector, academia and civil society will convene to discuss international cooperation, share best practices and lessons learned, and facilitate open, transparent and inclusive discussions on artificial intelligence governance” [17]. The defining verb is “discuss”, not “negotiate”, “regulate”, or “decide”. The United Nations has been careful to describe the Dialogue as a platform for discussion and cooperation rather than as a conventional negotiating forum, and this self-description must be taken seriously. A platform for discussion can shape norms, build trust, socialise governments into shared language, and surface disagreement; it cannot, by itself, produce binding rules. To attribute regulatory authority to the Dialogue that it does not possess would be a category error, and it is one that enthusiastic commentary sometimes commits.

The institutional test for the Dialogue, therefore, cannot be whether it produces a treaty—it was not designed to. The test is whether, operating within its actual mandate, it can shift the distribution of governing capability among its participants: whether it can raise the epistemic floor of weaker states, surface and legitimise alternative normative framings, create pathways into technical standard-setting for under-represented actors, and discipline the concentrated technological and commercial power that operates around it. These are the measurable criteria by which a discussion platform can nonetheless be judged, and they are the criteria this paper applies in Section 10.

3.2 Architecture: co-chairs, secretariat and the role of ITU

The Dialogue is co-chaired by two ambassadors: Egriselda López, Permanent Representative of El Salvador to the United Nations, and Rein Tammsaar, Permanent Representative of Estonia to the United Nations [17][18]. The co-chair formula is itself a statement. El Salvador is a small Central American developing country with limited domestic AI capacity; Estonia is a small, technologically advanced European state and a digital-government exemplar. The pairing pairs Global South with Global North, small state with small state, and developmental modesty with digital sophistication. It is a formula designed to pre-empt the critique that AI governance is being captured by the technologically powerful. It is also a formula that, by placing small states

at the helm, leaves the large powers structurally outside the chair—which is both a strength (independence) and a limitation (the chairs do not control the powers whose behaviour would most need to be governed).

The secretariat function is shared. The Dialogue was convened in Geneva at the Palexpo alongside the WSIS Forum and the ITU's AI for Good Global Summit, and the ITU played a central role in the practical organisation of the event [19]. The Dialogue is facilitated by a Joint Secretariat comprising the International Telecommunication Union (ITU), the United Nations Educational, Scientific and Cultural Organization (UNESCO), the UN Office for Digital and Emerging Technologies (ODET), and the Executive Office of the Secretary-General (EOSG); the inaugural 2026 session was coordinated by ITU and UNESCO [17]. This arrangement combines universal intergovernmental mandate with technical and convening capacity, while also bringing together institutions with different organisational cultures and constituencies. That institutional plurality is an asset, but it creates a governance question of its own: whether the Dialogue can preserve a genuinely universal political frame while benefiting from the technical expertise and stakeholder reach of the specialised agencies supporting it.

Three further architectural features deserve note. First, the Dialogue was designed to be annual, with its second session scheduled for 3–4 May 2027 in New York, back-to-back with the eleventh Multi-stakeholder Forum on Science, Technology and Innovation for the Sustainable Development Goals [20]. The alternation between Geneva and New York is itself a signal: Geneva situates the Dialogue in the technical and standards ecosystem, New York situates it in the political and developmental one. Second, the Dialogue was paired from inception with the Independent International Scientific Panel on Artificial Intelligence, co-chaired by Yoshua Bengio and Maria Ressa, whose Preliminary Report was published on 1 July 2026 and presented at the Dialogue as a shared scientific evidence base [21][22]. Third, the General Assembly established a participation-support mechanism to facilitate the attendance of developing-country delegates, with the co-chairs encouraging voluntary financial and in-kind contributions [23]. This mechanism recognises that universal participation carries material costs; it should not be conflated with the separate Global Fund for AI proposed by the High-level Advisory Body.

3.3 Participation: who was in the room

The United Nations' post-session account provides the most precise official participation figures: the first annual Dialogue brought together more than 3,000 participants, including delegations from 163 countries, six Heads of State and Government, more than 60 ministers, and more than 1,000 stakeholder entities [17]. A separate European External Action Service account reported a broader figure of representatives from more than 170 countries and over 4,200 participants [24]. These figures should not be treated as interchangeable measures: they appear to reflect different counting conventions. For the principal analysis, this paper uses the UN's post-session figures because they provide the clearest institutional breakdown. The UN does not, however, publish in the same account a complete verified breakdown of state participation by regional group, developing-country status, SIDS status or LDC status. The defensible conclusion is therefore that the Dialogue achieved broad and geographically extensive participation, while the depth and distribution of substantive representation remain more difficult to quantify.

What can be said with confidence is that the participation design was explicitly multi-stakeholder, modelled on the WSIS tradition rather than on a purely intergovernmental conference. The programme included a high-level segment, thematic sessions, side events, and formal interventions from governments and stakeholders delivered under an inscription-based format in three-minute slots (or five minutes for group statements) [26]. This format has well-known strengths and weaknesses. Its strength is that it maximises the number of voices heard and signals inclusivity. Its weakness is that three-minute interventions from a long queue prioritise breadth of access over sustained deliberation, making them poorly suited to the development of shared positions, the negotiation of trade-offs, or the exercise of meaningful agenda-setting

by those at the back of the queue. The format allocates airtime; it does not allocate influence. The distinction matters.

3.4 Agenda-setting

The agenda of the inaugural session was organised around thematic clusters, with cluster co-chairs reporting back in plenary [26]. The substantive framing drew on the Preliminary Report of the Scientific Panel, which organised its findings across seven domains: AI science, advances and trajectories; societal applications in science, health, education and agriculture; economic implications; security, systems and environmental implications; human rights, information and democracy; cultural and individual flourishing, autonomy and child safety; and management, governance and reliability [22]. The alignment between the Panel's domains and the Dialogue's thematic structure is analytically significant, but the agenda was not simply imposed by the Panel. The Dialogue followed an eight-month preparatory process involving thematic, regional and virtual consultations, and the UN reports that 1,532 unique written submissions from organisations and individuals informed the agenda and deliberations [17]. The more defensible conclusion is therefore that the Scientific Panel materially shaped the evidence and framing of the conversation, while the agenda itself was co-produced through consultations, submissions, Member State inputs and stakeholder engagement.

Agenda-setting is the most underestimated form of power in international institutions. To choose what is on the table is to choose what can be decided, and to choose the framing of what is on the table is to shape the range of acceptable answers. If the Dialogue's agenda is set by the Scientific Panel's framing, and the Panel's framing is set by its composition, then the question of who chose the Panel becomes a question about who set the Dialogue's agenda. The General Assembly resolution provided the mandate; the Secretary-General appointed the members; the co-chairs and secretariat organised the sessions. At each stage, the locus of agenda-setting authority was upstream of the Dialogue's participants. This is not unusual in multilateral institutions, but it does mean that the Dialogue's universalism in participation does not extend to universalism in agenda-setting. The room is universal; the agenda is curated.

4. Who Was Represented—and How?

Representation is a more demanding concept than participation, and it is the concept on which the legitimacy of a universal platform ultimately rests. To participate is to be present; to be represented is to have one's interests, perspectives and capacities meaningfully carried into the deliberation. The two can coincide, but they need not. A state may be present in the room and structurally absent from the substance; a civil-society organisation may have a speaking slot and no path to influence the conclusions; a region may have many flags in the plenary and no voice in the framing. This section examines the question of who was represented at the Geneva Dialogue, and how, across four dimensions: governmental representation, non-governmental participation, Global South representation, and the representation of affected communities.

4.1 Governmental representation

The governmental composition of the Dialogue reflected the universal membership of the United Nations, but with the asymmetries of capacity that characterise multilateral diplomacy more generally. All UN Member States were entitled to participate; not all were equally able to. The participation-support mechanism was established precisely to address the cost barrier for developing countries [23], but travel and per-diem support cannot, on its own, build the substantive capacity to engage with frontier AI policy questions across multiple technical domains in a two-day format. The difference between a delegation that includes a minister, a digital-policy official, a technical regulator and advisers, and a delegation that consists of a single diplomat

covering multiple files, is a difference in substantive representational capacity even when both delegations are formally present.

The pattern of which states invest in AI diplomacy is itself revealing. A growing number of states have published national AI strategies—more than 80 countries between January 2017 and June 2025, according to research synthesised by the World Bank [27]—but the depth, funding and regulatory maturity behind those strategies varies enormously. China, Brazil, the Republic of Korea, Saudi Arabia, Türkiye and others have substantial national strategies; many low-income countries have strategy documents without the implementing institutions or budgets to make them operative. The OECD's Digital Government Outlook 2026 found that nearly all OECD countries have an AI-in-government strategy and that 30 of 36 (83%) have at least one institution responsible for governing AI [28]; comparable institutional density in the Global South is far lower. Governmental representation at the Dialogue, in other words, was numerically universal but substantively graded, and the gradient tracked the underlying institutional capacity that the Dialogue is, in principle, supposed to help correct.

4.2 Non-governmental participation

The Dialogue was explicitly multi-stakeholder, and the private sector, academia and civil society were all represented. The design placed major technology firms, academic researchers and civil-society organisations in the same deliberative space as governments. This is a genuine and important feature: it avoids the pretence that AI is a purely intergovernmental matter when the substantive power over AI is held substantially outside governments. But the multi-stakeholder design introduces its own asymmetries. Major technology firms that develop frontier models may possess analytical capacity, technical depth and lobbying resources far greater than those available to many participating states and civil-society organisations. A multi-stakeholder platform that places a frontier-model developer and a small developing-country delegation in the same speaking format does not thereby equalise their influence; it may, in some respects, formalise the asymmetry by granting the firm a legitimacy-conferring seat alongside sovereign states.

The civil-society read-out published after the session highlighted both the value of the access and its limits, and advanced recommendations to strengthen substantive inclusion of civil-society and affected-community perspectives in future sessions [25]. The recurring concern in such read-outs across the AI governance space is that civil society is admitted to the room but largely confined to the margins of the agenda—invited to speak on human rights and inclusion, less often invited to shape the technical and economic framings where the distributional questions are actually decided. If the Geneva Dialogue reproduces this pattern, its multi-stakeholderism will be real but partial: universal access to the room, uneven access to the framing.

4.3 Global South representation: a quantitative and qualitative reading

The Global South was represented at the Dialogue through its states, through the co-chairmanship of El Salvador, and through the participation of developing-country civil-society and academic actors. The qualitative question is whether this representation was numerical, substantive, or both. The paper is cautious here because, as noted, the organisers have not published a verified statistical breakdown of state participation by region, developing-country status, or vulnerability classification. What can be said is that the design of the Dialogue—universal membership, a participation support fund, a Global South co-chair—was explicitly engineered to secure numerical presence, and that numerical presence appears to have been broadly achieved.

Numerical presence, however, is the easier problem. The harder problem is substantive representation, and it has three components. First, the framing problem: if the agenda is pre-structured by a scientific panel and a set of thematic clusters largely derived from the priorities of the existing AI governance conversation, then

Global South representation operates within a frame that may not reflect Global South priorities—which include, characteristically, not only the safety of frontier models but the extraction of data, the bias of imported systems, the displacement of labour, the concentration of compute, and the asymmetric distribution of AI’s benefits and harms. Second, the capacity problem: substantive representation requires the ability to engage across seven technical domains in a two-day format, which in turn requires domestic analytical capacity that most developing countries have not yet built. Third, the follow-through problem: representation in a discussion platform is meaningful only if it connects to implementation—to national regulation, to regional standard-setting, to procurement decisions—and that connection depends on institutional capacity that the Dialogue itself does not create.

The Scientific Panel’s preliminary report is candid about the consequence. It warns that the Global South is “disproportionately exposed to the misuse of AI due to limited capacity for mitigation and limited [governance infrastructure]” [29]. This is an important admission, and it bears directly on the question of representation: if the Global South is disproportionately exposed to AI harm and disproportionately limited in mitigation capacity, then its numerical representation in a discussion platform does not, by itself, shift the underlying distribution of harm or capacity. Representation becomes meaningful only when it is coupled to a redistribution of the capabilities that determine exposure and mitigation. This is the core of the distinction between participation and governing agency, developed in the next section.

4.4 Affected communities and the missing voices

Any honest accounting of representation must acknowledge who was not meaningfully represented even under a universalist design. The people most exposed to AI-mediated harm—workers displaced by automation, communities subject to biometric surveillance, populations misidentified by algorithmic welfare systems, users of low-resource languages underserved by commercial models, children exposed to AI-generated manipulation—were represented, if at all, through intermediaries: civil-society organisations, a few governments, and the Panel’s attention to human rights and child safety. They were not, in any direct sense, governing agents in the Dialogue. This is not a criticism unique to the Geneva Dialogue; it is a structural feature of intergovernmental platforms. But it is a feature that a paper concerned with who gets to govern AI must register. A universal platform that represents states, firms and accredited organisations does not thereby represent the people on whom AI acts. The gap between representation-of-states and representation-of-affected-people is one of the deepest in contemporary global governance, and it is a gap the Dialogue’s design does not close.

5. Participation versus Governing Agency

The distinction between participation and governing agency is the analytical hinge of this paper, and it is the contribution AUN seeks to make to the emerging debate over the Geneva Dialogue. The distinction is not original in the abstract—students of international institutions have long observed that presence is not power—but it has been under-applied in the AI governance discussion, where the rhetoric of “inclusive” and “multi-stakeholder” platforms can substitute for a harder analysis of who can actually do what. This section sets out seven dimensions of governing power, develops the AUN concept of “AI Governance Agency” as a bundle of capabilities, and uses both to read what the Geneva Dialogue can and cannot do.

5.1 The seven dimensions of governing power

A rigorous assessment of governing power should examine seven distinct dimensions: participation, representation, influence, agenda-setting, decision-making power, implementation capacity, and technological power. These dimensions are analytically related but neither cumulative nor strictly sequential.

An actor may be strong in one dimension and weak in another, and its overall governing agency is better understood as a multidimensional profile rather than as a single scalar hierarchy. This distinction matters because technological power, for example, does not require formal participation in a diplomatic forum, while regulatory or agenda-setting power can be substantial even where frontier technological capacity is limited.

First, participation is the capacity to be present in the deliberation. It is the dimension at which the Geneva Dialogue most clearly succeeds. A seat at the table is a necessary but insufficient condition of governing agency: it creates an opportunity for influence but does not itself confer representation, agenda-setting power, decision authority, implementation capacity or technological control. Universal participation is a real institutional achievement; it does not by itself redistribute the underlying capabilities required to exercise those other forms of power.

Second, representation is the capacity to have one's interests, perspectives and constraints meaningfully carried into the deliberation, by an actor who understands them and can articulate them in the technical register of the conversation. Representation requires analytical capacity at home and diplomatic capacity abroad. A delegation that cannot deploy either is present but not represented, and the difference is invisible to a headcount.

Third, influence is the capacity to alter the positions of others in the deliberation. Influence flows from a combination of argumentative weight, coalition-building, reputation, and the credible linkage of the deliberation to decisions that others care about. Influence is what converts a speaking slot into an outcome.

Fourth, agenda-setting is the capacity to determine what is on the table and how it is framed. As argued in Section 3, agenda-setting is the most underestimated form of power in international institutions, because it operates before deliberation begins and shapes the range of acceptable answers. The Dialogue's agenda is curated upstream of its participants.

Fifth, decision-making power is the capacity to author or block binding outputs. The Geneva Dialogue, by its own self-description, does not possess this. It cannot issue regulations, impose standards, allocate resources, or create binding law. The decision-making power over AI remains dispersed across national regulators, regional bodies (notably the EU), standard-setting organisations, and the major firms themselves.

Sixth, implementation capacity is the capacity to translate governance into practice—to test models, certify systems, enforce rules, build infrastructure, train regulators, and evaluate outcomes. Evidence across international development and AI-readiness research suggests that implementation capacity is unevenly distributed and is particularly constrained in many lower-capacity states. This is the dimension at which the gap between nominal and effective governance can become most visible.

Seventh, technological power is the capacity to build, control and shape underlying technological assets—compute, semiconductors, data, models, cloud infrastructure, and talent. This dimension is highly concentrated. An actor with limited technological power may still exercise substantial regulatory, market or standard-setting influence over AI deployed within its jurisdiction, but it is less able to control the upstream infrastructure and technologies on which that deployment depends.

The central analytical claim of this paper is that international discussion of AI governance—and, in particular, some commentary around the Geneva Dialogue—can operate as if the seven dimensions were interchangeable. Universal participation may be treated as evidence of universal representation, influence, agenda-setting, decision-making, implementation and technological power. It is not. The dimensions are distinct, and the gaps between them can be large. A platform that secures universal participation without changing the distribution of the other capabilities can therefore produce broad inclusion in a system whose

underlying governing power remains concentrated. That is the situation the Geneva Dialogue inherits, and it is the situation against which it should be evaluated.

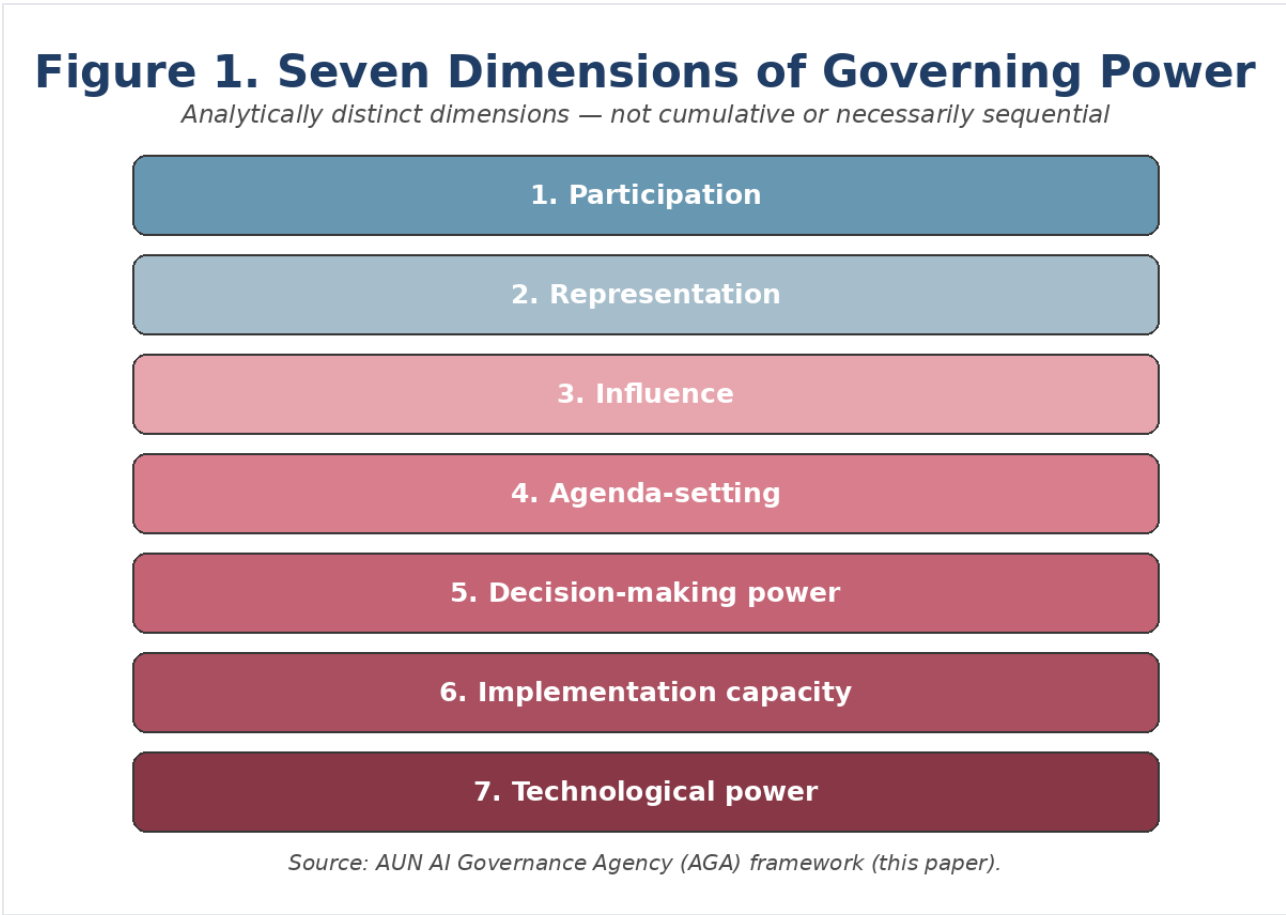


Figure 1. Seven dimensions of governing power: participation, representation, influence, agenda-setting, decision-making, implementation capacity and technological power. The dimensions are analytically distinct and not cumulative or necessarily sequential.

5.2 The AUN concept of AI Governance Agency

To make the distinction operationally useful, this paper proposes the AUN concept of “AI Governance Agency” (AGA), defined as the substantive capacity of a state or society to shape the development, deployment and oversight of artificial intelligence in ways that reflect its own priorities, evidence and constraints. AI Governance Agency is not a single attribute; it is a bundle of seven capabilities that cut across, rather than map one-to-one onto, the seven dimensions of governing power identified above. The seven capabilities are summarised in Table 1 and discussed in turn.

Capability	What it means	Underlying assets
Epistemic agency	Capacity to produce, validate and contest the evidence base for AI governance	Research institutions, peer-reviewed output, data, evaluation labs, scientific personnel

Capability	What it means	Underlying assets
Normative agency	Capacity to define governance priorities and shape the norms that frame AI	Policy capacity, diplomatic weight, legal scholarship, moral authority, coalition leadership
Standard-setting agency	Capacity to participate in and shape technical standards	Standards bodies membership, technical experts, industry participation, procurement leverage
Infrastructural agency	Capacity to control compute, data, models and cloud	Compute, semiconductors, data assets, cloud infrastructure, frontier model development
Regulatory agency	Capacity to enact, enforce and update binding rules	Legislation, regulatory authorities, enforcement capacity, legal remedies, market size
Implementation agency	Capacity to deploy, evaluate and adapt AI in practice	Talent, digital government, procurement, sandbox capacity, public-sector adoption
Accountability agency	Capacity to challenge, remedy and seek redress against harmful AI	Judicial access, civil society, audit capacity, media freedom, affected-community voice

Table 1. The AUN AI Governance Agency (AGA) framework: seven capabilities and their underlying assets.

Epistemic agency is the capacity to produce, validate and contest the evidence base on which AI governance rests. Without it, a state must accept the scientific framing supplied by others, which is another way of saying that others set its agenda. The Independent International Scientific Panel is, in part, an attempt to pool epistemic agency at the international level; whether it redistributes epistemic agency, or merely concentrates it differently, is examined in Section 8.

Normative agency is the capacity to define governance priorities and shape the norms that frame AI. The dominant normative framings of contemporary AI governance—risk-based, safety-first, innovation-friendly, human-rights-anchored—were substantially developed in a small number of jurisdictions. States that adopt these framings wholesale may gain legitimacy and interoperability; they may also import priorities that do not reflect their own circumstances. Normative agency is the capacity to choose which framings to adopt, which to adapt, and which to reject.

Standard-setting agency is the capacity to participate in and shape the technical standards that make AI governable in practice. Standards are where governance meets engineering. A state that does not participate in ISO/IEC SC 42, in IEEE working groups, or in the work of its own national standards body, is a state that has delegated its technical governance to those who do. Participation in standard-setting requires a technical cadre that most states have not yet built, and the Geneva Dialogue does not, by itself, create it.

Infrastructural agency is the capacity to control compute, data, models and cloud—the material substrate of AI. This is the most unevenly distributed capability of all, and it is examined in detail in Section 6. A state without strong infrastructural agency can still regulate AI on its territory, influence market access and procurement, and participate in standard-setting, but it has less control over the upstream infrastructure and technologies on which deployment depends.

Regulatory agency is the capacity to enact, enforce and update binding rules. The EU has built it through the AI Act; a handful of other jurisdictions have built parts of it; most have not. Regulatory agency without implementation agency is a paper tiger, but implementation agency without regulatory agency is technical capacity without democratic direction.

Implementation agency is the capacity to deploy, evaluate and adapt AI in practice—to run sandboxes, test systems, certify models, train regulators, and learn from deployment. This is the point at which governance becomes operational, and it is a dimension in which capability gaps can be especially consequential. A strategy document without implementation capacity is a declaration, not effective governance.

Accountability agency is the capacity to challenge, remedy and seek redress against harmful AI. It requires courts that understand AI, regulators that can investigate it, civil society that can document its harms, media that can report on it, and affected communities that can be heard. Accountability agency is the ultimate test of whether governance serves the governed; it is also the capability most dependent on institutions that lie outside the AI field entirely—the rule of law, judicial independence, civic space.

The AUN framework’s analytical contribution is twofold. First, it disaggregates “governing AI” into seven distinct capabilities, making it possible to ask, of any actor, not whether it “governs AI” in the abstract but which capabilities it possesses and where those capabilities are weak. Second, it makes visible the gradient structure of governing agency: actors can be strong in one or two capabilities and weak in others, and the overall distribution of AI governance power is better described by capability profiles than by a binary North/South or developed/developing split. The Geneva Dialogue, read through this framework, allocates participation broadly but does not, by itself, allocate the other capabilities of governing agency. Whether it can be designed to strengthen them is the question taken up in Section 11.

5.3 What the Geneva Dialogue can and cannot decide

A consequence of the seven-dimension analysis is clarity about what the Dialogue can and cannot do. It cannot issue binding rules because it is not a negotiating forum and lacks decision-making authority. It cannot itself allocate compute or finance infrastructure, and it cannot certify models or enforce standards. What it can do—within its mandate as a discussion platform—is shape norms, surface disagreement, socialise governments into shared language, build epistemic capacity through the Scientific Panel, and create conditions under which other institutions (national regulators, regional bodies, standards organisations) may act in more interoperable ways. These are real functions, and they are not trivial. But they are functions of a discussion platform, not of a governing institution, and the distinction between influence, agenda-setting, coordination and binding decision-making remains central.

It follows that the Dialogue’s success cannot be measured by the metrics appropriate to a negotiating forum. Attendance, declarations, and diplomatic warmth are necessary but insufficient indicators. The metrics appropriate to a discussion platform are subtler: whether it raises the epistemic floor of weaker states, whether it surfaces and legitimises alternative framings, whether it creates durable pathways into technical standard-setting, whether it builds implementation capacity, whether it disciplines concentrated technological power, and whether it sustains continuity across sessions. These are the criteria developed in Section 10, and they are the criteria by which the Dialogue must be judged if judgement is to be honest.

6. The Global Distribution of AI Capabilities and Governance Power

If the Geneva Dialogue allocates participation universally but allocates higher governing capabilities not at all, the question becomes: how are those higher capabilities actually distributed? This section maps the global distribution of the structural capabilities that underlie AI governance—compute, semiconductors, data, models, cloud, talent, standards, financing and regulatory capacity—drawing on the best available evidence before the publication cutoff. The mapping is not offered as a polemic against concentration; some concentration is intrinsic to the economics of the sector. It is offered as the empirical ground on which any judgement about “inclusive” AI governance must rest. The question is not who has access to AI, but who has the capabilities necessary to govern it.

6.1 Compute and cloud infrastructure

Compute is a foundational input of AI and a major locus of concentration. The United States hosts more data centres than any other country: by April 2026, one industry count placed the number at approximately 4,423 facilities, followed by the United Kingdom and Germany in the hundreds [30]. The 2026 Stanford AI Index similarly reports that the United States hosts more than ten times as many data centres as any other country, illustrating the same underlying concentration [38]. This concentration reflects advantages in capital availability, hyperscaler presence, energy and infrastructure. The global cloud-infrastructure market is likewise concentrated among a small number of hyperscalers; in Q2 2026, Amazon, Microsoft and Google together accounted for 63% of worldwide enterprise spending on cloud infrastructure services [31]. Cloud concentration matters for governance because the choices of a small number of firms about what models to host, what content policies to apply, and what jurisdictions to serve shape the AI experience of billions of users, and because states without domestic cloud capacity must govern AI through infrastructure they do not fully control.

The energy and capital intensity of frontier compute is widening the gap rather than narrowing it. Goldman Sachs Research forecast in early 2025 that global data-centre power demand would rise by approximately 50% by 2027 and by as much as 165% by the end of the decade [32]. The capital, power and water requirements of frontier AI compute are increasingly beyond the reach of all but the largest firms and a small number of states, with the consequence that the infrastructure of AI is concentrating even as the user base of AI is universalising. This is the central paradox of contemporary AI governance: more people than ever use AI, while fewer actors than ever build it.

6.2 Semiconductors: the deepest choke point

The semiconductor supply chain is the deepest choke point in the AI capability stack and the clearest illustration of how unevenly technological power is distributed. Taiwan accounts for more than 60% of global foundry revenue and more than 90% of leading-edge chip manufacturing, with the Taiwan Semiconductor Manufacturing Company (TSMC) dominant and US firms such as Apple, NVIDIA and AMD relying heavily on its advanced manufacturing capabilities [33][34]. Advanced-node manufacturing is concentrated in Taiwan and, to a lesser extent, Korea (Samsung); the extreme ultraviolet (EUV) lithography equipment required for leading-edge fabrication is supplied by ASML, headquartered in the Netherlands; design of advanced AI accelerators is led by NVIDIA, with AMD and others competing; and advanced electronic-design-automation software is concentrated in a small number of firms. The global semiconductor market reached approximately USD 795.6 billion in 2025 [35]. WSTS’s Q2 2026 update, incorporating actual first-half results, put the full-year 2026 calculation at approximately USD 1.655 trillion; AI infrastructure remained a major driver, alongside other semiconductor demand [36].

The geopolitics of this concentration is now the single most consequential variable in the global distribution of AI capability. The United States has, since October 2022, imposed successive rounds of export controls restricting the sale of advanced AI accelerators to China, with further tightening in 2023 and 2024, and the

controls' extraterritorial reach has shaped the global market for advanced chips [37]. China has invested heavily in domestic alternatives through SMIC, Huawei and others, with documented progress in mature nodes and reported efforts at leading-edge manufacturing under constraint [34]. The point for this paper is not the rights and wrongs of any particular control regime; it is that the underlying distribution of semiconductor capability is now the object of explicit geoeconomic strategy by the two largest powers, and that this strategy is conducted almost entirely outside the universalist frame of the Geneva Dialogue. A platform that discusses AI governance while the choke point of AI capability is being actively contested by great powers through export controls and industrial policy is a platform that discusses a downstream variable while an upstream variable is being set by actors not at the table.

6.3 Models, data and talent

Frontier model development is similarly concentrated. The Stanford HAI AI Index 2026 documents that the production of notable machine learning models remains heavily dominated by the United States, with China and a small number of European and other actors following [38]. The economics of frontier model development—large compute, large data, large teams—produce a high and rising bar to entry, with the consequence that the set of actors capable of training frontier models is small and largely unchanged in composition. Open-weight and smaller models have widened the set of actors capable of fine-tuning and deploying, which is a real and important democratisation of use, but it is a democratisation of use rather than of frontier capability.

Data is concentrated by language. The largest training corpora are overwhelmingly in English and a small number of other high-resource languages, with the consequence that models perform best in the languages of the data-rich and worst in the languages of the data-poor. This is not merely a question of technical performance; it is a question of epistemic agency. A society whose language is poorly represented in training data is a society whose concerns are poorly represented in the models that mediate its information environment, and whose regulators must govern systems that were not built with their linguistic or cultural context in mind.

Talent is concentrated by geography and by firm. The AI Index 2026 reports that Singapore led in the share of job postings mentioning AI skills (4.69%), followed by Hong Kong, Luxembourg and Spain, while the United States remained a major absolute concentration of AI talent [40]. The same report finds that the number of AI researchers and developers moving to the United States has fallen sharply since 2017, indicating that historical talent concentration remains substantial even as recent inflows have slowed [38]. Compensation differentials, compute access and agglomeration effects reinforce the concentration of frontier research capacity. Talent concentration remains a direct constraint on epistemic and implementation agency: a state with limited domestic access to advanced AI expertise may depend on external institutions to develop the evidence and technical capacity needed for regulation.

6.4 Financing and investment

Financing is the most visible indicator of the AI capability divide and, because it is well measured, the most quotable. The Stanford HAI AI Index 2026 reports that United States private AI investment reached approximately USD 285.9 billion in 2025, more than twenty-three times the approximately USD 12.4 billion invested in China, although the authors note that private-investment figures likely understate China's total AI spending given its government guidance funds [38][39]. In generative AI specifically, US investment exceeded the combined total of China and Europe by a wide margin. Globally, corporate investment in AI more than doubled in 2025, and a growing number of countries—including Japan, South Korea and Italy—passed national AI laws, with more than half of newly adopted national legislative measures concerning AI in that year [38].

These figures should be read with care. Investment is an input, not an output, and the relationship between dollars spent and governance capacity secured is not linear. But investment figures are the clearest available proxy for the resources devoted to building AI capability, and the disparity they reveal is stark. If governing AI requires the resources to build, test and evaluate AI systems, and the resources to build those systems are concentrated in two jurisdictions, then the resources to govern them are, at the margin, also concentrated. The Global Fund for AI proposed by the High-level Advisory Body—intended, in its words, to “put a floor under the AI divide” [15]—was not established by the General Assembly resolution, and its absence is one of the clearest divergences between the Advisory Body’s blueprint and what was actually built.

6.5 Standards and regulatory capacity

Standards and regulatory capacity complete the picture. The OECD AI Policy Observatory tracks more than 900 national AI policies and initiatives, but the distribution of regulatory maturity is highly uneven, with the EU’s AI Act and a handful of national frameworks representing most of the operative regulatory capacity [8] [28]. The Oxford Insights Government AI Readiness Index 2025 ranks 195 countries on their capacity to harness AI for public benefit, and its results consistently show a strong correlation between readiness and income level, with the top of the index occupied by high-income economies and the bottom by low-income and vulnerable states [41]. The World Bank’s synthesis of national AI strategies identified more than 80 countries that had published a strategy by mid-2025, but publication of a strategy and possession of the institutions, budgets and personnel to implement it are different things [27].

The pattern across all nine capabilities—compute, semiconductors, data, models, cloud, talent, standards, financing and regulatory capacity—is consistent: each is concentrated, the concentration is reinforced by agglomeration and network effects, and the gap is widening in some dimensions even as it narrows in others (notably access and use). This is the structural AI governance divide. It is not a binary divide between North and South—China is a frontier actor in compute and models; Singapore, the Republic of Korea, the United Arab Emirates, Saudi Arabia and others have built substantial capacity; Rwanda, India, Brazil and others have built meaningful national strategies. It is a multi-dimensional gradient, and the AUN AGA framework is designed to read it as such. But the gradient is steep, and at the top of each dimension the actors are few. The Geneva Dialogue inherits this gradient; it does not flatten it.

Figure 2. Structural AI Governance Divide

Qualitative synthesis of nine structural capability inputs discussed in Section 6

Not a statistical index; labels summarise the concentration pattern described in the text.

STRUCTURAL INPUT	PATTERN	STRUCTURAL INPUT	PATTERN
Compute	High concentration	AI talent	High concentration
Semiconductors	Very high concentration	Financing	Very high concentration
Data	High / uneven	Standards	Concentrated expertise
Frontier models	Very high concentration	Regulatory capacity	Uneven / concentrated
Cloud	High concentration		

Interpretive note: qualitative synthesis only; no invented numerical scores or country ranking.

Figure 2. Structural AI governance divide: a qualitative synthesis of nine capability inputs discussed in Section 6. The figure is conceptual and non-quantitative; it does not assign empirical scores or estimate a population-weighted Global South average.

7. Global South Representation and Institutional Agency

The Global South is the principal subject of the inclusivity claims made for the Geneva Dialogue, and it is the test on which those claims must be examined. This section argues two propositions that are sometimes treated as in tension but are not. First, the Global South is not a passive recipient of AI governance: its actors are exercising leadership, building domestic and regional policy, and proposing alternative governance approaches. Second, the Global South is not homogeneous: its states differ in national interest, technological capability, regulatory maturity and geopolitical alignment, and a serious analysis must not romanticise it. Both propositions bear on the question of who gets to govern AI, and both must be held together.

7.1 Global South leadership: where agency is already being exercised

The image of the Global South as a passive recipient of AI governance is empirically incomplete and analytically unhelpful. Across multiple dimensions, Global South actors are already exercising agency. At the national level, Rwanda adopted its National AI Policy in April 2023, one of the earliest comprehensive strategies on the African continent, with support from The Future Society [42]; India has built a substantial digital public infrastructure and AI mission; Brazil has developed national AI policy and data-protection enforcement; the United Arab Emirates and Saudi Arabia have invested heavily in AI capacity and host major model developers; Singapore has built one of the world's most digitally sophisticated governments and an AI Safety Institute; the Republic of Korea and Japan have hosted and shaped the AI safety summitry. These are not recipients; they are actors, and several of them have built capabilities that exceed those of many Northern states.

At the regional level, the African Union has advanced a continental approach to AI; ASEAN has developed an AI governance guide; and the G20, under South Africa's 2025 presidency, established a dedicated AI Task Force, placing a Global South state at the helm of the principal economic coordinating forum during a formative period [11][12]. At the international level, the co-chairmanship of the Geneva Dialogue by El

Salvador is a concrete instance of Global South leadership in the universalist architecture; the co-chairmanship of the Scientific Panel by Maria Ressa of the Philippines is another. The presence of Kenya in the inaugural International Network of AI Safety Institutes, alongside the US, UK, EU, France, Japan, Korea, Singapore, Canada and Australia, is a third: a Global South state in the technical safety architecture from its founding [43].

These examples matter because they establish that Global South agency is not a future aspiration but a present fact, and that any framework that treats the Global South as a homogeneous bloc of capacity-deficient recipients is empirically false and politically counter-productive. The Geneva Dialogue did not create Global South agency; it created a space in which existing Global South agency could be more visible. The risk is that the visibility will be mistaken for the underlying capability, and that the Dialogue's universalism will be cited as evidence of a levelling that has not occurred.

7.2 Differences within the Global South: against homogenisation

To say that Global South agency exists is not to say that it is uniform. The Global South is a political category, not an analytical one, and the differences among its members are consequential for governance. Four axes of difference deserve attention. First, technological capability: there is a large gap between China (a frontier actor in compute, models and semiconductors) and an LDC with no domestic AI industry, and grouping both as "Global South" obscures more than it reveals. Second, regulatory maturity: some Global South states have operative AI regulation; many have strategy documents without institutions; the variance within the category is wider than the variance between the category's average and the Global North's average. Third, geopolitical alignment: Global South states are variously aligned with the US, with China, with a non-aligned position, or with regional coalitions, and these alignments shape their positions on export controls, on standard-setting, and on the role of the United Nations. Fourth, priorities: a middle-income Latin American state, a SIDS concerned with sea-level rise, an African state concerned with agricultural productivity, and a Gulf state investing in frontier models have different AI governance priorities, and a universal platform that treats them as a bloc will serve none of them well.

The implication for the Geneva Dialogue is that Global South representation must be read as the representation of diverse actors with diverse interests, not as the representation of a single position. The Dialogue's design, with its universal membership and its multi-stakeholder format, is well suited to surface this diversity; it is less well suited to resolve it, because resolution requires the kind of negotiation the Dialogue is not mandated to conduct. The risk is that the Dialogue produces a lowest-common-denominator universalism that smooths over the real differences among Global South actors and presents the result as a shared Global South position—which would be neither an accurate representation of the Global South nor a useful input to governance.

7.3 Structural vulnerability and the special case of SIDS and LDCs

Within the Global South, a particular subset of states—small island developing states (SIDS), least developed countries (LDCs), and other structurally vulnerable economies—warrants specific attention because their exposure to AI is disproportionate and their capacity to govern it is most constrained. SIDS face AI-related risks that include the disruption of tourism and services labour markets, the bias of imported systems against small and low-resource-language populations, and the use of AI in climate modelling and disaster response that they cannot independently validate. LDCs face risks of data extraction, of dependence on imported models, and of regulatory regimes designed elsewhere being applied to circumstances they did not shape. The participation support fund established for the Dialogue recognises the cost barrier to presence [23]; it does not, by itself, address the capacity barrier to substantive engagement, which is the binding constraint for these states.

The Geneva Dialogue's universalism is, for these states, a double-edged inheritance. It secures them a seat; it does not secure them a voice proportionate to their exposure, because voice scales with capacity and their capacity is most constrained. The AUN framework's contribution is to make this asymmetry explicit and to connect it to a design question: what would a Dialogue look like if it were designed to raise the governing agency of the most constrained states, rather than to record their presence? That design question is taken up in Section 11.

7.4 Alternative governance approaches from the Global South

A final dimension of Global South agency is the generation of alternative governance approaches. Several Global South actors have proposed framings that differ from the dominant risk-and-safety paradigm. These include emphases on data sovereignty, on the rights of affected communities, on the distributional consequences of AI for labour and development, on the role of the public sector in AI provision, and on the relationship between AI and sustainable development. These alternative framings are not uniform—they differ across regions and actors—but they share a tendency to locate AI governance within a development and justice frame rather than a safety and competitiveness frame. The Geneva Dialogue's agenda, structured substantially around the Scientific Panel's seven domains, makes some room for these framings but does not centre them. Whether future sessions can create space for genuinely alternative approaches, or whether the Dialogue will consolidate a single dominant framing in the name of interoperability, is an open question on which the inclusivity of the platform will partly depend.

8. The Role of Science, Standards and Technical Capacity

Science and standards occupy a privileged position in the Geneva Dialogue's design. The Dialogue is explicitly paired with the Independent International Scientific Panel, and the Panel's outputs materially informed the evidence and framing of the inaugural discussions. This pairing reflects a deliberate institutional choice: to build the universalist governance layer on a shared scientific foundation, modelled on the IPCC, rather than on a shared political negotiating text. This section assesses whether the Scientific Panel can meaningfully democratise the evidence base for AI governance, examines the technical standard-setting landscape, and considers the role of technical capacity in determining who can actually govern AI.

8.1 The Independent International Scientific Panel on AI

The Independent International Scientific Panel on Artificial Intelligence was established by the same General Assembly resolution that established the Dialogue, and its preliminary report was published on 1 July 2026, five days before the Dialogue opened [22]. It is co-chaired by Yoshua Bengio of Canada, one of the founders of deep learning, and Maria Ressa of the Philippines, a Nobel Peace Prize laureate and journalist, a pairing that itself signals an ambition to hold scientific authority and human-rights insight together [21][22]. The preliminary report organises its findings across seven domains—AI science and trajectories; societal applications; economic implications; security, systems and environment; human rights, information and democracy; cultural and individual flourishing; and management, governance and reliability [22]—and the Panel's next annual report is intended to inform the second Dialogue in May 2027.

The Panel's preliminary report has real strengths, and contemporary analysis has been generous in acknowledging them. It recognises the structural concentration of AI power, insists that AI is neither inherently good nor bad, and is candid about the disproportionate exposure of the Global South to AI misuse owing to limited mitigation capacity [29][44]. Its framing of risks—spanning human rights, information integrity, child safety, security and the environment—is broad enough to accommodate the concerns of diverse actors, and its explicit design as a shared evidence base for governments and for the Dialogue is a

serious attempt to lower the cost of evidence-based engagement for states that cannot maintain their own equivalent scientific assessments.

Whether the Panel can meaningfully democratise the evidence base, however, depends on five criteria that deserve honest scrutiny. The first is independence: the Panel must be able to assess the state of AI without deferring to the firms whose systems it describes, and its funding and composition must insulate it from capture. The second is geographic diversity: a panel whose active authorship is concentrated in a small number of jurisdictions will, however inadvertently, produce a framing that reflects those jurisdictions' concerns. The third is scientific authority: the Panel must carry sufficient standing that its findings are taken seriously by the technical community, not merely by the diplomatic one. The fourth is accessibility: a report that cannot be read, in usable form, by a developing-country regulator is a report that does not democratise the evidence base in any operational sense. The fifth is policy translation: the Panel must convert scientific findings into inputs that officials can act on, without overstepping into advocacy that would compromise its scientific standing.

On each of these criteria, the preliminary report is promising but incomplete. The Panel's co-chairmanship by Bengio and Ressa addresses scientific authority and human-rights grounding simultaneously; the inclusion of members from across regions addresses geographic diversity in composition, though the distribution of active authorship is harder to assess from outside. Independence is partly structural (the Panel reports to the General Assembly, not to firms) and partly cultural (the norms of the scientific community), and the early record is encouraging. Accessibility and policy translation are the criteria on which the Panel's long-term contribution will most be tested: a 2026 preliminary report is a beginning, not a settled evidence base, and the Panel's capacity to produce shorter, region-specific, and operationally usable derivatives will determine whether it raises the epistemic floor of weaker states or merely produces an authoritative document that those states cannot afford to use. The Chatham House assessment of September 2025 captured the broader dynamic well: the UN's new AI governance architecture is "mostly powerless but, if implemented effectively, could set important global agendas on AI" [45]. The Panel is the principal mechanism by which that agenda-setting could be made substantive.

8.2 Technical standard-setting: the hidden layer of governance

Technical standard-setting is the hidden layer of AI governance and, in operational terms, often the most consequential. Standards determine what counts as a safe system, what a management process must contain, how conformity is assessed, and what buyers can procure. The International Organization for Standardization and the International Electrotechnical Commission, through subcommittee SC 42, have produced a growing family of AI standards, most prominently ISO/IEC 42001 on AI management systems, which is fast becoming the de facto reference for organisational AI governance. The IEEE maintains its own portfolio of AI-related standards. National standards bodies—NIST in the United States with its AI Risk Management Framework, the British Standards Institution, and others—generate reference standards that travel internationally through procurement, certification and supply-chain requirements.

The political economy of standard-setting is distinctive. Standards are written in technical language by technical experts, often in meetings that receive little public or diplomatic attention, and their content is then adopted by procurement officials and regulators who treat them as authoritative. This means that the substantive content of AI governance—what a risk assessment must contain, how bias is measured, what transparency is required—is substantially set in standards bodies rather than in the more visible diplomatic forums. A state without experts in SC 42, in IEEE working groups, or in its own national standards body is a state that has delegated a large share of its technical governance to others, often without knowing it. The High-level Advisory Body recognised this in proposing an AI Standards Exchange as one of its seven recommendations, intended to bring together standards bodies, firms, civil society and the scientific panel to

make standard-setting more transparent and more inclusive [14][15]. That Exchange was not established by the General Assembly resolution, and its absence is a second significant divergence between the blueprint and the architecture as built.

The Geneva Dialogue's relationship to standard-setting is therefore indirect and incomplete. The Dialogue can discuss standards; it can legitimise attention to them; it can encourage its participants to invest in standard-setting capacity. It cannot, as a discussion platform, write standards, and it does not, as designed, formally link to the standards bodies that do. A genuine integration of the Dialogue with the standard-setting system—for example, through a standing interface between the Dialogue and SC 42, or through the AI Standards Exchange the Advisory Body proposed—would be one of the most consequential design improvements available. Without it, the Dialogue risks governing the visible layer of AI governance while the hidden layer continues to be governed by those already in the room.

8.3 Technical capacity and the right to govern

Underlying both the scientific and the standard-setting functions is a deeper condition: technical capacity. A state can participate in a dialogue, accept a scientific report, and adopt a standard, but if it cannot test models, evaluate systems, certify conformity, investigate incidents and train regulators, it cannot govern AI in any operative sense. Technical capacity is an important condition for effective implementation and for some forms of epistemic, standard-setting, regulatory and infrastructural agency. The Geneva Dialogue does not, by itself, create technical capacity; that is the work of national investment, regional cooperation, and the capacity-building mechanisms that were proposed but not built.

The High-level Advisory Body's proposed AI Capacity Development Network and Global Fund for AI were, in this respect, the most important of its seven recommendations—the two that would have built the capabilities on which the other five could have operated [15]. Their non-establishment is the single largest gap between the architecture as designed in *Governing AI for Humanity* and the architecture as built in resolution A/RES/79/325. The Geneva Dialogue is left, in effect, to govern a capability divide it has no instruments to close. This is not a failure of the Dialogue itself; it is a feature of the political compromise that produced it. But it is a feature that any honest assessment must register, and it is a feature that the recommendations in Section 12 seek to address.

Figure 3. Architecture as Built vs. Architecture as Designed

What the High-level Advisory Body proposed, and what resolution A/RES/79/325 established

Governing AI for Humanity (HLAB, Sept 2024) — 7 recommendations

- ✓ 1. International Scientific Panel on AI
- ✓ 2. Policy Dialogue on AI Governance
- ✗ 3. AI Standards Exchange
- ✗ 4. AI Capacity Development Network
- ✗ 5. Global Fund for AI
- ✗ 6. Global AI Data Framework
- ✗ 7. AI Office (UN Secretariat)

As built by A/RES/79/325 (Aug 2025)

- ✓ Scientific Panel on AI — established
- ✓ Global Dialogue on AI Governance — established
- ✗ AI Standards Exchange — NOT established
- ✗ Capacity Development Network — NOT established
- ✗ Global Fund for AI — NOT established
- ✗ Global AI Data Framework — NOT established
- ✗ AI Office — NOT established

Interpretive point: the Panel and Dialogue were established; five proposed mechanisms with stronger resource, capacity, standard-setting or institutional-distribution effects were not established.

Source: this paper, comparing HLAB (2024) and UNGA A/RES/79/325 (2025).

Figure 3. Architecture as designed (HLAB, 2024) versus architecture as built (UNGA A/RES/79/325, 2025): selective uptake of the seven recommendations. Five proposed mechanisms with stronger resource, capacity, standard-setting or institutional-distribution effects were not established.

9. Fragmentation of the International AI Governance Landscape

The Geneva Dialogue enters an AI governance landscape that is already crowded and partly competitive, and the question of whether it reduces fragmentation or adds another layer to it is one of the most important judgements to be made about its institutional contribution. This section maps the principal mechanisms, considers their overlap and competition, and assesses the Dialogue's positioning among them.

9.1 The crowded field

The international AI governance landscape in 2026 contains, at minimum, the following elements. At the universal level: the UN General Assembly's resolutions on AI (including resolution 78/241 of March 2024 on safe, secure and trustworthy AI); the Global Digital Compact; the Geneva Dialogue and its sister Scientific Panel; UNESCO's Recommendation on the Ethics of AI, applicable to all 194 member states [6][7]; the ITU's AI for Good and WSIS processes; and the human-rights machinery of the Office of the High Commissioner for Human Rights. At the regional and club level: the OECD AI Principles and AI Policy Observatory [8]; the Council of Europe Framework Convention on AI, the first international legally binding treaty specifically devoted to AI governance, not yet in force as of 11 September 2026 [9][10]; the EU AI Act; the G7 Hiroshima Process Code of Conduct; and the G20 AI Task Force under South Africa's 2025 presidency [11][12]. At the technical level: ISO/IEC SC 42 and ISO/IEC 42001; the IEEE; and national standards bodies. At the safety and summit level: the Bletchley Declaration, the Seoul Summit, the Paris AI Action Summit of February 2025, and the International Network of AI Safety Institutes (the US, UK, EU, France, Japan, Korea, Singapore, Canada, Australia and Kenya) launched in San Francisco in November 2024 [43][46]. At the industry level: the Frontier

Model Forum, the voluntary commitments of frontier developers, and the safety institutes that, in 2025, were reconfigured—the UK's becoming the AI Security Institute and the US's becoming the Center for AI Standards and Innovation [46].

This is not a comprehensive list, but it suffices to make the structural point: there is no shortage of AI governance mechanisms. There is, rather, a surplus, and the surplus is not well coordinated. Each mechanism has its own membership, its own mandate, its own pace, and its own institutional culture. The OECD is a club of mostly high-income democracies; the Council of Europe treaty is regional but open to non-member signatories; the EU AI Act is a regulation with extraterritorial reach; the G7 and G20 are leader-level groups with rotating memberships; the safety institutes are a technical network of a small number of states; the standards bodies are technical and industry-heavy; and the Geneva Dialogue is universal but non-binding. The result is a regime complex—a set of overlapping, partly hierarchical, partly horizontal institutions—rather than a single coherent regime.

9.2 Does the Dialogue reduce fragmentation or add a layer?

Against this background, the Geneva Dialogue can be read in two opposing ways. The optimistic reading is that the Dialogue is the universalist hub that the regime complex has been missing: a platform at the centre of the system, around which the other mechanisms can coordinate, and through which the states not represented in the clubs (OECD, G7, the safety institute network) can engage with the agenda those clubs have set. On this reading, the Dialogue reduces fragmentation by providing a universal convening point. The pessimistic reading is that the Dialogue is a further addition to an already crowded field, that it duplicates functions already performed by UNESCO, the ITU and the OECD, and that, lacking binding authority, it will be bypassed by the actors (the EU, the US, the major firms) whose decisions actually move AI. On this reading, the Dialogue adds a layer.

Which reading is more accurate depends on how the Dialogue develops over its next several sessions. Three conditions would have to be met for the optimistic reading to hold. First, the Dialogue would have to establish itself as the principal universal forum for AI governance, displacing or absorbing competing universalist claims—a claim that UNESCO, with its 194-member ethics recommendation, might contest. Second, the Dialogue would have to develop genuine interfaces with the binding and technical layers of the regime complex—the EU AI Act, the Council of Europe treaty, the standards bodies, the safety institute network—so that discussion in Geneva feeds into decision elsewhere and decision elsewhere is informed by discussion in Geneva. Third, the Dialogue would have to maintain its inclusivity without collapsing into a lowest-common-denominator universalism that satisfies no one. If these conditions are met, the Dialogue reduces fragmentation; if they are not, it adds a layer.

The early evidence is mixed. The coordination with the ITU through the Geneva venue, the pairing with the Scientific Panel, and the alternation with New York are consistent with the optimistic reading. The absence of formal interfaces with the EU AI Act, the Council of Europe treaty and standards bodies, and the absence of the Standards Exchange and Capacity Network, are consistent with the more cautious reading. The Co-Chairs' Summary, released in September 2026, explicitly identifies coherence, complementarity and interoperability among existing initiatives as areas of emerging convergence, but it does not itself create formal institutional interfaces. The honest assessment is therefore that the Dialogue is a candidate hub whose institutional capacity remains to be demonstrated. The CIC NYU brief of January 2026 proposed an integrated, sequenced, UN-anchored roadmap for the Dialogue precisely because deliberate sequencing would be needed to prevent another fragmented layer [47].

9.3 The US question: governance under superpower dissent

One further feature of the landscape deserves explicit treatment: the attitude of the United States. At a UN Security Council open debate on 24 September 2025, the day before the high-level launch of the Dialogue, the United States stated plainly that it “totally reject[s] all efforts by international bodies to assert centralized control and global governance of AI” [48][49]. This is a striking position from one of the two principal AI powers, and it bears directly on the Dialogue’s prospects. A universalist platform built on consensus is weakened when a superpower with the largest AI industry publicly opposes the project of global AI governance. The Dialogue does not require US endorsement to exist, but its capacity to set global agendas is constrained by the unwillingness of the largest actor to be set. This is the institutional expression of the deeper asymmetry this paper has tracked: the actors whose behaviour most needs to be governed are not the actors whose participation is most easily secured, and a platform built on universal participation will always struggle to discipline those who attend but do not consent.

The US position should not be read as a reason to abandon universalist AI governance; it should be read as a reason to be clear-eyed about what universalist platforms can and cannot do under superpower dissent. They can build norms, they can socialise states, they can produce evidence, they can include the excluded; they cannot, without the largest actor, produce binding rules that the largest actor will follow. The Dialogue’s value in this environment is precisely its capacity to do what does not require US consent—to raise the epistemic floor of the rest of the world, to legitimise alternative framings, to build coalitions—rather than to do what does. The framing matters: if the Dialogue is judged by its capacity to produce a globally binding regime, it will fail; if it is judged by its capacity to shift the distribution of governing agency among the willing, it may yet succeed.

10. The Geneva Dialogue’s Institutional Strengths and Limitations

Assessing the Dialogue requires a set of criteria appropriate to what it is—a discussion platform, not a negotiating forum. This section applies an institutional test organised around nine measurable criteria: influence, inclusiveness, agenda-setting, evidence quality, capacity-building, interoperability, implementation, accountability, and continuity. The criteria are applied with a clear understanding that the Dialogue has held only one session and that judgement must be provisional; the purpose is to establish a framework for evaluation that can be applied to future sessions as well as to the inaugural one.

10.1 Strengths

The Dialogue’s principal strengths are four. First, its universal mandate gives it a legitimacy that no club, region, or technical body possesses: it is a UN platform in which every Member State is entitled to participate. Second, its pairing with the Scientific Panel gives it a shared evidence base of a kind that most international platforms lack at inception, and the Panel’s design as an independent global scientific body is a serious attempt to pool epistemic agency internationally. Third, its multi-stakeholder format places governments, firms, academics and civil society in the same deliberative space, which is appropriate to a technology whose substantive power is widely distributed across these actors. Fourth, its co-chair formula—El Salvador and Estonia—provides geographic and institutional diversity, while its annual cadence, alternating between Geneva and New York, provides a structure for sustained engagement.

10.2 Limitations

The Dialogue’s limitations are the mirror image of its strengths, and they are five. First, its mandate is bounded by design: it can discuss, share and facilitate, but it is not a negotiating forum and does not itself

exercise binding decision-making authority. Second, its agenda-setting is distributed across the preparatory process: the Scientific Panel shapes the evidence base, while the secretariat, Member States, consultations and stakeholder submissions shape the agenda and programme. Universal participation in deliberation therefore does not necessarily translate into equal influence over framing. Third, its multi-stakeholder format exposes asymmetries of capacity between major firms and small states, and the speaking format can be poorly suited to the development of shared positions. Fourth, its design does not include the capacity-building and resource-transfer mechanisms proposed by the High-level Advisory Body, such as a Global Fund for AI and Capacity Development Network. Fifth, its relationship to the binding and technical layers of the regime complex remains relatively light. The September 2026 Co-Chairs' Summary calls for greater coherence, complementarity and interoperability among initiatives, but no formal interface with the EU AI Act, the Council of Europe treaty or standards bodies is established by the Dialogue itself [50].

10.3 The institutional test: nine criteria applied

Applied to the inaugural session and the design as built, the nine criteria yield the following provisional assessment, summarised in Table 2.

Criterion	Provisional assessment	Principal constraint
Influence	Low to moderate; indirect, via norms and the Scientific Panel	Non-binding mandate; superpower dissent
Inclusiveness	High numerically; uneven substantively	Capacity asymmetries; framing upstream
Agenda-setting	Shared but upstream; Co-Chairs, secretariat, consultations, submissions and the Scientific Panel shape the agenda and framing	Participants do not independently control the agenda they deliberate
Evidence quality	Promising (Scientific Panel), unproven over time	Geographic diversity of authorship; accessibility
Capacity-building	Not yet delivered; instruments absent	No Capacity Development Network or Global Fund established by A/RES/79/325
Interoperability	Potential not yet realised	No formal institutional interface established with EU AI Act, treaty or standards bodies
Implementation	Outside the Dialogue's mandate	Depends on national and regional institutions
Accountability	Outside the Dialogue's mandate	Depends on courts, regulators, civil society

Criterion	Provisional assessment	Principal constraint
Continuity	Established (annual, Geneva/New York)	Sustained political support and financing

Table 2. Provisional institutional assessment of the Geneva Dialogue against nine criteria.

The pattern is clear. The Dialogue is strongest on the dimensions that universalism can directly deliver—broad participation and continuity—and weaker on dimensions that require decision authority, resource transfer, formal linkage or sustained capability. The Co-Chairs’ Summary released on 8 September 2026 is consistent with the paper’s central diagnosis: it identifies persistent divides in infrastructure, computing power, data, skills, and technical and institutional capacity, while stressing that capacity—not access alone—must be built. Because the Summary records emerging convergence rather than a negotiated outcome, it is treated here as corroborating post-session evidence, not as proof of institutional impact [50].

10.4 Study limitations and inference boundaries

This study has five principal limitations. First, it analyses one institutional episode—the inaugural 2026 Dialogue—so it cannot establish how the mechanism will perform over time. Second, the paper uses capability indicators such as investment, compute, talent, standards participation and regulatory capacity as proxies for governing agency; these are analytically informative but are not equivalent to a direct measure of political influence. Third, the UN has not published a complete verified breakdown of participants by region, income group, SIDS/LDC status or other vulnerability classifications, which limits quantitative claims about representational equality. Fourth, some international AI governance mechanisms use different definitions, counting conventions and reporting periods, so cross-source comparisons are interpreted cautiously rather than mechanically pooled. Fifth, the AGA framework is conceptual and has not yet undergone independent empirical validation. These limitations bound the strength of the paper’s inferences: the findings support a structured institutional diagnosis and a set of testable propositions, not a causal estimate of the Dialogue’s effect on global AI governance.

The appropriate next step is longitudinal validation. Future sessions can test whether participation produces measurable changes in the seven AGA capabilities, whether the Dialogue creates durable interfaces with standards and regulatory institutions, whether capacity-building reaches under-resourced states, and whether its recommendations or discussions demonstrably alter decisions elsewhere in the governance regime complex. This is also the basis on which the framework can be falsified, refined or rejected.

11. An AUN Framework for Genuinely Inclusive AI Governance

The preceding sections establish a diagnosis: the Geneva Dialogue has built a universalist deliberative layer that succeeds at participation but does not, by itself, create the capabilities required for effective governing agency. The question this section addresses is constructive: what would an AI governance system look like if meaningful governing agency—not merely participation—were the design principle? The answer proposed here is an AUN framework that takes the seven AGA capabilities as its design target and treats the Dialogue not as a stand-alone platform but as the convening layer of a wider, capability-oriented architecture. The framework is offered as a contribution to the design conversation, not as a blueprint to be adopted

wholesale; its purpose is to make the design question explicit and to give future sessions of the Dialogue and its participants a vocabulary in which to ask it.

11.1 From universal participation to universal agency: the design inversion

The standard design logic of inclusive international governance is to secure universal participation first and to hope that capability follows. The logic is understandable—presence is achievable, capability is not—but it inverts the actual relationship between the two. Participation without capability can become primarily legitimating rather than substantively empowering when the consequential decisions are made elsewhere by actors with greater capability. The AUN framework proposes an inversion of the standard logic: to design for the distribution of capability first, and to treat participation as the consequence rather than the precondition of inclusive governance. On this inverted logic, the question is not “how do we get everyone into the room?” but “what would have to be true for everyone in the room to matter?”

The inverted logic has three design implications. First, the Dialogue’s success would be measured not by attendance but by the change in the capability profiles of its participants across the seven AGA dimensions over time. Second, the Dialogue’s agenda would be designed to surface and build capability, not merely to record positions, which would mean a different format from the three-minute-intervention queue. Third, the Dialogue would be connected to the capability-building instruments—a Capacity Network, a Global Fund, a Standards Exchange—that turn discussion into capacity. None of these requires the Dialogue to become a negotiating forum; all of them require it to become more than a discussion platform. The inversion is from a platform that records the world as it is to a platform that changes the distribution of the capabilities with which the world governs AI.

11.2 The four pillars of a capability-oriented architecture

Drawing the design implications together, the AUN framework proposes four pillars of a capability-oriented architecture built around, but not limited to, the Geneva Dialogue. The pillars correspond to the capabilities the Dialogue cannot create on its own and the linkages it has not yet built.

Pillar 1—Epistemic infrastructure. Anchored on the Independent International Scientific Panel but extended through a network of regional and national scientific assessments, regional hubs, and operationally usable derivatives (short briefs, translated summaries, policy-ready syntheses) that lower the cost of evidence-based engagement for states that cannot maintain their own assessments. The test is whether a regulator in a low-capacity state can, within hours, obtain an authoritative and usable read on an AI risk relevant to her jurisdiction. If she cannot, the epistemic infrastructure has not democratized the evidence base; it has produced an authoritative document she cannot use.

Pillar 2—Standard-setting interface. A standing interface between the Dialogue and the technical standard-setting system (ISO/IEC SC 42, IEEE, national standards bodies), in the form the High-level Advisory Body’s proposed AI Standards Exchange envisaged, with funded participation for developing-country experts and with a mandate to translate diplomatic concerns into standard-setting inputs and to translate standards into diplomatic understanding. The test is whether a concern raised in Geneva can travel into a standards working group, and whether a standard adopted in a working group can be understood and contested in Geneva. Without such an interface, the Dialogue governs the visible layer while the hidden layer is governed elsewhere.

Pillar 3—Capability transfer. The establishment, in some form, of the capacity-building instruments the Advisory Body proposed—an AI Capacity Development Network and a Global Fund for AI—so that participation connects to capability through funded regional and national capacity-building, model

evaluation facilities accessible to under-resourced states, and regulatory training. The test is whether a state that attends the Dialogue in year n has measurably greater implementation agency in year $n+3$. If the Dialogue cannot point to such changes, its inclusivity is procedural rather than substantive.

Pillar 4—Interoperability and accountability interfaces. Formal, light-touch interfaces between the Dialogue and the binding and technical layers of the regime complex—the EU AI Act, the Council of Europe treaty, the safety institute network, and the UNESCO Recommendation—so that discussion in the universal platform informs decision in the binding and technical ones, and so that the experience of the binding and technical layers informs the universal discussion. The test is whether the Dialogue becomes, over time, the place where the fragmentation of the regime complex is made visible and partly managed, rather than a further fragment of it.

The four pillars do not require the Dialogue to become a governing institution; they require it to become the convening and connecting layer of a wider capability-oriented architecture. This is a modest and achievable ambition, and it is the ambition that follows most directly from the diagnosis in Sections 5 through 10. It is also the ambition that would let the Dialogue answer, in its own practice, the question that gives this paper its title.

11.3 The capability gradient as design constraint

The AUN framework's final design implication is a recognition that capability is distributed on a gradient, not in a binary, and that a capability-oriented architecture must therefore be gradient-aware. A single programme for "the Global South" will underserve the most capable Global South actors (China, India, Brazil, Singapore, the Gulf states) and overwhelm the least capable. A gradient-aware design would differentiate: frontier-capable states might contribute to the Scientific Panel and to standard-setting; middle-capability states might host regional hubs and build implementation infrastructure; structurally vulnerable states might receive concentrated capacity investment and dedicated epistemic support. The differentiation is not a hierarchy of worth; it is a matching of instrument to need, and it is the only design that can avoid the simultaneous failure of over- and under-support that binary designs produce.

Gradient-awareness also implies that the AUN AGA framework should be used diagnostically by the Dialogue itself. A periodic capability profile of participating states—not as a ranking but as a self-assessment of strengths and gaps across the seven capabilities—would let the Dialogue track whether its sessions are shifting the distribution of agency, and would let member states identify, peer-to-peer, where cooperation would raise capability fastest. The framework this paper proposes is, in this sense, a tool for the Dialogue's own accountability, not merely an external assessment of it.

12. Policy Architecture and Recommendations

The recommendations that follow are concrete and addressed to specific actors at five levels: global, regional, national, technical and scientific, and civil-society. They are designed to be implementable within the existing institutional architecture—they do not require the Geneva Dialogue to become a negotiating forum, and they do not depend on the agreement of actors who have publicly declined to participate in global governance. They are sequenced: the global recommendations create the frame, the regional and national recommendations build the capability, the technical recommendations connect the layers, and the civil-society recommendations hold the system to account. Each recommendation is linked to the AGA capability it would strengthen.

12.1 Global-level recommendations

Recommendation G1—Publish a verified participation census. The co-chairs and secretariat of the Geneva Dialogue should publish, before the second session, a verified statistical breakdown of state and stakeholder participation in the inaugural session, by regional group, developing-country status, SIDS and LDC status, and stakeholder category. A platform whose legitimacy rests on inclusivity must be able to document it, and the absence of such a census is an accountability gap that should be closed at low cost. (Strengthens inclusiveness and accountability.)

Recommendation G2—Establish the AI Standards Exchange. The General Assembly should establish the AI Standards Exchange proposed in Governing AI for Humanity, as a standing interface between the Dialogue, the Scientific Panel, ISO/IEC SC 42, the IEEE, national standards bodies, firms and civil society, with funded participation for developing-country experts. This is the single highest-leverage design improvement available, because it connects the universalist layer to the hidden layer where much of AI governance is actually set. (Strengthens standard-setting agency and interoperability.)

Recommendation G3—Establish a modest Global Fund for AI capacity. The General Assembly should establish a modest, initially voluntary, Global Fund for AI capacity, modelled on the Advisory Body's proposal, focused on the lowest-capability states and on the implementation and epistemic capabilities that the Dialogue cannot otherwise build. The fund should be independent in governance, transparent in operation, and focused on capability rather than on attendance. (Strengthens implementation and epistemic agency.)

Recommendation G4—Make the Scientific Panel's outputs operationally usable. The Scientific Panel should commit to producing, alongside its annual report, shorter regional and thematic briefs, translated and policy-ready, that a developing-country regulator can use within hours rather than days. The Panel's democratising effect depends on accessibility, and accessibility is a function of format as much as of content. (Strengthens epistemic agency.)

Recommendation G5—Institutionalise interfaces with binding and technical instruments. The Dialogue should establish light-touch, formal interfaces with the EU AI Act regime, the Council of Europe Framework Convention, the International Network of AI Safety Institutes, and the UNESCO Recommendation, so that the universal discussion informs and is informed by the binding and technical layers. (Strengthens interoperability and influence.)

12.2 Regional-level recommendations

Recommendation R1—Build regional epistemic and evaluation hubs. Regional organisations (the African Union, ASEAN, the OAS, the Arab states' structures, and the Pacific Islands Forum) should establish or strengthen regional hubs for AI evidence, model evaluation and regulatory cooperation, hosted where capacity exists and networked to the Scientific Panel. Regional hubs are the most efficient scale for capacity-building and the most credible scale for the expression of regional priorities. (Strengthens epistemic, standard-setting and implementation agency.)

Recommendation R2—Develop regional regulatory interoperability. Regional organisations should develop interoperability frameworks among their members' AI regulations, so that a state adopting a national rule can do so in a way that is compatible with its neighbours'. The EU AI Act demonstrates the value of a regional regulatory bloc; other regions can build interoperability without uniformity. (Strengthens regulatory and standard-setting agency.)

Recommendation R3—Use the G20 and regional summits to feed the Dialogue. The G20, the G7, and regional summits should be used as feeders into the Geneva Dialogue, with their outputs explicitly carried into the universal discussion, so that the clubs do not set the agenda in parallel to the universal platform. (Strengthens agenda-setting and interoperability.)

12.3 National-level recommendations

Recommendation N1—Invest in implementation, not just strategy. States with published national AI strategies but limited implementation institutions should prioritise the latter—regulatory authorities, sandbox capacity, model evaluation, regulator training—over the production of further strategy documents. A strategy without implementation is a declaration; the AGA framework prioritises the capability that makes strategy operative. (Strengthens implementation and regulatory agency.)

Recommendation N2—Fund participation in technical standard-setting. States should fund sustained participation of their experts in ISO/IEC SC 42, the IEEE and their own national standards bodies, treating standard-setting as core AI governance rather than as a technical sideline. The hidden layer of governance is governed by those who show up to write the standards. (Strengthens standard-setting agency.)

Recommendation N3—Build accountability institutions. States should invest in the institutions of accountability agency—courts with AI competence, regulators with investigation capacity, independent audit, and civic space for civil-society documentation of harm—because these are the institutions through which affected people become governing agents rather than subjects. (Strengthens accountability agency.)

12.4 Technical and scientific recommendations

Recommendation T1—Open the standards process. Standards bodies should make their working processes more transparent and more accessible to non-industry stakeholders, with funded participation for under-represented states and civil-society experts. Standards are too important to be governed only by those who can afford to attend. (Strengthens standard-setting agency and inclusiveness.)

Recommendation T2—Build shared evaluation infrastructure. The International Network of AI Safety Institutes and the Scientific Panel should cooperate to build shared, accessible model-evaluation infrastructure that under-resourced states can use, so that the capacity to test frontier models is not confined to a small number of states. (Strengthens implementation and epistemic agency.)

The deeper point, and the one AUN most wishes to register, is that the governance of transformative technology is decided less in the rooms where it is discussed than in the upstream spaces where its evidence, standards, infrastructure and implementation are built. A global dialogue is a necessary condition of inclusive AI governance; it is not a sufficient one. Sufficient conditions require the redistribution of the capabilities with which governing is done. There is, however, an early sign that the institutional architecture may begin to develop the linkages identified in this paper: on 10 September 2026, the President of the eighty-first session of the General Assembly described efforts to strengthen regular communication among the Scientific Panel, the Dialogue's Co-Chairs, the Global Digital Compact review process, interested Member States and technology companies ahead of the 2027 Dialogue [51]. That development should be read as an institutional signal, not as evidence that interoperability has already been achieved. Whether the Geneva Dialogue becomes a beginning of inclusive governance, or only a beginning of inclusive discussion, will depend on whether such coordination is converted into durable, measurable capability and accountability mechanisms.

12.5 Civil-society recommendations

Recommendation C1—Fund civil-society capacity for AI governance. Donors should fund sustained civil-society capacity to document AI harms, represent affected communities, and engage with the Dialogue and the standard-setting system. Civil society is the principal conduit through which affected people become visible to governance. (Strengthens accountability agency.)

Recommendation C2—Create an affected-communities interface. The Dialogue should establish a formal interface for affected communities—workers, communities subject to surveillance, users of low-resource languages, children—so that the people on whom AI acts are not represented only through intermediaries. Universal participation of states does not amount to representation of affected people; a dedicated interface begins to close that gap. (Strengthens representation and accountability agency.)

Recommendation C3—Independent monitoring of the Dialogue. An independent civil-society consortium should monitor the Dialogue against the nine institutional criteria set out in Section 10, publishing an annual assessment that holds the platform to its own claims. A platform whose legitimacy rests on inclusivity should be assessed on whether its inclusivity is substantive, and that assessment should be independent of the platform itself. (Strengthens accountability and continuity.)

13. Conclusion

The question this paper set out to answer was: who gets to govern artificial intelligence? The Geneva Dialogue was the case through which the question was examined, but the answer does not lie in the Dialogue alone. It lies in the distribution of the capabilities that make governing possible, and in the institutional design that connects participation to those capabilities.

The evidence assembled here supports, with qualifications, the proposition with which the paper began. A central problem of global AI governance is not simply that some countries are underrepresented in international discussions. It is that the capabilities required to shape AI's rules, evidence, infrastructure, standards and implementation remain unevenly distributed, so that universal participation does not automatically produce universal governing agency. The Geneva Dialogue has created a universal deliberative layer, but it does not by itself redistribute representation, influence, agenda-setting, decision-making, implementation or technological power. The seven capabilities of the AUN AI Governance Agency framework—epistemic, normative, standard-setting, infrastructural, regulatory, implementation and accountability—are distributed unevenly, and the actors with the strongest capability profiles remain relatively few. The Dialogue inherits this distribution; it does not automatically flatten it.

The qualifications matter. The proposition is not that universalism is worthless; it is that universal participation is necessary but insufficient for substantive governing agency. The proposition is not that the Dialogue is a failure; it is that it is a candidate hub whose contribution depends on whether emerging commitments to coherence, interoperability and capacity are converted into durable institutional linkages. The proposition is not that the Global South is powerless; it is that Global South agency is real but uneven, and that a universal platform that treats it as uniform will serve no one. The proposition is not anti-technology or anti-Western; it is institutional, and it is directed at the gap between what platforms claim and what capabilities allow.

So, who gets to govern AI? At present, AI is governed, in the operative sense, by a small set of actors: the states that control the choke points of compute and semiconductors; the firms that build frontier models and cloud infrastructure; the standards bodies that write the technical rules; the regulators (disproportionately in a few jurisdictions) that issue binding law; and the scientific institutions that set the evidence base. The Geneva Dialogue, for all its universalist ambition, does not yet alter this set; it adds a universal deliberative layer on top of it. The question is whether the Dialogue can be designed to alter the set—to raise the governing agency of those presently outside it, to discipline the concentrated power presently within it, and to connect the universal discussion to the binding and technical decisions that actually move AI.

The answer this paper offers is that it can, but only if the design is inverted from a logic of universal participation to a logic of universal agency. That inversion requires the four pillars set out in Section 11—an epistemic infrastructure, a standard-setting interface, a capability transfer mechanism, and interoperability and accountability interfaces—and the concrete recommendations set out in Section 12. None of these requires the Dialogue to become a negotiating forum; all of them require it to become more than a discussion platform. The test is whether, in the sessions ahead, the Dialogue can point to measurable changes in the capability profiles of its participating states and in the behaviour of the concentrated powers around it. If it can, it will have answered its own question, and the answer will be that AI is governed, increasingly, by those who were once only present. If it cannot, the Dialogue will remain what it is at its inauguration: a universal layer on a concentrated architecture, and a largely legitimating layer of inclusion in a system whose governing capabilities remain unevenly distributed.

The deeper point, and the one AUN most wishes to register, is that the governance of transformative technology is decided less in the rooms where it is discussed than in the upstream spaces where its evidence, standards, infrastructure and implementation are built. A global dialogue is a necessary condition of inclusive AI governance; it is not a sufficient one. Sufficient conditions require the redistribution of the capabilities with which governing is done. The Geneva Dialogue's inaugural session was a serious and creditable beginning; whether it becomes a beginning of inclusive governance, or only a beginning of inclusive discussion, depends on the willingness of its member states and its secretariat to design for agency rather than for attendance. That willingness, more than the communiqués of any single session, will determine who gets to govern artificial intelligence in the years ahead.

References

Sources are listed in the order of first in-text citation. Where a source was consulted online, the direct source is identified. This manuscript was editorially verified against sources available through 15 September 2026; the evidence cutoff is 14 September 2026. The evidence cutoff is the date-lock for substantive claims, while the 15 September verification records the final pre-publication check. The paper does not attribute to any source a claim that the source does not support.

- [1] United Nations, "Geneva / Global Dialogue on AI Governance Opening," UNifeed Media, 6–7 July 2026.
- [2] United Nations, Secretary-General António Guterres, address to the inaugural Global Dialogue on AI Governance, Geneva, 6 July 2026 (via UNifeed).
- [3] UNESCO, "UN Global Dialogue opens with urgent call for safe and inclusive AI that benefits all," 6 July 2026.
- [4] United Nations, Global Digital Compact / Pact for the Future, September 2024; and UN, "AI Panel and Dialogue," Global Digital Compact portal: "Resolution A/RES/79/325 Adopted. On 26 August 2025, the United Nations General Assembly established the Independent International Scientific Panel on AI ..."
- [5] Digital Watch Observatory (DiploFoundation), "UNGA adopts terms of reference for AI Scientific Panel and Global Dialogue," 27 August 2025.
- [6] UNESCO, Recommendation on the Ethics of Artificial Intelligence, adopted November 2021; UNESCO, "Recommendation on the Ethics of Artificial Intelligence," 2024.
- [7] United Nations News, "193 countries adopt first-ever global agreement on the Ethics of Artificial Intelligence," 25 November 2021.

- [8] OECD, OECD AI Policy Observatory (over 1,000 AI policies and initiatives); OECD AI Principles (2019, updated 2024).
- [9] Council of Europe, Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, opened for signature 5 September 2024.
- [10] Council of Europe Treaty Office, Chart of Signatures and Ratifications, CETS No. 225, status as of 12 September 2026 (entry into force requires five ratifications including at least three Council of Europe member States; the Convention had not yet entered into force).
- [11] Atlantic Council, "The G20 is moving forward on global AI governance—and ...," 2 December 2025.
- [12] G20 (South Africa 2025 presidency), AI Task Force on Artificial Intelligence, Data Governance and Innovation for Sustainable Development; see also SAI20, "AI Analysis | South Africa 2025."
- [13] United Nations, Secretary-General's High-level Advisory Body on Artificial Intelligence, Governing AI for Humanity: Final Report, September 2024.
- [14] American National Standards Institute (ANSI), "UN Releases Proposed Framework for Global AI Governance," 23 September 2024 (Governing AI for Humanity, AI Standards Exchange recommendation).
- [15] United Nations, Governing AI for Humanity, September 2024, "Recommendation 5: Global fund for AI ... to put a floor under the AI divide"; Global Policy / Digital Rights, "Governing AI For Humanity: Our thoughts on the Final Report," 19 September 2024.
- [16] United Nations, Global Digital Compact, adopted at the Summit of the Future, September 2024 (part of the Pact for the Future, A/RES/79/1).
- [17] United Nations, "Global Dialogue on AI Governance," official portal.
- [18] Center on International Cooperation, New York University, "Guidance for the New Global Dialogue on AI Governance," January 2026 (confirms co-chairs Egriselda López and Rein Tammsaar).
- [19] International Telecommunication Union (ITU), "Global Dialogue on AI Governance, Geneva, 6–7 July," 2 June 2026 (convening alongside WSIS Forum 2026 and AI for Good).
- [20] United Nations, "Save the Date: Global Dialogue on AI Governance 2027," 3–4 May 2027, New York, back-to-back with the 11th Multi-stakeholder STI Forum.
- [21] Mila / UN Web TV, "The UN Global Dialogue on AI Governance Explained," interview with co-chairs of the Scientific Panel on AI, Yoshua Bengio and Maria Ressa; UN Web TV, "Launch of the preliminary report by the Independent International Scientific Panel on AI," 1 July 2026.
- [22] United Nations, "Preliminary Report | Independent International Scientific Panel on AI," published 1 July 2026 (seven domains).
- [23] United Nations, "Participation Support for the AI Dialogue (Geneva 2026)," with co-chairs' letter encouraging voluntary contributions.
- [24] European External Action Service (EEAS), "Advancing Global Cooperation on Artificial Intelligence," 16 July 2026 (reporting representatives from more than 170 countries and over 4,200 participants).
- [25] Global Partners Digital (Digital Rights), "Civil society read-out report: 1st United Nations Global Dialogue on AI Governance," 27 July 2026.

- [26] United Nations, "Programme (2026) | Global Dialogue on AI Governance" (high-level segment, thematic sessions, three-minute interventions, closing plenary).
- [27] R. Radu et al., "The Global Landscape of National AI Strategies," World Bank (more than 80 countries publishing national AI strategies between January 2017 and June 2025).
- [28] OECD, Digital Government Outlook 2026 (30 of 36 OECD countries, 83%, have at least one institution responsible for governing AI).
- [29] Global Issues / IISD, "UN Artificial Intelligence Panel Launches Report Ahead of ...," 2 July 2026 (Global South disproportionately exposed to AI misuse due to limited mitigation capacity).
- [30] Statista, "Data centers worldwide by territory, 2026," 25 June 2026 (approximately 4,423 data centres in the United States as of 14 April 2026).
- [31] Synergy Research Group, "Q2 Cloud Market Passes \$143 Billion; Highest Growth Rate in Eight Years," 30 July 2026 (Amazon 28%, Microsoft 20%, Google 15% of worldwide cloud infrastructure service spending in Q2 2026; combined 63%).
- [32] Goldman Sachs Research, "AI to drive 165% increase in data center power demand by ...," 4 February 2025 (50% rise by 2027, up to 165% by end of decade).
- [33] US International Trade Administration, "Taiwan — Semiconductors including chip design for AI," Country Commercial Guide (Taiwan: over 60% of global foundry revenue, over 90% of leading-edge chip manufacturing; TSMC dominant; Apple, NVIDIA, AMD reliance).
- [34] Mordor Intelligence, Taiwan Semiconductor Market Size, Share & 2025–2030 Outlook (TSMC mass production of 2 nm in late 2025; >90% of leading-edge chips).
- [35] Semiconductor Industry Association (SIA), "2026 State of the Industry Report: Historic Growth Amid Intensifying Global Competition," 27 July 2026 (global semiconductor sales of USD 795.6 billion in 2025).
- [36] World Semiconductor Trade Statistics (WSTS), "Global Semiconductor Market records exceptional growth in Q2 2026," 2026 (full-year 2026 calculation of approximately USD 1.655 trillion based on actual Q2 results and the Spring 2026 forecast assumptions).
- [37] US export controls on advanced semiconductors to China: initial measures October 2022, tightened October 2023 and subsequently; documented in industry and policy reporting (e.g., TaiwanPlus/Reuters coverage, October 2022).
- [38] Stanford HAI, The 2026 AI Index Report, including the Research & Development, Responsible AI and Economy chapters (US private AI investment, model development, compute/data-centre concentration, talent flows, and language-performance disparities).
- [39] Stanford HAI, The 2026 AI Index Report, Economy chapter (US generative AI investment exceeds combined China and Europe); IEEE Spectrum, "The Stanford AI Index Report 2026."
- [40] Lightcast / Stanford HAI, The 2026 AI Index Report (Singapore 4.7% of postings mentioning AI skills, followed by Hong Kong 3.5%, Luxembourg 3.4%, Spain 3.3%).
- [41] Oxford Insights, Government AI Readiness Index 2025 (ranks 195 countries).

[42] The Future Society, "Cabinet of Rwanda approves National AI Policy," 29 May 2023 (Rwanda National AI Policy approved 20 April 2023).

[43] ANSI, "U.S. Launches International AI Safety Network with Global Partners," 25 November 2024 (inaugural members: US, Australia, Canada, EU, France, Japan, Kenya, Republic of Korea, Singapore, UK); CSIS, "The AI Safety Institute International Network."

[44] Tech Policy Press, "The UN Scientific Panel on AI's Preliminary Report ...," 8 July 2026 (report recognises structural concentration of AI power; AI neither inherently good nor bad).

[45] Chatham House, "Can the UN's new AI governance efforts weather the AI race?," 18 September 2025 ("mostly powerless but, if implemented effectively, could set important global agendas on AI").

[46] UK AI Security Institute, "Fourth progress report" (official note confirming the UK renamed the AI Safety Institute as the AI Security Institute on 14 February 2025); US National Institute of Standards and Technology (NIST), "U.S. AI Safety Institute Establishes New U.S. Government Taskforce...," updated to note that AISI was re-established as the Center for AI Standards and Innovation (CAISI) in June 2025.

[47] T. Camelli and J. de Mooij, Center on International Cooperation (NYU), "Guidance for the New Global Dialogue on AI Governance," January 2026 (integrated, sequenced, UN-anchored roadmap).

[48] US Mission to the United Nations, "Remarks at the Security Council's Open Debate on Artificial Intelligence," 24 September 2025 ("We totally reject all efforts by international bodies to assert centralized control and global governance of AI").

[49] Center for Strategic and International Studies (CSIS), "What the UN Global Dialogue on AI Governance Reveals About Global Power Shifts," 7 October 2025 (US opposition at the Security Council, 24 September 2025).

[50] United Nations Office of Digital and Emerging Technologies (ODET), "UN Global Dialogue on AI Governance – Co-Chairs' Summary," 8 September 2026.

[51] General Assembly of the United Nations, President of the 81st session, "Speech on multilateral AI governance at the Ambassadorial Lunch," 10 September 2026.

Methodological and publication note. This manuscript is an original research work. The final editorial verification point is 15 September 2026, using an evidence cutoff of 14 September 2026. Time-sensitive institutional, participation, treaty and market information were rechecked for publication. Where official statistical breakdowns of Geneva Dialogue participation by regional group, developing-country status or vulnerability classification are not published, the paper says so rather than substituting estimated figures. Quantitative indicators of AI capability concentration are drawn from the Stanford HAI AI Index 2026, OECD, World Bank, UN and industry sources; each figure is cited to its source. The AUN AI Governance Agency framework is an original analytical contribution of this paper and is offered for independent scrutiny, refinement and subsequent empirical testing.