Overview of India's AI Governance Framework

Below are ten points that summarise India's overall approach to AI governance:

The goal is to encourage innovation and adoption, while protecting individuals and society from the risk of harm caused by the development or use of AI. An effective governance framework is one which balances these twin objectives. India's approach in general is to govern the applications of AI by empowering the relevant sectoral regulators, and not to regulate the underlying technology itself.

A balanced, agile, flexible, and pro-innovation approach to AI governance is best suited to India's goals. The primary goal at this stage is to leverage AI for economic growth, inclusive development, resilience and global competitiveness. Given India's talent advantage, the wide adoption of AI across sectors can result in productivity gains, which can drive economic growth and create jobs. Further, AI-based applications, with multilingual and voice-based support, are being deployed in agriculture, healthcare, education, disaster management, law, and finance are enabling digital inclusion and creating real positive impact. A balanced framework would help maximise these benefits, while retaining the regulatory agility and flexibility to intervene and mitigate risks as and when they emerge.

Governance frameworks should boost awareness, infrastructure, investments and overall domestic capacity. Key sectors such as pharmaceuticals, telecommunications, manufacturing, media and social sectors hold significant potential for AI adoption, but to realize this potential requires a governance framework to enhance awareness, infrastructure, and investments. Initiatives like IndiaAI Mission are steps toward fostering AI adoption. Expanding domestic capacity while accelerating responsible adoption across sectors is critical to advancing India's goals of inclusive growth and global competitiveness.

Mitigating the risks of AI to individuals and society is a key pillar of the governance framework. In general, the risks of AI include malicious use (e.g. misrepresentation through deepfakes), algorithmic discrimination, lack of transparency, systemic risks and threats to national security. These risks are either created or exacerbated by AI. An India-specific risk assessment framework, based on empirical evidence of harm, is critical. Further, industry-led compliance efforts and a combination of different accountability models are useful to mitigate harm.

Existing regulations can be applied to address many of the risks. Existing laws (for e.g. on information technology, data protection, consumer protection and statutory civil and criminal codes, etc.), can be used to govern AI applications. Therefore, at this stage, a separate law to regulate AI is not needed given the current assessment of risks. However, timely and consistent enforcement of applicable laws is required to build trust and mitigate harm.

Legal amendments may be considered to encourage innovation and address gaps. Existing laws on copyright may need to be amended, for example, to enable the large-scale training of AI models, while ensuring adequate protections for copyright holders and data principals. Rules for how digital platforms are classified should also be updated to better describe the unique functions, obligations, and liability regime applicable to different actors in the AI value chain. Similarly, if existing regulations are unable to tackle the emerging risks to individuals, then additional rights or obligations may be introduced. For example, data portability rights could be adopted to give individuals more control over their data.

Voluntary measures can help mitigate emerging risks. Voluntary frameworks, if proactively adopted in the form of principles, commitments, or standards, can help build trust. The goal of this approach is to enhance trust and safety without introducing burdensome regulations during the nascent stage of ecosystem development. As the industry matures, some baseline measures may be converted into mandatory requirements, which will be enforced by sectoral regulators.

Techno-legal approaches can be applied to support specific policy objectives. Techno-legal solutions can be effective tools of governance. They can be used to give effect to established policy through verifiable methods. While traditional approaches to governance have focused primarily on regulatory instruments, effective AI governance could benefit from technology-enabled solutions in areas such as content authentication, privacy preservation, and bias mitigation.

Transparency about the AI value chain can promote accountability. The AI value chain comprises various actors (developers, deployers and users), operating at different layers of the technology stack (data center, models, applications), performing dynamic functions (training, customisation, distribution, etc.) through complex inter-personal relationships. Many aspects of these technical and organisational relationships are dynamic and not fully understood by regulators . Greater transparency about the technical and organisational aspects of AI development and deployment will help regulators design governance mechanisms that are targeted, proportionate and effective.

A ‘whole of government’ approach is required to coordinate policy actions and prepare for future AI development. Given the cross-sectoral nature of AI, the constraints on regulatory capacity , and the absence of a nodal regulator for emerging technologies, India’s AI governance framework would benefit from a coordinated institutional effort, wherein key agencies, sectoral regulators, and standard setting bodies are involved in the formulation and implementation of policy frameworks to give effect to the objectives of such AI governance frameworks.

Part 1: Key Principles

The Committee recommends that India's AI governance framework be guided by certain principles or 'sutras', applicable across sectors and technologies.

A useful set of principles in this regard has been published by a committee set up by the Reserve Bank of India (RBI) in August 2025. The committee to develop a Framework for Responsible and Ethical Enablement of Artificial Intelligence ("FREE-AI Committee") recommends seven principles or sutras to guide AI development and risk mitigation in the financial sector.

These principles have been suitably adapted below to ensure they have cross-sectoral applicability, are technology-neutral, and align with this Committee's recommendations.

01 Trust is the Foundation

Trust is essential to support innovation, adoption, and progress, as well as risk mitigation. Without trust, the benefits of artificial intelligence will not be realised at scale. Trust must be embedded across the value chain – i.e. in the underlying technology, the organisations building these tools, the institutions responsible for supervision, and the trust that individuals will use these tools responsibly. Therefore, trust is the foundational principle that guides all AI development and deployment in India.

02 People First

AI governance frameworks should be human-centric. That means AI systems should be designed and deployed in ways that empower individuals and reflect the value systems of the people for whom the technology is built to serve. From a governance perspective, a people-first approach means that humans should, as far as possible, have final control over AI systems, and human oversight is essential to maintain accountability. A people-first approach also prioritises human capacity development, ethical safeguards, trust and safety.

03 Innovation over Restraint

AI-led innovation is a pathway to achieving national goals, such as socio-economic development, global competitiveness, and resilience. Therefore, AI governance frameworks should actively encourage adoption and serve as a catalyst for impactful innovation. That said, innovation should be carried out responsibly and should aim to maximise overall benefit while reducing potential harm. All other things being equal, responsible innovation should be prioritised over cautionary restraint.

04 Fairness and Equity

A key goal of India's approach to AI governance is to promote inclusive development. Therefore, AI systems should be designed and tested to ensure that outcomes are fair, unbiased, and do not discriminate against anyone, including those from marginalised communities. AI should be leveraged to promote inclusive development while mitigating risks of exclusion, bias, and discrimination.

05 Accountability

To ensure that India AI's ecosystem progresses based on trust, AI developers and deployers should remain visible and accountable. Accountability should be clearly assigned based on the function performed, risk of harm, and due diligence conditions imposed. Accountability may be ensured through a variety of policy, technical and market-led mechanisms.

06 Understandable by Design

Understandability is fundamental to building trust and should be a core design feature, not an afterthought. Though AI systems are probabilistic, they must have clear explanations and disclosures to help users and regulators understand how the system works, what it means for the user, and the likely outcomes intended by the entities deploying them, to the extent technically feasible.

07 Safety, Resilience and Sustainability

AI systems should be designed with safeguards to minimise risks of harm and should be robust and resilient. These systems should have capabilities to detect anomalies and provide early warnings to limit harmful outcomes. AI development efforts should be environmentally responsible and resource-efficient, and the adoption of smaller, resource-efficient 'lightweight' models should be encouraged.

Part 2: Issues & Recommendations

Using these seven principles or *sutras* as guidance, the Committee recommends an approach to AI governance that fosters innovation, adoption, and scientific progress, while proposing measures to mitigate the risks to individuals and communities.

Effective governance includes not just regulation, but other forms of policy engagement, including education, infrastructure development, diplomacy, and institution building. Therefore, the Committee has made its recommendations across the following **six pillars:**

2.1 Infrastructure

The goal of India's AI governance framework is to promote innovation, adoption, diffusion, and advancement of AI while mitigating risks to society. The government is taking significant strides to achieve this goal through the India AI Mission, which across seven pillars, is building the infrastructural backbone for large-scale adoption. As on August 31, 2025, some highlights include:

Compute: Over 38,231 GPUs are being made available to startups, researchers and developers at subsidised rates.[iv] A secure GPU cluster is also being constructed to house 3,000 next-generation GPUs for sovereign and strategic applications.[v]

Datasets: AIKosh has onboarded 1,500 datasets and 217 AI models from 34 entities across 20 sectors.[vi] It provides permission-based access, allowing contributors to retain control over data usage while facilitating AI development.

Foundation Models: Four startups are being supported in the first phase to develop India's sovereign models.[vii] They will receive credits and funding covering up to 25% of compute costs, provided through a mix of grants (40%) and equity (60%).[viii]

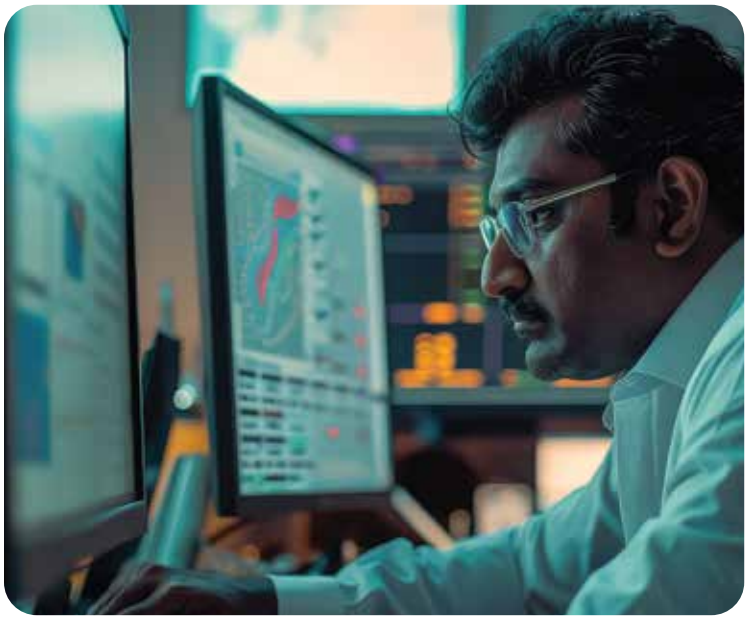
Applications: The India AI Application Development Initiative (IADI) has taken 30 sectoral applications to the prototyping stage across different sectors.[ix]

To ensure the continued development and adoption of AI in India, the Committee recommends empowering the India AI mission, line ministries, sectoral regulators and state governments to implement such initiatives focused on enablement.

Further, to accelerate AI adoption among MSMEs, the government should provide targeted incentives and financing support, including tax rebates on certified solutions, AI-linked loans through SIDBI and Mudra, and subsidised access to GPUs. This will help lower the cost of adoption, if supported by sector-specific AI toolkits and pre-built starter packs tailored to industries like textiles, retail, logistics, and food processing.

Data & Compute Access

While technology-mature sectors such as telecom, media, pharmaceuticals, and manufacturing are scaling AI rapidly, adoption remains uneven in agriculture, education, healthcare, and public services due to lack of adequate infrastructure and access to resources. Urban centres demonstrate higher maturity,[x] while rural and under-resourced areas lag, highlighting the need for more inclusive strategies.[xi]



Expanding access to data and compute is essential for scaling adoption.[xii] Market incentives should be introduced to encourage public and private entities to contribute to existing platforms like AIKosh, the Open Government Data Platform, and the National Data and Analytics Platform. Further, appropriate data governance frameworks must be developed to support the sharing of anonymised data, data stewardship, and data sovereignty.

Moreover, providing access to foundational resources, such as data and computing resources, is critical to mitigate the risks of AI. For example, in order to evaluate the fairness of AI systems in the Indian context, developers need access to reliable and representative datasets in the form of standardised 'evaluation datasets.' Similarly, access to computing resources is necessary to perform safety evaluations and to test and validate the effectiveness of guardrails implemented by developers and deployers at population scale.

Digital Public Infrastructure

The transformational potential of AI in sectors such as agriculture, healthcare, education, and governance positions it as a critical enabler of socio-economic development. AI can serve these goals by building on Digital Public Infrastructure (DPI).

For countries in the Global South, with limited access to AI infrastructure and resources, the cost of deploying AI solutions at scale can be prohibitively high. DPI offers a unique pathway to adoption. Features such as identity databases, data exchanges, authentication capabilities, and payment systems can be harnessed to design AI-led solutions that are scalable, affordable, and tailored to local needs, which can support widespread adoption.[xiii]

Leveraging DPI can also ensure that AI solutions are embedded with principles of privacy, transparency, interoperability, and security by design, which are key pillars of AI governance.[xiv]

Therefore, the Committee recommends that India's AI governance strategy support greater adoption by focusing on three enablers: expanding access to high-quality and representative datasets, providing affordable and reliable access to computing resources, and integrating AI with Digital Public Infrastructure (DPI).

The Committee also recommends that special schemes be designed with the specific goal of encouraging investments at all levels of the AI value chain. It is only when India is perceived as a hub for AI that innovation can be catalysed through private entrepreneurship.

Recommendations

- Empower the India AI mission, line ministries, sectoral regulators and state governments to increase AI adoption through initiatives on infrastructure development and increasing access to data and computing resources.
- Increase data availability, sharing, and usability for AI development and adoption with robust data portability standards and data governance frameworks.
- Encourage the use of locally relevant datasets to support the creation of culturally representative models and applications.
- Promote access to reliable evaluation datasets and compute infrastructure for AI development and deployments, and to conduct safety testing and evaluations.
- Integrate AI with Digital Public Infrastructure (DPI) to promote scalability, interoperability and inclusivity.

2.2 Capacity Building

India has initiated various capacity building initiatives, such as the India AI FutureSkills, FutureSkills PRIME, and other higher education programs in AI. These efforts are currently supporting more than 500 PhD fellows, 8,000 undergraduates, and 5,000 postgraduates.[xv]

These efforts need to be significantly expanded to enhance AI adoption, address existing inequalities, and for inclusive development – a key goal of India's AI governance framework.

Small businesses and ordinary citizens need both access and exposure to AI's capabilities. In the public sector, officials and regulators often do not have the technical grounding to evaluate AI procurements, manage risks, or oversee responsible deployment.[xvi]

Therefore, the Committee recommends specific initiatives around education, skilling and training to build trust, empower people and increase adoption, which are key principles or sutras guiding India's overall approach to AI governance.

Recommendations

- Increase societal trust and public awareness about the risks and capabilities of AI through regular training programs and publicity campaigns.
- Conduct training programs for government officials, regulators and civil servants to understand AI technology developments, to manage public procurements effectively, and to encourage the responsible use of AI in the public sector.
- Develop the capacity of law enforcement agencies (LEAs), police, cybercrime units, and prosecutors to detect, investigate and resolve AI-enabled crimes.
- Expand capacity building initiatives to achieve deeper penetration of AI into tier-2 and tier-3 cities, and in vocational institutes.

2.3 Policy & Regulation

The overarching goal of India's AI governance framework is to encourage innovation, adoption and technological progress, while ensuring that actors in the AI value chain are mitigating risks to individuals and society. In that respect, the Committee has reviewed the current legal framework and suggested areas where regulatory intervention is necessary.

Applicability of existing laws

In recommending a suitable regulatory approach, the Committee has paid close attention to the existing system of laws and regulations in India, comprising constitutional provisions, statutory laws, rules, regulations, and guidelines. This includes laws and regulations across domains such as information technology, data protection, intellectual property, competition law, media law, employment law, consumer law, criminal law, amongst others.

The Committee's current assessment is that many of the risks emerging from AI can be addressed through existing laws. For example, the use of deepfakes to impersonate individuals can be regulated by provisions under the Information Technology Act and the Bharatiya Nyaya Sanhita; and the use of personal data without user consent to train AI models is governed by the Digital Personal Data Protection Act. The Annexure to this report contains examples of how existing laws can be applied to deal with other AI harms.

At the same time, there is an urgent need to conduct a comprehensive review of relevant laws to identify regulatory gaps in relation to AI systems. For example, the Pre-Conception and Pre-Natal Diagnostic Techniques (PC-PNDT) Act should be reviewed from the perspective of AI models being used to analyse radiology images, which could be misused to determine the sex of a foetus and enable unlawful sex selection. In priority sectors such as finance, where such analysis is already underway,[xvii] regulatory gaps should be quickly identified and plugged in with targeted legal amendments and regulations.

Ongoing deliberations

There are a few domains in which deliberations are already underway to study regulatory issues relating to AI governance and potential gaps. Some of these engagements are by way of inter-ministerial consultations, rulemaking under newly adopted laws, and expert committees. In this section, the Committee outlines a few such areas.

(a) Classification and Liability

The Information Technology Act, 2000 (IT Act) is the primary legislation that deals with the classification of digital platforms, their obligations under law, and related liability.

The IT Act, given that it was drafted more than two decades ago, requires an update in relation to how digital entities are classified, specifically in the context of AI systems. For example, there is a need to define clearly the roles of various actors in the AI value chain (developer, deployer, users, etc.) and how they will be governed under current definitions ('intermediary', 'publisher', 'computer system', etc.). At present, the term intermediary is broadly defined to mean any entity that "on behalf of another person receives, stores or transmits [an electronic record] or provides any service with respect to such record". Under current laws, it includes telecom service providers, search engines and even cyber cafes.[xviii] However, there is a need to provide clarity, especially with regard to how this definition would apply to modern AI systems, some of which generate data based on user prompts or even autonomously, and which refine their outputs through continuous learning.

Another important question is how liability should be apportioned across the AI value chain. Under Section 79 of the IT Act, legal immunity is available to intermediaries for unlawful third-party content, provided they do not initiate the transmission of data, select the recipient of the data or modify it. It appears that such legal immunity would not be applicable to many types of AI systems that generate or modify content. Further, the liability of AI developers and deployers who fail to observe due diligence obligations under the IT Act also needs further deliberations.

Therefore, the Committee is of the view that the IT Act should be suitably amended to ensure that India's legal framework is clear on how AI systems are classified, what their obligations are, and how liability may be imposed.

(b) Data Protection

The Digital Personal Data Protection Act (DPDP Act) which governs the collection and processing of all digital personal data in India, was adopted by Parliament in August 2023 and will be in force once draft rules to implement various aspects of the law are notified. Even as the rulemaking process for the DPDP Act is underway, new questions have emerged about the impact of data protection regulations on AI development and risk mitigation.

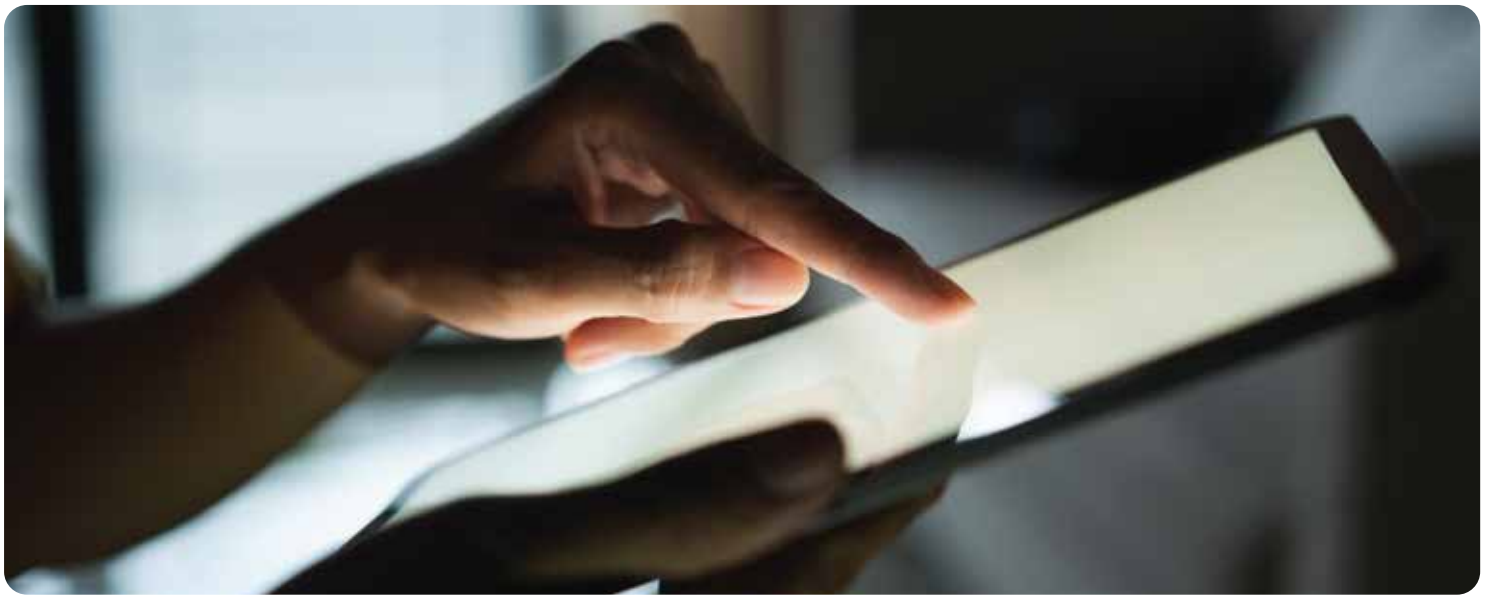
Key issues include for example, the scope and applicability of exemptions available for the training of AI models on publicly available personal data;[xix] whether the principles of collection and purpose limitation are compatible with how modern AI systems operate;[xx] the role of 'consent managers' in AI workflows and the value of dynamic and contextual notices in a world of multi-modal AI and ambient computing;[xxi] the scope of the research & 'legitimate use' exception for AI development;[xxii] and various other issues.

The Committee believes that resolving these issues are central to a robust AI governance framework. Further, some of the issues raised above may require legislative amendments to take effect, and the Committee recommends a detailed review by relevant bodies such as the AI Governance Group, which this committee has suggested establishing.

(c) Content Authentication

Generative AI technologies, including image, video, and music generation tools offer significant opportunities for creativity, human expression, access to knowledge and innovation. At the same time, the risks of misuse are significant. The creation and distribution of deepfakes and other unlawful material, such as child sexual abuse material (CSAM) and non-consensual images ('revenge porn'), have the potential to cause serious harm, especially to vulnerable groups.[xxiii] India's AI governance framework should therefore preserve the benefits of these technologies while addressing their misuse.

In this context, the Committee has examined the issue of content authentication and provenance, i.e. the determination of whether or not any piece of information was generated or modified by an AI system.



A popular method for content authentication is the use of watermarks. Such labels and other unique identifiers can be used to authenticate whether or not any piece of information was generated or modified by an AI systems.[xxiv]

This principle of using unique identifiers for content authentication and provenance is embedded in existing industry standards such as the Coalition for Content Provenance & Authenticity (C2PA).[xxv]

A related issue is content traceability, i.e. tracing the origin of a particular piece of content generated or modified by AI. Various forensic tools and attribution methods currently exist for this purpose (for e.g. watermarking to trace the origin of AI-generated content, dataset provenance tools to identify training data sources in copyright infringement cases, attribution methods to determine if harmful content originated from a specific AI model).[xxvi] Such attribution tools have potential utility for both content authentication and provenance. At the same time, their inherent limitations must also be examined (for e.g. the ability of malicious actors to bypass these safeguards and risks to citizen privacy).[xxvii]

The issue of harmful deepfakes is a growing menace to society and immediate action is required. Therefore, it is recommended to set up a committee of experts with representatives from government, industry, academia and standard-setting bodies to develop global standards around content authentication and provenance. These standards, governance frameworks and technical measures may be presented in standard-setting bodies and subjected to rigorous testing to ensure that these measures are effective.

In parallel, it is recommended that the proposed AI Governance Group (AIGG), with support from the Technology & Policy Expert Committee (TPEC), described later in this report, should review the regulatory framework in India applicable to content authentication and make recommendations to relevant agencies, such as MeitY, including the use of appropriate techno-legal solutions and additional legal measures if necessary in order to tackle the problem of AI-generated deepfakes in India.

(d) Copyright

Copyright is a contested issue in AI governance, particularly in relation to generative AI systems. Public consultations on this topic have yielded strong and divergent views from technology companies, news publishers, content creators and civil society on the issue of how legal frameworks can protect creative labour without stifling innovation.[xxviii]

Following the publication of the draft report on 'AI Governance Guidelines Development' published in January, 2025, the Department for Promotion of Industry and Internal Trade (DPIIT) established a committee in April, 2025 to deliberate on this issue.[xxix] The DPIIT committee's mandate includes examining the legality of using copyrighted work in AI training and its implications, evaluating the copyrightability of works produced by generative AI systems, and reviewing international practice to propose a balanced copyright framework suited to India's needs.

As part of its deliberations, this Committee has specifically examined the implications of using copyrighted materials in the training and development of AI models.

According to Section 52 of the Indian Copyright Act, limited 'fair dealing' exceptions apply for private or personal use, including research. These exceptions are restricted to non-commercial use and do not extend to organisational or institutional research. As a result, they may not cover many types of modern AI training.

Based on current practice, AI models are often trained on large collections of publicly available data to improve accuracy and relevance of the model, and to promote inclusivity. Various lawsuits have been filed claiming that such practices constitute infringement based on the limited exception provided under Indian copyright law.[xxx]

Globally, some groups are in support of a 'Text and Data Mining' (TDM) exception to enable AI development. Some jurisdictions, such as the EU, Japan, Singapore and the UK have adopted this approach in varying capacities.[xxxi] This Committee is of the view that the committee set up by DPIIT for this purpose may consider a balanced approach, which enables Text and Data Mining, with the objective of fostering innovation and enabling provisions to protect the rights of copyright holders.

The Committee awaits the DPIIT committee's detailed recommendations on these issues.

Global diplomacy on AI governance

Given the strategic importance of technology in protecting national security and sovereignty, AI governance is a critical element of foreign diplomacy. This is clearly demonstrated in the centrality of international AI governance in various national AI strategies (see for example, the US 'AI Action Plan'[xxxii] and China's 'Global AI Governance Action Plan').[xxxiii]

The Committee is of the view that India's balanced approach to AI governance could benefit countries in the Global South, i.e. a majority of the world's population.

AI governance should therefore be integrated into India's strategic engagements and foreign policy. India should continue its participation in multilateral AI governance forums, such as the G20, UN, OECD, and deliver tangible outcomes as host of the 'AI Impact Summit' in February 2026.

Foresight on AI governance

The pace of progress in AI makes it challenging for regulation to keep up. For example, highly autonomous 'AI agents' are demonstrating new capabilities, such as self-directed action and multi-agent collaboration, which may require us to rethink our current approaches to governance.

Potential risks also include autonomous AI-to-AI communication and coordination. Advanced AI systems may create covert protocols or collaborate with each other in ways that amplify security concerns, run disinformation campaigns, and cause disruptive loss of control. Governance frameworks must therefore have clear monitoring standards, audit trails, and ensure that human-in-the-loop mechanisms are in place at critical decision points. This is explained in more detail in the next section under mitigating loss of control.

The Committee recommends that governance frameworks should be future looking, flexible and agile, such that they enable periodic reviews and reassessments.

As the ecosystem in India matures, the Committee recommends undertaking foresight research, policy planning, and simulation exercises to anticipate future issues and demands so that policy and regulation can be adapted accordingly.

Recommendations

- Develop governance frameworks that are balanced, agile, flexible, and principle-based, and enable monitoring and recalibration based on feedback.
- Review the current legal framework to evaluate risks and regulatory gaps.
- Consider targeted legislative amendments to encourage innovation (for eg. in copyright and data protection) and to clarify issues around classification and liability.
- Develop common standards and benchmarks to achieve regulatory objectives (e.g., on content authentication, data integrity, cybersecurity, fairness, etc.).
- Establish a committee of international experts from government, industry, academia and standard-setting bodies to develop global standards around content authentication, with a focus on certifying information as genuine.
- The proposed AI Governance Group (AIGG), with support from the Technology & Policy Expert Committee (TPEC) should examine issues of content authentication in detail and issue appropriate guidelines.
- Create regulatory sandboxes to enable the development of cutting-edge technologies in constrained environments affording reasonable legal immunities, provided these tests produce evidence with published details of what was tested, guardrails applied, risks observed, etc.
- Support strategic engagements and foreign diplomacy in national, regional and multilateral forums to further India's interests on AI governance issues.
- Conduct horizon-scanning and scenario planning analysis to anticipate future developments in AI that may require policy or regulatory responses.

2.4 Risk Mitigation

Risk mitigation is the act of translating policy and regulatory principles into practical safeguards to mitigate the possibility of harm. This part of the report sets out different ways in which AI systems can be transparent, fair, and accountable, with particular emphasis on risk assessment and mitigation frameworks that are best suited for India's unique context.

Risk Assessment

The Committee recognises that because AI systems are probabilistic, generative, agentic, and adaptive by design, they have the potential to cause harm to individuals and society by either creating new risks or exacerbating existing ones.

Several efforts are underway to measure, evaluate, and classify the risks of AI, and develop frameworks based on the nature and probability of harm. Based on a review of available literature, the Committee outlines the following main categories of risks.[xxxiv]

01 **Malicious uses,** for example misinformation involving the distribution of harmful AI-generated content (deep fakes), trojan attacks using AI tools, model or data poisoning, adversarial inputs in critical infrastructure etc.

02 **Bias and discrimination,** such as the use of inaccurate data to make a decision about future employment, which may result in loss of opportunity or livelihood.

03 **Transparency failures** from the lack of adequate disclosures, for example the use of personal data to develop an AI system without the individual's consent.

04 **Systemic risks,** including disruptions in the AI value chain due to market concentration, geopolitical instability, and regulatory changes.

05 **Loss of control** over AI systems, which could disrupt public order and safety.

06 **National security,** for example AI-facilitated disinformation campaigns, cyberattacks on critical infrastructure and the use of lethal autonomous weapons that threaten public safety and national sovereignty including in relation to counter-terrorism efforts and maintenance of border security.

Beyond these categories, there is a special need to **protect vulnerable groups from the risks of AI.** Children face risks from AI recommendation engines not just through exposure to harmful content, but through the way algorithms exploit their developing brains by prioritising engagement over well-being.[xxxv] These create harm to the long-term mental development and well-being of children.[xxxvi] Given the large number of children in India and the increasing usage of AI tools and applications by children, India could lead the efforts towards building techno-legal solutions to address issues of child safety. Similarly, women face the brunt of AI-generated deepfakes, sometimes referred to as 'revenge porn', even as the harmful creation and distribution of such content remains an acute challenge.

Therefore, the Committee recommends that a suitable risk assessment and classification framework be developed for India that accounts for its unique social, economic, and cultural context, on the basis of which appropriate risk mitigation measures can be deployed.

Incident Reporting

The OECD defines an AI incident as an event, circumstance, or series of events where the development, use, or malfunction of one or more AI systems directly or indirectly leads to a specific harm. These harms include injury to health, disruption of critical infrastructure, human rights violations, or damage to property, communities, or the environment.

To understand AI-related risks in the Indian context, there is a need to collect empirical data about the harms caused by AI.[xxxvii] Based on global best practices, the Committee suggests creating a national database of 'AI incidents', which gives policymakers insights into the real-world risks and harms posed by AI systems—for example, what types of harm are being caused by AI, how does AI contribute to the harm, when does it usually takes place, what are its main causes, etc.—which will inform the development of appropriate risk assessment and classification frameworks for India.

The database should be a national-level centralised system that has the ability to query and collect data from smaller, local databases in a federated manner. Local databases may be set up and maintained by authorised entities or sectoral regulators, provided they follow a standard schema to enable structured data collection and interoperability.[xxxviii]

Such databases are also useful from a national security perspective. They must be expanded to include classified threat intelligence involving incidents of AI-enabled disinformation, cyberattacks, and hybrid threats, provided that such information is securely communicated and stored. Existing incident reporting mechanism, such as those operated by the Indian Computer Emergency Response Team (CERT-In) should be leveraged to monitor vulnerabilities in AI systems across critical sectors and support the development of AI-driven threat detection tools (e.g., anomaly detection, deepfake detection) to counter AI-enabled disinformation. Law enforcement agencies (LEAs) may also collaborate with the AI Safety Institute (AISI) and Technology and Policy Expert Committee (TPEC) to determine how such incident reports can be used to develop risk frameworks that apply to sensitive sectors and protection of critical infrastructure, such as telecom networks, energy grids and nuclear plants.

These incident reporting systems should be designed to encourage participation from public and private organisations, sectoral regulators, and individuals, enabling effective analysis of trends across sectors.

Organisations should be encouraged to report incidents themselves, through protocols that protect confidentiality. The database should be set up in a way that encourages reporting cases without the threat of penalties, with the goal of identifying harms, assessing its impact, and mitigating harm through a multi-stakeholder approach.

Over time, a structured feedback loop should be created: reports feed into threat analysis, which helps policymakers identify emerging risks, understand patterns of harm, and strengthen oversight.[xxxix] This process will also build a culture of accountability.

Voluntary Frameworks

The Committee believes that voluntary measures can serve as an important layer of risk mitigation in India's AI governance framework. While not legally binding, they support norms development, create accountability, and inform future regulatory choices.[xl]

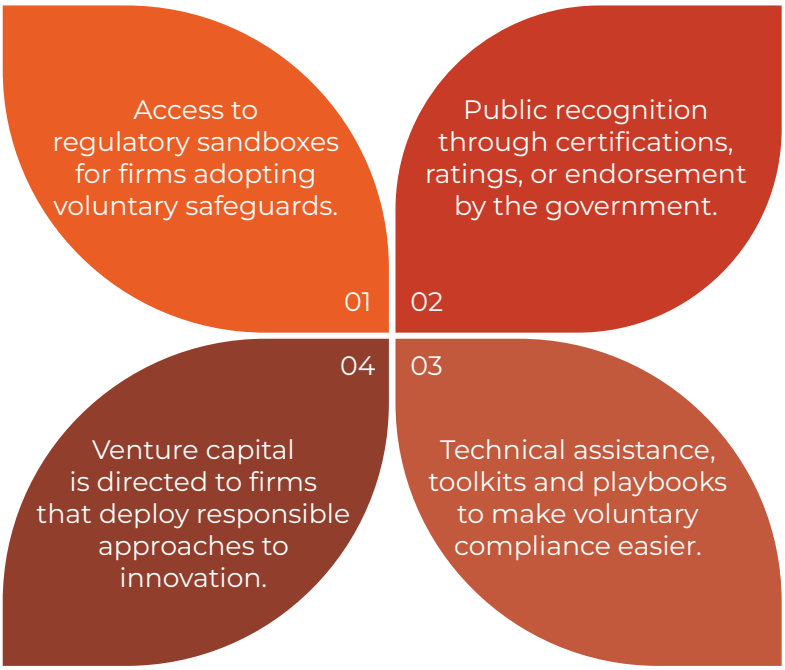
Voluntary measures typically take the form of industry codes of practice, technical standards and self-certifications. Their essential features include optionality, flexibility, adaptability, and lack of legal enforceability or punitive action. The table included in Annexure 5 describes various types of voluntary frameworks relevant for India.

Such voluntary measures align well with the proposed pro-innovation approach, allowing responsible innovation to emerge without compliance-heavy regulations.[xli] They offer the agility to respond quickly as risks evolve, and the flexibility to adapt to India's diverse social and cultural context. Over time, such measures can also provide the evidence base for binding rules, ensuring that governance remains rooted in real-world experience. Therefore, it is important that the evidence they generate should be in a format that regulators and common users can understand, and their impact should be studied on an ongoing basis.

As the industry matures, some of these voluntary measures may be converted into mandatory baseline requirements, which can be enforced by the relevant regulatory bodies.

While voluntary measures are useful in a variety of contexts, they should also be proportionate to the risk of harm. Low-risk applications may require only basic commitments such as transparency reporting and grievance mechanisms, whereas high-risk applications in sensitive sectors such as health or finance may require additional safeguards.

Finally, to ensure that voluntary measures are adopted at scale, the Committee recommends some financial, reputational, technical, and regulatory incentives, for example:



The Committee recommends that voluntary measures be adopted to mitigate the risks of AI. Part 4 of this report contains indicative guidelines for industry and regulators in this regard.

Techno-legal approach

A techno-legal approach to governance uses technology architectures to embed legal requirements directly into system design.[xlii] It is both a design philosophy and family of architectures that makes regulatory principles automatically enforceable in practice.

In a techno-legal approach, specific policy measures are codified and embedded directly into the underlying system through technical standards and protocols. To the extent that it is possible to use technology measures to give effect to regulatory principles, it supports 'compliance-by-design'. In other words, "digital architecture enforces what law requires".[xliii]

A techno-legal approach is also useful to enable innovation at scale while mitigating risks to individuals and society. For India, that means using digital public infrastructure (DPI) like UPI and Aadhar to reach billions of users, with built-in privacy, accountability, and auditability by design. A techno-legal approach helps reduce administrative burden through automated, standardised mechanisms, making risk management more effective and scalable.

These techno-legal measures tend to be most effective when they have been previously tested. Examples of where such measures are useful include content authentication and provenance, privacy-preserving tools for AI development, and transparency in automated decisions that have an impact on life or livelihood. As a general rules, such measures should be adopted in situations where there is a clear regulatory objective to be met (for eg. data protection or non-discrimination) and the measures are likely to have a positive impact on a large number of people.

For these reasons, the Committee encourages the development and use of techno-legal measures to buttress existing policy choices, regulatory instruments, and voluntary measures outlined in the AI governance framework for India.

DEPA for AI Training

One example of how a techno-legal approach can potentially be applied towards AI governance is 'DEPA for AI Training'.

The Data Empowerment and Protection Architecture (DEPA), developed in India and originally deployed in the financial sector, provides a techno-legal system for permission-based data sharing through consent tokens.[xliv]

At its core, DEPA enables techno-legal regulation by codifying legal requirements into technology architecture, ensuring compliance by design, and integrating data protection principles into digital public infrastructure.

Modifying the DEPA for AI Inference architecture and applying it to the development cycle (e.g. DEPA for AI Training) is an example of a techno-legal approach with both opportunities and challenges.[xlv] It supports privacy-preserving mechanisms at the input stage of AI model training and makes the use of personal data for AI training more transparent and auditable. On the other hand, the use of privacy-enhancing technologies can result in a loss of performance on certain benchmarks, which could impact their utility. Further, the DEPA for Training architecture also has a limited role to play in governing downstream AI impacts once the model is trained. These tradeoffs must be examined before adopting these approaches.

Therefore, there is a need for complementary measures for effective AI governance, including:

Algorithmic auditing to detect bias and unfairness.

Transparency frameworks for explainability and accountability.

Sector-specific regulations for sensitive and high-risk AI use cases.

Thus, the DEPA for AI Training approach can act as an enabler, ensuring trust and inclusivity in India's AI ecosystem. Yet, it must sit within a wider AI governance architecture, combining techno-legal tools with ethical, regulatory, and institutional oversight.

Mitigating Loss of Control

AI systems, by design, can evolve in ways that are difficult to fully predict, creating the risk of losing control.[xlvi] To mitigate these risks, the Committee emphasises the need for appropriate mechanisms to retain control and prevent harm. This includes building, where appropriate, human-in-the-loop mechanisms at critical decision points, ensuring that AI outputs can be reviewed, overridden, or supplemented by human judgment before they cause harm. This is consistent with the 'People First' sutra referenced earlier in this report.

In some contexts, such as high-velocity algorithmic trading, direct human oversight is ineffective, given the speed at which they operate. In such cases, safeguards such as circuit breakers, automated checks, or system-level constraints should be considered.

Especially in critical sectors, regular monitoring and testing, audit trails, and reporting protocols should be implemented. The aim is to ensure that AI systems remain within defined bounds, that risks are detected early, and that appropriate risk mitigation interventions are adopted, whether human or otherwise.

Recommendations

- Develop a risk assessment and classification framework that is customised for India's local context, and accounts for risks to vulnerable groups.
- Establish a robust AI incidents mechanism to encourage individuals and organisations to report harm and create a feedback loop to track and analyse risks.
- Encourage the adoption of voluntary frameworks to mitigate risks through principles, commitments, standards, audits, and appropriate incentives.
- Guide the development and deployment of AI systems that are transparent, fair, open, non-discriminatory, explainable, and secure by design.
- Use techno-legal measures, where appropriate, to buttress existing policy choices, regulatory instruments, and voluntary measures.
- Require human oversight and other safeguards to mitigate loss of control risks especially in sensitive sectors involving critical infrastructure.

2.5 Accountability

Accountability, being one of the seven *sutras,* is the backbone of AI governance. In practice, accountability must be secured through a combination of formal mechanisms, grounded in enforcing laws, and other market mechanisms. What matters is that firms feel meaningful pressure to comply, that regulators have an understanding of how firms are complying with the law, and that liability is imposed in a clear, proportionate, and consistent manner.

Legal Enforcement

The Committee notes that many of the risks associated with AI can be addressed under existing laws. However, their effectiveness depends on predictable and timely enforcement.

Therefore, regulators must ensure that in situations where the use of an AI system has resulted in the violation of any law, or where the developer or deployer of an AI system has failed to satisfy their obligations under applicable laws, the applicable legal provisions may be enforced in order to deter repeated offences and to prevent future harm.

To support organisational compliance, clear guidance is essential. Institutions such as the AI Safety Institute, referenced in the next section of this report, should provide guidance notes, model codes, or master circulars clarifying how existing laws apply to AI development and deployment. Such guidance will reduce uncertainty for industry actors, promote voluntary compliance, and ensure that enforcement is proactive rather than reactive.

Accountability Mechanisms

Since voluntary frameworks lack legal enforceability, there is a need to adopt alternative mechanisms that can ensure accountability by creating practical checks at both the organisational and industry level.[xlvii] These mechanisms rely on peer pressure, reputational incentives, and institutional oversight.

Transparency reports: Firms publish red-teaming results, impact assessments, or risk mitigation steps, enabling public and peer scrutiny.

Self-certifications: Firms validate their results through auditors or standards bodies.

Internal policies: Organisations update their service terms to reflect commitments.

Committee hearings: Regulators and parliamentary bodies probe firms on their voluntary compliance efforts.

Peer monitoring: Competitors and civil society observe and report violations.

Techno-legal measures: Compliance is built into system design.

Together, these mechanisms seek to promote voluntary compliance as a first step, following which binding legal enforcement may be necessary. MeitY may publish a schedule to ensure compliance with these measures in the next 9-12 months.

Liability

The Committee is in favour of a graded liability system for AI systems where responsibility is proportional to the function performed, the level of risk anticipated, and the degree to which due diligence is undertaken. This approach ensures that accountability is meaningful without stifling innovation. The Committee recommends the following approach in this regard:

- Clarify how different entities in the AI value chain (e.g. developers, deployers, end-users) are governed under existing regulations, such as the IT Act.
- Recommend principles for attributing liability and responsibility for the concerned entities that is proportionate to their function and the risk of harm (for e.g. transparency reporting, audits, grievance redressal).
- Developing suitable accountability frameworks to mitigate harm.

In addition, the Committee recognises that AI systems are inherently probabilistic and may generate unexpected outcomes, which cause harm, despite reasonable precautions. It notes the recommendation of the RBI's FREE-AI Committee in this regard calling for a 'tolerant' stance in the financial sector towards 'first time/one-off aberrations'. While it is the prerogative of sectoral regulators to pursue enforcement strategies that may be useful in a particular domain, the Committee would like to emphasise that rule of law is paramount and that enforcement strategies should focus on prevention of harm while allowing space for responsible innovation.

Grievance redressal

The Committee recommends that organisations deploying AI systems should establish accessible and effective grievance redressal mechanisms as part of their accountability obligations. Such mechanisms should be designed to make it easy and reliable for individuals to report harms or concerns, without fear of retaliation or undue burden.[xlviii]

Organisations should adopt a proactive approach, ensuring that redressal channels are clearly visible, available in multiple languages and formats, and responsive within reasonable timelines. Feedback received through these channels should be systematically analysed and integrated into product improvements, creating a loop between user experience and risk mitigation. These grievance redressal systems should also be separate from the AI Incidents Database that the Committee has recommended.

Transparency

Accountability cannot exist without transparency. Regulators need to see and understand how AI systems are designed, which actors are involved, the relationship between different actors, and the flow of resources (data, compute) through the different stages of development and deployment—also referred to as the "AI value chain".[xlix]

The Committee is of the view that increasing transparency about the technical, economic, and organisational aspects that guide the development and deployment of AI systems are foundational for designing effective, proportionate, and targeted governance mechanisms, and therefore suitable frameworks may be explored to better understand the AI value chain.

Recommendations

- Clarify how different entities in the AI value chain (for example, developers, deployers, end-users) are governed under existing regulations, such as the IT Act.
- Impose obligations for each of these entities that are proportionate to their function and the risk of harm (for example, transparency reporting, content removal, grievance redressal, transparency, and legal assistance).
- Ensure laws are complied with through timely and consistent enforcement.
- Mandate grievance redressal mechanisms with adequate feedback loops.
- Provide guidance on how existing laws will be enforced in relation to AI systems (for eg. a master circular with a list of applicable regulations to support compliance).
- Develop accountability mechanisms that would support voluntary compliance to mitigate harm (for example, self-certifications, peer monitoring, third party audits).
- Increase transparency of the AI value chain so regulators have an understanding.

2.6 Institutions

India's AI governance framework would benefit from a coordinated effort, in which all line ministries, sectoral regulators, standards bodies and other public institutions work together to develop and implement AI policy. This is known as the "whole-of-government" approach.

To implement this approach, the Committee recommends the following:

The relevant sectoral agencies and regulators should take the lead in monitoring harms, providing guidance and enforcing regulations in their respective domains. For example, the Ministry of Finance and Reserve Bank of India (RBI) would be responsible for implementing the AI governance framework in the financial sector.

MeitY, as the nodal ministry, is responsible for overall adoption and regulation of AI systems. Its role is to promote innovation and adoption of AI technologies, while providing regulatory guidance in collaboration with bodies such as the AI Safety Institute (AISI) and the Indian - Computer Emergency Response Team (CERT-In).

A new body called the 'AI Governance Group' (AIGG) should be set up to coordinate policy on AI governance across all ministries. It should be a small, permanent and effective inter-agency body responsible for overall policy development and coordination on AI governance in India. It should be supported by a Technology & Policy Expert Committee (TPEC), which will advise the group on strategy and implementation. Further details of the proposed AIGG and TPEC are provided below.

A. AI Governance Group (AIGG)

The Committee recommends the creation of an AI Governance Group (AIGG) to develop and oversee India's position and strategy on AI governance.

The AI Governance Group should be a small and effective decision-making body, with a broad mandate on AI policy and governance in India.

Key functions of the AI Governance Group are suggested as follows:

- Coordinate policy across ministries, departments and sectoral regulators, and oversee cross-sectoral governance issues
- Review existing mechanisms and issue guidelines to ensure that firms are held accountable for compliance with local laws.
- Oversee national initiatives on AI governance across the public and private sector.
- Promote responsible AI innovation and beneficial deployment of AI in key sectors.
- Study the emerging risks of AI, regulatory gaps, and need for legal amendments.

It is suggested that representatives from the following institutions be a part of the group:

Suggested composition (illustrative and subject to periodic reviews)

Chair	• Principal Scientific Adviser (PSA)
Government agencies	• Ministry of Electronics and Information Technology • Ministry of Home Affairs • Ministry of External Affairs • Department of Science & Technology • Department of Telecommunications
Regulators	• Telecom Regulatory Authority of India (TRAI) • Competition Commission of India (CCI) • Data Protection Board (DPB) • Sectoral regulators and bodies such as the Reserve Bank of India (RBI), Securities and Exchange Board of India SEBI, Indian Council of Medical Research (ICMR), University Grants Commission (UGC), etc.
Advisory bodies	• NITI Aayog • Office of Principal Scientific Advisor

Technology & Policy Expert Committee (TPEC)

A Technology & Policy Expert Committee (TPEC) should be set up by MeitY, comprising a small group of experts with experience in domains such as:

- Research and development in frontier technologies
- Engineering, machine learning, data science, etc.
- Law and public policy with a focus on emerging technologies
- Public administration, including current and former government officials
- National security, including law enforcement experts

The TPEC's primary goal is to provide expertise to the AI Governance Group (AIGG) and enable it to perform its functions effectively. It will brief the AIGG on matters of national importance in relation to AI policy and governance, including with respect to:

- New and emerging capabilities of AI
- Potential risks and regulatory gaps
- Global developments in AI policy and governance
- India's diplomatic engagements on AI governance

B. AI Safety Institute

The recently established AI Safety Institute (AISI) should act as the main body responsible for guiding the safe and trusted development and use of AI in India.

The AISI should be involved in research, risk assessment, and capacity-building. It should test and evaluate AI systems for risks and provide advice to policymakers and industry actors on issues of AI safety. Further, the ongoing work under the IndiaAI mission to support the development of technical solutions to address issues relating to machine unlearning, bias mitigation, privacy-enhancing tools, explainable AI, etc. should also continue.[1]

The AISI should also anchor India's participation in global forums and facilitate collaborations, such as in the International Network of AI Safety Institutes.

The AISI can operate on a hub-and-spoke model and should be supported by a dedicated secretariat for research, drafting, and capacity building.

Key functions of the AISI are suggested as follows:

- Coordinate with agencies, sectoral regulators, and public bodies on AI safety issues.
- Analysis the emerging risks of AI and potential regulatory gaps.
- Develop draft guidelines, codes, standards, respective evaluation metrics and testing frameworks in collaboration with relevant agencies and sectoral regulators.
- Provide practical advice to support voluntary efforts to mitigate risks.
- Conduct forecasting research on the potential impact of AI and issues in online safety, privacy, data governance, labour, and competition.
- Promote the adoption of AI safety tools in areas such bias mitigation, fairness testing, and explainability, through platforms, APIs and open access tools.
- Foster public-private partnerships to develop tools that can support law enforcement and enhance trust and transparency.
- Conduct training programs on AI safety to build awareness and institutional capacity.
- Represent India in international forums such as the Network of AI Safety Institutes, ensuring that India's perspectives on scale, diversity and inclusion are reflected.
- Support the TPEC and AIGG by providing risk assessments, updates on industry compliance and policy recommendations.

Recommendations

- Establish an AI Governance Group to coordinate overall policy development and align AI governance frameworks with national priorities and strategic objectives.
- Constitute a Technology & Policy Expert Committee to provide expert inputs to the AI Governance Group on matters of national and international importance relating to AI governance.
- Provide adequate resources to the IndiaAI Safety Institute to conduct research, develop draft standards and their evaluation metrics and testing methods and benchmarks, collaborate with international bodies, national standard making bodies and provide technical guidance to regulators and industry.

Part 3: Action Plan

The Action Plan below identifies outcomes mapped to short, medium and long-term timelines.

Timelines	Action Items	Expected Outcomes
Short-term	◆ Establish the AI Governance Group (AIGG) as a permanent high-level policy making body. ◆ Constitute the Technology & Policy Expert Committee (TPEC) to support the AIGG. ◆ Develop India-specific AI risk assessment and classification frameworks with sectoral inputs. ◆ Conduct regulatory gap analysis and suggest appropriate legal amendments and rules. ◆ Adopt voluntary frameworks to promote responsible innovation and mitigate risks. ◆ Publish a master circular with applicable regulations and best practices to support compliance. ◆ Prepare the groundwork for AI incidents database and grievance redressal mechanisms. ◆ Develop clear liability regimes across the AI value chain. ◆ Expand access to foundational infrastructure including data, compute and models. ◆ Launch public awareness and training programs for citizens and regulators on AI capabilities and risks. ◆ Operationalise Safe and Trusted tools in areas such as bias mitigation, privacy-enhancing tools, deepfake detection, etc.	◆ Strong institutions to coordinate AI governance. ◆ Frameworks for risk classification and mitigation customised for the Indian context. ◆ Culture of voluntary industry compliance. ◆ Understanding of regulatory gaps and needs. ◆ Infrastructure in place for incident reporting and grievance redressal. ◆ Improved societal trust and literacy on AI.

Timelines	Action Items	Expected Outcomes
Medium-term	◆ Publish common standards (e.g. content authentication, data integrity, fairness, cybersecurity). ◆ Operationalise national AI incidents database with localised reporting and feedback loops. ◆ Amend laws, as may be needed, to address regulatory gaps ◆ Pilot regulatory sandboxes in high-risk domains ◆ Support the integration of Digital Public Infrastructure (DPI) with AI with policy enablers.	◆ Mature, standardised governance framework. ◆ Safe experimentation environment for innovation. ◆ Broader adoption of DPI-enabled AI systems ◆ Easier compliance through guidance and updated laws. ◆ Effective grievance redressal for citizens.
Long-term	◆ Continuously review and monitor the governance framework and activities under this Action Plan ◆ Adopt new laws to account for emerging risks and capabilities. ◆ Expand global diplomatic engagement and contribute to standards development. ◆ Conduct horizon-scanning & scenario planning to prepare for future risks and opportunities.	◆ Mature, balanced and agile legal framework. ◆ International credibility in AI governance leadership. ◆ Effective accountability system for AI harms. ◆ Future-ready governance system for emerging risks.

Institutional framework for AI Governance in India (illustrative)

An institutional framework to implement the AI governance guidelines is suggested below, mapping key agencies, sectoral regulators, advisory bodies to key functions.[li]

	Key Institution	Key functions
Inter-Ministerial body	• AI Governance Group (AIGG)	• Overall policy formulation coordination of AI governance in India across all agencies
Nodal ministry	• Ministry of Electronics and Information Technology (MeitY)	• Nodal ministry responsible for AI governance in India
Government agencies	• Ministry of Home Affairs (MHA) • Ministry of External Affairs (MEA) • Ministry of Agriculture • Ministry of Education • Ministry of Healthcare • Department of Science and Technology (DST) • Department of Telecommunications (DoT) • Department for Promotion of Industry and Internal Trade (DPIIT) • Indian - Computer Emergency Response Team (CERT-In) • Grievance Appellate Committee (GAC)	• Responsible for AI governance in their respective domains • Issuing sector-specific rules and regulations • Enforcing applicable laws and regulations in these domains • Supervising compliance efforts and legal mandates for domain-specific applications. • Handling grievances in their respective domains • Monitoring of AI-driven disinformation, cybersecurity analysis and attribution. • Responsible for India's diplomatic engagements on AI governance
Advisory bodies	• AI Safety Institute (AISI) • Technology & Policy Expert Committee (TPEC) • National Institution for Transforming India (NITI Aayog) • Office of the Principal Scientific Advisor (PSA)	• Supporting the AI Governance Group with regular briefings and strategic advice on AI governance

	Key Institution	Key functions
Sectoral regulators and bodies	• Reserve Bank of India (RBI) • Securities and Exchange Board of India (SEBI) • Insurance Regulatory and Development Authority of India (IRDAI) • Telecom Regulatory Authority of India (TRAI) • Indian Council for Medical Research (ICMR) • National Health Authority (NHA)	• Issuing sector-specific rules and regulations • Enforcing applicable laws and regulations in these domains • Supervising compliance efforts and legal mandates for domain-specific applications • Handling grievances in their respective domains
Standards bodies	• Bureau of Indian Standards (BIS) • Telecommunication Engineering Centre (TEC)	• Developing standards in relation to AI risk taxonomies, certification standards, etc. • Engagement with global tandard setting bodies • Standardising testing, assessment, evaluation and validation procedures

Part 4: Practical Guidelines for Industry & Regulators

Guidelines for industry

The Committee recommends that any person involved in developing or deploying AI systems in India should be guided by the following:

- Comply with all Indian laws and regulations, including but not limited to laws relating to information technology, data protection, copyright, consumer protection, offences against women, children, and other vulnerable groups that may apply to AI systems.
- Demonstrate compliance with applicable laws and regulations when called upon to do so by relevant agencies or sectoral regulators.
- Adopt voluntary measures (principles, codes, and standards), including with respect to privacy and security; fairness, inclusivity; non-discrimination; transparency; and other technical and organisational measures.
- Create a grievance redressal mechanism to enable reporting of AI-related harms and ensure resolution of such issues within a reasonable timeframe.
- Publish transparency reports that evaluate the risk of harm to individuals and society in the Indian context. If they contain any sensitive or proprietary information, the reports should be shared confidentially with relevant regulators.
- Explore the use of techno-legal solutions to mitigate the risks of AI, including privacy-enhancing technologies, machine unlearning capabilities, algorithmic auditing systems, and automated bias detection mechanisms.

Guidelines for regulators

The Committee suggests the following principles to guide policy formulation and implementation by various agencies and sectoral regulators in their respective domains:

- The twin goals of any proposed AI governance framework is to support innovation, adoption and the distribution of the technology's benefits to society, while ensuring that potential risks can be addressed through policy instruments.
- Governance frameworks should be flexible and agile, such that it enables periodic reviews, monitoring, and recalibration based on stakeholder feedback.
- When using policy instruments to mitigate risks, regulators should prioritise those where there is real and present harm or a threat to life, livelihood or well-being.
- Proposed AI governance frameworks should avoid compliance-heavy requirements (for example, mandatory approvals, licensing conditions, etc.) unless deemed necessary.
- The appropriate regulator or agency should determine which type of policy instrument is the most useful, relevant, and least burdensome to achieve the desired objective (for example, industry codes, technical standards, advisories, binding rules).
- Regulators should encourage the use of techno-legal approaches to meet policy objectives around privacy, cybersecurity, fairness, transparency, etc. where such policy measures have already been put in place.

Glossary

Sl. No.	Term	Description
01	**Accountability**	The obligation of individuals or organizations to account for their actions, accept responsibility, and disclose results transparently through specific means and criteria.
02	**Adversarial Input Attacks**	Deliberate changes to input data intended to mislead AI models into incorrect decisions or predictions.
03	**Agentic AI**	Highly autonomous system that senses and responds to its environment and takes actions to achieve its goals.
04	**AI Incident**	An event where an AI system malfunctions, produces unintended outcomes, or behaves unpredictably, potentially causing harm or violating legal rights.
05	**AI Safety Institute**	An institution under India AI Mission promoting safe, secure, and trustworthy AI innovation by coordinating research and collaboration across academia, industry, startups, and government.
06	**Algorithmic Trading**	Automated rule-based trading where decisions are made by computer models.
07	**Artificial Intelligence**	An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

08	**Auditability**	The ability to inspect and verify system processes and decisions.
09	**Behaviour Audit**	Evaluating AI decisions in real-world settings for ethical and legal alignment.
10	**Bias**	Systematic difference in treatment of certain objects, people or groups in comparison to others.
11	**Data Minimisation**	Collecting only as much personal data as is necessary to achieve a specific purpose.
12	**Data Poisoning**	Manipulating training data to corrupt AI/ML models.
13	**Deepfake**	AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful.
14	**Equity**	Fair treatment of individuals.
15	**Explainability**	Property of an AI system to express important factors influencing the AI system results in a way that humans can understand.
16	**Fairness**	Ensuring AI decisions are free from harmful bias or discrimination.
17	**Federated Learning**	Federated learning is a decentralized approach to training machine learning (ML) models. Each node across a distributed network trains a global model using its local data, with a central server aggregating node updates to improve the global model.
18	**Foundation Models**	Large AI models trained on vast datasets for general tasks.

19	**Generative AI**	Models that generate text, images, or other content.
20	**GPU (Graphics Processing Unit)**	A co-processor designed to accelerate graphics and image processing, and specialized tasks in Machine Learning and Deep Learning involving heavy matrix operations.
21	**Human in the loop/ Human-allied AI**	Involving human expertise in the AI lifecycle particularly during training and deployment to actively improve system performance & reliability.
22	**Large Language Models**	Foundation models capable of understanding and generating natural language.
23	**Machine Learning**	A process of optimizing model parameters through computational techniques, such that the model's behaviour reflects the data or experience.
24	**Model Bias**	Systematic errors in a model arising from erroneous assumptions during the modelling process, that cause it to consistently make incorrect or skewed predictions.
25	**Red Teaming**	An exercise, reflecting real-world conditions, that is conducted as a simulated adversarial attempt to compromise organizational missions and/or business processes to provide a comprehensive assessment of the security capability of the information system and organization.
26	**Small Language Models**	AI models, smaller in scope and scale, capable of processing, understanding & generating natural language content, audio, video, etc.
27	**Transparency**	Making information about an AI system available to relevant stakeholders in an accessible and understandable manner, to the extent technically feasible.
28	**Understandability**	Ease with which users comprehend AI operations and outputs.

Annexures

- Background of the Drafting Committee
- Overview of global AI governance frameworks
- Overview of current laws in India applicable to AI systems
- Applicability of existing laws in India to AI harms
- Types of voluntary frameworks for AI risk mitigation
- Standards published/under development by BIS

Annexure 1: Background of the Drafting Committee

Constitution of the Committee

The Government of India set up a high-level advisory group in 2023 under the chairmanship of the Principal Scientific Advisor (PSA) to examine various issues relating to AI. The committee under PSA, after extensive deliberations, set up a sub-committee on AI governance, that included Prof Balaraman Ravindran, Debjani Ghosh and Sharad Sharma. The subcommittee prepared a draft report which was published by MeitY for public feedback. More than 2,500 submissions were received from government bodies, academic institutions, think tanks, industry associations, private sector organisations, and individual stakeholders. A drafting committee was formed (Committee) to review stakeholder feedback and has prepared this report on the AI governance framework.

Terms of Reference of the Committee

The Terms of Reference of the subcommittee set up by the PSA and the drafting committee constituted by MeitY,

- To recommend a governance framework that promotes innovation and adoption of AI in India while mitigating the risks to individuals and society.
- To present a rationale for India's approach to AI governance based on local factors.
- To create a foundation of trust so that future development of AI promotes long-term growth, resilience and sustainability of India's digital ecosystem.
- To provide a set of practical guidelines for industry to promote ease of doing business and global competitiveness of Indian firms.
- To provide guiding principles for sectoral agencies and regulators to make informed decisions with respect to AI governance in their respective domains.

Members of the Committee

The Committee constituted by the Ministry of Electronics and Information Technology (MeitY) in July 2025 to draft this report comprises the following members:

Name & Affiliation	Designation
Balaraman Ravindran, Professor, IIT Madras	Chairman
Abhishek Singh, Additional Secretary, MeitY	Member
Debjani Ghosh, Distinguished Fellow, NITI Aayog	Member
Kalika Bali, Advisor, Safe and Trusted AI, IndiaAI	Member
Rahul Matthan, Partner, Trilegal	Member
Amlan Mohanty, Non-Resident Fellow, NITI Aayog	Lead Writer
Sharad Sharma, Co-founder, iSPIRT	Member
Kavita Bhatia, Scientist G, MeitY & COO, IndiaAI	Member
Abhishek Aggarwal, Scientist D, MeitY	Member
Avinash Agarwal, DDG(IR), DoT	Invitee Member
Shreeppriya Gopalakrishnan, DGM, IndiaAI	Member Convenor

Annexure 2: Overview of Global AI governance frameworks

Jurisdiction	Summary of Approach
Australia	Ongoing deliberations on a government whitepaper titled **"Safe and Responsible AI in Australia"**, proposing mandatory guardrails to regulate AI in high-risk settings and general-purpose AI models.
Brazil	Proposals for a new AI law (Bill No. 2,338/2023) that promotes secure, reliable AI systems, categorizing them by risk and imposing various compliance requirements.
Canada	Published the **draft Artificial Intelligence and Data Act (AIDA)** that focuses on responsible AI use, consumer protection, and fair competition. The law is still at the parliamentary review stage.
China	Technology-specific regulations aimed at specific issues, including algorithmic recommendations and generative AI. Various national standards for AI systems and 'Labeling Rules' have also been introduced to enhance the security and governance of generative AI.
European Union	Statutory framework in the form of the **Artificial Intelligence Act** that categorizes systems by risk levels, imposes stringent requirements on high-risk applications, and aims for transparency and accountability.
Japan	Adopted the law on **Promotion of AI-Related Technologies** in May 2025. It establishes an AI Strategy Center and implements non-binding guidelines to promote innovation and adoption. The framework emphasizes voluntary compliance and international cooperation.
Singapore	Voluntary, use-case based approach that emphasizes a sectoral approach based on governance frameworks. It has released a **draft Model AI Governance Framework for Generative AI** to address emerging risks and provide guidance for safety evaluations. It has developed practical testing methods such as Veritas and AI Verify, which allow organisations to evaluate fairness and transparency in real use cases.
United Kingdom	Context-based and cross sectoral framework that focuses on core principles (safety, transparency, fairness, accountability, contestability) that will be implemented by existing sectoral regulators.

United States of America	A pro-innovation approach that emphasises innovation, infrastructure development and international diplomacy to promote American leadership and global competitiveness. Voluntary commitments, such as the NIST AI Risk Management Framework, and some executive orders relating to AI governance are applicable.
South Korea	Adopted the **Basic Act on the Development of Artificial Intelligence and Establishment of Trust.** The Act adopts a risk-based approach focusing on high-impact AI systems and generative AI transparency requirements, with moderate enforcement through administrative fines.
New Zealand	Developed the **Algorithm Charter for Aotearoa New Zealand** in 2020 which applies specifically to public sector algorithmic decisions, establishing six commitments for fair, ethical, and transparent government algorithm use. The framework emphasizes human oversight and Māori data sovereignty considerations.
Israel	**"Artificial Intelligence Regulations and Ethics"** encourages "responsible AI innovation in the private sector" through a principled-based, sector-specific regulatory approach using 'soft' tools, such as non-binding ethical principles and voluntary standards.
South Africa	**National AI Policy Framework** establishes twelve strategic pillars for responsible AI development. The framework emphasizes human-centered AI, addressing socioeconomic disparities through talent development, digital infrastructure, and ethical governance.

Annexure 3: Overview of current laws in India relevant to AI systems (Illustrative)

Below is an illustrative list of statutes and regulations in India that may be applicable to the development, deployment and use of AI systems.

Information Technology Act, 2000 (IT Act):
The IT Act remains the backbone of India's digital regulation. Section 66D addresses cheating by personation using computer resources, applicable to AI-generated impersonations and deepfakes. Section 79, along with the 2021 Intermediary Guidelines, places due diligence obligations on online platforms, requiring active monitoring and takedown of unlawful AI-generated content, including misinformation and harmful deepfakes.

Bharatiya Nyaya Sanhita, 2023 (BNS):
In addition to the IT Act, certain harms/cybercrimes perpetuated by AI could also fall under the BNS. For instance, identity theft and cheating by personation are offences under Section 319(2) (cheating by personation), section 336(1) and 336(2) (forgery for the purpose of cheating), section 294 and 296 (selling/circulating/distributing obscene objects), and section 356(1) (causing harm to reputation/defamation).

Digital Personal Data Protection Act, 2023 (DPDP Act):
The DPDP Act introduces obligations of consent, purpose limitation, and data minimisation that have direct bearing on AI model training and deployment. It prohibits processing of personal data without consent, requires safeguards against misuse of sensitive data, and empowers the Data Protection Board to investigate harms caused by misuse of AI-driven profiling. These provisions create accountability pathways for AI developers and deployers handling personal data at scale.

Consumer Protection Act, 2019 (CPA):
The CPA protects consumers against unfair trade practices, misleading advertisements, and deficiency of service. Its provisions can be invoked where AI-enabled systems mis-sell financial products, misrepresent the capabilities of AI-driven health devices, or cause consumer harm through opaque algorithms in e-commerce. The Central Consumer Protection Authority is empowered to order corrective advertising or levy penalties on misleading AI claims, including advanced forms of dark patterns.

Sectoral legislations:
Sector-specific legislations such as the Telecommunications Act, 2023, under which rules are being notified in areas such as cybersecurity, critical infrastructure, and incident reporting also strengthen the implementation of AI governance principles.

AI-specific guidelines:
Sectoral regulators and technical bodies have been adapting their mandates to address AI-specific risks, issuing frameworks on cybersecurity, fairness, robustness, and ethical safeguards. These initiatives reflect the operational realities of each domain: financial stability in banking, integrity in securities markets, safety and reliability in telecom, and accountability in healthcare. Collectively, they demonstrate how India's oversight architecture is evolving in practice.

Reserve Bank of India (RBI):
RBI's regulatory architecture on technology risk has progressively expanded to cover AI. The *Cybersecurity Framework for Banks (2016)* established board-approved cyber policies, continuous monitoring, incident reporting, and resilience planning, all of which extend to AI-enabled services.

The Digital Lending Guidelines (2022) require transparency, consent, and accountability in automated decision-making, and are now expected to incorporate disclosure obligations for AI-driven credit scoring and fairness audits.

Building on these foundations, the *Framework for Responsible, Explainable and Ethical AI (FREE-AI) Committee Report (2025)* sets out detailed AI-specific measures: adoption of board-approved AI policies covering governance, lifecycle management, vendor oversight, and annual review; integration of AI-specific threats such as adversarial attacks and model poisoning into cybersecurity protocols; and the creation of a tiered incident reporting system for AI failures, including bias, explainability gaps, and unintended outcomes.

Securities and Exchange Board of India (SEBI):
SEBI's *Cybersecurity and Cyber Resilience Framework* requires market infrastructure institutions and intermediaries to maintain security operation centres, conduct vulnerability assessments, and submit compliance reports. AI-driven trading algorithms and surveillance systems fall under this framework, linking automation to accountability for market integrity. SEBI has also released a consultation paper on *"Guidelines for responsible usage of AI/ML In Indian Securities Markets"* in June, 2025.

Insurance Regulatory and Development Authority of India (IRDAI):
IRDAI mandates insurers and intermediaries to comply with its *Guidelines on Information and Cyber Security* for Insurers, with direct implications for AI-driven underwriting, claims management, and fraud detection.

Telecommunication Engineering Centre (TEC):
TEC has issued a Voluntary Standard for Fairness Assessment and Rating of AI Systems, covering bias detection and mitigation, and is developing a Standard for Assessing & Rating Robustness of AI Systems in Telecom Networks and Digital Infrastructure. TEC has also published a Draft Standard for the Schema and Taxonomy of an AI Incident Database in Telecommunications and Critical Digital Infrastructure. These standards provide structured pathways for trustworthy AI assessment focusing on fairness, robustness, and incident reporting in areas like critical infrastructure, network optimisation, and service quality management.

Indian Council of Medical Research (ICMR):
The *Ethical Guidelines for Application of AI in Biomedical Research and Healthcare* set expectations for safety, transparency, accountability, fairness, and human oversight. They require bias audits, independent ethics review, data quality checks, and delineation of responsibility between developers and healthcare providers.