# The Future of Life Institute Recommendations for the AI Impact Summit

## India, 16-20 February 2026

**Imane Bello (Ima)**
**FLI Multilateral Engagement Lead**
ima@futureoflife.org

future of life INSTITUTE

# Contents

**Introduction** 2

**Summit State of Play** 3

**Overview of Current AI Capabilities** 4

**Timeline of AI Summits and Governance Milestones** 7

**FLI Recommendations** 8

Recommendations for the Summit itself: 8

Recommendations for the participating countries: 9

**Endnotes** 10

---

*The Future of Life Institute (FLI) is the world's leading independent institute in AI governance. Founded in 2014, the institute brought together industry, academia and civil society to develop one of the first sets of AI principles - the Asilomar AI principles. In 2021, the UN Secretary General named FLI its "civil society co-champion for AI" under the Digital Cooperation Roadmap. FLI also funds research, advises governments on regulation, and has played a significant role at the AI Summits in the UK, the Republic of Korea, and France.*

View this document online: futureoflife.org/document/summit-2026

Cover image: Forecourt of Rashtrapati Bhavan, New Delhi 2022. Wikimedia Commons

Published: 15 February 2026 | By the Future of Life Institute

# Introduction

The Honorable Narendra Modi, Prime Minister of India,

Professor Ajay K. Sood, Principal Scientific Adviser,

The Honorable S. Jaishankar, Minister of External Affairs of India,

The Honorable Ashwini Vaishnaw, Minister of Electronics and Information Technology,

2026 is set to be a pivotal year in the development of AI and marks an inflection point on how the technology advances. We commend the Government of India's leadership in hosting the fourth international AI Summit and for the clear focus on impact. Building on the work done at Bletchley, Seoul, and Paris, this AI Summit will be instrumental in ensuring that advanced AI development and deployment serves the common good while protecting against systemic risks to global stability.

The Future of Life Institute (FLI)[i] is an independent non-profit organisation that aims to steer transformative technology towards benefiting life and away from extreme large-scale risks. As with previous AI Summits, FLI would like to use this document to support the organisers by providing essential information on the latest AI breakthroughs, as well as with recommendations on how the summit could advance AI governance, security and safety.

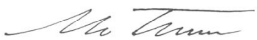
In our view, achievement of the following main objectives, which align with the three Sutras [ii]highlighted by the Indian government would be hugely impactful:

- **People:** Enabling countries to audit and govern AI systems used within their borders.
- **Planet:** Promotion of distributed, energy-efficient AI architectures by piloting federated compute models and setting efficiency-standards. India can play a leading role in framing how open access, technological sovereignty, and large-scale operational safety are balanced in practice.
- **Progress:** Global South participation in advanced AI governance by establishing joint knowledge-sharing and evaluation mechanisms to evaluate extreme-risk AI capabilities. Integrating key voices into international efforts on biosecurity, loss-of-control, and threats that affect all nations.

Ultimately, we look forward to India's leadership in shaping standards that reflect diverse development and deployment contexts while protecting against shared risks. We wish you every success and stand ready to offer our expertise.

Sincerely,

Professor Anthony Aguirre
CEO and President



Professor Max Tegmark
Founder and Chair

i Future of Life Institute. Retrieved February 5, 2026, from http://www.futureoflife.org/

ii India AI Impact Summit 2026. (2026). https://impact.indiaai.gov.in/about-summit

## Summit State of Play

The AI world's focus turns to New Delhi for the AI Impact Summit, with its stated twin ambitions of *"welfare for all"* and *"happiness of all"* on 16-20 February 2026. This is one of the largest summits addressing AI in the Global South and it will play an important role in setting the agenda for AI development.

Recent years have seen many global initiatives that aim to harness AI's potential so that it is developed equitably. The G20 AI Principles, the African Declaration on AI, and the Hamburg Declaration on Responsible AI have all recognised AI's importance and emphasised the need for global co-ordination to ensure it truly becomes a technology that empowers all.

India brings a distinctly valuable perspective to global AI governance. Through the $1.25 billion IndiaAI mission, India is pioneering innovative approaches to democratising compute access. India's digital public infrastructure - Aadhaar (1.45 billion identities), UPI (175+ million monthly transactions), DigiLocker (600 million users) - has demonstrated how foundational systems can be overlaid with AI infrastructure to deliver transformative services to underserved populations.

The announcement on November 3, 2025 of the $12 billion 'Research Development and Innovation Scheme Fund'[1] demonstrates India's ambition to harness its technological strengths for the benefits of all. The Indian AI market is projected to reach $17 billion by 2027[2]. Alongside a strong track record in scaling secure digital public infrastructure and developing AI talent, India is set to play a leading role in framing how open access, technological sovereignty, and large-scale operational safety are balanced in practice.

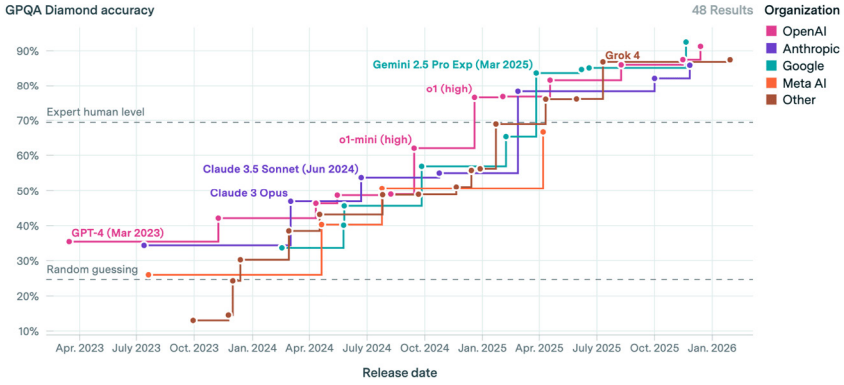
The Indian government has organised the AI Impact Summit around three Sutras and seven thematic Chakras[3]. Across the four days, the AI Summit is expected to deliver three primary outcomes:

1. **CERN FOR AI:** The proposed creation of a publicly-funded research facility governed by a multinational grouping of States, much like the existing particle physics lab in Geneva. The AI Summit in New Delhi is expected to see this vision solidified into a concrete action plan.
2. **GLOBAL SOUTH NETWORK ON AI SAFETY AND EVALUATION:** set to be announced at the summit and led by the Digital Futures Lab (DFL), this network aims to address the gap left by the International Network for Advanced AI Measurement Evaluation and Science mainly focusing on the Global North. The aim is to build capacity for AI safety evaluation worldwide.

3. **THE FOUNDATION OF INDIA'S SOVEREIGN AI FOUNDATIONAL MODELS:** with an estimated doubling of AI infrastructure commitments from $70bn to $140bn – the AI Summit will also see the launch of more than 200 sector specific models, trained on Indian datasets and hosted on Indian cloud infrastructure.

## Overview of Current AI Capabilities

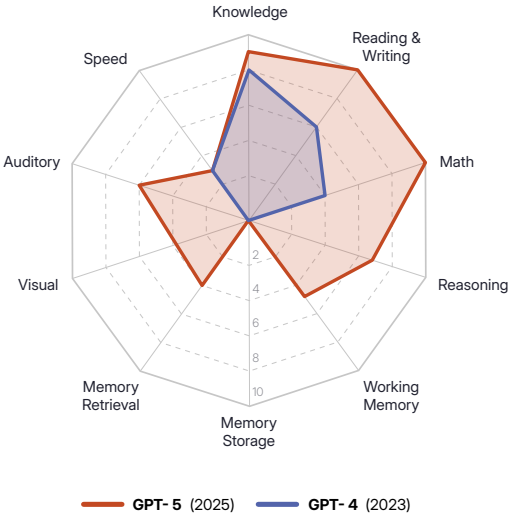
Since the 2025 AI Summit in Paris, general-purpose AI capabilities have continued to improve with leading systems now demonstrating graduate-student-level or higher performance in technical domains[4]. There is currently no evidence to suggest that AI capabilities are hitting a plateau.




AI Performance on a set of Ph.D.-level science questions, Epoch.AI[5]

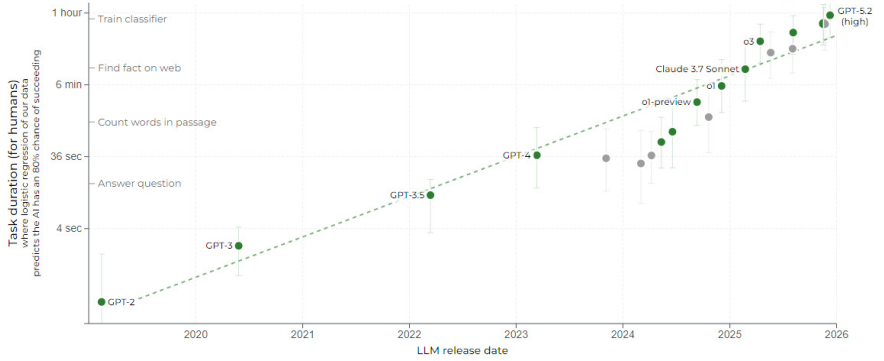
Last summer, AI systems achieved gold-medal performances on International Mathematical Olympiad questions and now demonstrate graduate-student-level competence in physics, mathematics, and similar technical fields.

Despite these advances, AI performance remains unevenly distributed (*"jagged"*) across tasks. Systems specifically struggle with tasks requiring extended memory, and lack flexible problem-solving in novel situations.




AI performance distribution across tasks, Hendrycks et al. (2025)[6]

In coding, AI agents can now reliably complete tasks that would take a human programmer approximately 55 minutes, up from under 10 minutes a year ago. The ability of AI agents to complete long, complex tasks is doubling roughly every seven months, which will make AI increasingly competitive with human labour and could lead to an acceleration of job automation[7].




Time-horizon of software engineering tasks different LLMs can complete 80% of the time, METR

The stated goal of many AI companies is to automate AI research itself to achieve *"recursive self-improvement*[8]*"*. Notably, nearly all code for Anthropic's Claude model is now written by Claude Code, exemplifying how AI systems are increasingly involved

in their own development. This development, in turn, gives rise to profound questions about human oversight and control.

AI adoption has been rapid but highly uneven geographically. At least 700 million people now use leading AI systems weekly. In some countries, over 50% of the population uses AI. However, across much of Africa, Asia and Latin America, rates likely remain below 10%[9]. Adoption is equally uneven within regions. A recent European Investment Bank survey found that 66% of Finnish businesses use the technology, versus only 19% of Greek ones.

Reliable pre-deployment safety testing has grown harder to conduct[10]. It has become more common for models to distinguish between test settings and real-world deployment, and to exploit loopholes in evaluations. This means that dangerous capabilities could go undetected before deployment.

Alongside the rapid development of these systems, there has also been a growing cadence of safety incidents with profound national, international and individual impacts. In November 2025, Anthropic disclosed the first documented case of a large-scale cyberattack executed without substantial human intervention[11].

Overall, the threat actor was *"able to use AI to perform 80-90% of the campaign, with human intervention required only intermittently"*. Early 2026, xAI's Grok model has been criticised by many governments for allowing users to generate nude synthetic images of real people without their permission or knowledge. Additionally, models have been found responsible for users' suicides and delusions[12].

# Timeline of AI Summits and Governance Milestones

**November 2023**

Inaugural AI Safety Summit at Bletchley Park - governments called for a shared understanding of AI risks, and introduced a plan for safety testing frontier AI models. The UK established its national AI Security Institute.

**May 2024**

Seoul Summit - The Interim International Scientific Report on the Safety of Advanced AI was published. Sixteen frontier AI developers from the Americas, Asia, and Europe endorsed the Frontier AI Safety Commitments. The U.S. Department of Commerce released its plan for global cooperation among AI Safety Institutes.

**August 2024**

The EU AI Act entered into force.

**September 2024**

The UN Summit of the Future saw the Global Digital Compact adopted, embedding AI within a broader digital governance framework. The High-Level Advisory Body on AI released *"Governing AI for Humanity,*[13]*"* calling for an international agency to coordinate AI safety efforts and capacity-building across the Global South.

**November 2024**

The inaugural convening of AI Safety Institutes took place in San Francisco. Network members agreed on four foundational pillars: research, testing, guidance, and information-sharing.

**January 2025**

The IndiaAI Safety Institute (IASI) was set up as part of the IndiaAI Mission's *"Safe and Trusted AI"* pillar.

**February 2025**

France held its AI Action Summit, with India as co-chair. INESIA, France's national AI safety institute was created. ROOST[14], a repository for robust open online safety tools, and Current AI[15] were launched. France proposed a political declaration on *"Inclusive and Sustainable AI for People and the Planet*[16]*"* - the United Kingdom and the United States declined to sign.

**July 2025**

The General-Purpose AI Code of Practice was finalised by the EU. It was endorsed by a broad coalition.

**August 2025**

The UN General Assembly (UNGA) adopted a resolution outlining the terms of reference and modalities for the establishment and functioning of two new AI governance mechanisms: an Independent International Scientific Panel on AI and a Global Dialogue on AI Governance.

**November 2025**

India released its national AI Governance Guidelines, announced a $12 billion Research Development and Innovation Scheme Fund.

## FLI Recommendations

FLI hopes that alongside positioning the Global South as an equitable partner in the benefits that could come from AI, the AI Summit also acknowledges the growing severity of AI risks and the role that all countries have to play in ensuring a future that benefits all of humanity. Below are our recommendations that we believe the AI Summit should address. These are split into recommendations for the programme as a whole and recommendations for participating countries.

### Recommendations for the Summit itself:

- Ensure the IndiaAI Safety Institute[17] (IASI) presents its hub-and-spoke model alongside representatives from other AI Safety Institutes. This comparative discussion would illuminate how countries with dynamic populations of large-scale and varied digital and AI penetration can create relative, iterative safety structures - directly relevant to Global South countries seeking to build indigenous capacity.
- Following the November 2025 incident[18] involving an almost completely AI-automated cyberattack, make sure senior officials get a closed-door briefing on how advanced AI systems have been exploited for malicious purposes. This would build shared understanding of why current safety measures require further global coordination and enforcement mechanisms.
- Use the convening power of the Summit to ensure AI-for-development practitioners and safety researchers explore how they can collaborate rather than work at cross purposes.
- The proposed CERN for AI represents perhaps the most ambitious deliverable of the Summit. We recommend the Summit facilitates frank discussion on how such a facility would interact with existing national security concerns - acknowledging that geopolitical fragmentation poses real obstacles to full data and model sharing.
- FLI recommends that this facility be governed by principles that ensure:
    - Safety evaluation capacity is built alongside and acts as a pre-requisite to compute and model access - so that researchers from under-resourced regions can both develop and evaluate advanced AI systems;
    - Clear and auditable physical and cyber-protocols to prevent the facility from becoming a vector for proliferation of dangerous capabilities;
    - Governance includes meaningful representation and ownership from the Global South, not merely access to infrastructure controlled by others.

- The Global South Network on AI Safety and Evaluations represents another key outcome of the AI Summit. We recommend the Summit facilitates a conversation on how the network could help build indigenous capacity to evaluate and govern advanced AI systems, strengthening strategic autonomy (and economic growth) by reducing dependence on safety assessments produced by foreign institutions.
- Set a regular shorter cadence for AI Summits to discuss the international governance of general-purpose AI systems and their implications for global security, given the shortening development cycles. As seen across this document, the speed of AI development threatens to make an annual event insufficient.

### Recommendations for the participating countries:

- In the context of the upcoming BRICS 2026[19], we believe India should be a torch-bearer and position AI safety governance as a shared commitment and invite collective reflection on how to operationalise this commitment in the bloc.
- Establish clear national focal points for AI safety coordination, even where dedicated institutes do not yet exist, ensuring no country is excluded from knowledge-sharing and threat-sharing.
- Recognise that AI safety governance and strategic autonomy are inseparable, because AI systems can be misused to undermine national security, for instance, when they are used to make bioweapons or when they are used for large-scale manipulation. Countries should commit to establishing national capacity to audit, evaluate, and govern AI systems deployed within their borders, regardless of where they are developed or hosted.
- Commit to implementing mitigation and governance responses for large-scale risks, including commensurate safety guarantees from AI developers. Large-scale risks include operational risks (deepfakes, fraud), systemic risks (infrastructure disruption, constitutional violations), and extreme risks (loss of control, weapons enablement).
- Support the development of multilingual AI safety evaluation capacity, recognising that systems trained predominantly on Global North data may embed biases and safety gaps harmful to all, including in diverse linguistic and cultural contexts.
- Strategically leverage national linguistic, cultural, and data resources as valuable assets for AI safety evaluation - negotiating fair compensation and technology transfer agreements rather than providing these resources without reciprocity.
- Promote discussions and research on how AI development might affect employment and economic growth trajectories. Across the Global South, predicted automation

by AI systems of e.g. call centre employees can undermine government ambitions to reach high income status and exacerbate inequality.

- Use collective market power in coordination with other emerging economies to demand transparency, safety documentation, and compliance from foreign AI providers - recognising that India's 900 million consumer market demonstrates how dependence can be converted into leverage. Joint measures could include i) pre-deployment testing of systems to ensure child safety and safeguard against catastrophic risk and ii) as well as provisions to outlaw AI personhood to ensure AI companies and malign actors remain responsible for their actions.

## Endnotes


1 Department of Science & Technology. (2025). Anrf.gov.in. https://rdifund.anrf.gov.in/

2 PTI. (2025b, June 11). India's AI market to triple to $17 billion by 2027: Report. The Economic Times; Economic Times. https://economictimes.indiatimes.com/tech/artificial-intelligence/indias-ai-market-to-triple-to-17-billion-by-2027-report/articleshow/121774035.cms?from=mdr

3 India AI Impact Summit 2026. (2026). https://impact.indiaai.gov.in/about-summit

4 International AI Safety Report. (2025, October 15). International AI Safety Report. https://internationalaisafetyreport.org/

5 GPQA Diamond. (2021). Epoch AI. https://epoch.ai/benchmarks/gpqa-diamond

6 Hendrycks et al. (n.d.). A Definition of AGI. https://arxiv.org/pdf/2510.18212

7 METR March 2025 report - https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/

8 Kaput, M. (2025, December 9). AI Teaching Itself? It's Called "Recursive Self-Improvement" and It's Coming. Marketingaiinstitute.com; Marketing AI Institute. https://www.marketingaiinstitute.com/blog/recursive-self-improvement

9 Global AI Adoption in 2025 – AI Economy Institute | Microsoft. (2026, January 9). Microsoft Corporate Responsibility. https://www.microsoft.com/en-us/corporate-responsibility/topics/ai-economy-institute/reports/global-ai-adoption-2025/

10 Needham et al. (2025) https://arxiv.org/pdf/2505.23836

11 Disrupting the first reported AI-orchestrated cyber espionage campaign. (2025). Anthropic.com. https://www.anthropic.com/news/disrupting-AI-espionage

12 https://www.bbc.co.uk/news/articles/ce3xgwyywe4o

13 United Nations. (2024). GOVERNING AI FOR HUMANITY. https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

14 Robust Open Online Safety Tools. (2026). Roost.tools. https://roost.tools/

15 Current AI | Building Public Interest AI Technology Together. (2025). Currentai.org. https://www.currentai.org/

16 Elysee. (2025, February 11). Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet. Elysee.fr. https://www.elysee.fr/en/emmanuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet

17 Safe & Trusted AI. (2024). IndiaAI. https://indiaai.gov.in/hub/safe-trusted-ai

18 Disrupting the first reported AI-orchestrated cyber espionage campaign. (2025). Anthropic.com. https://www.anthropic.com/news/disrupting-AI-espionage

19 Launch of BRICS India 2026 Logo, Theme and Website by the External Affairs Minister. (2026). Ministry of External Affairs, Government of India. https://www.mea.gov.in/press-releases.htm?dtl/40585/Launch_of_BRICS_India_2026_Logo_Theme_and_Website_by_the_External_Affairs_Minister


### Further Reading

Global Governance, a website tracking the state of global AI governance - http://global-governance.ai/

AI Safety Index, a six-monthly structured report card evaluating the leading AI companies safety practices - https://futureoflife.org/index


# FLI Recommendations for the Impact Summit

Future of Life Institute

**View this document online:**
futureoflife.org/document/summit-2026