Submission to the Call for Written Input for the UN Global Dialogue on AI Governance
April 30, 2026

Submitting organisations:

ILINA Program, University of Cape Town AI Initiative, AI Safety South Africa, Action Lab Africa

Priorities

In your opinion, what outcomes would make the first Global Dialogue on AI Governance a success?

Naming and framing frontier and extreme risks such as loss of control, as concerns that are shared by all member states. The dialogue should acknowledge that Global South populations are not insulated from these risks because they are not home to the frontier labs.

The Dialogue should center a discussion on enabling countries worldwide to develop the capacity to study what AI models can do and translate technical findings into actionable policy decisions. Concrete commitments are needed across AI development and deployment to strengthen Global South AI governing capacity to evaluate models, monitor them once they are deployed and report incidents. For evaluation, this means investing in compute resources in the Global South to build assessment capabilities. For post-deployment monitoring, periodic reporting procedures that capture real-world impacts in diverse contexts need to be established. For incident-reporting, building on these monitoring procedures will help determine collective mitigation measures for the types of incidents reported.

Drawing on growing international consensus, the Dialogue should place a discussion of AI red lines at its core. Red lines represent non-negotiable prohibitions on AI behaviours and uses that are too dangerous to permit under any circumstances. They serve as a critical backstop on developments that must not proceed, regardless of commercial incentives or national interests. Red lines fall into two categories: limits on what AI systems should not do such as enabling weapons of mass destruction or autonomous self-replication, and limits on how AI should not be used such as mass surveillance, electoral interference, or manipulation of vulnerable populations. The Dialogue should ensure that the process of defining red lines is not left solely

to powerful states. This will allow for harms disproportionately experienced by Global South nations to be considered when defining enforceable international AI red lines.

From your perspective, which of the following thematic areas identified by the General Assembly Resolution 79/325 for the AI Dialogue reflect your priorities for urgent action and active engagement by your entity? Please select up to 4 priorities. Briefly explain your selection.

Each of the organisations listed above are African-led with a focus on AI governance. Our missions and output map directly onto these 3 priorities. We treat these three areas as interconnected because they reinforce each other in practice:

Safe, Secure and Trustworthy AI is a precondition for public legitimacy and for managing extreme risks as systems become more capable.

Capacity-building determines whether Global South countries can participate in identifying and evaluating context-dependent risks that would otherwise go undetected, unmitigated, and ungoverned. This participation matters because AI safety cannot be guaranteed by frontier developers and institutions in the Global North alone. Countries that cannot evaluate what AI systems are doing within their own contexts cannot ensure those systems are safe for their populations, or for the world. These risks are not contained because harms that originate in a local context do not stay there. Closing the capacity gap is therefore a prerequisite for AI being genuinely safe wherever it is deployed.

Attention to social, cultural and linguistic implications determines whether AI systems actually benefit the majority of the world's population rather than a narrow slice of it. Whether AI systems are economically beneficial to that same majority is equally at stake. Large-scale worker displacement matters as an AI safety concern. The asymmetric impact of automation on economies with large informal and low-skilled workforces makes this a disproportionate threat across the Global South, and a risk that goes ungoverned precisely because the countries most exposed lack the capacity to evaluate it.

Impact of AI Governance

How are the governance gaps and related developments/advances in the thematic areas you selected above affecting your country, region, or sector? Please highlight the most significant challenges and opportunities.

The current state of global AI safety, where evaluations, monitoring, and governance are concentrated in a handful of institutions and contexts, leaves everyone, including African and Global South people, exposed to significant risks. When AI systems are deployed globally but evaluated narrowly, critical failures go undetected. Historically, when extreme risks materialise, Global South populations suffer most severely.

The Global South is primarily a deployment site for frontier AI systems built and evaluated elsewhere. Existing safety evaluations do not test for local languages or contexts, meaning the risks go unmeasured and Global South governments must absorb risks they cannot measure. Failures that go undetected in under-evaluated contexts propagate through globally deployed frontier models.

The dominant capacity-building agenda for the Global South focuses on enabling adoption through infrastructure, connectivity and digital skills. To enable governance of frontier AI, the Global South also needs the capacity and expertise to run safety evaluations, monitor AI impacts and track incidents. This is structurally difficult to address because:

- Running frontier evaluations imposes a "safety tax" of roughly $6000 monthly that most institutions cannot absorb. Without this capacity, context-specific risk pathways, arising from low-resource languages, sparse data, and constrained infrastructure, go undetected by Western-designed evaluations.
- Incident tracking and post-deployment monitoring require standing institutions with continuous mandates and funding. Without institutional homes for this work, serious incidents go unreported.

Organisations like ILINA are already doing AI safety work in Global South contexts but remain under-resourced relative to the scale of the challenge. The Dialogue represents an opportunity to change this. With dedicated funding and institutional recognition, this work could scale meaningfully and fill critical gaps in global AI safety coverage. Some Global South governments already possess decision-making capacity that is underutilised and with political commitment and targeted investment, can be leveraged and strengthened quickly.

International Cooperation on AI Governance

What role can the AI Dialogue play in advancing international cooperation on AI governance?

Its contribution lies in two functions that follow from structural features no competing forum combines: universal UN membership, a multi-stakeholder mandate that seats industry, civil society, and academia as participants rather than observers, institutional lightness that allows it to convene other bodies without being perceived as a rival, annual cadence, and an Independent International Scientific Panel that provides scientific authority independent of the political forum itself. Together, these features give the Dialogue the reach to legitimise and the credibility to build shared understanding.

Legitimising technical standards, evaluations methods and safety frameworks produced in other fora. Tabling, discussing and acknowledging methods and standards at the Dialogue can accelerate their adoption and implementation. This could be through their inclusion in the Co-chairs' summary, incorporation in the Panel's annual report or conversation in thematic discussions.

Building a shared understanding of AI risks and governance approaches. The Dialogue's multi-stakeholder design creates a forum where governments, industry, civil society and academia can: table competing claims about the AI risks to prioritise and the right regulatory approach for AI; discuss and debate contradictions between different AI frameworks; and work towards common positions. Over annual cycles, this iterative exchange can generate the shared language and mutual understanding that binding agreements eventually require.

What are some of the existing initiatives, partnerships, or mechanisms that the AI Dialogue should build upon or connect with, and what added value could the AI Dialogue bring?

The International Network for Advanced AI Measurement, Evaluation and Science has moved from announcement to formal convening to joint testing of frontier models in just 18 months, a pace rarely achieved in international governance. Its model of lightly-structured, technical-first coordination between independent national institutions is worth backing and building on. At the same time, Kenya remains the sole African and developing country in the Network. The Global Dialogue can help democratise the work of the Network, opening it to a broader swathe of countries. The Global Dialogue can also offer a forum for a two-way exchange between the Network and non-members, enabling sharing of best practices by the Network and more safety perspectives from non-member countries.

The EvalEval Coalition works on improving the state of evaluations. AI model evaluations are a core part of AI governance, yet the field of evaluations still faces many challenges around methodology, reproducibility and coverage. The Global Dialogue can highlight the Coalition's work and provide a space for member states and stakeholders to discuss the state of evaluations and what improvements are needed.

Work on benefit sharing offers a further building block for the Dialogue. A recently proposed framework from the Centre for the Governance of AI operationalises benefit sharing across three pillars: redistribution of AI-generated economic gains to countries that would otherwise be left behind, transfer of the technologies and technical expertise needed to build indigenous AI capacity, and governance arrangements that ensure expanded access does not outpace safety. Yet despite this conceptual progress, benefit sharing remains one of the least developed building blocks of international AI governance. The Global Dialogue is well placed to move benefit sharing from a research question to a negotiating agenda item, and to ensure that the countries with the most to gain from equitable distribution have a say in shaping frameworks.

Inclusive Participation

How can different stakeholders contribute to the AI Dialogue? Please share recommendations for the format and structure of the AI Dialogue.

Structure thematic sessions to center both Global South research and Global South concerns. It is common in these forums to have Global South government officials describing regional

challenges as representative of the region. We recommend that the experts involved in the scene-setting interventions also include Global South researchers presenting original work.

Reserve dedicated time during the thematic sessions for regional groups to present co-ordinated positions based on their regional consultations prior to the Global Dialogue. This will also further the goal of building a shared understanding of risks and consolidating findings and inconsistencies in priorities across different regions represented by the member states.

Invite contributions from non-UN member state stakeholders to allow their views to have a chance at influencing the Co-chair's summary and the subsequent briefs from the Independent International Scientific Panel. The Dialogue could structure an accreditation process to filter the number of written submissions and this could supplement the current stakeholder consultations once the exact themes and sessions are determined.

Which voices, communities, or perspectives are currently underrepresented in global discussions on AI governance? How could they be included?

The most significant underrepresentation in global AI governance is that of Global South countries and communities, the populations least involved in designing AI systems, least represented in evaluating them, and most exposed to their failures.

At the intergovernmental level, the forums that set the tone for global AI governance exclude most of the world structurally. The G7 represents 7 of 193 UN member states, the Bletchley AI Safety Summit invited 28 countries, the majority of which were wealthy and Western, and the OECD, which produced the most widely adopted AI principles, counts only 38 members, none from Sub-Saharan Africa.

At the technical level, African and Global South researchers and safety evaluators are largely absent from frontier AI evaluations meaning that context-specific risks remain systematically invisible.

At the civil society level, organisations working on AI governance in Global South contexts are chronically under-resourced and under-cited relative to Global North organisations.

Inclusion should be designed with this disparity as the starting assumption. Concretely, the Dialogue should:

- **Maintain a standing database of Global South organisations and individuals working on AI governance.** They should be actively consulted in planning all Dialogue activities.
- **Treat Global South representation as a quorum condition.** If relevant Global South organisations or individuals are absent from a convening or decision-making session or there is evidence that no concrete steps were taken to enable their participation, the quorum will not have been met for that activity.

Sessions failing this condition must be reconvened before their conclusions can be adopted as outputs of the Dialogue.

What innovative engagement formats could most effectively foster meaningful and dynamic engagement during the AI Dialogue?

Most multilateral convenings assume familiarity with parallel initiatives, governance approaches, and priority risks, yet participants arrive with very different baselines. We suggest these four components are incorporated into the format of the Dialogue to address this.

- **Landscape mapping:** A short, structured exercise in which participants map what they know about existing initiatives, approaches to safety and governance across the thematic areas, and the main risks they prioritise. Results can be collected in real time & displayed as a shared baseline for stakeholders.
- **Priority setting:** Participants rank the risks, governance gaps, and intervention areas they consider most urgent. Dialogue organisers can use structured polling to produce a live, transparent ordering of priorities that the Co-Chairs' summary can reference.
- **Divergence mapping:** Prioritisation results are broken down by region and stakeholder type, surfacing where governments, industry, civil society, and Global South and Global North participants diverge. These divergences can indicate to the Independent International Scientific Panel which issues most need bridging work.
- **Annual evidence record:** The synthesised landscape and priority ordering are published as an annex to each Co-Chairs' summary, creating a year-on-year record of how shared understanding and priorities shift.

This four part format operationalises the Dialogue's mandate by establishing a common baseline at the start of each session; addressing briefing asymmetries across wealthier and under-resourced delegations; and giving the Scientific Panel a trackable evidence base built on participants' own input.

Good Practices and Policy Approaches

Please share examples of policies, practices, platforms, or approaches that promote effective AI governance or offer concrete solutions to addressing its challenges.

A promising approach to effective AI governance is frontier AI auditing. This model advocates for rigorous, third-party assessments of safety and security practices at leading AI companies, directly addressing core governance challenges: coordination, capacity, trustworthiness, and accountability.

- First, this model improves coordination by establishing shared standards and evaluation frameworks across governments and private firms, reducing fragmentation in oversight.

- Secondly, it strengthens capacity by combining technical expertise from industry with regulatory authority from governments, accounting for limitations of either acting alone.
- Third,trustworthiness is enhanced through independent verification, helping policymakers, researchers and the public gain confidence.
- Lastly, it reinforces accountability by creating mechanisms for external scrutiny and consequences for non-compliance.

If implemented inclusively, particularly with the involvement of all member states from the initial design, the likely outcome will be greater transparency in AI development. However, a major limitation is speed: building international consensus, legal frameworks, and auditing infrastructure will likely take a considerable amount of time. Despite this, frontier AI auditing represents a concrete, scalable solution with strong potential to advance responsible AI governance.