

General vision for the outcome of the 2025 French AI Action Summit

Concordia AI has three main recommendations for the AI Action Summit, corresponding to the three categories articulated by President Macron on advancing AI science, solutions, and standards:

1. **Science:** Refine the scientific assessment process on AI; accelerate research on key technical AI safety and trustworthiness problems.
2. **Solutions:** Recognize AI for Sustainable Development Goals (SDGs) and AI safety as global public goods, ensuring their wide and equitable distribution across society.
3. **Standards:** Strengthen company transparency commitments, establish international red lines and early warning indicators for AI development and misuse, and develop shared standards for AI safety and trustworthiness evaluations.

(I) The AI Action Summit should agree on next steps for institutionalizing international scientific assessments of AI; the Summit should broker research partnerships and establish a global fund to advance AI safety and trustworthiness.

The Summit can host deliberations on how to develop a sound institutional footing for assessment and horizon scanning efforts by building on existing efforts like the International Scientific Report on the Safety of Advanced AI and the UN Global Digital Compact's (GDC) planned Independent International Scientific Panel on AI. Balancing considerations of legitimacy, agility, and rigor, we recommend housing this effort under UN auspices, leveraging the panel suggested by the GDC. The structure should include two distinct tracks: a scientific track producing technical assessments (with separate reports focusing on opportunities with the SDGs and risks), and a policy track developing recommendations based on these findings.¹ While governments would be able to engage with policy recommendations, they would not be able to modify technical findings, following an acknowledgment rather than approval process. The Summit could secure agreement on this institutional structure and commitment to implement it over 6-12 months.

¹ 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](#).

The Summit should broker new institutional partnerships between universities and research centers worldwide and establish a ‘Global AI Safety and Trust Fund’ to support international academic collaboration. These projects would address a critical gap — AI safety constitutes only about 2% of current AI research despite growing international calls for enhanced collaboration on safety, ethics, and societal impact.² New joint research centers and personnel exchanges between universities and/or research institutes would improve expertise around the world. A global fund in the range of US\$100 million would be the largest such fund for AI safety and trust, which could support ambitious research efforts on AI alignment, evaluation, ethics, and more. These initiatives would substantially advance global cooperation without requiring binding government agreements.³

(2) The France AI Action Summit should release a comprehensive mapping of underresourced AI public goods and ensure strong Global South representation.

These recommendations aim to address two critical gaps in global AI development: the underproduction of AI public goods and the exclusion of Global South voices. With 118 countries absent from key AI governance initiatives and fewer than 30 national governments at the UK AI Safety Summit, there is substantial room for the AI Action Summit to expand invitations to Global South countries.⁴

AI public goods — non-excludable and non-rivalrous resources that are typically underproduced by the private sector — include essential tools like open datasets for underresourced languages and open-source evaluation frameworks. The lack of linguistic and cultural diversity in AI development not only limits access but also undermines trust, as models lacking robust multilingual testing may fail to

² Emerging Technology Observatory on [AI safety](#). For more on the need for global academic cooperation, see [The Manhattan Declaration on Inclusive Global Scientific Understanding of AI](#), which was co-chaired by Turing Award Winner Yoshua Bengio and former White House OSTP Acting Director Alondra Nelson, and signed by Concordia AI CEO Brian Tse.

³ One example of such institutional partnerships is the Global Partnership on AI’s [Expert Support Centers](#). Similarly, the UN High-Level Advisory Body (HLAB) on AI has proposed a ‘Global Fund for AI’ that would include safety and governance funding, see [Governing AI for Humanity](#). The fund could include contributions from stakeholders such as national governments, technology companies, and philanthropists. A US\$100 million fund would be the largest such fund for AI safety and trust, compared for instance to the Frontier Model Forum’s US\$10 million+ [AI Safety Fund](#). Similar efforts to improve pandemic resilience through a World Bank “[Pandemic Fund](#)” have already raised US\$2 billion.

⁴ UN High-Level Advisory Body on AI, [Governing AI for Humanity](#).

perform safely across different contexts.⁵ The Summit should map gaps in the production of AI public goods and announce an action plan of targeted projects to be completed within 6-12 months, advancing recent UN General Assembly resolutions on bridging AI divides and ensuring responsible development.⁶

(3) The Summit should publish a platform for monitoring corporate AI safety and trust commitments; it should also create a working group to discuss shared standards for AI safety and trustworthiness evaluations.

The Summit should create a public website to track company adherence to voluntary commitments and commission independent, third-party ratings of compliance. Such measures would strengthen the requirement in the Seoul Frontier AI Safety Commitments for companies to publish safety frameworks at the AI Action Summit.⁷ The Summit should also seek commitments among countries to implement minimum transparency standards for AI companies domestically and share some of this information in an international database.

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.⁸ 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 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.⁹

⁵ [Open Letter on the Safety of Multilingual AI](#).

⁶ [Enhancing international cooperation on capacity-building of artificial intelligence : draft resolution](#); [Seizing the opportunities of safe, secure and trustworthy artificial intelligence systems for sustainable development](#).

⁷ [Frontier AI Safety Commitments, AI Seoul Summit 2024](#).

⁸ 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 autonomously cyberattacks and AI assisting in developing weapons of mass destruction, as per the [IDAIS-Beijing](#) statement.

⁹ For additional context on these three pillars, see [Concordia AI at the AI Seoul Summit](#).