State of AI Governance, 2024

A report analysing AI governance measures across countries, companies, and multistakeholder gatherings

Bharath Reddy, Rijesh Panicker, Sridhar Krishna, Arindam Goswami, Anwesha Sen, Adya Madhavan

A report analysing AI governance measures across countries, companies, and multistakeholder gatherings

Contents

1 About the State of AI Governance Report

The rapid progression in artificial intelligence capabilities has far-reaching impacts. It has the potential to significantly boost economic productivity, disrupt labour markets, and alter the balance of power between countries. The widespread diffusion of this general-purpose technology presents significant governance challenges that are amplified by the ongoing geopolitical competition for AI dominance.

It is against this backdrop that the Takshashila Institution, an independent centre for research and education in public policy, presents its inaugural State of AI Governance Report. This report provides a systematic comparative analysis of AI governance approaches across different countries, revealing their strategic priorities. Further, the effectiveness of corporate self-regulation initiatives, and the progress of multistakeholder collaborative efforts are also analysed. It concludes by offering predictions in these areas in the coming year.

This annual report will track key developments, analyse trends and offer informed predictions for the AI governance environment. It is intended to provide policymakers, analysts, and interested citizens with insights that help navigate the evolving AI governance landscape. You can learn more about our work at takshashila.org.in/

2 Executive Summary

This report analyses AI governance in three contexts – countries, companies and multistakeholder gatherings.

2.1 Countries

- Being at the forefront of AI innovation, the US favours a pro-market regulatory environment while prioritising geopolitical considerations to maintain its competitive advantage.
- The EU has focused on creating a comprehensive regulatory framework that prioritises transparency, accountability, and the protection of individual rights.
- China's approach prioritises national security and favours heavy state control in enforcing regulations and driving innovation.
- India has opted for a light-touch regulatory environment while investing in developing Indigenous AI models for Indian use cases.

2.2 Companies

- Many companies are proactively establishing principles, guardrails and transparency and disclosure norms that guide how they build or use AI. These initiatives sometimes go beyond what is strictly expected by the regulatory environment in which they operate and are intended to build trust with users.
- However, reporting on AI governance efforts by companies is not standardised, and the details of specific initiatives vary considerably. It is also unclear to what extent these efforts involve meaningful external scrutiny. The report analyses the governance initiatives of a few companies operating at different parts of the AI value chain.

2.3 Multistakeholder Gatherings

- Various multistakeholder gatherings, such as the AI Summits and the Global Partnership on AI, have been formed to raise awareness and coordinate AI governance efforts among different countries.
- Although most of these groupings do not have binding commitments or backing from all members (for instance, the US and EU refusing refused to sign the declaration on inclusive and sustainable

AI at the AI Action Summit in February 2025). However, they serve as a platform to highlight important concerns and drive convergence in AI governance efforts.

Finally, the report ends with some predictions on what we can expect in AI governance in the coming year.

3   Timeline of AI Governance Events

A timeline of significant AI governance events across countries, companies, and multistakeholder groupings is presented below. The timeline focuses only on AI Governance events and does not list milestones related to advancements in the technology.

Figure 1: Timeline of AI Governance Events

4   Analysis of AI Governance Measures Across Countries

AI governance measures often address multiple objectives. These include ensuring transparency and accountability, promoting innovation, addressing geopolitical considerations, enabling state capacity to implement the measures, and promoting societal well-being.

The authors have analysed and compared the country-specific AI governance measures across these different criteria. There is some subjectivity in this comparative analysis, but the authors feel it is a useful representation of how countries are pursuing these different AI governance priorities.

The United States of America, the European Union, China, and India are selected as countries/regions for comparison. These have been chosen for their significant role in influencing the path of innovation, governance or adoption of AI.

Figure 2: Comparison of AI Governance Measures Across Countries

A description of the different criteria is provided below:

- **Transparency and Accountability**: Assesses the extent to which governance frameworks attempt to promote transparency and accountability. This includes examining governance instruments such as evaluation and disclosure requirements, licensing requirements, penalties for non-compliance, and grievance redressal mechanisms.

- **Promotion of Innovation**: Evaluates how regulations foster innovation by creating an enabling environment. This includes examining the quantum of funding for AI infrastructure, restrictions on market participation, education and skilling initiatives and maturity of the R&D ecosystem.

- **Geopolitical Considerations**: This assesses the extent to which policy decisions address geopolitical priorities. It includes assessing whether a state can secure access to the building blocks of AI and deny access to other countries. Relevant policy measures include export controls, investments in domestic infrastructure, promotion of open-source technologies, and policies that reduce vulnerabilities in the value chain.

- **Societal Wellbeing**: Assesses how regulations address broader societal concerns, such as protecting individuals from risks and harms from the adoption of AI in various sectors, and reducing environmental costs associated with AI.

- **State Capacity to Govern**: An estimation of the financial resources, institutional frameworks, and skilled human capital being created to enforce compliance with AI regulations effectively.

4.1   Analysis of AI Governance Measures in the U.S.

- The US regulatory approach utilizes existing regulatory capacity, supplementing it with AI-specific considerations.
- Federal AI regulation in the US involves executive orders setting goals, agency rulemaking enacting specifics (such as by the NIST, DOE, DOJ and HHS), and independent agency actions within their sectors.
- Several US states have implemented their own AI-specific regulations focusing on consumer protection, privacy, transparency, and algorithmic discrimination.

- President Trump's Executive Order on Removing Barriers to American Leadership in AI has revoked President Biden's EO 14110, that might lead to changes in federal AI regulations.

Figure 3: Analysis of AI Governance Measures in the U.S.

4.1.1   Transparency & Accountability

- Compared to the EU, the US currently scores lower on transparency and accountability in AI governance partly due to the revocation of EO 14110 that had mandated disclosure for large foundational models and tasked NIST with developing safety testing standards.
- The OMB Act mandates risk management standards for AI use in identified high-risk government systems. The DOE is evaluating tools for identifying AI model risks in critical areas like nuclear and biological threats.

4.1.2   Promotion of Innovation

- The US strongly focuses on promoting AI innovation, with President Trump's executive order aiming to remove barriers to global AI dominance.
- Some initiatives from EO 14110 are expected to continue, such as NIST's NAIRR project and the OMB mandate for open-sourcing government AI models and data by default.
- The DOE will likely continue streamlining approvals for allied AI infrastructure like power and data centers.
- NIST's risk management framework and the US Patent Office's guidance on AI patentability will drive standardization and innovation.

4.1.3   Geopolitical Considerations

- Geopolitics is a significant focus in US AI regulation due to its leadership and competition with China.

- The US aims to maintain a competitive edge by controlling access to AI chips through the AI diffusion framework.
- A monitoring regime is in place for large AI model training runs on US Infrastructure as a Service (IaaS), including reporting by international customers.
- The US Treasury prohibits certain financial transactions involving high-risk AI systems.
- Security agencies are tasked with identifying and mitigating vulnerabilities across the AI supply chain.

4.1.4 Societal Wellbeing

- Several US states have enacted AI-specific regulations to prevent algorithmic discrimination, ensure disclosure, and label AI-generated content.
- State privacy laws have been updated to address the use of personal information and risks related to AI profiling and consent.
- Federal regulations from OMB and HHS establish standards for AI use in delivering public benefits and within government.

4.1.5 State Capacity to Govern

- The revocation of EO 14110 creates uncertainty around current US AI governance rules.
- Regulations focusing on building AI infrastructure, capital flows, and talent are expected to remain relatively stable.
- Regulations concerning AI governance capacity building, transparency, and auditability are likely to be reviewed and potentially rolled back.

4.2 Analysis of AI Governance Measures in the E.U.

- The EU AI Act establishes comprehensive regulatory guidelines for AI systems operating within the European Union.

- It interacts with other regulations like the GDPR, DMA, DSA, Chips Act, and Cyber Resilience Act, that collectively influence the operations entities involved in the manufacture, deployment, import or distribution of AI systems.

Figure 4: Analysis of AI Governance Measures in the E.U

4.2.1   Transparency & Accountability

- The EU has the most comprehensive measures for AI transparency and accountability, including risk tiering, evaluations, disclosure, licensing, penalties, and input controls.
- AI systems in the EU are classified into unacceptable, high, limited, and minimal risk categories, with corresponding compliance obligations.

4.2.2   Promotion of Innovation

- While compliance with the comprehensive regulations creates hurdles, there are provisions for promoting standards and regulatory sandboxes to facilitate adherence.
- The InvestAI initiative aims to mobilise $216 bn for open and collaborative development of complex AI models in Europe.
- The EU AI Act includes carve-outs to lessen the impact of penalties on SMEs and startups.

4.2.3   Geopolitical Considerations

- Strategic autonomy and technology sovereignty have gained prominence in EU AI policy discussions.
- The InvestAI initiative funds the open development of complex AI models, including AI gigafactories for data centers and the creation of datasets.
- The EU Chips Act involves significant investment to enhance competitiveness and resilience in the semiconductor industry.

- The EU AI Act includes exceptions for AI models released under free and open-source licenses.

4.2.4   Societal Wellbeing

- Among the regions compared in this report, the EU has the most comprehensive measures to address societal wellbeing through various governance instruments such as tiering, evaluation and performance requirements, disclosure mandates, licensing and certifications, penalties and input controls
- Voluntary codes of conduct are encouraged for assessing and minimising the environmental impact of AI systems.

4.2.5   State Capacity to Govern

- The EU AI Act mandates the creation of governance institutions at both the EU and member state levels for implementation.
- This includes the establishment of the EU AI Office, advised by a scientific panel of independent experts.
- Member states are required to establish their own institutions to enforce compliance with the EU AI Act regulations.

4.3   Analysis of AI Governance Measures in China

- China's AI governance framework is a comprehensive, state-driven approach across multiple dimensions.
- The system prioritizes centralized control through proactive regulation requiring algorithm transparency, security assessments, ethics reviews, and content monitoring aligned with state objectives.
- China's governance model emphasizes public-private collaboration for AI innovation along with strategic state investments.

- China's governance strategy strongly reflects national security priorities through data sovereignty measures, indigenous computing development and military-civil fusion initiatives.
- Massive financial investments such as $140 bn from the Bank of China and $184 bn in government VC funding are announced for AI firms.

Figure 5: Analysis of AI Governance Measures in China

4.3.1 Transparency & Accountability

- China's state-enforced AI regulations demonstrate a preference for proactive, centralised governance.
- Chinese regulations mandate algorithm filing, transparency, security assessments, and ethics reviews before AI deployment.
- Watermarking for AI-generated content and complaint redressal mechanisms are also required in China.

4.3.2 Promotion of Innovation

- China's governance mechanisms focus on state-driven public-private collaboration and strategic infrastructure development for innovation.
- There is a strong emphasis on AI talent cultivation and public-private research partnerships in China, including innovation hubs supported by AI-focused education guidelines and university-affiliated research centers.
- Guidelines reinforce an open AI ecosystem, promoting resource-sharing, interoperability, and common technical standards.
- Large-scale infrastructure projects ensure regional computing power distribution and provide tax incentives for AI infrastructure investments.

4.3.3 Geopolitical Considerations

- Policies emphasise technological self-sufficiency through measures requiring data sovereignty, AI security, indigenous computing infrastructure development and military-civil fusion.
- Strict data localization measures are imposed, requiring government approval for cross-border transfers of "important data."
- Mandatory algorithmic filing and real-time content monitoring ensure AI-generated content aligns with state narratives.
- Real-name verification requirements under cybersecurity laws address information control and foreign influence concerns.

4.3.4 Societal Wellbeing

- A state-led approach focuses on AI ethics, data privacy, content moderation, and digital fairness.
- Explicit user consent for personal data usage and mandatory opt-out options for personalised recommendations are required.
- Algorithmic price discrimination is banned, and watermarking/labeling of AI-generated content is mandated.
- Explicit user consent is required for use of voices and images in synthetic media.

4.3.5 State Capacity to Govern

- China's capacity to govern AI development is demonstrated through massive financial investments and centralised infrastructure planning.
- The government directs significant capital into priority AI sectors at an unprecedented scale.
- Local governments have also dedicated billions to AI research and infrastructure, reflecting a decentralised but state-coordinated funding approach.
- Strong oversight is provided by government institutions like the Cyberspace Administration of China (CAC).

- The state has the capacity to fund and manage large-scale computing networks for AI development.

4.4 Analysis of AI Governance Measures in India

- India's AI governance is characterised by limited regulations, with sector-specific guidelines in finance and health.
- The government is promoting AI innovation through the IndiaAI Mission addressing priority use cases such as agriculture, education and healthcare.
- Some policy interventions aim to address vulnerabilities related to access to AI resources.
- There is a lack of comprehensive regulations in India for managing the broad risks associated with AI.The draft AI Governance Guidelines propose a Technical Secretariat to enhance state capacity to govern AI, but its implementation is uncertain.

Figure 6: Analysis of AI Governance Measures in India

4.4.1 Transparency & Accountability

- India has limited cross-sector regulations. The finance and health sectors in India have more specific regulations and AI safety guidelines.The RBI has formed a panel to review AI regulation in finance, identify risks, and recommend a compliance framework. Similarly, ICMR Ethical Guidelines provide principles for AI development and deployment in healthcare.
- The AI Advisory (2024) recommends labeling deepfake and synthetic content, and draft guidelines suggest voluntary incident reporting.

4.4.2 Promotion of Innovation

- The IndiaAI Mission has allocated over $1.2 bn to develop AI models, datasets, compute, and education.

- Efforts are underway to enable access to compute resources and datasets through portals like IndiaAI Compute and AIKosha.
- While aiming to encourage startups and MSMEs, the implementation of these innovation initiatives may face inefficiencies.

4.4.3 Geopolitical Considerations

- The US AI Diffusion Framework's export controls could potentially restrict India's access to computing resources and advanced AI models.
- India is attempting to mitigate these risks by promoting indigenous AI models and domestic computing clusters.

4.4.4 Societal Wellbeing

- Sector-specific guidelines from ICMR and RBI, along with the AI Advisory (2024), offer a voluntary ethical framework for societal well-being in AI.
- There are currently no overarching regulations in India to address the diverse risks posed by AI technologies.

4.4.5 State Capacity to Govern

- MeitY's draft AI Governance Guidelines propose a Technical Secretariat to monitor and mitigate AI risks and harms in real-time, serving as a bridge between the industry and policymakers.
- As these are draft guidelines, it remains to be seen how state capacity will be enhanced to implement them.

5   Analysis of AI Governance Measures Across Companies

Companies are proactively adopting AI governance measures. These measures include developing AI principles, implementing risk mitigation strategies, enhancing transparency and establishing governance structures.

However, there is currently a lack of standardization in how companies report on their AI governance efforts. The specific details of company AI governance initiatives vary considerably.

The extent of meaningful external scrutiny of company AI governance efforts is unclear. It is also challenging to understand how companies are operationalizing their principles, measuring the effectiveness of their risk mitigation strategies, and ensuring accountability.

Microsoft, Google, OpenAI, Mistral, Anthropic, Amazon, Accenture, and Deloitte are selected for the comparative analysis. These companies operate across different stages of the AI value chain, including big technology platforms, AI model developers, and technology services firms.

Figure 7: Analysis of AI Governance Measures Across Companies

5.1   Principles

The extent to which the organisation's responsible AI policy is articulated and identifies the principles it seeks to adhere to. A comparison of the articulation of principles on responsible AI development and use by different companies is provided below.

- **Microsoft**: In 2018, six ethical principles were adopted: accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness. These principles have been translated into corporate policies, including a standard for engineering teams.

- **Google**: First published in 2018 and updated recently, the principles are bold innovation, responsible development and deployment, and collaborative progress.

- **OpenAI**: Committed to building safe and beneficial AGI. Specific principles are not detailed. Its usage policy prohibits uses such as biometric profiling, deceptive techniques, harm to children, etc.

- **Mistral**: Mistral's responsible AI policy emphasises key principles such as neutrality, empowerment, building trust and minimising potential harm and misuse. Furthermore, its terms of use emphasise data privacy and security and restrict access to individuals under 13.

- **Anthropic**: Anthropic emphasizes proportional safeguards that scale with the potential risks of AI systems. Their approach is inspired by biosafety standards, ensuring that safety measures align with the capabilities of their models.

- **Amazon**: AWS has introduced an updated Responsible AI Policy in January, 2025, which supplements the AWS Acceptable Use Policy and AWS Service Terms. This policy outlines a set of prohibited actions, and includes commitments to developing safe, fair, and accurate AI and machine learning services.

- **Accenture**: Accenture's four pillars of responsible AI implementation are: Organizational, Operational, Technical, and Reputational.

- **Deloitte**: Uses a Trustworthy AI framework to help organizations develop ethical safeguards across seven key dimensions: transparent and explainable, fair and impartial, robust and reliable, respectful of privacy, safe and secure, responsible, and accountable.

5.2 Risk Mitigation

The extent to which policy identifies the potential risks and lists actions to mitigate against those risks. The risk mitigation efforts by different companies are listed below:

- **Microsoft**: AI risk management is part of the company's enterprise risk management system. They also prioritise development of open-access responsible AI tools to map, measure and manage risks.

- **Google**: Work with external organizations to establish boundaries and reduce the risks of abuse. Google has developed a frontier safety framework to prepare for risks from frontier models. They have also published a generative AI toolkit and people + AI guidebook that guide AI development and deployment.

- **OpenAI**: Investments in research to inform regulation – including techniques for assessing potentially dangerous capabilities. Their red-teaming and safety procedures are publicly disclosed.

- **Mistral**: Includes measures that address data privacy and security. It prohibits use for illegal activities, hate and discrimination, misinformation, and professional advice. They have a zero-tolerance policy regarding child sexual abuse material.

- **Anthropic**: They employ AI Safety Level (ASL) Standards to address catastrophic risks, such as misuse or unintended autonomous actions. These standards include rigorous testing and safeguards before deploying advanced models.

- **Amazon**: Prohibits use for disinformation, deception, violation of privacy, harm or abuse of minors. Also prohibits AI/ML services for circumvention of safety filters and use in a weapon without human intervention.

- **Accenture**: Ongoing testing of AI for human impact, fairness, explainability, transparency, accuracy, and safety. Employing use of state-of-the-art responsible AI tools and technologies to mitigate problems.

- **Deloitte**: Using tools to test and monitor AI models.

5.3 Transparency and Reporting

The extent to which governance frameworks attempt to promote transparency and accountability. The measures by companies that promote transparency and accountability are listed below.

- **Microsoft**: Transparency reports are published with recent reports aligned with the NIST

framework. Provide documentation on the intended use, limitations and risks of AI technologies to consumers.

- **Google**: Transparency reports are published with recent reports aligned with the NIST framework. Gemini app, Google Cloud and Google Workspace are ISO/IEC 42001 certified. Engaging with governments, civil society, and academia on information sharing, establishing common standards, and promoting best practices for AI safety.

- **OpenAI**: Transparency reports and system cards for different products detail the red-teaming and safety procedures.

- **Mistral**: Publishes documentation for its AI models with details on their capabilities and limitations. Provides channels for reporting incidents, contact information for inquiries, and communities for users.

- **Anthropic**: Anthropic commits to sharing updates on their governance practices and lessons learned. They aim to maintain public trust by openly communicating their safety measures and policies.

- **Amazon**: Provides details on how AI systems are deployed, monitored and managed during their development and operations. Mentions openly sharing development choices, including data sources and algorithms.

- **Accenture**: Public disclosure of AI systems' capabilities, limitations, and suitable uses, addressing both security and societal risks.

- **Deloitte**: Recommends engaging stakeholders in AI governance by defining clear roles and responsibilities, documenting accountabilities, establishing expectations for AI ethics and trust, and empowering employees to voice concerns and act ethically.

5.4   Governance Structure

An estimation of the financial resources, institutional frameworks, and skilled human capital made available to enforce compliance with AI regulations effectively. A comparison of the governance structure of different companies is provided below.

- **Microsoft**: Aether is an internal advisory body set up to focus on AI ethics and responsible AI practices within the company. In 2023, over 350 employees worked on responsible AI, developing best practices for building safe, secure, and transparent AI systems designed to benefit society.

- **Google**: The AGI Safety Council and the Responsibility and Safety Council oversee alignment with AI principles. Google forced AI researcher Dr. Gebru to resign after she resisted orders to halt research indicating that speech technology, like Google's, could negatively impact marginalized groups. Others also resigned in response to how Dr. Gebru was treated.

- **OpenAI**: Specific governance structures are not publicly available. Concrete governance practices are specifically tailored to highly capable foundation models. The company structure has also seen a lot of churn, changing from a non-profit to a "capped profit" structure to enable the pursuit of AGI. Former employees have also raised concerns about the safety practices at the company.

- **Mistral**: Does not provide any details on the governance structure.

- **Anthropic**: Their governance framework is iterative and adaptable, incorporating lessons from high-consequence industries. It includes internal evaluations and external inputs to refine their policies.

- **Amazon**: Does not provide any details on the governance structure.

- **Accenture**: Establishment of transparent governance structures across domains with defined roles, expectations, and accountability. Creation cross-domain ethics committees. Has establishment a Chief Responsible AI officer role.

- **Deloitte**: Evaluating roles and responsibilities and implementing change management and training.

5.5   Third-Party Oversight

The willingness to subject itself to third-party oversight. Examines policies that encourage third-party oversight to identify and report risks they might have overlooked.

- **Microsoft**: Bug bounty programs are used to incentivise external discovery and reporting of issues and vulnerabilities. They are also building external red teaming capacity to enable third-party oversight, especially of sensitive capabilities of highly capable models.

- **Google**: Google Cloud AI achieved a "mature" rating in a third-party evaluation. Encourages third-party discovery and reporting of issues and vulnerabilities. Works with industry peers and standards-setting bodies towards efforts such as developing technical frameworks to help users distinguish synthetic media.

- **OpenAI**: Bounty systems to encourage responsible disclosure of weaknesses and vulnerabilities.

- **Mistral**: Its policies do not explicitly seek third-party oversight, but its commitment to transparency indicates a willingness to be open about its AI practices which could extend to third-party oversight.

- **Anthropic**: Anthropic collaborates with external experts for adversarial testing and red-teaming of their models. This ensures that their systems meet stringent safety and security standards.

- **Amazon**: While AWS is committed to developing safe, fair and accurate AI/ML services and provides tools and guidance to assist in this, it does not talk about third-party oversight.

- **Accenture**: No mention of third-party oversight.

- **Deloitte**: Offers the Omnia Trustworthy AI module, which provides guidelines and guardrails for designing, developing, deploying, and operating ethical AI solutions. Does not mention using

third party oversight over their own AI solutions.

6   Analysis of AI Governance Measures Across Multistakeholder Groupings

Various multistakeholder gatherings, including the AI Summits and the Global Partnership on AI, have been established to raise awareness and coordinate international AI governance efforts.

While state-level efforts have tended to focus on innovation and geopolitics, multistakeholder gatherings highlight broader societal concerns arising from the rapid development of advanced AI.

Most gatherings do not have legally binding commitments or backing from all members (for instance, the US and EU refusing refused to sign the declaration on inclusive and sustainable AI at the AI Action Summit in February 2025).

Achieving alignment or convergence on AI regulations through these platforms can simplify compliance for multinational technology companies.

The analysis in this section focuses on the membership composition, guiding principles, and recent developments in these gatherings.

6.1   The Organization for Economic Co-operation and Development

Membership: - OECD has 38 member countries committed to democracy, collaborating on addressing global policy changes, and is not an AI-specific body.

Principles and areas of focus: - OECD promotes inclusive growth, human-centric values, transparency and explainability, robustness and accountability of AI systems. - The OECD AI Principles are the first intergovernmental standard on AI.

6.2   Global Partnership on AI

Membership: - GPAI has 44 member countries, including the US, EU, UK, Japan, and India.

Principles and areas of focus: - GPAI promotes the responsible development of AI grounded in human rights, inclusion, diversity, innovation, and economic growth - Areas of focus include responsible AI, data governance, the future of work, and innovation and commercialisation.

Developments: - As of 2024, GPAI and the OECD formally joined forces to combine their work on AI and implement human-centric, safe, secure, and trustworthy AI. The two bodies are committed to implementing the OECD Recommendation on Artificial Intelligence.

6.3   AI Governance Alliance

Membership: - The AI governance alliance is a global initiative launched by the World Economic Forum. The alliance has over 603 members from more than 500 organisations globally.

Principles and areas of focus: - The principles of the AI Governance Alliance include responsible and ethical AI, inclusivity, transparency, international collaboration and multistakeholder engagement. - The areas of focus include safe systems and technologies, responsible applications and transformation, resilient governance and regulation.

6.4   AI Summits

Membership: - The AI summits are a series of international conferences addressing the challenges and opportunities presented by AI. Participants include heads of state and major companies such as Meta and DeepMind.

Principles and areas of focus: - Each AI summit has set its own agenda, but some common principles are ethical AI development, safety and security, transparency and accountability, and international collaboration. - The AI Summits have been held thrice since their inception. The first

summit, focussing on AI safety, was held at Bletchley Park in the UK in 2023. The second summit was held in Seoul, South Korea, in 2024. - The third event, the AI Action Summit was held in Paris in February 2025 and was attended by representatives from more than 100 countries. While 58 countries, including France, China and India, signed a joint declaration, the US and UK refused to sign the declaration on inclusive and sustainable AI. - The agenda has evolved from existential risks and global cooperation at Bletchley, to risk management frameworks and company commitments at Seoul to an action oriented focus on public interest, sustainability and global governance at Paris.

Sources: AI Action Summit, AI Seoul Summit,ǎ Bletchley Declaration, AI Governance Alliance

6.5   United Nations

Membership: - The UN is an international organisation committed to global peace and security, with 193 member states, including almost all internationally-recognised sovereign states. The safe development of AI is one of their many areas of work.

Principles and areas of focus: - Some of their core principles include doing no harm. AI applications should have a clear purpose, fairness and non-discrimination, safety and security to prevent misuse and harm, responsibility and accountability. - The UN Secretary-General is convening a multi-stakeholder High-level Advisory Body on AI to study and provide recommendations for the international governance of AI. - Other efforts include convening global dialogues, developing standards and building capacity.

6.6   United Nations Educational, Scientific and Cultural Organization

Membership: - UNESCO is a specialised agency of the UN with 194 member states and 12 associate member states

Principles and areas of focus: - The UNESCO general conference adopted the recommendation on the ethics of artificial intelligence – the first global standard on AI ethics principles aligned with the UN's principles on AI. - Areas of focus include developing an AI Readiness Assessment Methodology, facilitating policy dialogues and capacity building initiatives.

Sources:UN Ethical AI Principles, UN AI Advisory Body, UNESCO Recommendations on Ethics of AIǎ

7   Predictions

Confidence	Region	Prediction
High	Global	Compute thresholds for enforcing regulations will no longer be relevant. The effectiveness of these thresholds might be challenged as a measure of capability as inference computing begins to scale and smaller models become more efficient. The US and EU have 10^{26} and 10^{25} flops as training compute thresholds for enforcing certain regulations.
High	Global	Investments in sovereign cloud infrastructure will increase, driven by geopolitical considerations.
High	China, EU	Open-source and open-weight models will continue to be pushed by China and EU as a pathway to strategic autonomy and technology leadership. DeepSeek and Mistral will remain open-weight or open-source.

Confidence	Region	Prediction
High	India	The compute capacity created under the IndiaAI mission aimed towards incentivising startups and building indigenous models will be underutilised. This is due to lack of demand that meets the criteria to qualify for the subsidies as well as due to friction in the bureaucratic process involved.
Moderate	Global	AI governance regulations at the state level will continue to prioritise innovation over encouraging transparency, accountability, and societal well-being. In other words geopolitical considerations will trump protection of individual rights as a governance priority.
Moderate	US	US chip restrictions on China will not escalate further. This is because Deep Seek makes owning newer chips less relevant.

Confidence	Region	Prediction
Moderate	US	Federal laws focused on monitoring AI safety and federal agency assessment of AIs for discrimination and bias will be made defunct or watered down significantly. By the end of the year, AI safety guardrails will be driven by private firms.
Moderate	EU	EU's comprehensive regulatory framework, including penalties for non-compliance, would result in a few companies not releasing their AI models/features in the EU. This might lead to a milder enforcement of the regulations. As per the declared timeline, rules on notified bodies, general purpose AI models, governance, confidentiality, and penalties start to apply from August, 2025.
Moderate	China	The US Diffusion Framework will not stop state-of-the-art AI models coming out of China, at least not in the next year. This is because Deep Seek makes owning newer chips less relevant, and China has built up an overcapacity of data centres over the past few years.

Confidence	Region	Prediction
Moderate	India	Regulatory focus will be on protecting the information ecosystem from content deemed harmful for the government or Indian society.
Moderate	Corporate	Governance of AI within companies will become a bigger requirement as governments firm up on their positions regarding AI. 'Chief Responsible AI Officer' will be a new role at companies seeking to deploy AI solutions at scale, whose duty it will be to ensure AI is deployed in a manner that will, at the very least, protect them from litigation.

Confidence reflects how likely the authors believe their prediction will be accurate in the coming year.

8   Acronyms

AI	Artificial Intelligence
CAC	Cyberspace Administration of China

DMA	Digital Markets Act
DOE	Department of Energy
DOJ	Department of Justice
DPDPA	Digital Personal Data Protection Act
DSA	Digital Services Act
EO	Executive Order
EU	European Union
GDPR	General Data Protection Regulation
GPAI	Global Partnership on AI
HHS	Health and Human Services
IaaS	Infrastructure as a Service
ICMR	Indian Council of Medical Research
MeitY	Ministry of Electronics and Information Technology
MSMEs	Micro, Small and Medium Enterprises
NAIRR	National Artificial Intelligence Research Resource
NIST	National Institute of Standards and Technology
OECD	Organization for Economic Co-operation and Development
OMB	Office of Management and Budget

RBI	Reserve Bank of India
SMEs	Small and Medium-sized Enterprises
UN	United Nations
UNESCO	United Nations Educational, Scientific and Cultural Organization
US	United States of America
VC	Venture Capital

9   Acknowledgements and Disclosure

Endnotes

Notes

TAKSHASHILA
INSTITUTION

The Takshashila Institution is an independent centre for research and education in public policy. It is a non-partisan, non-profit organisation that advocates the values of freedom, openness, tolerance, pluralism, and responsible citizenship. It seeks to transform India through better public policies, bridging the governance gap by developing better public servants, civil society leaders, professionals, and informed citizens.

Takshashila creates change by connecting good people, to good ideas and good networks. It produces independent policy research in a number of areas of governance, it grooms civic leaders through its online education programmes and engages in public discourse through its publications and digital media.