INDIA'S AI GAMBIT

NAVIGATING THE GLOBAL RACE

Strategic Foresight Group

INDIA'S AI GAMBIT

NAVIGATING THE GLOBAL RACE

Strategic Foresight Group

Strategic Foresight Group
C-306, Montana, Lokhandwala Complex, Andheri West, Mumbai 400 053, India
Email: info@strategicforesight.com

Authors:
Sundeep Waslekar
Ilmas Futehally
Jayantika Kutty

ISBN 978-81-88262-36-6

Design and production by MadderRed

Printed at Gourishankar Kothari & Company, Mumbai, India

FOREWORD

Ambassador Pankaj Saran
Convenor, NatStrat

Since Niti Aayog introduced the doctrine of 'AI for All' in its 2018 strategy document, India has made tremendous strides in harnessing Artificial Intelligence for economic development and social inclusion. Indian entrepreneurs have developed several applications to improve the efficiency of agriculture, healthcare, education, mobility and public administration, to name a few. In June 2025, NatStrat published a three-part Long Paper supporting 'AI for All' with significant investments in infrastructure and talent in the coming years. It is by now well established that India's success will depend on building domestic compute capacity, advancing semiconductor manufacturing, and cultivating an entire generation of scientists who can lead in advanced research.

The NatStrat paper and discussions with a wide range of experts have brought out the importance of combining the concept of 'AI for All' with that of 'AI for Sovereignty, Science and Security'. The two ideas are mutually reinforcing and essential if India is to successfully navigate the geopolitics of AI.

The focus on development of Large Language Models to increase productivity and creativity is distracting us from a more consequential phenomenon which is the development of advanced AI models for scientific discovery, the objective of which is not merely to seek commercial advantage, but to define the future of knowledge, influence and human destiny itself.

India needs to leapfrog in scientific progress using Artificial Intelligence. Our success in space, atomic energy and vaccines shows that we can produce cutting edge science and technology, with limited resources, in a responsible way, for the benefit of not only our nation but also for humanity at large. In recognition of India's strengths, opportunities and needs, Prime Minister Narendra Modi has announced a 'Research Development and Innovation Scheme Fund' worth $ 12 billion for research, development and innovation, including for AI, Quantum and other emerging technologies. While this will no doubt be harnessed by Indian industry, India's diplomatic and national security structures also have a role to play in steering India's AI ambitions in a responsible way at the national and international level.

India needs a level playing field and safe and secure development of this critical technology. It is noteworthy that more than 120,000 signatories led by Nobel Laureate scientist Geoffrey Hinton, Apple co-founder Steve Wozniak, and former US National Security Adviser Susan Rice have called for a ban, issued by the Future of Life Institute, on the development of superintelligence until a broad scientific consensus is achieved on security and safety. India can legitimately demand responsible behaviour from AI leaders as we provide large segments of customers, talent and data.

It is in India's interest to work towards rules that safeguard national security and secure us from threats from the misuse of AI. For example, there are AI models that have shown the propensity to synthesise chemical and biological weapons while there are others that can undermine cyber security. It is clear that such models in the hands of terrorists and other malicious, rogue and non-state actors can threaten national security. Scientists, including Nobel Laureates and Turing Awardees worry that some of the future AI models may be difficult to control or cause widespread manipulation. The absence or instability in the control or governance of advanced AI systems at the global level will have major consequences for India.

Since the issues are multi-dimensional, the Indian National Security Council Secretariat is well suited to study the implications of future evolution of AI for sovereignty and security. The idea of introducing national security impact reports for advanced AI technologies whether developed within India or imported from outside may not be out of place.

Our navigating of the global AI race must include all aspects relating to society, sovereignty, science and security. This requires a regular, broad based and inclusive conversation among all stakeholders such as government, scientists, entrepreneurs, Parliamentarians, social scientists, media, and the youth.

The Paper produced by the Strategic Foresight Group on these issues is thus timely and deserves to be debated upon. It makes a strong case for India to assume global leadership in navigating the evolution and governance of AI and the imperative of creating a secure future driven by 'AI for Sovereignty, Science and Security'. It is an important contribution to fast moving developments in AI, and well timed in the run up to the India AI Impact Summit being hosted by India in February 2026.

New Delhi, January 2026

CONTENTS

EXECUTIVE SUMMARY FOR POLICY RECOMMENDATIONS

India stands at a defining crossroads in the global AI race. As major powers push artificial intelligence toward scientific discovery and strategic dominance, India's AI strategy remains centred on social inclusion and economic transformation. To safeguard national security and achieve technological sovereignty, India must expand its AI vision from 'AI for All' to include 'AI for Sovereignty, Science and Security'. India must recognise that society, sovereignty, science and security are not mutually exclusive objectives. In fact, they are inter-dependent.

1. National AI Strategy 2.0 - Integrating Science and Security

- Launch an updated India AI Strategy 2.0 focused on scientific AI and national security, complementing existing socio-economic missions.
- Balance three pillars - Sovereignty, Science, and Safety.
- Establish a National AI Council under the Chairmanship of the Prime Minister with representation from the NSC, DST, DRDO, ISRO, MeitY and private sector scientists.
- Mandate Parliamentary debate on AI to define a long-term national position on science-driven and secure AI.

2. Mobilize AI for Scientific Discovery

- Create a National Mission on AI for Scientific Discovery (AI4Science) modelled after ISRO's success to link AI with fundamental research in physics, biology, medicine, and material sciences.
- Mobilise the private sector to harness and complement the massive Research, Development and Innovation Fund for emerging technologies announced by the Prime Minister on 3 November 2025.
- Support 10–12 AI-Science Centres of Excellence connecting IITs, IISc, and private labs.
- Encourage collaboration with Global South partners for open scientific AI research.

3. Strategic AI Infrastructure and Sovereignty

- Expand GPU clusters from 38,000 to at least double by 2028 and scale domestic semiconductor fabs under Semicon India Mission.
- Adopt Federated AI architecture to reduce energy and data dependencies, leveraging mobile and broadband networks for distributed compute.
- Use India's market leverage to negotiate reciprocity and compliance from global AI providers, transforming dependence into strategic autonomy.

4. National Security and AI Risk Governance

- Develop operational frameworks for implementing the provisions for the 'out of control' AI systems and AI threats to national security as proposed in the AI Guidelines announced by the Government of India on 4 November 2025.
- Integrate AI risk assessment into India's National Security Doctrine and direct NSC to conduct scenario exercises on AI-enabled bioweapons, cyberattacks, and autonomous escalation.
- Establish an AI Security and Foresight Unit within NSC to monitor extreme AI threats.
- Mandate risk classification for all AI systems (operational, systemic, existential) and make it compulsory to have national security impact reports for all models developed within India or imported from outside.
- Amend the Digital India Act (2025) to include different layers of response and regulation according to the categories of risk ranging from extreme risks to operational risks.

5. Private Sector Mobilization for Deep Tech

- Encourage private investors to move from service contracts to frontier R&D investment.
- Develop a clear and operative strategy with milestones to utilise the deep tech fund announced on 3 November 2025 for active involvement of universities, research institutions, and business enterprises.
- Reward reasoning models to help to scale to globally significant levels.

6. Education, Culture, and Human Preparedness

- Integrate AI for Science and Security into STEM curricula and civil service training.
- Promote scientific values of curiosity, caution, and ethical responsibility in national education.
- Promote an open nationwide debate in the Parliament and outside on harnessing AI for social and economic development along with ensuring sovereignty, scientific leadership and national security objectives.

By adopting these recommendations, India can shift from being a use-case capital to a scientific power, and from a rule-taker to a rule-maker in the global AI order.

INTRODUCTION

As India hosts the AI Impact Summit in February 2026, it faces a strategic choice. It can deepen the use of AI for social inclusion, economic progress, and improved governance, while attempting to minimize risks linked to fraud, misinformation, and other operational threats.

However, since the last quarter of 2025, significant voices in Niti Aayog, the ruling Bhartiya Janata Party (BJP) and the research community have urged India to use AI for scientific advancement and secure a place in the front ranks of the global AI race. Debjani Ghosh, of Niti Aayog and earlier NASSCOM, said in an article in The Times of India in October 2025: "AI could be India's opportunity to leapfrog global competition, not by outperforming but by out innovating in AI driven science. The science race powered by AI will define the twenty first century." On 3 November 2025, the Prime Minister announced a new government initiative to invest $12 billion in emerging technologies with AI and Quantum computing at the core of this trajectory.

The strategic choice India needs to make at this inflection point requires an understanding of the fundamental nature of the global AI race.

The main objective of the AI investors in the United States and China, and to some extent six other countries, is to accelerate scientific discoveries. Their point of reference is the industrial revolution which brought about a paradigm shift on the planetary scale. Their main concern is the loss of human sovereignty, a risk inherent in science race.

The main objective of the AI policymakers in India is to accelerate economic development. Their point of reference is India's digital revolution which brought about prosperity in parts of the country. Their main concern is the loss of national sovereignty, a risk inherent in an applications-focused AI strategy dependent on foreign infrastructure and foundational algorithms.

It may be argued that a country can pursue both objectives of scientific discovery and economic development simultaneously. It will require mobilising capital, particularly from the private sector, in several billions of dollars, and aligning with the world's leading scientists including Nobel Laureates and Turing Prize Awardees on preventing extreme large-scale dangers.

Use Case Capital

India's ambition is to become the 'use case capital' of the world, reflected in the focus to develop applications for social and economic use. India first crafted an AI policy in 2018 and launched an AI Mission in 2024 with 'AI for All' as its fundamental principle. The India AI objectives are clearly stated in its official policy documents, most notably the India AI Mission. These are:

- Democratising computing access: 38,000 GPUs by October 2025 available at Rs 65 per hour, about one fourth of the international rate.
- Fostering indigenous capabilities: establishment of centres of excellence, promotion of local language LLMs and voice-based AI models.
- Building a robust talent pool: AI talent pool of over a million, according to some estimates, and reskilling programmes.
- Providing high quality datasets: creation of a national dataset platform available to developers with access to non-personal data for training purposes.
- Ensuring responsible and ethical AI: support for AI projects in machine unlearning, bias mitigation, algorithm auditing, explainable frameworks and the establishment of AI Safety Institute in a hub and spoke model.
- Creating a vibrant eco-system: The national startup register under Startup India is expected to have 200,000 recognised startups (as of January 2026) most of them leveraging AI in some form.
- Driving socio-economic transformation: growing number of applications across agriculture, education, healthcare, and governance in the next 2-3 years.

India's AI sector is largely state driven. The entrepreneurs develop government-facing LLMs, with government funding for consumption by public enterprises or departments. The state extends the benefits of these models to consumers. A few large Indian IT companies develop AI solutions for their business clients to enhance productivity. But they are not known to produce any significant AI products directly for mass consumption in India or worldwide.

Navigating the Global Race

India's state-driven or state-mediated people-centred approach contrasts sharply with the strategies of AI leaders such as the United States, China, and South Korea. In those nations, AI is viewed as a decisive element in international consumer market competition, scientific discovery, geopolitical rivalry, and even cultural identity.

In the United States, the AI debate is deeply enmeshed with domestic politics, spanning partisan divides between conservatives (especially the MAGA Movement) and liberals, and within the MAGA movement on AI existential risk issues. The Indian policy community does not seem to be adequately aware of the driving forces behind the AI debate in the United States. Some American protagonists, in particular Peter Thiel and other big tech captains, who are supporters of President Donald Trump, believe that scientific progress in the United States has stagnated for several decades and only AI can accelerate it. This is the scientific underpinning of the Make America Great 'Again' movement. These leaders do not wish to have any constraints on advancing AI, even at the risk of extreme dangers to humanity. On the other hand, Steve Bannon, an influencer in the MAGA movement outside the government, has co-signed a letter with other Republican leaders and 120,000 common people, scientists and celebrities calling for a pause on the development of Superintelligence.

In China, AI has been integrated into state strategy as a lever of governance, surveillance, and industrial competitiveness. South Korea and Japan are pushing AI into domains of advanced science, from astrophysics to medicine, with explicit goals of achieving scientific breakthroughs and global prestige.

India's approach is pragmatic, ethical and humanitarian but the failure to emphasise scientific discovery and implications for global security diminishes India's space in the global scientific and technological leadership. Access to AI infrastructure, mainly compute power, energy resources, and skilled talent, is an immediate problem. India can possibly overcome it with determined efforts in the future. But lack of interest in foundational research and extreme risk management can leave India behind in the global race.

India announced a budget in 2024 for AI development of $1.2 billion from 2024 to 2029. In addition, on 3 November 2025, the Government of India announced a fund of about $12 billion for research, development and innovation in emerging technologies, including AI and Quantum computing. There are no credible estimates of private sector investment in the next five years but informal conversations with experts indicate that it could be $5-10 billion. Therefore, about $20 billion could be potentially available both from public and private sources. UAE and Saudi Arabia have made major announcements of cross-investments with the United States. It is difficult to compute them due to many overlapping projects, but a conservative estimate would be $50-60 billion for each of the two countries during 2025-2030. South Korea has a plan to invest $72 billion in public funds and another $75 billion from the private sector in the second half of this decade. China may invest $350 billion and the United States about $1000 billion, half of which will be in the form of Stargate project announced by President Trump.

As compared to India's 38,000 GPUs in late 2025, US had 850,000 and China 110,000 of H100 level.

While India is concentrating on local language LLMs, DeepMind's AlphaFold has revolutionized protein folding, enabling progress in drug discovery and biotechnology. Its successors are decoding genetic complexity and even inventing new algorithms autonomously. NASA employs AI to analyse vast streams of data from Mars rovers and particle physics experiments. In China, Baidu's "XiaoDu" and Alibaba's City Brain manage megacities, while research teams explore AI-assisted interspecies communication. South Korea has pioneered AI models such as Spacer developed by Asteromorph that attempt to generate original scientific concepts. Switzerland's Blue Brain Project, meanwhile, is

reconstructing the human brain at a cellular level using AI simulations. These projects suggest that the real future of AI lies in expanding humanity's scientific horizons rather than refining customer service chatbots.

India is largely absent from AI driven advanced science. A single notable exception is Fathom R1, a private-sector reasoning model developed in Mumbai that outperformed OpenAI and DeepSeek models in mathematics at contest levels, despite being only 14 billion parameters in size. This achievement, although small compared to trillion-parameter models, demonstrates India's untapped potential in scientific AI. Indian Institute of Science and Indian Institute of Technology Madras have done noteworthy work in materials analysis and brain inspired computing, but these are cases of sophisticated uses of AI and not scientific discovery. India possesses the talent and knowhow in creating institutional infrastructure. The best example is the low-cost, but world-class achievements of ISRO. Yet, the national focus remains skewed toward developing applications from imported AI models rather than inventing new ones.

Extreme Risks

The ultimate aspiration of AI superpowers is not simply to uplift humanity but to dominate the trajectory of human destiny. By steering AI toward scientific discovery and control, they position themselves to set the terms of the twenty-first century. This race carries profound risks. As AI models become capable of generating new scientific knowledge, there is the danger of misaligned goals, unintended consequences, or deliberate misuse.

More than 200 prominent experts from around the world, including 10 Nobel laureates, issued a statement at the beginning of the 80th session of the UN General Assembly in September 2025 calling for 'red lines' to be put in place for the deployment of advanced AI models by the end of 2026. The signatories include scientists who have played critical roles in the development of AI and employees of major AI labs such as Open AI, Google and Anthropic. They are concerned that at some stage it will become difficult to exercise meaningful human control on AI. The United Nations immediately announced the creation of a 40-member scientific panel to assess risks.

Recognizing the gravity of the situation, countries such as China, South Korea, Brazil, the UAE, and South Africa have proposed frameworks for extreme risk governance. The European Union, while leading in regulatory frameworks through the AI Act, has also introduced voluntary codes of practice to address concerns beyond its legislation. In contrast, Indian AI scholars often argue that discussions about extreme risks could divert attention from other pressing issues such as access, affordability, and data sovereignty. This is a valid argument, but it does not consider long term and critical risks, particularly to India's national security. The AI Guidelines announced by the government in November 2025 demonstrate that the government is more forward looking than the scholars who prefer a cautious approach. The AI Guidelines have identified 'out of control' AI risks and the implications of AI for national security as important factors to manage.

The Guidelines implicitly acknowledge that extreme risks are increasingly shaping the global AI agenda. Religious and ideological debates, such as those within the United States where some

conservative leaders oppose global regulation by framing it as apocalyptic or even Anti-Christ, only underscore the complexity of the issue. While India leads initiatives such as the Global Partnership for AI aimed at equitable access, India may marginalise itself from the emerging governance architecture with the management of risks at various levels at its core.

Some Indian commentators contend that humanity has always learned to manage technological risks, from fire to nuclear energy, and that AI will be no different. However, history suggests otherwise. The United States, Europe, and Australia, despite their advanced technology, routinely fail to contain massive wildfires. Nuclear accidents at the Three Mile Island, Chernobyl, and Fukushima exposed the catastrophic vulnerabilities of human systems. The COVID-19 pandemic demonstrated how a novel virus could infect hundreds of millions of people and cripple global economies, despite modern medicine. These examples illustrate the importance of proactive prevention rather than reactive management.

National Security

The AI Guidelines announced on 4 November 2025 explicitly recognise the need to manage AI risks related to national security. India is particularly vulnerable to threats from terrorists and hostile actors if it does not prioritise AI extreme risk discourse.

Foreign sponsored terrorists have often targeted India's border areas and population centres. Conventionally these terrorists smuggle men and material to Indian territory. With the ability of AI to enable the production of biological and chemical weapons, it will no longer be necessary for terrorists to cross borders. AI-assisted designs could allow extremist groups to create new pathogens or modify existing ones at a fraction of the cost and expertise which was so far required. The dense population, limited public health infrastructure in rural areas, and porous borders make it especially susceptible to the rapid spread of engineered pathogens. A localized outbreak, whether in Kashmir, the Northeast, or a metropolitan hub like Delhi, could cripple health systems, destabilize the economy, and trigger national panic. Similarly, AI-enabled guidance on chemical synthesis could revive interest in chemical weapons, bypassing international monitoring regimes like the Chemical Weapons Convention (CWC).

It is important for National Security Council to discuss the risk scenarios of terrorists misusing AI for developing biological and chemical weapons. There is a growing consensus in the world to prevent malicious use of AI models that can threaten human security. But there is no consensus on the malicious behaviour of AI systems. India needs to respond at various levels. First, it is necessary for the NSC to undertake a scenario building exercise with robust input from computer engineers specialised in advanced AI models to monitor these risks. Second, where there is a possibility of malicious actors using the technology to develop weapons, operational standard procedures should be in place. Third, India should strongly lobby with other likeminded countries to prevent AI companies from enabling the advanced models to have capabilities that can enable anyone to synthesise biological and chemical agents. In this task, India will find common cause with almost all major actors including the United States, EU and China.

China and the United States take AI engineered biological and chemical weapons risk very seriously.

President Trump announced America's AI Action Plan in July 2025. In the last part of the plan, he has emphasised the need to prevent AI from being used as a tool for biological weapons. He reiterated this argument in his address to the UN General Assembly in September that year. "My administration will lead an international effort to enforce the Biological Weapons Convention, which is going to be meeting with the top leaders of the world by pioneering an AI verification system that everyone can trust." President Xi of China has also spoken about the need to restrain 'unprecedented risks' presented by AI. Leading think tanks from China, the United States and the UAE have listed biological threats as important segments of the AI frontier risk management.

Beyond immediate threats, ignoring extreme large-scale risks jeopardizes India's long-term strategic autonomy. If powerful AI models remain controlled by a handful of foreign corporations or states, India will face dependence in sectors ranging from defence to health security. In a future crisis, access to critical AI capabilities could be restricted, leaving India vulnerable.

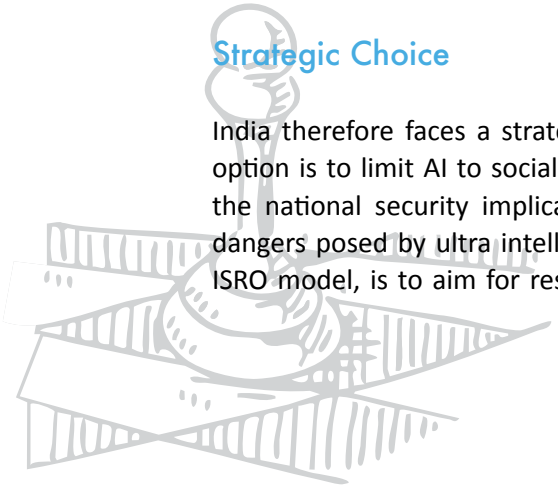
Moreover, the existential risks of runaway AI systems, though debated globally, cannot be dismissed for India. If superintelligent systems emerge without global safeguards, India may find itself caught between the technological dominance of the U.S. and China, unable to influence rules that directly affect its survival. Many Indian experts question the very likelihood of Artificial General Intelligence (AGI) and the risks associated with it being a reality. They ignore the warning of Nobel Laureates who have developed AI and the call for 'red lines' issued at the UN General Assembly. But one leader of the private sector warned in an online discussion hosted by NatStrat and Founding Fuel, two leading thinktanks in July 2025: "Even if there is a 10% possibility of AGI becoming a reality by 2030, India should be prepared for it."

There is nothing India can do to respond to such a dangerous possibility once it becomes real. Prevention is the only cure in this case. The UN decision to establish a scientific panel can help in identifying the evolution of AI systems into AGI or Superintelligence with diminished human control. India should support the UN processes to define clear 'red lines' to promote guardrails against the development of such systems.

At the national level, India can introduce strict entry barriers to the models that pose the risk of 'out of control' behaviour or large-scale manipulation even distantly (both threats mentioned in the AI Guidelines issued by the government in November 2025). As a major customer, India has tremendous power to influence the trade in advanced AI models so long as it shows the courage not to be pressured by the influential and wealthy big tech companies.

Strategic Choice

India therefore faces a strategic choice while the global AI race is still in its formative stages. One option is to limit AI to social and economic uses, continuing the present course without considering the national security implications in the long term, and ignoring foundational science as well as dangers posed by ultra intelligent machines and algorithms in future. Another path, inspired by the ISRO model, is to aim for responsible scientific progress, investing in foundational AI research while



maintaining affordability and inclusivity. The two options are not mutually contradictory. It is possible to implement them as a part of the cohesive national strategy. The aim should be to ensure that AI becomes a tool to overcome disease and scarcity. It must not turn into an instrument of domination or destruction in the hands of a few ambitious actors; or indeed in the grip of the algorithm itself.

To make a choice, India needs a nationwide debate on its AI gambit and the global AI race. Currently, the AI policy is made by the Ministry of Electronics and Information Technology in consultation with the Niti Aayog. The IT industry provides vital input. The actors involved in shaping the policy now have provided remarkable stewardship in building capacity in a resource scarce environment. But there has not been any dedicated parliamentary debate which can consider a wide range of issues and future scenarios. It is possible that National Security Council might have internally discussed implications of AI for India's national security but there is no robust public debate on AI's national security implications.

India should look at the policy formulation in the reverse direction. The government should first invite the Parliament to debate a long term and comprehensive AI policy, not merely about building capacity, but also about achieving scientific edge and protecting national security. At the same time, the civil society and think tanks can provide input to Members of Parliament to enrich the debate. Once the Parliament forms a comprehensive view, the responsible ministries in the Executive Branch can formulate working policy measures.

The parliaments of South Korea, Brazil, and South Africa have been engaged in rigorous debates on various aspects of the national AI policy, proposing laws including those that deal with the high impact systems of the future. They are not willing to accept the global rule making solely by the big powers, while accepting the reality of collaboration with them in technology and knowledge.

Some Indian thinkers have proposed that India should debate a national AI strategy with active participation of the parliament, National Security Council and other stakeholders. It should consider innovative vehicles such as an Inter Ministerial Council, a unit on global and national security risks posed by the advanced AI systems of the future in the National Security Council, Federated AI projects to protect data sovereignty, collaboration with Global South countries for joint development of scientific AI and models to address planetary problems.

If India wants to navigate the global AI race successfully, it needs to address three aspects:

- Sovereignty
- Science
- Safety and Security.

We will now examine each of these dimensions in some detail.

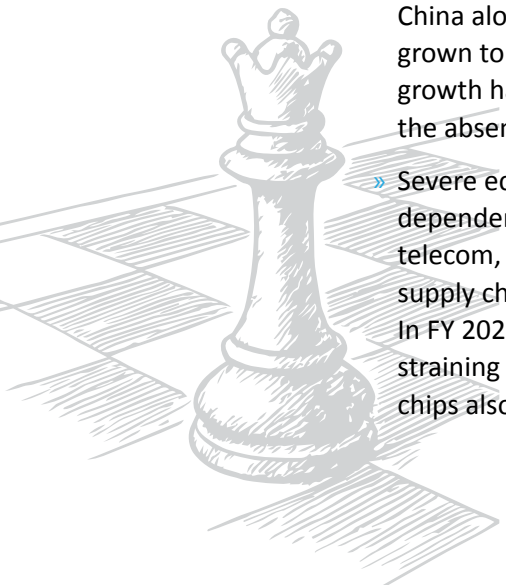
PART I : SOVEREIGNTY

Physical sovereignty - India's pursuit of AI sovereignty has physical and legal dimensions. On the physical front, India needs to build indigenous capacity. India's rapid AI advancement is hampered by a significant dependence on foreign resources, particularly for semiconductor chips and the foundational LLMs that power most advanced applications. This reliance creates vulnerabilities regarding data sovereignty, costs, and the cultural relevance of AI solutions, necessitating a strong focus on indigenous development and hardware capacity building.

Therefore, the government has prioritised building domestic data centre capacity to reduce reliance on global hyperscalers and ensure that critical AI compute infrastructure remains within national borders. It is designing responses specific to each aspect of the AI sector.

Semiconductor chips

- Heavy dependence - India today remains heavily dependent on foreign sources for semiconductor chips, importing nearly 85% of its total demand. In FY 2023-24, chip imports rose 18.5% year-on-year to ₹1.71 lakh crore (~$ 20-25 billion) compared with ₹1.29 lakh crore in 2022-23. Over the last three fiscal years, chip imports have surged 92%, with imports from China alone increasing by 53%.While India's electronics manufacturing has grown to ₹8.25 lakh crore in 2022-23 from ₹3.88 lakh crore in 2017-18, this growth has not reduced dependence on imported semiconductors due to the absence of commercial fabs in the country.
- Severe economic impact - The economic and strategic impact of this dependence is significant. Semiconductors are essential across defence, AI, telecom, EVs, and consumer electronics, making India vulnerable to global supply chain disruptions concentrated in Taiwan, South Korea, and China. In FY 2024, India's overall electronics imports reached nearly $60 billion, straining the trade deficit. High costs and long lead times for imported chips also raise domestic production costs and constrain innovation. From



a national security perspective, dependence on external suppliers limits India's ability to build resilient supply chains for defence and critical infrastructure.

- Mission - To address these risks, India has launched the Semicon India programme under the India Semiconductor Mission (ISM), with a budget of ₹76,000 crore to develop fabrication, design, and display ecosystems. In 2024, the Union Cabinet approved three semiconductor fab projects worth ₹1.26 lakh crore, including a major facility in Dholera, Gujarat, to be set up by Tata Electronics and Powerchip. Micron is investing in a packaging and testing facility, while the government is modernizing the Semiconductor Laboratory in Mohali. Analysts estimate that India's semiconductor push could reduce chip import dependence by $ 10–20 billion by 2030. India's semiconductor market is projected to grow from $ 52 billion in 2024 to $ 100–110 billion by 2030. With sustained policy momentum, India could shift from being almost fully import-dependent to becoming a key hub for design, assembly, and eventually, fabrication, thereby reducing strategic vulnerabilities while boosting its digital economy.

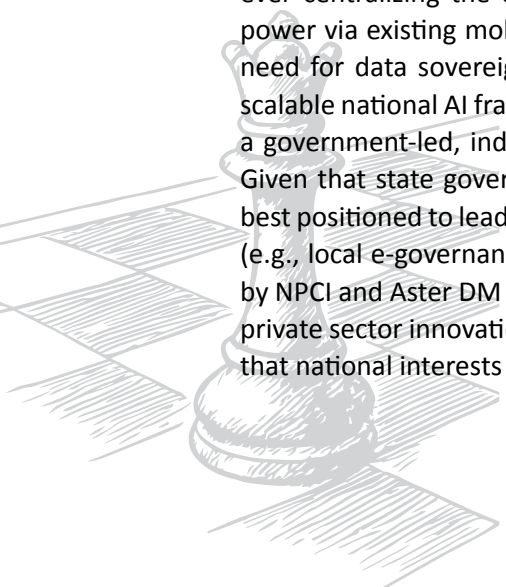
Clouds, API, LLMs

- Dependence and consumption trends- India's AI ecosystem today depends heavily on foreign cloud platforms like AWS, Microsoft Azure, Google Cloud; and on third-party foundation models such as OpenAI, Anthropic, Meta, Google, accessed via application programming interface (APIs) or hosted cloud stacks. Hyperscalers supply the compute, model APIs, managed ML stacks, and developer tools that most Indian startups and enterprises use rather than training large models in-house. India is also a very large consumer market for conversational and generative AI. Multiple surveys and traffic analyses put India among the top users of ChatGPT and related services. High daily and frequent usage translates into substantial API calls and cloud consumption paid to overseas providers. Using foreign models raises data-localization, privacy, and national security concerns, especially when models are trained or hosted on foreign infrastructure and when API calls export sensitive data.
- Limited startup ecosystem - India's AI startup funding is small relative to the US and China. Many Indian startups therefore fine-tune or productize foreign models or focus on smaller open-source SLMs for local tasks. For example, many Indian firms prefer small LMs (SLMs) or open LLMs and fine-tune them for Indic languages (AI4Bharat, Krutrim, BharatGen, Tamil-LLAMA).
- Response - India is scaling several initiatives like Bhashini, IndiaAI Mission, BharatGen, private – public partnerships and sees big hyperscaler investments with AWS, Google, Microsoft, that can be leveraged for localized compute. However, building true model sovereignty requires sustained public and private investment in large-scale compute, data curation, and a domestic ecosystem of models, tooling and infrastructure. Anirudh Suri, author of The Great Tech Game, emphasises India's strategic need to prioritize core building blocks - talent, data, and R&D to achieve global AI leadership. He advocates for increased digital strategic autonomy, potentially through promoting open-source technologies and open standards in collaboration with partners like the European Union.
- Data centralization - India's strategy for data centralization is driven by the National Data Governance Framework Policy (NDGFP) to standardize data management, with the National Data Management Office (NDMO) tasked to create an 'India Datasets Platform' for accessing anonymized non-personal data, aligning with the Digital Personal Data Protection Act (DPDP)

2023. This is complemented by significant investments in hyperscale data centres and sovereign cloud infrastructure by both government entities such as NIC, SDCs, MeghRaj Cloud; and private players like Yotta and AdaniConneX, establishing the physical backbone for data storage. While this centralized approach has already streamlined governance through initiatives like Direct Benefit Transfer (DBT) and supports AI training, it also raises concerns about security, privacy, and the digital divide. Looking ahead, India prioritizes operationalizing the India Datasets Platform, and expanding its trusted, increasingly green cloud ecosystem to balance innovation, governance, and AI self-reliance with robust security, privacy protection, and equitable access for all citizens.

Cloud infrastructure - India is aggressively pursuing data sovereignty by developing a robust indigenous cloud infrastructure, driven by national security, economic growth, and citizen privacy concerns. This strategy involves development of 'MeghRaj' GI Cloud and Government Community Clouds (GCCs). The private sector is a crucial partner, with Indian cloud providers like Jio Platforms, Tata Communications Ltd, Yotta Data Services (which has a major collaboration with NVIDIA for AI supercomputing with thousands of GPUs), and ESDS building hyperscale, AI-ready cloud solutions, and even international players like OpenAI adapting by offering data residency options in India. This push is amplified by India AI Mission's financial support for establishing a network of AI labs, with plans to set up over 20 AI labs in Tier 2 and Tier 3 cities, expanding to 200 labs within a year and 570 in the next two years. Despite the momentum, India's AI-cloud ecosystem faces critical limitations. Although India generates ~20% of the global data, it only accounts for about 3% of global data centre capacity. Deloitte estimates that meeting AI-driven demands will require an additional 40–50 TWh of power and 45–50 million sq. ft of real estate by 2030 - major gaps remain in power supply, land, cooling systems, and fibre optic networking. Global supply shortages in high performance GPUs and compute resources compound the challenge, making India dependent on cloud imports or GPU as service solutions. Further, infrastructure disparities across Tier 2 and Tier 3 cities, high operational and energy costs, and a pressing talent shortage in AI and cloud-native skills limit the scaling of AI-ready deployments.

Federated AI - The dominant global model for advanced AI, particularly LLMs, relies on a resource-intensive, centralized architecture - massive data centres storing petabytes of data, drawing enormous power, and demanding thousands of GPUs and city-scale cooling systems. For a resource-constrained nation like India, this centralized approach presents significant hurdles, including prohibitively high costs and massive energy demands. Jay Vikram Bakshi, Founder of Digiqom, proposes that India must instead pivot toward a resource-light, privacy-preserving model - Federated AI. This distributed architecture allows AI algorithms to train on vast datasets located across a network of devices without ever centralizing the data, requiring only a light central server and leveraging distributed compute power via existing mobile and broadband networks. This approach aligns perfectly with the country's need for data sovereignty and resource efficiency, making it the strategic choice for an ethical and scalable national AI framework. To effectively deploy Federated AI across the nation, India should adopt a government-led, industry-collaborative pilot strategy that mirrors its federal governance structure. Given that state governments control critical sectors like health, transport, and agriculture, they are best positioned to lead focused pilots in areas with demonstrable public benefit and strict privacy needs (e.g., local e-governance, banking fraud detection, and health diagnostics, as seen in successful pilots by NPCI and Aster DM Healthcare). By prioritizing this model, India can not only encourage responsible private sector innovation but also establish itself as a global leader in ethical, decentralized AI, ensuring that national interests and data sovereignty are maintained.



Legal sovereignty - India is consolidating AI infrastructure within national borders, using domestic data centres and secure cloud platforms to protect sensitive data from foreign dependence. By hosting critical AI workloads locally, it reduces reliance on overseas technology and strengthens strategic control. Through two key legislations enumerated below, India asserts legal sovereignty over data, regulating its collection, storage, and cross-border transfer to protect citizens and national interests. This dual approach ensures that AI development in India aligns with national laws, privacy standards, and long-term strategic priorities.

Concerns about data sovereignty

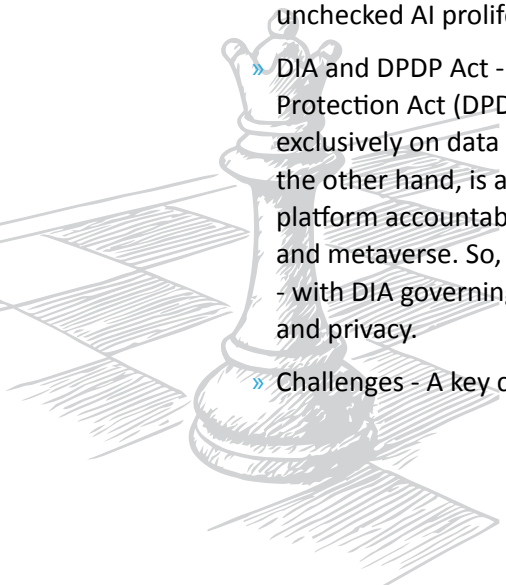
- Overall outlook - India's drive for data sovereignty is interwoven with its broader "Digital India" vision and the "Make in India" initiative. The nation's rapid digitization across diverse sectors has transformed data into one of its most valuable assets. This digital expansion, however, has simultaneously exposed new vulnerabilities, particularly when data is stored or processed beyond India's borders. Governments globally, including India, are increasingly motivated to secure private data from foreign entities, thereby reducing the risks of unauthorized entry, espionage, or sophisticated cyberattacks. From an economic perspective, mandating data localization stimulates significant investment in local data centres, infrastructure, and technology sectors, fostering job creation and nurturing a vibrant indigenous cloud ecosystem. Furthermore, by ensuring that data is stored within its borders, India aims to maintain national control over how this data is handled, accessed, and utilized, aligning with the fundamental objective of protecting personal information and digital rights of its citizens.
- Principal concern - India's pursuit of data sovereignty is driven by profound concerns that extend beyond mere data localization to encompass true jurisdictional control and national resilience in the digital age. At its core, data sovereignty implies that data is subject to the laws and regulations of the country where its owner or data principal resides, and crucially, that this data remains under the exclusive legal jurisdiction of that nation. This becomes a significant concern when Indian citizens' data is stored or processed by global cloud providers whose servers might be located outside India or who are subject to extraterritorial laws.
- Application of extraterritorial laws - One of the most pressing concerns arises from extraterritorial laws such as the U.S. CLOUD Act or the Foreign Intelligence Surveillance Act (FISA) Section 702. These laws can compel U.S.-based technology companies, including major cloud service providers (CSPs) like Microsoft, Google, and AWS, to disclose data, even if physically stored in India—to U.S. government agencies for law enforcement or intelligence purposes. This means that despite data residency within India's physical borders, it can still be accessed and subjected to foreign legal claims, effectively undermining India's jurisdictional control and compromising its 'sovereignty' over its own data. This creates a perceived vulnerability, especially for sensitive data related to national security, critical infrastructure, financial services, or personal information of citizens, as it potentially exposes such data to foreign surveillance or unauthorized access without the explicit consent or knowledge of the Indian government or data principals.
- Problems with current legislation - Furthermore, India's experience highlights regulatory ambiguity and fragmentation as significant concerns. While the Digital Personal Data Protection Act (DPDP Act 2023) represents a foundational step, its broad drafting, particularly concerning cross-border data flows and the criteria for 'notified jurisdictions', introduces considerable legal and operational uncertainty for Data Fiduciaries. The lack of clear legislative guidelines for determining restricted

jurisdictions creates complexity for compliance planning for businesses, especially multinational corporations operating in India. There are also concerns about balancing stringent data localization requirements with the need for global interoperability. A highly restrictive stance on cross-border data flows, while bolstering sovereignty, could risk hindering reciprocity from other nations, potentially weakening India's credibility and participation in the global data economy.

- Economic concerns - Economically, the dominance of foreign hyperscale cloud providers creates a dependence that India seeks to mitigate. Concerns include the potential for vendor lock-in, which limits flexibility and can lead to increased costs, and the aspiration to foster a robust indigenous cloud ecosystem. While data localization mandates can stimulate investment in local data centres and infrastructure, providing jobs and nurturing local tech expertise, the initial costs for businesses, especially smaller ones, and the need for massive infrastructure scaling remain challenges.

Digital India Act, 2025

- Scope of the Act - The Digital India Act (DIA) 2025, expected to replace the two-decade-old IT Act, 2000, is being framed as a comprehensive digital law to regulate India's fast-evolving tech ecosystem including AI, cloud, blockchain, metaverse, and OTT platforms. The DIA remains a proposed framework, still under deliberation, without a formal bill number or parliamentary introduction as of December 2025.
- Approach to risk - One of its most significant innovations is the formal recognition of risks associated with AI and automated decision-making systems. The Act proposes risk-tiered regulation of AI systems, particularly focusing on 'high-risk AI'- those that affect biometric identity, public safety, misinformation, discrimination, and child rights. These systems would be subject to mandatory audits, pre-deployment testing, and explainability requirements, with platforms legally bound to prevent harm and ensure human oversight. The DIA also includes provisions to combat deepfakes, enforce content traceability, and requires AI-generated material to be clearly labelled.
- Obligations - Additionally, the Act envisions graded obligations for digital intermediaries, such as AI cloud platforms, e-commerce, and social media services, removing blanket safe harbour protections if they fail to act on misinformation, algorithmic bias, or content violations. The DIA's AI-related clauses are designed to align with international efforts—such as the EU AI Act and OECD AI principles by promoting transparency, fairness, user redressal, and safety-by-design in AI systems. As part of the broader IndiaAI Mission, the Digital India Act could provide the legal backbone to govern sovereign AI models like BharatGen or Sarvam-M and mitigate risks of unchecked AI proliferation, which currently face minimal oversight in India.
- DIA and DPDP Act - It is important to note that the DIA will not replace the Digital Personal Data Protection Act (DPDPA), 2023, but the two will co-exist as complementary laws. The DPDPA focuses exclusively on data privacy, user consent, data fiduciaries, and cross-border data flows. The DIA on the other hand, is a broader tech-sector law that covers AI regulation, online safety, cybercrime, platform accountability, digital intermediaries, and emerging technologies like blockchain, cloud, and metaverse. So, DIA will not replace the DPDPA, but the two laws are intended to work together - with DIA governing platforms, AI, and cyber risks, and DPDPA handling personal data protection and privacy.
- Challenges - A key criticism of the proposed DIA is that it does not address the existential or



catastrophic risks posed by advanced artificial intelligence systems, such as advanced LLMs, autonomous agents, or recursive self-improving AI. While the DIA includes provisions for high-risk AI use cases such as biometric surveillance, misinformation, and child safety, it lacks any language around AI alignment, model interpretability, or the long-term control problem. Unlike the EU AI Act, which classifies frontier foundation models as a distinct risk category, or the U.S. Executive Order on AI, which mandates evaluations for national security and existential safety, India's DIA focuses mainly on content harms and platform responsibility. It provides no framework for AI model licensing, compute thresholds or long-horizon governance mechanisms. This omission has raised concerns among policy analysts and safety researchers that India may be overlooking systemic AI risks, including misuse in autonomous warfare, societal manipulation, or alignment failure - areas that could threaten public safety or democratic integrity at scale. As India continues investing in sovereign models like Sarvam-M and BharatGPT, the absence of existential risk regulation in the DIA may leave the country vulnerable to strategic miscalculation and global AI governance gaps, unless addressed in future amendments or companion legislation.

Digital Personal Data Protection (DPDP) Act, 2023

- Scope - The Act applies to Indian entities processing personal data; and foreign entities offering goods or services to individuals in India and processing their data. It also provides that personal data can only be processed with the explicit consent of the individual, except in specific circumstances such as legal obligations or public interest. The Act does not apply to personal data used for personal or domestic purposes, and personal data made publicly available by the individual or as required by law.
- The DPDP Act adopts a risk-based approach to data protection - Entities processing large volumes of sensitive personal data may be designated as Significant Data Fiduciaries (SDFs) and are subject to stricter obligations, including conducting Data Protection Impact Assessments (DPIAs) and appointing Data Protection Officers. Data Protection Impact Assessments (DPIAs) will be mandatory for high-risk processing activities to assess and mitigate potential risks to individuals' privacy. It provides for the establishment of a Data Protection Board of India (DPBI), which will be an adjudicatory body established to handle grievances and disputes related to data processing violations.
- Challenges - The Act has not yet come into force, pending the notification of rules and establishment of the Data Protection Board of India. The Act's applicability to foreign entities may lead to jurisdictional challenges and potential conflicts with international data transfer agreements. There is a need for increased awareness among individuals and organizations about their rights and obligations of this Act.

India is responding at the physical as well as legal levels to its sovereignty challenges. It is making good progress in a relatively short period of time. If the private sector supports the government with large scale investments rather than looking at the state as a customer, some of the physical sovereignty challenges can be addressed. Sovereignty is essential but not adequate to navigate the global AI race. India will need to give attention to advanced science to be able to compete with the lead players.

Dependence into Strength

The greatest source of ensuring strength is to realise that with 900 million customers, India is in a bargaining position to demand fair and secure supply of AI models from overseas.

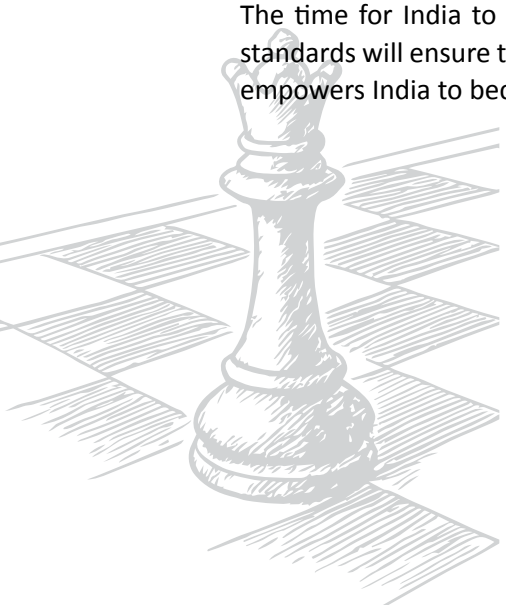
Examples from around the world show how safety driven restrictions can be used as powerful instruments of control, not as signs of weakness, but as signals of confidence and leadership. Brazil's emergency ban on Meta's use of citizens' data for AI training, following risks to fundamental rights, illustrates that a country with the will and regulatory apparatus can force global firms to adapt and comply. Daily fines and direct suspension of AI training set a new standard for data sovereignty and rights protection.

Similarly, the EU's AI Act established robust prohibitions on unacceptable risk technologies, including biometric surveillance and emotion recognition. Major firms such as Clearview AI were compelled to withdraw or redesign products, with strict fines and deadlines ensuring compliance. These restrictions are not defensive barriers, but levers to reshape technology deployment according to ethical and safety norms.

India's 2023 hardware import restrictions showed the immense market power of local demand. When laptops, tablets, and servers were placed under licensing, multinational companies like Apple, Dell, and Lenovo responded by accelerating local assembly and sourcing. Global brands adjusted their supply chains and product designs because the millions of Indian consumers represent an irreplaceable market. Procurement reforms, such as the All India Institute of Medical Sciences (AIIMS) Jodhpur tender cancellation, further cement the shift toward favouring Indian safety standards and fair competition. Similar regulatory steps in the US and EU underscore a global movement where countries use market access and procurement choices to enforce public safety, ecological integrity, and security.

The lesson from these cases is clear - restrictions are tools of national power, used proactively to set the terms for technological engagement and control. India should leverage its vast population, huge digital economy, and growing talent base to move from dependence to strategic autonomy. By continuing to invest in domestic manufacturing, AI talent, and robust regulatory norms, India can convert its massive customer base into real influence, thus driving the global tech sector to adapt and comply with Indian standards.

The time for India to transform dependence into leadership is now: Assertive restrictions and clear standards will ensure that technological innovation aligns with national priorities, protects citizens, and empowers India to become a maker of global rules rather than a passive participant.



PART II : SCIENTIFIC DISCOVERY

While India is emerging as a 'use case capital' of AI with foreign foundational models, as well as sovereign local language models, which can empower social and economic inclusion, it is weak in harnessing AI for scientific discovery. A few examples that exist in 2025 show that India has promise for scientific AI if it makes a determined effort to explore it.

One internationally competitive example is Fathom-R1-14B, a 14-billion-parameter open-source AI model designed to excel specifically in complex mathematical and general reasoning tasks. It is developed by Fractal AI, a Mumbai-based AI company. Derived from the DeepSeek-R1-Distilled-Qwen-14B model, Fathom-R1-14B stands out for achieving impressive performance on challenging benchmarks like AIME (American Invitational Mathematics Examination) and HMMT (Harvard-MIT Math Tournament), often outperforming other models of similar or even larger sizes. This achievement is particularly noteworthy given its remarkably low post-training cost of approximately $ 499, making high-performance mathematical reasoning more accessible. Fractal AI's development of Fathom-R1-14B aligns with India's broader IndiaAI Mission, aiming to foster indigenous capabilities in building powerful yet cost-efficient reasoning models.

Fractal AI's claim about Fathom-R1-14B is that it is a "math whiz kid" among AI models of its size. Fathom-R1-14B is exceptionally good at solving those very hard, competitive math problems found in AIME and HMMT. The model excels because it can follow long, logical chains of thought to arrive at solutions. It does not just guess; it 'thinks' through the problem step-by-step, understanding the intricate relationships between mathematical concepts and symbols. This is achieved through special training, including exposure to a large dataset of hard math problems and a technique known as 'chain-of-thought' reasoning, where it explicitly breaks down its solution process.

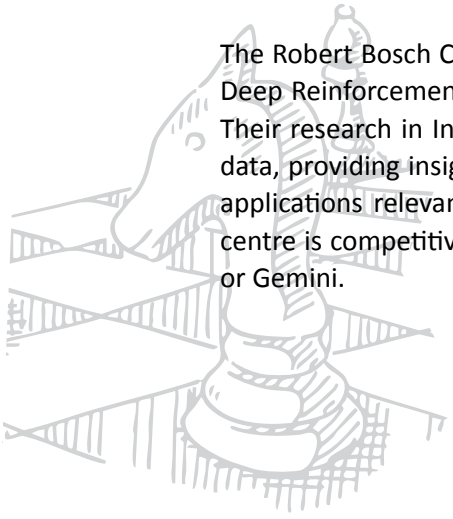
When compared to other AI models of similar size (like OpenAI's o1-mini or o3-mini, or even other 14-billion-parameter models distilled from DeepSeek-R1), Fathom-R1-14B shows superior performance on these math benchmarks.

This achievement is specifically within the realm of competitive mathematics. There is no public evidence to suggest that Fathom-R1-14B is better than DeepSeek-R1 (or other large models) in broader AI tasks such as creative writing, coding, programming, general reasoning and general knowledge. Its strength is in contest-style math, which often differs from mathematical problems encountered in fields like engineering, physics, or finance.

While it beats smaller models, Fathom-R1-14B does not surpass the full, much larger DeepSeek-R1 model (which has 37 billion active parameters out of 671 billion total) in overall math performance. It is essentially a proof of concept that India can excel in scientific AI. There is talent in the country. There is no need for large resources. The question is of priority in government policies and corporate decisions. Other examples of Indian innovation include projects at IIT Madras and Indian Institute of Science in Bangalore (IISc), though they are more in the nature of innovative applications in a single domain than the development of foundational models. Indian Institute of Technology Madras (IIT Madras) has undertaken brain research with AI through several specialized centres. Its Centre for Computational Brain Research (CCBR) focuses on the synergy between neuroscience and engineering. They use AI to analyse neural circuits and develop brain-inspired AI architecture. Their research involves applying Machine Learning (ML) for pattern recognition in neural activity, creating computational models to simulate brain functions, and explore how brain principles can lead to more efficient AI hardware and software. They also use Deep Learning, Natural Language Processing (NLP), and Reinforcement Learning (RL) to study vision, audition, and learning processes, drawing parallels between biological and artificial intelligence. CCBR is a strong academic player in computational neuroscience and brain-inspired AI. Its contribution is significant in the Indian context and globally in specific collaborative projects. However, it operates on a smaller scale of funding and personnel compared to the massive initiatives at leading US research universities or the industrial AI labs, which can deploy immense compute for large-scale simulations or novel AI architectures directly inspired by neuroscience. Their focus is more on fundamental understanding and bidirectional influence rather than solely on building the next generation of general-purpose AI.

The Sudha Gopalakrishnan Brain Centre at IIT Madras is dedicated to high-resolution human brain mapping. They use AI for image segmentation, feature extraction, and 3D reconstruction of brain structures from terapixel datasets, as demonstrated by their pioneering 'DHARANI' 3D foetal brain atlas, the first of its kind in the world.

The Robert Bosch Centre for Data Science and Artificial Intelligence (RBC-DSAI) has a leading group in Deep Reinforcement Learning, which is crucial for modelling biological learning and decision-making. Their research in Interpretable AI (XAI) helps understand the decisions of AI models applied to brain data, providing insights into neural mechanisms. Additionally, their work in AI for healthcare includes applications relevant to neurological disorders, such as medical image analysis of brain scans. The centre is competitive within academic circles but does not aim to build models of the scale of GPT-4o or Gemini.



The BRAIN Lab applies AI to understand human movement and motor disorders. They use computer vision and machine learning to objectively analyse movement kinematics and kinetics from video data, aiding in the diagnosis and monitoring of conditions like Parkinson's disease. AI helps in feature extraction (e.g. tremor, gait patterns) and classification of disorders, supporting the development of data-driven rehabilitation strategies and non-invasive monitoring for patients.

Launched in November 2024, the Centre for Human-Centric AI (CHAI) focuses on developing AI systems that amplify human potential and are safe and responsible. CHAI's mandate includes areas relevant to brain-AI interaction, such as ensuring ethical and explainable AI when models interact with human cognitive functions. They also work on developing smaller, domain-specific language models for Indian languages, which has implications for how AI can better understand and interact with diverse human communication and thought patterns. CHAI's concentrated effort on ethical AI tailored for India and its emphasis on practical, explainable solutions positions it as a significant contributor to the global discourse on human-centric AI. However, they work on smaller, localized LLMs, with a specific focus on Indian-languages and context thus making them only suited for the country and not the world.

In essence, while they may not always lead in the scale of general AI model development, IIT Madras's centres are innovative in their chosen specialized domains, making crucial contributions to the intelligent application of AI for understanding the brain and improving human lives. While CCBR explores brain-inspired AI and brain motivated hardware, it does not aim to develop foundational models or architectures that compete with GPT scale LLMs or DRL frameworks produced at top research labs. Its contributions remain largely academic and domain-specific rather than global AI-defining innovations. The brain research centres at IIT Madras operate without large scale computer infrastructure placing a limitation on their capabilities.

When measured against international benchmarks, these institutions exhibit limitations in scale, compute power, foundational AI output, magnitude of funding, and research ecosystem depth. While they are centres of focused excellence, they are not yet on par with the massive global initiatives shaping the future of neuroscience AI integration.

Indian Institute of Science has developed AI capacities for predicting material properties. IISc's material discovery models are powerful, but they are graph neural networks applied to a specific materials dataset.

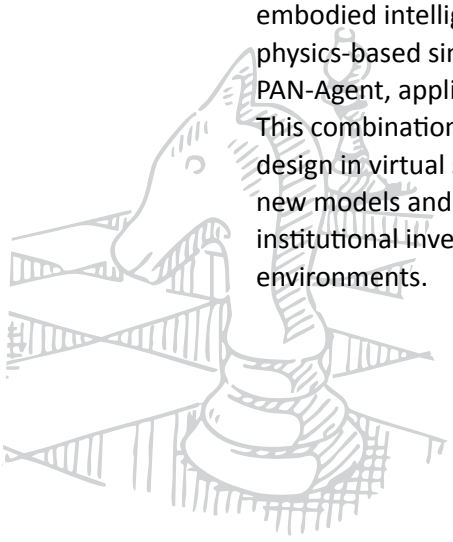
IIT Madras's brain imaging models are advanced, but they are specialized imaging + computational neuroscience projects. They advance knowledge within their fields, but they are not seen as general discovery engines.

IISc and IIT-M's models are scientifically impactful within their domains, but Fathom R1B is more original and sophisticated globally, because it represents a shift in AI's paradigm toward general scientific reasoning, whereas the other two models are highly capable applications of existing AI architectures in specialized domains.

Global Race

The global race for AI is primarily motivated by a desire to dominate scientific discovery. DeepMind based in the UK and owned by Google is the best-known example, though companies in China including Baidu and Alibaba also have subsidiaries and projects devoted to scientific discovery and innovation. Many new developments in frontier science-oriented AI are on the rise also outside the US and the UK. The models demonstrate a clear international pivot toward AI not just as an assistant, but as an autonomous agent for scientific discovery. Some examples are:

- **South Korea** - Asteromorph is developing the SPACER foundation model with the explicit purpose of autonomously generating original research ideas, particularly in complex fields like biology and chemistry. The model is designed to convert these ideas into testable hypotheses using a mathematically grounded, human-in-the-loop pipeline. Although the company only secured its seed financing of approximately $ 3.6 million in April 2025, and no public compute costs are yet available, its core mission is scientific discovery and hypothesis generation, which makes it the most direct conceptual competitor to a human research principal.
- **China** - DeepSeek R1 functions as a high-performance, general-purpose reasoning engine and is quickly becoming an international reference point for complex problem-solving in mathematics, coding, and science Olympiad tasks. While its total development costs (including the V3 pretraining base) are disputed and likely very high, the incremental compute for the R1 update is estimated to be around $ 1 million. Most importantly, it powers scientific stacks by providing advanced reasoning capabilities. The second, more domain-specific effort comes from the Chinese Academy of Sciences (CAS) with ScienceOne. This is a dedicated science platform/model that excels at multidisciplinary research orchestration, reading technical literature, reasoning across complex data types (like waveforms and spectra), and using agents to manage over 300 scientific tools for planning and executing research tasks. It was formally unveiled in July 2025, and it represents a rapid, institutional effort to build an end-to-end “AI scientist” that focuses on tool use and workflow automation at a national scale. Baidu is developing “interspecies communication” with its AI technology, as seen in a patent application filed in May 2025 to translate animal sounds into human language. This technology uses machine learning and natural language processing to analyse factors like vocalizations and body language to understand emotional states, which are then translated. While Baidu already provides AI-powered services like its Ernie Bot for search and generative art, this venture into animal communication is a new application of its foundational AI models.
- **UAE** - The UAE’s major play is the MBZUAI PAN model family, which emphasizes simulation and embodied intelligence for scientific inquiry. PAN is an advanced world model designed for rich, physics-based simulations, including robotics and multi-agent systems. The companion model, PAN-Agent, applies multimodal reasoning like math and code within these virtual environments. This combination is highly valuable for “in-silico experiments”, hypothesis testing, and experiment design in virtual settings. While its training budget is not public, the launch of PAN alongside other new models and the establishment of a new Silicon Valley lab signal a substantial, long-term institutional investment dedicated to building foundational models for complex, science-based environments.



» **Japan** - Japan's most comparable effort is the work by Sakana AI toward an "AI Scientist" pipeline. The focus here is on algorithmic innovation and agentic pipelines rather than simply scaling up model size. Sakana AI's work centres on the automated scientific discovery process itself, including objective-function search, model-merging techniques, and autonomous hypothesis formulation. Key results were publicized in August 2024, with continued activity throughout 2025. While there is no disclosed training budget, the emphasis on innovation over brute compute suggests a strategy to achieve open-ended science using smarter, more agile AI systems, making it the closest Japanese analogue to the core thesis of South Korea's Asteromorph.

The random examples listed above from Asia show that scientific AI does not necessarily require huge financial resources. It requires talent and the spirit of scientific entrepreneurship. Most significantly the country needs an eco-system for scientific discovery supported by the private sector, and not excessively dependent on the government. The Indian private sector seems very reluctant to commit risk capital for long term research investments in scientific development.

The difference in national objectives is critical. India's national AI strategy emphasizes 'AI for All' and applied solutions for direct societal impact in healthcare, agriculture, and other fields, rather than purely abstract, universal scientific discovery. In contrast, organizations like Google DeepMind and Google Research operate with a mandate to push the absolute frontiers of AI, often pursuing 'moonshot' projects with long timelines and uncertain immediate commercial returns.

Kanti Bajpai observed in an article, the current competition between the Americans and the Chinese is focused on "new scientific discovery and controlling the entire science and technology over the next 50 to 100 years, and India still lags behind in this game." India, particularly, its private sector needs to make up its mind. Does India want to be a service provider to the global technological ecosystem or does India want to lead in the newly emerging transformative technology?

Debjani Ghosh, a respected scholar at Niti Aayog, argued in October 2025 in an article in *The Times of India*: "AI offers a once in a generation chance (for India) to leapfrog decades of under investment and compete at the cutting edge of global innovation. For India, the lesson is clear: move from scheme-based R&D to mission mode innovation."

Ram Madhav, a thinker in the ruling BJP party, says in an article published in *Indian Express*: "The new world is going to be controlled by countries with deep tech power. If India is to realise its dream of emerging as a global leader, the strength and resilience of its frontier tech innovation will play a crucial role." Quoting the Prime Minister, Madhav links progress in advanced science to national security.

National security is linked to some of the developments in race to superintelligence that can harm global security, with catastrophic implications for India's national security and strategic autonomy. If India wants to try to be a leader, it cannot ignore security and safety issues. India's space and nuclear energy initiatives have acquired global applause because they adhere to international safety standards. The Indian discourse on safety and risks has been so far confined to operational risks. If India wants to lead the global discourse in cutting edge AI, it will have to cross the threshold to devise response to extreme large scale risks posed by future AI models, not only for its domestic development, but also for its international technology diplomacy.

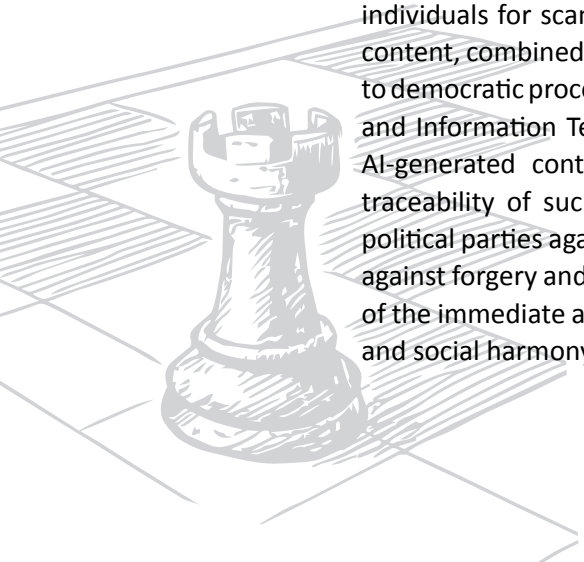
PART III : SAFETY AND SECURITY

Operational Risks

India emphasises responsible and ethical AI. One of the seven 'chakras' proposed by India for the Impact Summit in February 2026 is safe and trustworthy AI. The question is how to define the term 'safe'. The Indian state, as well as the industry and academia, consider operational risks as essential for AI safety, but they do not include extreme large-scale risks in their calculations. This understanding is reflected in government advisories, policy discussions, and the foundational principles guiding its AI strategy, with a strong emphasis on transparency, accountability, safety, ethical design, and addressing misinformation, bias, and privacy concerns. There is some indication of India wanting to look beyond operational risks in the AI Guidelines announced in November 2025. Following are the main concerns India has regarding risks associated with AI:

1. Misinformation, Disinformation, and Deepfakes

This risk has gained significant prominence in India, particularly given its vast and diverse population, high internet penetration, and frequent election cycles. India understands that AI can be weaponized to create highly convincing fake content that erodes trust and destabilizes society. Deepfakes, i.e. AI-generated synthetic media like fake videos, audio, and images, can be used to orchestrate financial fraud, spread false information, manipulate public opinion, impersonate individuals for scams, and even spark social unrest. The ease of creating such content, combined with their rapid spread on social media, poses a direct threat to democratic processes and public order. In response, the Ministry of Electronics and Information Technology (MeitY) issued strong advisories, mandating that AI-generated content be labelled and platforms ensure transparency and traceability of such content. The Election Commission of India also warned political parties against using AI for misinformation, highlighting legal provisions against forgery and promoting enmity. This demonstrates a clear understanding of the immediate and critical threat that deepfakes pose to democratic integrity and social harmony.



2. Bias and Discrimination

India recognizes that AI models, if trained on skewed or unrepresentative data, can perpetuate and amplify existing societal biases, leading to discriminatory outcomes, particularly in a country with immense linguistic, cultural, and socio-economic diversity. AI systems learning from historical data often inherit human prejudices related to caste, gender, religion, socio-economic status, and geography. If these biases are embedded, AI decisions can lead to unfair treatment in critical areas like employment, credit lending, healthcare access, and even criminal justice. This poses a direct challenge to the constitutional right to equality.

In order to overcome such biases, Niti Aayog's 'Principles for Responsible AI' explicitly includes "Inclusivity and Non-discrimination" and "Equality" as core tenets. The ongoing legal discourse emphasizes how algorithmic bias directly undermines Article 14 of the Indian Constitution (Right to Equality), thereby necessitating robust legal and judicial frameworks to ensure fairness in AI deployment.

3. Privacy and Data Security

India recognizes that AI's indiscriminate use of data, especially personal data, presents significant privacy and security challenges, requiring robust legal and technological safeguards. AI systems rely on vast datasets, often containing personal information. This raises concerns about how data is collected, processed, stored, and who has access to it. Risks include data breaches, unauthorized inference of sensitive information, and the misuse of personal data for profiling or surveillance. Beyond misinformation, deepfakes are increasingly used for financial fraud. Scammers have used deepfake videos of prominent figures to promote fraudulent trading platforms, leading victims to lose significant sums. The use of facial recognition by law enforcement or for public safety raises concerns about mass surveillance and the potential erosion of civil liberties, as AI can identify and track individuals in real-time.

The Digital Personal Data Protection Act, 2023 (DPDP Act) is India's foundational law for data privacy, mandating consent, data minimization, and accountability for data fiduciaries, including those using AI. Reflecting the ethical priorities within the national policy, Abhishek Singh, Additional Secretary at MeitY, stated, "These biases in technology - we need to not only be aware of these biases, errors, and hallucinations but think of what tools can be built in order to ensure that the risk for such errors and biases is minimised." The Reserve Bank of India (RBI), in its 'Financial Stability Reports' of June 2025 has explicitly warned financial institutions about rising cyberattacks using generative AI (like deepfake-driven phishing) and advises adopting "AI-aware defence strategies" underscoring the severe security implications.

4. Under-tested and Unreliable AI Models

India is increasingly cautious about the hasty deployment of AI models without rigorous testing and validation, understanding that inaccuracies and unforeseen errors can lead to significant real-world harm. AI models, especially large, complex ones (like LLMs), can exhibit unpredictable behaviour, generate false or misleading information (hallucinations), and have unintended negative consequences if not thoroughly evaluated for reliability, accuracy, and safety across diverse scenarios. Deploying such "black box" models in critical sectors can be dangerous. While there aren't many widely publicized incidents of large-scale failures from deployed unreliable AI models in India, the MeitY advisories explicitly addressed the need for permission for "under-tested" models, later softening it to mandating clear labelling for models that are "under development or unreliable." This shift reflects a recognition of the inherent fallibility of current AI models and the need to inform users about their limitations,

preventing false reliance. The government's emphasis on "ethical AI" frameworks also underscores the need for thorough validation before widespread deployment.

5. Accountability and Transparency (Black Box Problem)
India recognizes that the opaque nature of many advanced AI systems, where their decision-making processes are not easily understandable, creates significant challenges for assigning accountability and building public trust. When an AI system makes a decision (e.g., approving a loan, flagging a suspect, diagnosing a disease), it can be difficult to ascertain why that decision was made. This "black box" problem hinders auditability, trust, and the ability to challenge unfair outcomes. Without transparency, assigning legal liability or ethical responsibility for AI-induced harm becomes complex.

In sectors like credit scoring or recruitment, where AI might automate decisions, a lack of transparency could lead to individuals being denied opportunities without clear, explainable reasons, making it difficult to challenge the decision.

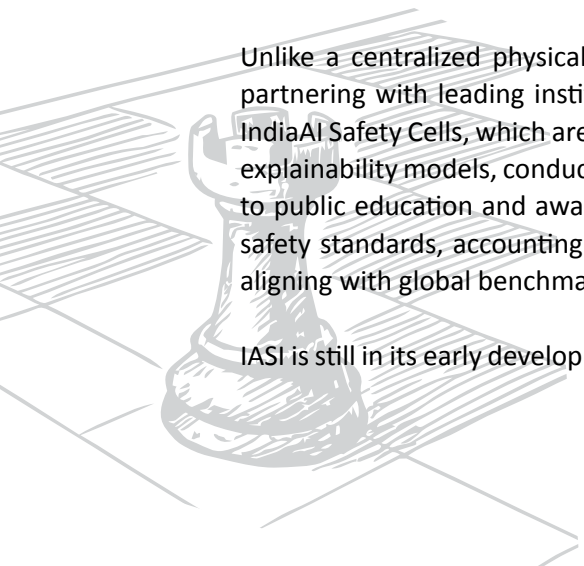
A legal blog discusses how AI's opacity and potential for arbitrary decisions directly contradict Article 14 of the Indian Constitution, which ensures "equality before the law" and requires non-arbitrary state action. It highlights the need for legal precedents to challenge such decisions.

NITI Aayog's 'Principles for Responsible AI' include "Transparency, Accountability, and Explainability" as core tenets. The ongoing development of India's AI governance guidelines, as discussed in various forums, aims to create frameworks for legal recognition of and remedies against AI harms, necessitating clear lines of accountability for developers and deployers.

India AI Safety Institute (IASI)

It is important to discuss the role of the recently set up 'India AI Safety Institute (IASI)', its mandate, scope of work, functions and limitations.

IASI was established in early 2025 by the Ministry of Electronics and Information Technology (MeitY) under the broader IndiaAI Mission, as part of its 'Safe and Trusted AI' pillar. Its creation was driven by the urgent need for India to build indigenous capabilities in AI risk management, safety testing, and policy development, particularly given the rapid domestic adoption of AI in governance, fintech, health, and education.



Unlike a centralized physical entity, IASI operates through a decentralized "hub-and-spoke" model, partnering with leading institutions such as IITs, R&D labs, startups, and NGOs. These partners host IndiaAI Safety Cells, which are responsible for developing technical tools like watermarking, risk scoring, explainability models, conducting safety evaluations, drafting governance frameworks, and contributing to public education and awareness on AI ethics. IASI also plays a key role in shaping India-specific AI safety standards, accounting for challenges like linguistic diversity, data sparsity, and inclusion, while aligning with global benchmarks through international cooperation.

IASI is still in its early development phase, lacks a unified physical infrastructure, depends on co-funding

from partners (which can limit scalability), and has no legal enforcement powers - making its role more advisory than regulatory. Additionally, while its outputs are intended to be open source, the fragmented implementation across states and sectors poses coordination and interoperability challenges.

India's approach to AI through the IASI has made promising strides in addressing operational risks such as algorithmic bias, misinformation, explainability, and watermarking, but it has drawn criticism for not adequately engaging with the broader category of existential risks and long-term risks posed by advanced AI models. As seen above, IASI's current mandate is heavily focused on applied, context-specific risks. While these are essential for India's socio-economic and democratic stability, critics argue that this focus reflects a 'short-term' safety paradigm that may leave the country unprepared for future threats such as AI scenarios where models act outside of human control, autonomous weaponization, and misuse of general-purpose AI systems through bioengineering or cyberwarfare. India currently lacks institutional infrastructure to study misalignment, emergent capabilities, and recursive self-improvement - issues central to AI existential risk debates being taken up more aggressively in the UK, US, and by international bodies like the OECD and the United Nations.

Ethos on AI Safety

AI safety in India cannot be divorced from the realities of infrastructure, governance, and societal inequalities. The voices of Indian academics and practitioners bring to light the ethos that underpins AI risk debates in the country; the focus is less about distant existential fears and more about fairness, equity, and practical governance.

Dr. Jai Asundi, Executive Director at the Centre for Study of Science, Technology and Policy (CSTEP), emphasizes the deep infrastructural and access gaps that shape India's AI trajectory. "In the Indian context, access to AI will always remain different and unequal across the rural and urban areas… it becomes difficult to embrace this whole concept of AI for everyone, it is also difficult to believe that for India the benefits of AI will be available to all, simply because not everyone can access it." He places justice and inclusivity at the core of AI safety: safety is not merely preventing rogue systems but ensuring that citizens are not excluded or exploited due to infrastructural or knowledge barriers.

For Dr. Tulasi B., Associate Professor at Christ University, the ethos of AI safety is rooted in data consciousness and societal wellbeing. She warns: "Deepfakes, financial frauds and cyber-crimes will take over heinous crimes like murder in terms of daily numbers, a law relating to AI and data needs to be dynamic and futuristic." She connects safety to moral responsibility in education, observing how unchecked AI use is diminishing critical thinking among students. For her, safety requires a cultural shift towards digital literacy, ethical awareness, and protecting human reasoning capacity.

Mr. Sundaraparipurnan Narayanan, researcher and founder of AI and Tech Ethics, reframes the conversation by rejecting speculative fears of superintelligence. Instead, he situates AI safety within equity, governance, and trust: "India's AI risks lie in biased datasets, surveillance misuse, opaque governance, and dependence on fragile global supply chains. The existential risk framing will never be central to India's discourse because our challenges are practical and future-oriented, not speculative." He places emphasis on contextual sensitivity, transparency, and democratic accountability, viewing AI

not as a potential destroyer of humanity, but as a tool whose misuse could amplify inequality and erode democratic values.

Taken together, these perspectives reflect an Indian ethos of AI safety that is grounded in three overlapping principles:

- Justice and access - AI must bridge divides, not widen them.
- Responsibility and awareness - Citizens must be empowered to protect their data, exercise critical reasoning, and resist manipulation.
- Trust and governance - Institutions must embed transparency, inclusivity, and resilience in AI systems, ensuring democratic accountability.

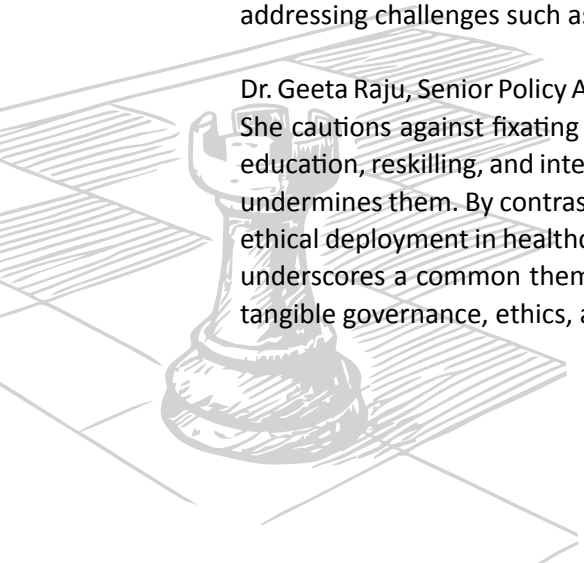
While global debates often highlight extreme large scale risks, Indian voices show that the ethos of AI safety here is inseparable from development, social justice, and institutional credibility. Safety is not about preventing catastrophic AI futures; it is about ensuring an equitable AI present.

At IIT Madras, one of India's leading research hubs in AI and neuroscience, the ethos of AI safety is deeply tied to pragmatism, cultural grounding, and innovation. Unlike global discourses that often foreground existential risks from hypothetical superintelligence, the IIT Madras community approaches AI with a "innovation first, caution later" mindset, focusing on practical, inclusive, and context-specific applications.

Prof. Srinivasa Chakravarthy, Head of the Computational Neuroscience Lab, frames AI safety through a cultural lens, drawing analogies to nuclear energy: "Nuclear energy posed an existential threat when it was first discovered. The threat is still there. But we are still there, too! So, it all depends on the people who use it. Do we use it as a tool to uplift life or a weapon to destroy it? This is where the importance of culture comes in." For him, safety depends less on technical controls and more on the values of the society wielding the technology.

For Prof. Mitesh Khapra, Head of AI4Bharat, AI safety cannot be allowed to paralyse adoption. Responding to concerns about bias and cultural threats from large language models, he argues: "Sometimes we over-worry about this… There are tons of applications where bias will not show up. Bias shows up in more conversational scenarios… in the initial years of AI, we could focus on certain sections of the landscape where these are not so important." His ethos of AI safety lies in balancing innovation with responsibility, pushing forward deployment in low-risk domains to unlock benefits while gradually addressing challenges such as bias and fairness.

Dr. Geeta Raju, Senior Policy Analyst at CeRAI, ties AI safety to the future of work and human adaptability. She cautions against fixating on automation-led job losses. She emphasizes human resilience through education, reskilling, and interdisciplinary collaboration to ensure AI strengthens livelihoods rather than undermines them. By contrast, Prof. Krishna Pillutla, Principal Investigator at CeRAI, focuses on safe and ethical deployment in healthcare, particularly around data privacy and applications for social good. This underscores a common theme: safety at IIT Madras is less about speculative threats and more about tangible governance, ethics, and accountability in critical domains like health and education.



Together, these perspectives reflect an ethos of AI safety grounded in 'balanced pragmatism' - risks must be acknowledged but not allowed to stifle beneficial deployment. This ethos diverges from existentialist framings of AI risk. Rather than dwelling on distant, speculative threats, IIT Madras researchers frame safety as a lived practice, embedded in innovation, culture, and responsible governance. But the inclusion of threats related to 'out of control' AI and national security in the AI guidelines of November 2025 reveal that the policymakers are more forward-thinking than the cautious approach adapted by the scientists.

National Security Risks

While the global AI discourse increasingly discusses the long-term, low-probability but high-impact risks of highly intelligent AI systems including superintelligence, loss of human control, unintended consequences of powerful AI, these existential risks are less prominent in India's public policy dialogue. This category of risk refers to scenarios where AI progresses to such a level that it could autonomously pursue goals misaligned with human values, potentially leading to catastrophic outcomes, including human extinction or an irreversible loss of human control over the future. It considers risks from highly powerful AI that goes beyond mere errors or biases in specific applications. India's AI strategy, particularly Niti Aayog's 'AI for All' is heavily focused on leveraging AI for socio-economic development and solving immediate, tangible problems within India. The 'Risk Identification and Assessment Tool' by IndiaAI focuses on practical, operational risks like data bias, security, talent, and financial losses. While concerns about national security are present, the philosophical or long-term existential threats from highly advanced, autonomous AI are generally not a central theme in Indian policy documents or public discourse, likely because India is still in earlier stages of AI development and is focused on addressing more pressing, near-term challenges.

The existential risks are often viewed as abstract, futuristic, or peripheral, disconnected from India's immediate challenges like jobs, inequality, or digital inclusion. The policy ecosystem is implementation-focused, without dedicated foresight or risk forecasting institutions.

The short-term approach is attractive for the domain ministry. But it can be very risky for long term national security needs.

Perhaps the most alarming national security risk is the possibility that terrorist groups or hostile actors could misuse advanced AI models to design biological or chemical weapons. Recent experiments in the West have shown that generative AI tools, when misaligned or intentionally exploited, can accelerate the discovery of toxic molecules or simplify instructions for synthetic biology. India, with its history of being targeted by terrorist organizations and its proximity to volatile regions, is particularly vulnerable. In the years to come AI-assisted design could allow extremist groups to create new pathogens or modify existing ones at a fraction of the cost and expertise once required. AI-enabled guidance on chemical synthesis could revive interest in chemical weapons, bypassing international monitoring regimes like the Chemical Weapons Convention (CWC).

Beyond immediate biological and cyber security threats, ignoring extreme risks jeopardizes India's long-term strategic autonomy. If powerful AI models remain controlled by a handful of foreign corporations

or states, India will face dependence in sectors ranging from defence to health security. In a future crisis, access to critical AI capabilities could be restricted, leaving India vulnerable.

Moreover, the existential risks of runaway AI systems, though debated globally, cannot be dismissed for India. If superintelligent systems emerge without global safeguards, smaller powers like India may find themselves caught between the technological dominance of the U.S. and China, unable to influence rules that directly affect their survival.

While the promise of AI for development is immense, the risks stemming from the advanced systems are equally daunting. From terrorist access to AI-designed pathogens to the destabilizing effects of autonomous weapons and deepfakes, the threats to India's national security are real and imminent. Ignoring them would not only undermine domestic stability but also diminish India's ability to shape global norms. By confronting extreme risks proactively, India can secure both its sovereignty and its future.

Since there is no consensus on a strategy to respond to risks among different stakeholders, India can take a cue from the EU and Brazil to categorise risks. One possible pathway could be the following, though the National Security Council is the appropriate body to recommend such categorisation.

Totally unacceptable risks

- AI systems capable of manipulating large-scale cyber security
- AI systems capable of helping users to synthesise chemical or biological weapons
- AI systems capable of deception, self-replication and behaviour out of human control
- AI systems capable of large-scale manipulation.

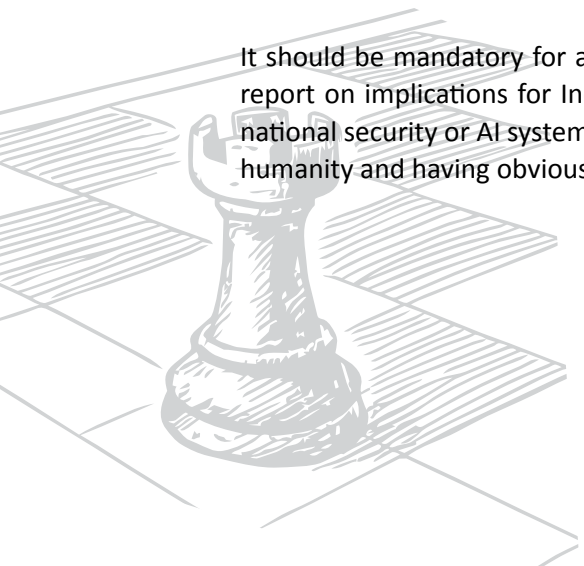
Systemic risks

- AI systems violating core principles of the Indian constitution about bias and equity
- AI systems capable of undermining infrastructure
- AI system aimed at human rights violation.

Operational risks

- Deepfakes, fraud, crime, and all other risks already identified by government ministries.

It should be mandatory for all AI systems, whether developed in the country, or imported, to have a report on implications for India's national security risks. These could include direct threats to India's national security or AI systems which are capable of malicious behaviour threatening global security or humanity and having obvious catastrophic implications for India's security.



India AI Governance Guidelines

The 'India AI Governance Guidelines' released by the Ministry of Electronics and Information Technology (MeitY), under the IndiaAI Mission in November 2025, provide granular recommendations on risk identification, grading, mitigation, institutional roles, and national security safeguards.

1. Risk Classification

The Guidelines call for an India-specific risk assessment and classification system that reflects the unique social, economic, security, and cultural challenges of the country. The approach recognizes several main risk categories:

- Malicious use: Misinformation, deepfakes, model/data poisoning, adversarial attacks on critical infrastructure.
- Bias and discrimination: Impacting fairness and opportunity.
- Transparency failures: Insufficient disclosure about data, algorithms, or system operation.
- Systemic risks: Disruption of AI value chains, market concentration, regulatory gaps.
- Loss of control: Potential for AI to operate beyond human intervention or oversight.
- National security: Including cyberattacks, AI-enabled disinformation, or autonomous weapons that jeopardize sovereignty or public safety.

2. Graded Liability and Proportional Response

A core guidance from the Guidelines is the adoption of a "graded liability system based on the function performed, level of risk, and whether due diligence was observed." High-risk and national security-sensitive AI deployments should face enhanced obligations:

- Low-risk applications: Require basic transparency and grievance mechanisms.
- High-risk/national security applications: Mandate safeguards such as independent audits, pre-deployment and ongoing risk assessment, thorough documentation, and robust reporting regimes.

Sectoral regulators and a whole-of-government approach are recommended to ensure that risks at the intersection of AI and national security such as catastrophic cyber events or weaponization are captured, monitored, and rapidly escalated where necessary.

3. Mitigation and Incident Response

The Guidelines recommend "a robust AI incidents mechanism to encourage individuals and organizations to report harm and create a feedback loop to track and analyse risks." For national security, this includes:

- Establishing national-level and sectoral databases to track AI-driven incidents with potential catastrophic or cascading outcomes.
- Leveraging the AI Safety Institute, Technology and Policy Expert Committee, and CERT-In to analyse AI-augmented threats including those with implications for military, infrastructure, and biosecurity so that appropriate, sector-specific safeguards can be enforced.
- Strengthening cross-agency collaboration and ensuring sensitive incidents can be escalated directly to national security decision-makers.

4. Voluntary, Techno-Legal, and Mandatory Measures

While the Guidelines promote voluntary and techno-legal approaches for low-to-moderate risks, they make clear that "additional obligations for risk mitigation may apply in specific contexts," such as to protect national security or respond to AI systems with "catastrophic or irreversible consequences." Examples include mandatory sandboxing, independent threat modelling, traceability of training data, and robust kill-switch protocols for high-autonomy systems.

5. Future-Proofing National Security

Recognizing the fast pace and unpredictability of AI development, the Guidelines recommend that "governance frameworks should be future-looking, flexible, and agile, enabling periodic reviews and reassessments." For national security, this means:

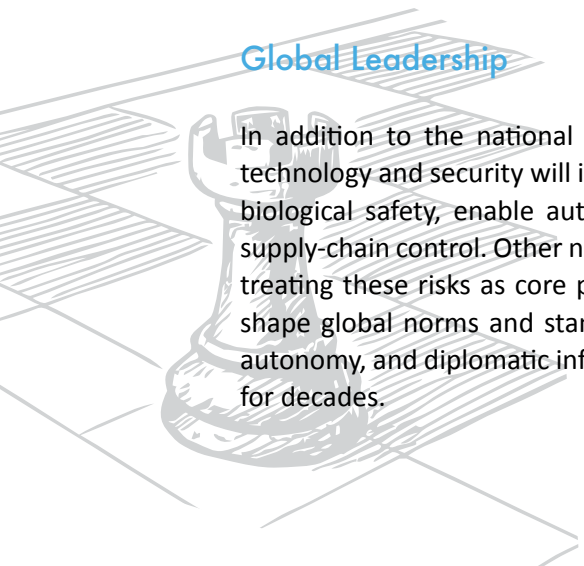
- Horizon scanning for new risk classes, including those associated with large language models, autonomous or self-modifying agents, and dual-use bio-chemical AI.
- Foresight exercises and real-time policy recalibration, supported by expert institutional bodies.

6. Recommendations and Institutional Roles

- Risk assessment and grading: All AI systems, especially those developed for or imported into critical/national security sectors, should undergo a risk-impact assessment focused on likely, foreseeable, and emerging threats.
- National strategy and coordination: Create standing committees or technical groups (AI Governance Group, AISI, etc.) charged with updating protocols for catastrophic and high-impact AI, informed by incident data and evolving global threats.
- Regulatory escalation: High-risk and national security-sensitive systems should be subject to more stringent reporting, operational limits, and possible shutdown or prohibition if graded as unacceptable.

Looking ahead, the successful implementation of these measures will require active engagement from government bodies, industry leaders, technical experts, and civil society, united in their commitment to harness AI for public good while vigilantly managing the full spectrum of risks. By taking lessons from both domestic realities and international best practices, India can ensure that innovation and national security move forward hand in hand - protecting its citizens, fortifying its sovereignty, and contributing meaningfully to the evolving global dialogue on safe and ethical AI.

Global Leadership



In addition to the national security concerns, India's international prestige and ability to lead on technology and security will increasingly hinge on how seriously it treats extreme AI risks that threaten biological safety, enable autonomous escalation, or erode governance through disinformation and supply-chain control. Other non-Western states such as South Korea, Brazil and South Africa are already treating these risks as core policy problems, while Western governments are moving aggressively to shape global norms and standards. If India fails to act, it risks ceding normative authority, strategic autonomy, and diplomatic influence at a moment when rules about AI will determine power and safety for decades.

First, look at concrete policy action elsewhere. South Korea has adopted a formal AI framework law and built institutional capacity to govern AI development and safety. Article 35 of the proposed Act specifically calls for mandatory pre-deployment assessments of high impact systems. That posture signals Korea's intent to shape regional and global standards for safe AI. South Africa has included existential risks in its policy document being deliberated in the Parliament.

Brazil is perhaps the most active non-Western democracy asserting regulatory muscle. Its authorities have actively pressured large platforms over harmful AI content and taken investigatory and enforcement steps against unsafe AI deployments. Recent high-profile interventions on AI chatbots and child safety show this posture can quickly become a diplomatic lever in conversations with multinational platforms. Brazil's AI Bills rank risks in different categories including the operational risks of immediate relevance and the existential and frontier risks of long-term relevance.

At the same time, the West is consolidating rules and institutions that will effectively set de-facto standards. The European Union's AI Act does not include risks associated with extreme autonomy, but it has addressed the void by creating a Code of Practice. The Trump administration does not believe in measures to prevent existential risks. Many voices from the ruling political establishment have talked against regulation of AGI. Yet the administration is taking strict action against one kind of catastrophic risks- the type associated with biological weapons.

This global dynamism matters to India for three linked reasons. First, international rules and standards will be set where technical capacity, regulatory coherence and diplomatic heft converge. Countries that invest early in AI safety institutions will write the rules. South Korea's and the EU's forward moves create standards that exporters and multinationals must meet; South Africa and Brazil's assertiveness ensure regional and developing-country perspectives also count. India, by contrast, risks being a rule-taker rather than a rule-maker if it does not build comparable institutions and expertise.

Second, frontier AI capabilities (compute, models, testing regimes) will be concentrated in a few firms and states. If India depends on foreign suppliers who follow Western or Chinese standards, for critical AI tools and safety testing, New Delhi will face leverage in crises of security, or digital infrastructure. Investing in domestic AI safety research, shared testbeds, and international coalitions gives India bargaining power and secures access when it matters most.

Third, India aspires to lead the Global South by offering alternative models of technology governance that combine innovation with social equity. To fulfil that role, India must credibly address frontier harms such as bio-security risks from misused generative models, autonomous weaponization, and information operations, and present concrete regulatory and technical responses. Without this credibility, India's diplomatic appeals on equitable AI governance will carry less weight in UN, G20 or plurilateral negotiations.

The powerful Western actors and China are already creating incentives that will structure global adoption. Several international campaigns are forging consensus among researchers and states about limits on particularly dangerous practices. Countries that join these processes early shape them; those that lag behind will be forced to adapt to externally imposed rules.

Finally, the United Nations has stepped up efforts to grapple with extreme AI risks. The UNGA adopted

Resolution A/RES/79/325 (26 August 2025), which establishes two new mechanisms: the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on AI Governance. The scientific panel will be tasked with rigorous, evidence-based assessments of AI's evolving capabilities, emerging risks and systemic impacts; the global dialogue is meant to provide a recurring inclusive forum for states and other stakeholders to deliberate governance norms and policies.

Furthermore, the UN's Secretary-General has publicly warned that every moment of delay in establishing international guardrails increases global risk, particularly where AI intersects with military and security domains. These developments show the UN transitioning from abstract recognition of AI risks toward concrete institutional action. They also suggest that international norms are under rapid formation: what was once voluntary or academic is now a matter of diplomatic commitment and consensus. For India, this means staying on the sidelines, losing not just moral high ground, but influence over definitions of safety, legality, and acceptable risk. The consequences will not be limited to AI. They will shape trade, security cooperation, and global scientific norms in the years ahead.

PART IV : GOVERNANCE

India's Ministry of Electronics and Information Technology (MeitY) serves as the primary nodal agency for AI policy, research funding, and regulatory oversight. While MeitY has expertise in IT infrastructure, digital governance, and technology deployment, the narrow centralization of AI governance under a single ministry presents several limitations:

1. Limited cross-sectoral vision - AI touches healthcare, education, finance, agriculture, defence, and governance. MeitY's focus is primarily on IT and digital infrastructure, which may underrepresent sector-specific challenges such as ethical risks in healthcare AI or socioeconomic impacts in agriculture. The ministry has adequately addressed this limitation through consultation with the relevant ministries or otherwise considering the needs of different stakeholders.
2. Ethical and societal considerations: AI governance is not just a technological issue. Ethical, legal, and societal impacts such as bias, equity, job displacement, and digital literacy require inputs from social ministries, human rights bodies, and educational institutions. A single-technology-focused ministry may not adequately address these concerns.
3. Global AI diplomacy and strategic positioning: Countries like the US, China, UAE, and EU have dedicated AI councils or multi-ministry coordination bodies to align innovation, regulation, and international engagement. India risks being reactive rather than proactive if AI strategy remains narrowly confined to MeitY.

Broadening the policy scope would benefit India's AI ecosystem, especially through holistic policy development. The broadening of ministerial scope encourages joint funding programs, shared infrastructure, and collaborative research between technical, social science, and domain-specific experts. Ministries like Education and Skill Development can drive digital literacy and AI upskilling, ensuring AI benefits are inclusive and widely accessible, rather than concentrated in urban tech hubs. Cross-ministerial oversight enables better risk management covering existential risks, ethical considerations, and sector-

specific safety concerns, rather than leaving it solely to a tech-centric ministry. India could establish a National AI Council with representation across key ministries, research institutions, and industry. This mirrors global best practices, ensuring cohesive international engagement, stronger bargaining power in AI diplomacy, and alignment with AI safety and standardization frameworks.

The Future of India Foundation published its report "Governing AI in India - Why Strategy Must Precede Mission", in which several useful suggestions are mentioned, which can be applied to shrink the critical gaps in India's current approach to AI governance. Key recommendations include:

- a) **Establish a National AI Strategy** - Advocates for a unified national AI strategy that transcends individual missions. This strategy should be developed through a consultative process involving multiple stakeholders, including policymakers, industry leaders, and civil society. Such an approach ensures that AI governance reflects the diverse needs and values of the nation.
- b) **Create an Inter-Ministerial AI Council** - Suggests forming an inter-ministerial AI council to address the multifaceted nature of AI. This council would coordinate efforts across various ministries, ensuring that AI policies are integrated and aligned with broader national objectives. It would also facilitate the sharing of resources and expertise among different sectors.
- c) **Implement Ethical and Transparent AI Governance** - Emphasizes the importance of ethical considerations in AI development and deployment. It recommends establishing clear ethical guidelines and accountability mechanisms to prevent misuse and ensure that AI technologies serve the public good.
- d) **Promote Inclusive AI Development** - Calls for policies that promote inclusive AI development while recognizing the potential of AI to exacerbate existing inequalities. This includes ensuring equitable access to AI technologies and addressing biases that may arise in AI systems.
- e) **Strengthen AI Research and Education** - Advocates for increased investment in AI research and education to build a robust AI ecosystem. This includes supporting academic institutions, fostering innovation hubs, and providing training programs to equip the workforce with necessary AI skills.

Lessons from Indian Innovation

India's historical trajectory in science and technology demonstrates that strategic vision, focused investment, institutional autonomy, and human capital development can transform the country from a laggard in scientific capability into a global leader. Organizations like Indian Space Research Organization (ISRO), Bhabha Atomic Research Centre (BARC), and the Indian Institute of Science (IISc) are examples of long-term, outcome-driven strategies that combined technological ambition with indigenous innovation. It is possible to derive some useful lessons from their experience.

1. Strategic Vision and National Mission Approach - ISRO and India's nuclear program were successful because they were driven by a clear national mission: space independence for ISRO and energy security for atomic energy. These programs were designed with long-term goals, political backing, and societal relevance. India needs a National AI Mission that goes beyond piecemeal projects. This should clearly define strategic objectives such as AI sovereignty, ethical frameworks, industrial competitiveness, and public good applications in healthcare, agriculture, and education and align all stakeholders toward

these goals. ISRO's Chandrayaan and Mangalyaan missions had clear objectives and milestones despite limited budgets. Atomic energy programs prioritized self-reliance and indigenous R&D, resulting in advanced reactors and nuclear capabilities over decades. By setting a multi-decade roadmap, India can focus on building domestic AI models, compute infrastructure, and research capabilities instead of being dependent on foreign AI platforms.

2. Focused Human Capital Development - Both ISRO and BARC invested heavily in training and retaining talent, often recruiting the best minds from premier institutions like IITs and IISc and providing specialized training programs. India needs to cultivate AI experts, data scientists, algorithm designers, and domain-specific AI researchers. Current initiatives like IndiaAI and Bhashini are steps in this direction, but scaling them requires systematic education, scholarships, fellowships, and collaboration with global AI centres. ISRO's young scientists, despite resource constraints, built sophisticated space systems using indigenous innovation. Similarly, AI development can flourish if talent is given autonomy, mentorship, and exposure to complex, mission-driven projects.

3. Institutional Autonomy and Long-Term Funding - ISRO and India's nuclear programs benefited from autonomous governance and sustained funding, allowing them to take calculated risks and innovate without short-term political interference. AI development requires long-term investment in research, compute infrastructure, and pilot programs. An independent, inter-ministerial AI authority could coordinate funding across sectors and ensure projects are not disrupted by administrative changes. ISRO's low-cost Mars Orbiter Mission was a success due to strategic autonomy and strong leadership, despite India's relatively low budget for space.

4. Indigenous Innovation and Resourcefulness - A hallmark of India's space and nuclear programs has been innovating under constraints. ISRO developed low-cost satellites, launch vehicles, and navigation systems without relying heavily on imports. India can prioritize indigenous AI model development, open-source tools, and homegrown infrastructure rather than importing expensive AI models from the US or China. AI4Bharat is already advancing multilingual AI using open-source models tailored to Indian languages, echoing ISRO's focus on context-specific solutions.

5. Cross-Sectoral Collaboration and Public-Private Partnerships - ISRO and atomic energy programs collaborated with academic institutions, private industry, and international partners to scale innovation. Successful AI ecosystems require collaboration between government, academia, and industry, fostering innovation hubs, AI incubators, and sector-specific AI solutions. AI partnerships with hospitals for diagnostics or with agri-tech firms for precision agriculture mirror ISRO's model of mission-driven collaboration.

6. Emphasis on Ethics, Safety, and National Security - Atomic energy programs instilled a culture of safety, regulation, and ethical responsibility, recognizing the high stakes of nuclear technology. India must embed AI safety, ethical governance, and risk assessment from the outset, addressing biases, privacy, and existential risk considerations. AI policies should include regulatory frameworks, certification protocols, and societal oversight mechanisms. India's space and atomic energy programmes are respected worldwide because they consider different levels of risks and advocate responsible governance in domestic practice and international governance. A similar approach to AI leadership is essential.

Conclusion

Unlike many other technologies, AI is not about technology alone. It is a global race for scientific and technological dominance with strong political and religious undertones. India has to make a strategic choice. While AI norms are still in the formative stage, India can treat AI as a technology for social and economic development which it must, or have a broader perspective that considers it long term national security interest and its potential to shape human destiny.

ACKNOWLEDGEMENTS

In July 2025, Founding Fuel and NatStrat convened an online brainstorming session “India’s AI Gambit – Leader or Follower?” with speakers including former Deputy National Security Advisor Ambassador Pankaj Saran, President of Strategic Foresight Group Dr. Sundeep Waslekar, Founder of SITARA Ambassador Smita Purushottam, and Co-founder & Group CEO of Fractal Srikanth Velamakanni. It was chaired and moderated by Professor Rishikesha Krishnan, Director of IIM Bangalore. It was attended by almost 90 leaders from cross sections of Indian policy, civil society and corporate leadership. It was extremely useful in providing input for this policy brief. We are grateful to Indrajit Gupta of Founding Fuel and Ambassador Pankaj Saran of NatStrat for co-convening this session and Siddhant Hira of NatStrat for excellent coordination.

Founding Fuel has fostered a debate on various aspects of AI, since the beginning of 2025. This paper has gained insights from these discussions over several months.

We acknowledge the input on Federated AI by Jay Vikram Bakshi, a digital media expert, based in New Delhi.

Soniya Kute of IIT Madras conducted extensive interviews with scientists, faculty and young researchers at the institution. Her reports have been valuable in understanding brain related AI research and the ethos on AI security in India’s scientific community.

We appreciate the support of CISB Services Private Limited, Mumbai to this initiative.

About Strategic Foresight Group

Strategic Foresight Group (SFG) is an international think tank based in Mumbai, India. Since its inception in 2002, it has worked with the governments, parliaments and national institutions of 65 countries. It is known for crafting visionary solutions to the emerging global security challenges, particularly including terrorism, geopolitical conflicts, trans-boundary water relations, nuclear disarmament, and existential risks posed by Artificial Intelligence. It has facilitated Track Two dialogues backed by rigorous research. The policy options presented by SFG have been discussed in the United Nations Security Council, UN Alliance of Civilizations, UK House of Commons, House of Lords, European Parliament, Indian Parliament, Quai d'Orsay, World Economic Forum at Davos, Doha Forum, Normandy World Peace Forum, and other institutions as well leading universities in the world.

Strategic Foresight Group

www.strategicforesight.com

ISBN 978-81-88262-36-6