AI Action Summit: Global AI Governance

1. What is the current state of global AI governance, including key actors, initiatives, and frameworks? What critical gaps exist in this ecosystem, and where do you see potential for synergy or unnecessary overlap?

To strengthen the global AI governance processes, the AI Action Summit should:

1. Significantly expand Global South participation compared to previous summits.
2. Launch a worldwide public opinion survey on AI development and safety issues.
3. Require long-term scenario planning (around 5+ years) in AI risk assessments, particularly for impacts on future generations.

(1) Expand participation of the Global South in the Summit.

The global AI summits have achieved important milestones – notably securing cooperation between China, the US, and EU through the Bletchley Declaration and promoting safety practices in companies around the world through the Seoul Frontier AI Commitments. However, legitimacy and effectiveness of the summits are limited since at least 118 countries, primarily from the Global South, remain excluded from major AI governance initiatives.[1] Expanding participation is essential not only for legitimacy but also to secure broader buy-in for development, safety, and trust measures.

(2) Prioritize understanding diverse global perspectives on AI by commissioning a global survey on AI development and trustworthiness.

While some studies exist, including Concordia AI's analysis of Chinese AI surveys, there is insufficient understanding of how global publics view AI development and safety.[2] Moreover, insights are heavily skewed towards Western populations. The Summit should commission a comparative and truly global survey, leveraging expertise from organizations like Missions Publiques (France) and the Center for

---

[1] UN High-Level Advisory Body on AI, Governing AI for Humanity.
[2] See Concordia AI's State of AI Safety in China pages 68-73.

International Security and Strategy (China). Results should be delivered within 12 months to inform future governance decisions.

(3) The Summit should commission exercises to develop future AI scenarios and ensure that such analysis is included in potential AI evaluations efforts under the UN or global AI summits.

An authoritative, international understanding of AI opportunities and risks should incorporate a long-term, future-generations perspective. However, current evaluations, like the AI Risk Global Pulse Check for the UN AI Advisory Body, typically look only one to two years ahead.[3] There is uncertainty around the appropriate time horizon for AI forecasting, and longer-term projections confront increased uncertainty. Nevertheless, the current short-term focus is insufficient given AI's potential multi-generational effects – from climate impact of large models to the potential for developing vastly more powerful systems. Drawing lessons from the International Panel on Climate Change's scenario planning, the Summit should commission analyses of AI scenarios beginning at least several years in the future.[4]

2. How can we best institutionalize and expand upon the International Scientific Report on the Safety of Advanced AI? What structures and processes are needed to ensure its ongoing relevance, objectivity, and impact on global policy decisions?

We recommend that the Summit agree upon the following institutional structure for the International Scientific Report on the Safety of Advanced AI (hereafter the "Report") and implement these changes over the subsequent six months.

1. House the Report under the United Nations as part of the Independent International Scientific Panel on AI.
2. Create differentiated layers for the panel: a scientific layer to produce technical reports on AI opportunities and AI risks, and a policy layer that produces a separate implications document and allows government engagement.

---

[3] UN High-Level Advisory Body on AI, Governing AI for Humanity.
[4] See IPCC and Carbon Brief.

(1) The Summit should declare that the Report will be housed under UN auspices, leveraging the Independent International Scientific Panel on AI proposed by the Global Digital Compact.

It is essential to provide the Report with a stronger institutional footing and legitimacy beyond the AI summit process and the signatories of the Bletchley Declaration. The Summit can host deliberations on balancing legitimacy, agility, and rigor in AI assessment and horizon scanning in the future institutional structure of the Report. Ultimately, UN imprimatur provides significantly stronger legitimacy than positioning the report under the global AI summits or moving under the network of AI safety institutes (which represents an even smaller group of countries than the global AI summits).

(2) The Independent International Scientific Panel on AI should implement a two-track structure: a scientific track producing technical assessments, and a policy track developing recommendations based on these findings.

The core scientific layer could comprise one group working on an opportunities/SDGs-focused report, and one group working on a narrow, risk-focused report. The scientific layer would produce detailed technical reports without requiring government approval. Meanwhile, the government interface layer could create a separate "Policy Implications Document," on which governments could provide feedback without modifying technical findings. The whole panel could be administered by the UN or OECD, analogous to the Intergovernmental Panel on Climate Change Technical Support Unit.[5] The panel should invest in stakeholder engagement to maintain buy-in through regular briefings and consultations with governments and technical workshops with national experts.[6]

---

[5] The IPCC is divided into three working groups and a task force, which is administered by a Technical Support Unit.

[6] Other views worth referencing include Carnegie Endowment for International Peace, The Future of International Scientific Assessments of AI's Risks and Centre for International Governance Innovation, Framework Convention on Global AI Challenges.

3. What are the potential benefits, challenges, and key considerations in establishing a global body for AI oversight? How should such a body be structured, governed, and empowered to effectively address international AI challenges?

The AI Action Summit should:
1. Declare in the joint statement that a World AI Organization is needed for AI governance.
2. Hold workshops on various innovative forms a World AI Organization could take.

(1) The Summit's joint statement should affirm that signatories support creating a World AI Organization.

The France AI Commission's report argues for creating a World AI Organization that "would share scientific findings on the workings and effects of AI, and define binding standards for AI systems and how they should be audited."[7] We agree that international coordination is becoming increasingly critical as companies plan to scale frontier AI systems 100-1000x (in terms of effective compute) in the next three to five years.[8] Without coordinated licensing and oversight, countries risk a regulatory race to the bottom, as seen with global corporate taxes. As these systems' potential for misuse and loss of control grows, we must establish global governance mechanisms before private companies' decisions irreversibly impact humanity's future.

(2) The Summit should begin a deliberative process for exploring the shape of this global body through public and private workshops.

There are many difficult tradeoffs to consider in how to create a World AI Organization and how to structure it. It may be impossible to firmly decide all of these issues at the outset, but kicking off a deliberative process is key to gradually forming an international consensus before being overtaken by events. For instance, the Paris Climate Accords set ambitious goals for limiting greenhouse gas emissions, which despite lacking strict enforcement mechanisms, established an international consensus and created positive pressures on states. One proposed approach is a Framework Convention on Global

---

[7] France Artificial Intelligence Commission, Our AI: Our Ambition for France.
[8] Epoch AI, Training Compute of Frontier AI Models Grows by 4-5x per Year.

AI Challenges, which would create an inclusive overarching convention under which specific protocols on public goods, power concentration, safety, and security could be developed.[9] While the exact structure of a global organization remains debatable, starting this process is crucial for building international consensus before more dangerous AI capabilities are developed.

4. How can we enhance the effectiveness and impact of global AI summits? What specific measures would ensure these events drive meaningful progress in AI governance and foster sustained international cooperation? Which countries could organize the next ones?

The AI Action Summit should build on previous successes of the global AI summits by:

1. Brokering agreement on international red lines regarding AI misuse and loss of control, and undertaking projects to harmonize global evaluation methods.
2. Announcing future summits in both China and the US to strengthen international cooperation.

The summit series has proven effective at bringing together major AI powers while maintaining diverse representation, particularly through including China alongside Western nations – a key advantage over other AI governance initiatives. This inclusive approach has already yielded important agreements on AI safety and trust, and also resulted in commitments from leading AI developers around the world on model safety, testing, and transparency.[10]

(1) The Summit should prioritize agreeing upon red lines for critical transnational threats to AI trustworthiness and implementing them through dialogues on evaluations.

As part of three pillars to address potential catastrophic risks, the Summit should first broker agreement on international red lines and early warning indicators regarding AI misuse and loss of control.[11] These risks are the most likely to garner widespread international agreement, and a light-touch approach, predicated on risk thresholds, can help prepare for the possibility of surprising, exponential changes. The

---

[9] Centre for International Governance Innovation, Framework Convention on Global AI Challenges.

[10] The Bletchley Declaration, Frontier AI Safety Commitments, AI Seoul Summit 2024.

[11] As Chinese Premier Li Qiang stated at the World Economic Forum: "there should be a red line in AI development, a red line that must not be crossed." Specific red lines could include autonomous cyberattacks and AI assisting in developing weapons of mass destruction, as per the IDAIS-Beijing statement.

second pillar is continuous AI safety testing for early warning indicators. The Summit should host at least one joint evaluation exercise among diverse countries to demonstrate the current state of warning indicators. It should also create an international working group to develop rigorous testing standards in the year after the Summit. The third pillar is developing a set of crisis management protocols that can be triggered if certain risk thresholds are crossed. This could include mandating further AI safety research, assurances, and human oversight until proven safe. There should be special attention and support for Global South countries in building resilience to risks.[12]

(2) Institutionalize the selection of future global AI summit hosts and consider holding summits in both China and the US.

The ability to involve China alongside Western countries is perhaps the biggest comparative advantage of the global AI summits. Holding meetings in both China and the US would deepen international understanding of Chinese perspectives on AI safety while mitigating geopolitical concerns on either side. The Summit should also establish systematic criteria for selecting future hosts, for instance, through an open call for proposals by participating countries to host a subsequent iteration.

5. How do we ensure that all countries can have a say in the governance of AI and enjoy the technologies' benefits, including those with limited access to AI resources?

To promote equitable global AI development and voice, the AI Action Summit should:
1. Launch a coordinated initiative to build high-quality datasets for underrepresented languages.
2. Create a Global AI Talent Program focused on developing expertise in the Global South.
3. Announce that countries support creation of a World AI Organization under UN auspices.

(1) The Summit should announce a collaboration among technology companies, governments, and civil society groups to improve AI training datasets in underrepresented languages.

---

[12] For additional context on these three pillars, see Concordia AI at the AI Seoul Summit.

These initiatives are essential for ensuring AI benefits extend beyond current technology hubs. High-quality local language data is crucial for foundation model performance, but must be collected with meaningful local participation and respect for data sovereignty.[13] Potential inspirations could include Project SEALD for Southeast Asian Languages, Chinese Peng Cheng Lab's dataset of languages in the Belt and Road Initiative, and the Maori Data Sovereignty Network.[14]

(2) The Summit should announce a Global AI Talent Program, a multi-stakeholder project to foster AI R&D, safety, and governance talent, especially from underrepresented groups.

Examples of impactful projects include educational initiatives such as the Smart Africa Digital Academy, incentivizing top AI companies to establish offices in the Global South, and financing AI-focused entrepreneurship.[15]

(3) A World AI Organization should be established under UN auspices to guarantee universal participation rights.

While subsequent specialized initiatives may involve smaller groups for efficiency, the overarching framework for global AI governance must remain fully inclusive and equitable, which is best ensured by the UN.[16] Meanwhile, subordinate efforts could be pursued by smaller groups of countries as needed, consistent with proposals such as a Framework Convention approach.[17]

---

[13] Rest of World, We tested ChatGPT in Bengali, Kurdish, and Tamil. It failed.

[14] See for more information about Project SEALD, Peng Cheng Lab, and the Maori Data Sovereignty Network.

[15] See Smart Africa Digital Academy.

[16] Additional suggestions on Global South voice in AI can be found in Oxford Martin School Voice and Access in AI: Global AI Majority Participation in Artificial Intelligence Development and Governance.

[17] Centre for International Governance Innovation, Framework Convention on Global AI Challenges.