To

The Ministry of Electronics and Information Technology

COMMENTS AND SUGGESTIONS ON

THE REPORT ON ARTIFICIAL INTELLIGENCE GOVERNANCE GUIDELINES DEVELOPMENT, 2025

FEBRUARY 2025

CENTRE FOR ADVANCED STUDIES IN CYBER LAW AND ARTIFICIAL INTELLIGENCE (CASCA)
RAJIV GANDHI NATIONAL UNIVERSITY OF LAW, PUNJAB

Sidhuwal, Bhadson Road, Patiala - 147001
Email ID: casca@rgnul.ac.in

Comments to The Ministry Of Electronics And Information Technology

The Report On Artificial Intelligence Governance Guidelines Development, 2025

Authors: Tanmay Durani, Sanskriti Bishnoi, Aadit Seth, Amishi Jain, Kunaal Hemnani, R. Dayasakthi, Raima, Sanskriti Koirala, Shoptorishi Dey, Swastika Saha Chowdhury, Vishwaroop Chatterjee, Aarav Singhal, Eknoor Kaur, Jacob, Tarush Saitia, Uday Gupta

Research Consultant: Dr. Ivneet Walia, Associate Professor of Law and Officiating Registrar, RGNUL

CENTRE FOR ADVANCED STUDIES IN CYBER LAW AND ARTIFICIAL INTELLIGENCE [CASCA] is a research-driven centre at RGNUL dedicated to advancing scholarly research and discourse in the field of Technology Law and Regulation. As a research centre of a leading institution in India, we are committed to promoting interdisciplinary research, fostering collaboration, and driving innovation in the fields of cyber law, artificial intelligence, and other allied areas.

For more information

Visit **cascargnul.com**

Disclaimer

Table of Contents

I. Introduction

1. The need for AI Governance

Artificial Intelligence ("**AI**") governance is the system of strategic framework of policies, designed to ensure that AI technology is used productively, ethically and securely to minimize risks.[1] This system promotes ethical and responsible use of AI, which bridges the gap between accountability and ethics. The governance system works in tandem with established guidelines and legal standards so that AI is compliant with regulations.

The need for AI governance arises from how AI functions for society and organizations both. The point of governance is to prevent unethical use and harm by building the key things for AI technology, i.e. compliance, transparency, trust and ethical use.[2] As AI becomes more deeply integrated into society, AI governance provides oversight to formulate a framework to ensure sustainable and ethical development of AI technologies. It otherwise poses risks of biased decision-making that go undetected when operated without transparency as AI logic is often hard to identify.

In consonance, AI governance is instrumental in increasing transparency and accountability. For this, it is essential to understand the domain in which the AI ecosystem operates, which is related to its computational capacity and the context of its operation in different sectors. Principles and guidelines, herein, prove of essence to ensure that AI is deployed with responsibility and constructive use as per the sector it functions in, for instance, healthcare or recruitment or academia.[3]

The guiding principles for AI governance should function consistent to the contemporary Indian framework. The United Nations' Report[4] on the irrefutable need for AI governance shows the gaps in how technology thrives off disparities and infringes on human rights.[5] It sheds light on the ethics of AI. There also exists a stressed need for equity in AI governance, which is imperative to be incorporated in India. To operationalize AI governance, the need of the hour is to provide a foundation for governing the AI ecosystem in consonance with contemporary standards. The key guiding principles[6] are:

I. **Transparency**: AI should disseminate meaningful information, such as interpretable policies for when AI is deployed in a sector and when users interact with the technologies, such as, online support centres for customer queries.

---

[1] Parashar P, "AI Governance: Ensuring Safe Adoption of AI Technologies" (*HCLTech*, October 31, 2024) <https://www.hcltech.com/trends-and-insights/ai-governance-ensuring-safe-adoption-ai-technologies>

[2] "Artificial Intelligence and Privacy – Issues and Challenges – Office of the Victorian Information Commissioner" <https://ovic.vic.gov.au/privacy/resources-for-organisations/artificial-intelligence-and-privacy-issues-and-challenges/>

[3] Appiah F, "What Is AI Governance? The Reasons Why It's so Important" (American Military University, July 15, 2024) <https://www.amu.apus.edu/area-of-study/information-technology/resources/what-is-ai-governance/>

[4] United Nations, "Governing AI for Humanity: Final Report" (2024)

[5] "'Irrefutable' Need for Global Regulation of AI: UN Experts" (UN News, September 19, 2024) <https://news.un.org/en/story/2024/09/1154541>

[6] OECD, https://www.oecd.org/en/topics/sub-issues/ai-principles.html

II. **Accountability**: Developers of each AI system owe responsibility to the society and legal system to adhere to user rights. Mechanisms exist to ensure clear accountability and attribute potential issues to which sector effectively.

III. **Safety, Reliability, and Robustness**: AI systems, by the virtue of being formed by technology and the increasing reliance on it, need to ensure resilience to inconsistencies or minimize scope of adverse outcomes. Safe AI environment includes regular monitoring of the technological landscape so that the intended functions are performed and adhered to.

IV. **Privacy and Security**: AI systems need to comply with data protection laws, and due regard to user privacy. This forms the core basis of AI governance as it poses a critical need to ensure the ethical use of AI.

V. **Fairness and Non-Discrimination**: With growing instances of AI recording bias and discrimination, it becomes imperative to promote fairness and inclusivity.[7] Prejudices prevalent in the societal landscape need not impede in the digital arena and undermine equality.

VI. **Human-Centered Values and 'Do No Harm'**: While the purpose AI serves is to ease human work, the governance must ensure that human intervention is still as needed to prevent undue reliance and ethical dilemmas are resolved. Adverse societal impacts that AI may create can only be resolved through the rule of law, which is apt AI governance.

VII. **Inclusive and Sustainable Innovation**: Benefits of innovation should be accrued by all, equitably. Achievement of SDGs is the primary goal which the globe is striving towards, and AI should leverage that instead of impeding on human success. Beneficial outcomes for all should be prioritized.

VIII. **Digital-by-Design Governance**: Appropriate techno-legal measures to operationalize AI governance is imperative so that AI systems are enhanced for compliance effectively.

The need for AI governance is not merely a regulatory necessity but a moral imperative to ensure that AI technologies serve humanity equitably and ethically. This research aims to draw a roadmap for how AI is developed and deployed in the technological landscape and how it can be instrumental in enhancing governance. Further, the need remains to ensure transparency and accountability in order to ensure acknowledgement and adherence to limitations and legal standards of the Indian jurisprudence. How AI incorporates decision-making skills while supporting Sustainable Development Goals ("**SDGs**") and to bridge the legal gaps, a coordinated government approach is needed, which will also ensure traceability of actors in the AI ecosystem.

[7] Reuters, "Racist, Sexist, Casteist: Is AI Bad News for India?" (The Hindu, September 11, 2023) <https://www.thehindu.com/sci-tech/technology/racist-sexist-casteist-is-ai-bad-news-for-india/article67294037.ece>

II. Foundations of AI Governance

2. AI-Development Lifecycle

A. Background

The AI life cycle comprises three distinct phases: Firstly, *AI development*, which is inclusive of problem formulation, data collection and model training. Secondly, *AI deployment* entails the integration of the developed model into operational systems; *AI diffusion* refers to its adoption and enhancement, as well as consideration of the impact and assessment of the model with reference to legal frameworks and ethics.

Deployers should bear in mind that the system lifecycle is not always rigid and can be a continuous progressive process of tailoring AI systems that have already been developed and deployed.[8] AI oversight should be developed into AI systems by its inherent design. In all the phases of the AI development lifecycle, deployers should make sure that ethical business objectives, governance and key stakeholders are properly identified and aligned with principals of justice, fairness and transparency.[9] The idea of rigorous testing throughout the AI lifecycle guarantees the model's performance and ability to adapt across the demographics and minimise risks such as biasness, misinformation, data privacy issues, etc. These adjustments make the AI have minismed risk associated with it. The difficulties outweigh the possibilities themselves when it comes to ethical considerations and bias mitigation at every stage of a project. Outlining the key measures to safeguard AI programs from potential misuse and mechanisms from threats and possible gaps guarantees the system's stability and reliability.

B. Ensuring AI Governance Across the AI Lifecycle - Recommendations

Few recommendations to ensure that AI governance is ensured throughout AI lifecycle are and its different phases are:

(1) AI Development - The key recommendation for the AI Development phase shall be to regulate the procedure of data collection. Data used for model training, testing, and validation should be adequately indicative to reduce risks of arbitrary bias. This can be accomplished through record of datasets procured and variable outcomes of the model across distinct target population sub-groups.[10] If programmers are supplying AI systems to deployers, they should provide a suitable disclosure or supporting record that outlines the types of data used to train the AI system, and how

[8] Pretorius C, "EU Employment Law and the AI Act: A Policy Brief Putting the Human Back in 'Human-Centric' Policy" (European Center for Populism Studies (ECPS) 2024) <https://doi.org/10.55271/pop0002>

[9] "What Is the AI Development Lifecycle?" (Palo Alto Networks) <https://www.paloaltonetworks.com/cyberpedia/ai-development-lifecycle>

[10] Camm WB, "Privacy Act and the Data Base: Implementation of the Privacy Act" (Defense Technical Information Center 1981) <https://doi.org/10.21236/adp001300>

they have controlled potential bias. In consonance, the Data Protection laws shall be embedded into the AI's algorithm from its initial development.[11]

(2) AI Deployment - The second stage of the lifecycle shall incorporate principles of Data Lineage to ensure that data can be tracked throughout its lifecycle, from its origin to its current destination.[12] This entails tracking every transformative step and link between data points to gain a clear picture of how data evolves. One approach to accomplish this is to preserve a data provenance record, which allows an organization to determine the quality of data based on its origin and following transformation, trace probable error sources, update data, and attribute data to sources. Data retention policies must be created to guarantee data is stored and disposed of appropriately.[13]

(3) AI Diffusion - An outcome analysis process should be established as deployers should document the AI system's objective and ensure that risks are managed through proper means. Identifying and analysing ethical obstacles is imperative for addressing their fundamental cause. During the outcome analysis sequence, corporate stakeholders should monitor the AI system's performance and ensure it meets the initial objectives with its interaction with humans.[14] Organizations can conduct acceptance tests to evaluate both functional and non-functional features, such as security and performance. Any unwanted obstacle arising out of the diffusion of the AI must be excluded from the AI program with due effect.

3. The AI Ecosystem: Roles and Responsibilities of Providers, Developers, Deployers, and End-Users

The network of participants who perform various functions ranging from the process of creation and implementation of AI solutions, all these actors combined are known as the AI ecosystem.

A. Role of Data Principals in AI Development

At the base are the data subjects whose information is fed into the AI systems; these data subjects are also called data principals. While training their AI algorithms, Meta gathers user interactions, posts and comments of billions of Facebook and Instagram users in such a scenario people who have no idea how their everyday data usage of social media contributes to advancement of AI systems of Meta[15]. Scale AI

---

[11] Rauh L and others, "Towards AI Lifecycle Management in Manufacturing Using the Asset Administration Shell (AAS)" (2022) 107 Procedia CIRP 576

[12] "Data Lineage: Making Artificial Intelligence Smarter" (*SAS India*) <https://www.sas.com/en_in/insights/articles/data-management/data-lineage--making-artificial-intelligence-smarter.html>

[13] Kiernan R, "Unveiling the Path: Why Data Lineage Is Crucial for Building Effective AI Products" (*Artefact*, November 21, 2024) <https://www.artefact.com/blog/unveiling-the-path-why-data-lineage-is-crucial-for-building-effective-ai-products/>

[14] Healey R, "ASEAN AI Governance: Shaping the Future of Southeast Asia" (*Formiti*, February 19, 2024) <https://formiti.com/navigating-the-asean-ai-governance-framework-a-pathway-to-responsible-innovation/>

[15] Jiménez J, 'Worried About Meta Using Your Instagram to Train Its A.I.? Here's What to Know.' (Worried About Meta Using Your Instagram to Train Its A.I.? Here's What to Know., 26 September 2024) <www.nytimes.com/article/meta-ai-scraping-policy.html> accessed 20 January 2025.

which provides relevant labelled data used to develop self-driving cars. This data then goes to the AI developers and model builders who use it to design AI systems through decisions in the training process[16]

B. Developers and LLMs

Data gathered from these data principles is used by AI developers and model builders, who create LLMs (large language models) based on this data then these LLMs (Claude is a LLM) in some cases many LLM's are integrated in one model (GPT is a family of many LLMs). Open AI has released a series of GPT models which have been developed by training them on more complex and diverse data.

C. Role of Deployers

Moving on to the next actor, AI deployers utilize pre – made basic models which are used for day-to-day work and combine it with their services through an API(Application Programming Interface). Language – learning platform Duolingo has been using GPT – 4 (developed by Open AI) to integrate conversational AI features into its language application[17]. Major healthcare companies such as Babylon Health in the US utilize AI models for patient admission and as well as medical help. These developers use foundational AI models and club them with the specialized services which makes them ready for their tailored uses. These AI deployers modify the basic model according to their own safety and performance requirements.

D. End Users in the AI Ecosystem

End users can be considered as the final and the most important actor in the AI ecosystem as after the release of GPT-3 into the public domain, the goal of development of AI largely has shifted from replacing human beings to assisting human beings. Even after the development of autonomous capabilities of the AI systems human interference is needed. Since the development of AI is still in nascent stages so the feedback mechanism is necessary, where end users constantly give active feedback is required. Users of Microsoft's Co- Pilot use AI help to write codes and then they give feedback that in turn help the developer ecosystem.

Now to summarize the roles of actors in a AI ecosystem, one could say that the data used for the development of AI is given by data providers (data principals), then this data is used and acted upon by developers who make LLMs (large language models) based on this data and then data deployers club basic, pre made data models with their services and make a final product for the use of end consumers then based on their user experience, ease of usage, usefulness and reliability end consumers give a feedback which helps in the improvement of the AI models.

---

[16] Korosec K, 'Scale AI releases free lidar data set to power self-driving car development | TechCrunch' (Scale AI releases free lidar data set to power self-driving car development | TechCrunch, 22 May 2020) <https://techcrunch.com/2020/05/22/scale-ai-releases-free-lidar-dataset-to-power-self-driving-car-development/> accessed 20 January 2025.

[17] Marr B, 'The Amazing Ways Duolingo Is Using AI And GPT-4' (The Amazing Ways Duolingo Is Using AI And GPT-4, 26 April 2023) <www.forbes.com/sites/bernardmarr/2023/04/28/the-amazing-ways-duolingo-is-using-ai-and-gpt-4/> accessed 20 January 2025.

E. Harmonizing Private and Public Roles in AI Governance

An AI governance ecosystem involves certain ecosystem members whose roles, capabilities, interests, and actions are specific to AI governance and who all affect the process of governance. AI neither affects a single actor nor is governed by one, but rather involves a variety of actors. Government, industry, civil society and academia etc. all play certain roles in AI governance on a national and global level.They together form a network of actors who have connected interests, cooperate and compete with each other in order to survive, resembling a biological ecosystem that gradually moves from random collection of elements to a more structured community. [18]

- Traditional Perspective - By a traditional view of AI governance, individual and specific AI systems are comparatively easy to control and limit in their risk potential. If one understands ecosystems as proprietary platform environments with high degree of compatibility that attracts a multitude of actors who build their activities on the platform environment (Apple, Microsoft etc.), then it must follow that AI must first be responsibly managed by their owners. This brings this area into the realm of private sector AI governance. Companies that use AI applications must be held accountable for their outputs and results. [19]
- Networked and connected view - These multifaceted, intertwined relationships between various actors in the AI ecosystem have its own strengths and weaknesses with regards to governance of AI systems. One of the understanding takes into consideration macro perspectives that include different actors, systems, and processes that integrate their AI activities. Such an understanding refers to the networked context of multiple actors that form a dynamic, complex and relational structure representing all transactions that are more effective together than individually. The challenge is the network of different AI systems and actors whose results are further used by other AI systems. Such an ecosystem creates a need for a broader public AI governance perspective. [20]

These are the two perspectives related to governance of an AI ecosystem. While one considers each actor as an individual entity and everyone liable for their independent actions and does not take into account the consequences of actions of one actor on another. The other considers the ecosystem as connected and interwoven and pays attention to how one's actions can be detrimental for others.

A viable approach to AI governance would combine individual accountability with mechanisms that foster collective responsibility across the entire AI ecosystem. For instance, the Ministry could institute clear standards and guidelines—backed by legislation—that delineate the duties and liabilities of each stakeholder group (data providers, developers, deployers, and end users). At the same time, these standards should be supplemented by collaborative structures such as cross-sector councils or working groups, where government bodies, industry representatives, civil society, and academic experts regularly meet to coordinate on shared risks and dependencies. In practice, this could entail transparent reporting

[18] Frederick Moore J, 'Predators and Prey: A New Ecology of Competition' [1999] Harvard Business Review 75.
[19] Bullock J and others (eds), The Oxford Handbook of AI Governance (Oxford University Press 2022) <http://dx.doi.org/10.1093/oxfordhb/9780197579329.001.0001> accessed 20 January 2025.
[20] Bullock J and others (eds), The Oxford Handbook of AI Governance (Oxford University Press 2022) <http://dx.doi.org/10.1093/oxfordhb/9780197579329.001.0001> accessed 20 January 2025.

obligations for AI developers, safe and privacy-centric data-sharing frameworks for data principals, joint scenario-planning by deployers to handle unforeseen AI outcomes, and ongoing feedback loops from end users to continuously refine the AI systems. By reinforcing both the autonomy of individual actors and the interdependence that characterizes the broader AI environment, this hybrid model would help ensure that each stakeholder remains directly accountable for their own actions while also participating in a cooperative, holistic governance ecosystem.

III. Core Principles of AI Governance

4. Ensuring Transparency in AI Systems

A. Meaning and Scope

Transparency is the cornerstone of efficient use of AI, with its exponential growth and rising popularity amongst the general public. Transparency of an AI is generally associated with public knowledge or awareness of the inner network of such systems, or the understanding of 'how' and 'why'. Increased usage of AI calls for better regulation and readability of its algorithm or awareness of the data through which the outcomes are configured. Hence, AI transparency largely refers to the developers sharing the logic and reasoning behind a model.[21] For instance, Adobe ensured transparency in its Firefly AI system by releasing the entire information of the images and data used online. Further, Microsoft in its machine learning model included a feature called 'model explainability' which complements the interpretability of the system. Thus, Transparency can be exhibited through methods or processes which depend upon the nature of the system.

Such a practice promotes trustworthiness, and allows the users to assess the outcomes received and check for any potential biases or errors. Ethical usage of AI and increased accountability of its developers are other advantages for clearly laid out AI models. Although ensuring transparency is crucial, it is a tricky endeavour for complex AI models, such as Generative AI. This leads to the conundrum of 'Black Box AI', where the functioning of an AI is unknown to the users. The intricate nature of the algorithms affects the interpretation of such models by non experts. It is an undisputed fact that transparency is significant for the long term sustenance of AI.

Transparency, being a comprehensive concept, encompasses various facets and requirements such as :

1. AI Explainability

Explainable Artificial Intelligence (XAI) is an important tool for ensuring ethical and legal use of AI in high stake situations to avoid any errors.[22] It may be defined as, "*set of processes and methods that allows human*

[21] Ibm, "AI Transparency" (*IBM*, December 19, 2024) <https://www.ibm.com/think/topics/ai-transparency> accessed January 19, 2025

[22] "What Is Explainable AI?" (*SEI Blog*, January 17, 2022) <https://insights.sei.cmu.edu/blog/what-is-explainable-ai/>

users to comprehend and trust the results and output created by machine learning algorithms."[23] It has been established through research that businesses are likely to benefit from increased explainability of AI. There are two approaches to increasing explainability; firstly, structuring such models to be based upon simplified algorithms and systems which are easily comprehensible; secondly, developing tools of assistance for such users.[24] Both the approaches pose several difficulties with highly developed models, as XAI involves being able to ascertain with sufficient accuracy the data or processes involved behind the outcomes or decisions of AI. Many players in the market have been criticized for faulty models, which hinder the explainability and further complicate the interpretation of the data used. For instance, image generators such as Midjourney have been criticised for the racial undertones in their depiction of working professionals as white men. An impressive transparency model has been adopted by Salesforce, which provides clear guidelines to be followed for ascertaining transparency.[25] These comprehensive guidelines lay equal emphasis on both accuracy and interpretability of data.

2. AI Data Governance

It refers to the regulation of data used by an AI system. This process is carried out throughout the life cycle of an AI model and involves ensuring the accuracy, reliability, validity along with appropriate and legal collection of the data for training AI models. The data must not only be relevantly procured but must be processed in an appropriate manner. Businesses and organisations can execute policies and form committees to ensure that guidelines such as quality control are complied with. This is a continuous process, which is carried out simultaneously with the development of an AI software.[26]

3. AI Accountability and Auditing

Accountability in terms of AI platforms largely refers to developing, deploying and utilizing it in a manner such that responsibility for its outcomes can be attributed to the concerned parties.[27] For instance, there is widespread use of AI in security systems in the banking or retail sector in the country. This has been made possible through applications similar to Lighthouse, which can detect every single movement across the designated property. Any such errors in these systems may lead to a security breach or unintended surveillance which can cause irreparable damages. Further, with the development of bomb-detecting AI

[23] "Explainable AI" (*IBM*, December 19, 2024) <https://www.ibm.com/think/topics/explainable-ai>

[24] Grennan L and others, "Why Businesses Need Explainable AI—and How to Deliver It" (*McKinsey & Company*, September 29, 2022) <https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it>

[25] Goldman P and Baxter K, "Generative AI: 5 Guidelines for Responsible Development" (*Salesforce*, May 14, 2024) <https://www.salesforce.com/news/stories/generative-ai-guidelines/>

[26] Chu D, "AI Data Governance" (*Secoda*, August 12, 2024) <https://www.secoda.co/blog/ai-data-governance>

[27] Carnegie Council for Ethics in International Affairs, "AI Accountability" (*Carnegie Council for Ethics in International Affairs*) <https://www.carnegiecouncil.org/explore-engage/key-terms/ai-accountability#:~:text=Definition%20&%20Introduction,explore%20the%20curated%20resources%20below.>

(as promised by Israeli startup UVeye), it becomes all the more essential to have clearly laid out guidelines to ascertain accountability.[28]

Transparency assumes great significance for ensuring accountability in the use of AI. However, this question largely remains unanswered due to the 'black box' nature of AI platforms and the multi-layered nature of the concept. While the initial responsibility may lie with the user operating the AI system, it trickles down to include multiple players such as data providers, vendors etc.[29]

AI Auditing is a process centered around the assessment for detecting whether an AI model is indulging in prohibited or illegal activities, which may give rise to risks. Lack of a standard auditing procedure or an accepted law further complicates this arena.[30] The New York City Local Law 144 or NYC Bias Audits, is one of the first laws aimed at auditing automated (employment) systems.[31]

Given the rapid integration of AI into various sectors, ensuring transparency is no longer a mere ethical consideration but a regulatory necessity. The challenges posed by complex AI models, such as the "black box" dilemma, highlight the urgent need for clear guidelines, robust auditing frameworks, and enforceable accountability mechanisms. Governments must take proactive steps to develop standardized regulations that mandate AI transparency, ensuring that businesses and developers prioritize explainability, data governance, and ethical AI deployment. Implementing structured AI policies—such as transparency requirements, independent audits, and liability frameworks—will foster public trust, safeguard user rights, and mitigate risks associated with AI systems. A well-regulated AI ecosystem will not only promote innovation and economic growth but also ensure responsible and fair AI usage, reinforcing national and global AI governance standards.

5. Accountability of AI Stakeholders: Upholding User Rights and Legal Standards

While AI holds the potential of global progress by provisioning easy access to solutions to people for various day-to-day life challenges, there also exists potential harm attached to it. The major concerns related to AI are disinformation, misinformation, privacy issues (invasion), discrimination, violation of one's legal and human rights, fraud, etc. Therefore, the requirement of AI governance in today's world arises. In order to ensure effective AI governance, a transparent collaboration of different stakeholders,

[28] Intelligent Automation Network, "8 Surprising Examples of AI in Security" (*Intelligent Automation Network*, August 29, 2023) <https://www.intelligentautomation.network/business-intelligence/news/8-surprising-things-powered-by-ai-security>
[29] Stevens J and Digital E, "AI Accountability: Who's Responsible When AI Goes Wrong?" (*Emerge Digital*, August 11, 2023) <https://emerge.digital/resources/ai-accountability-whos-responsible-when-ai-goes-wrong/#:~:text=Accountability%20in%20AI%20is%20crucial,and%20damage%20to%20business%20reputation.>
[30] Centraleyes, "What Is AI Auditing? Where to Start" <https://www.centraleyes.com/glossary/ai-auditing/#:~:text=AI%20audits%20determine%20whether%20an,and%2For%20introduces%20unacceptable%20risks.>
[31] "NYC Bias Audit Law: A Comprehensive Guide" <https://www.nycbiasaudit.com/>

like, developers, engineers, regulators, users, etc. is required.[32] This can be contributive to effective AI governance as it accommodates various points of view, arguments, perspectives and expertise on different issues being dealt with.

The Indian legal system suggests that any data that has been gained from someone must have been consented for the same, and that data should be further used in a fair, transparent, and lawful manner.[33] Similarly, AI developers and deployers can be seen as data fiduciaries. They have an obligation to ensure there is maintenance of confidentiality and no invasion of privacy, also that the users have been safeguarded against the high security risk of AI systems. The Indian legal system has also expanded with the DPDP Act by providing a right to the users regarding the data they have provided to the AI systems. The users are free to access the personal data they have shared, correct and erase it according to their will,[34] and are entitled to get to know how their data are being processed and learn more about the data fiduciaries who are dealing with their personal data, and any additional information that concerns the Data Principal.[35]

The case of *Moffatt v. Air Canada*[36] is one of the recent examples of misinformation generated by AI. This case also marks a critical governance principle of 'accountability'. The case clearly highlights the need of designing and deploying AI mechanisms in a way that ensures that the organisation takes actions for any sort of misleading information generated from their AI system.

There is also a constant threat of the AI algorithms being biased and discriminatory.[37] In the US, the body cameras were introduced as a mechanism to hold the police accountable for any misbehaviour on their part. To a surprise, the set algorithm was faulty in itself and resulted in being a trouble for the mass itself. The said algorithm routinely misidentified women and people of colour. The Body Camera Accountability Act was introduced that underlined the fact that ensuring security of a community at large does not have to come at an expense of personal freedom of people.

There have been many cases in the modern world that seek the application of good AI governance. In order to ensure that AI is leveraged for good, both AI developers and deployers must work together. While developing AI, the foundation of its creation must be in compliance with the guidelines laid out by law, and with the user's legal and fundamental rights. There should be risk evaluation at each step of AI development to ensure the safeguarding of its potential users. Adherence to any governing law or guidelines would help the organisations themselves when in the position to give any explanation, and it also ensures fairness. The AI developers should focus on fairness, transparency and adherence to law. As the developers are also the data fiduciaries, the last step after ensuring all of the above, can be to

[32] 'AI Accountability: Stakeholders in Responsible AI Practices' (10 September 2024, Lumenova) <https://www.lumenova.ai/blog/responsible-ai-accountability-stakeholder-engagement/>
[33] Digital Personal Data Protection Act 2023, s4; Digital Personal Data Protection Act 2023, s 6
[34] Digital Personal Data Protection Act 2023, s 12
[35] Digital Personal Data Protection Act 2023, s 11
[36] Moffatt v. Air Canada, 2024 BCCRT 149
[37] Kade Crockford, 'How is Face Surveillance Technology Racist?' (ACLU, 16 June 2020) <https://www.aclu.org/news/privacy-technology/how-is-face-recognition-surveillance-technology-racist>

demonstrate accountability by establishing guidelines to ensure liability for actions made by those involved in the AI system development.[38]

While deploying, first and foremost, the deployers need to be familiar with the likely security threats, its impacts and problems with the development of such AI systems. Deployers shall also be up to date with the errors and user complaints through active management and supervision. The deployers shall also be aware of the role and responsibilities, along with the liabilities of each stakeholder and also opt for some additional measures such as validating the algorithm before use, also time and again ensuring regular examination would help in detecting any problems and issues to be dealt with.

A. How can the stakeholders collaborate?

The collaboration of different stakeholders can ensure that the technical and ethical problems that a certain AI system holds get addressed. Different types of stakeholders can include developers, engineers, deployers as well as policymakers or government bodies, industry leaders, scholars, and community organisations.[39] While AI can simplify our lives, there are many potential risks as we are required to share personal data with such AI systems. So, in order to guarantee safe development and deployment of AI, there needs to be set standards for the same, on which various stakeholders have worked together.[40]

A basic start to the goal of effective AI governance could be to carry out Multi-Stakeholder Forums facilitating unbiased-open dialogues. This can help in discussion of challenges through different perspectives, and sharing insights on each matter along with aligning the possible and wanted course of action. Such dialogues should necessarily include scholars of the specific field. This can ensure a kind of legal and regulatory framework that aligns with the objectives of national security, right to privacy, data security, elimination of inequality, algorithmic fairness, etc.[41] The insights and proposals generated in these sessions will be pivotal not only in setting a precedent in critical debates and generating real momentum for responsible AI development, but also in spurring world leaders and organizations to action.[42]

While the outcome of the collaboration ensures balance between innovations, accountability, and trust, their role cannot be just limited to these. There should also be provision of transparency in communication channels. The stakeholders can further engage in regular reporting to the users with regards to the capabilities and faults in the AI system, and decision making procedure through collaborative dialogues between them.

---

[38] Anas Baig and Omer Imran Malik, 'How to Develop an Effective AI Governance Framework?' (Securiti, 10 November 2023) <https://securiti.ai/ai-governance-framework/>

[39] 'AI Accountability: Stakeholders in Responsible AI Practices' (10 September 2024, Lumenova) <https://www.lumenova.ai/blog/responsible-ai-accountability-stakeholder-engagement/>

[40] 'How to we build trust between humans and AI' (1 August 2019, World Economic Forum) <https://www.weforum.org/stories/2019/08/can-ai-develop-an-empathetic-bond-with-humanity/>

[41] 'Strategies to craft effective AI governance: Essential building blocks for nations' (14 October 2024, Access Partnership) <https://accesspartnership.com/effective-ai-governance-building-blocks/#:~:text=Achieving%20this%20requires%20the%20concerted,layer%2C%20and%20the%20acceleration%20layer.>

[42] 'Responsible AI governance can be achieved through multistakeholder collaboration' (14 November 2023, World Economic Forum) <https://www.weforum.org/stories/2023/11/ai-development-multistakeholder-governance/>

6. The Pillars of AI: Safety, Reliability, and Robustness

Building AI systems for virtually every aspect of life has become a new research agenda which is bound to be followed by the majority of the technical gurus of the world. Helpful, indeed, however it also makes the normal non-robot human more and more reliant on its prowess. While it is duly understood that generating better AI systems is the way of the future, the question is not whether we can reach the goal of making AI robust enough to solve real world problems in a humanized description[43], but if we can create an atmosphere wherein the use and creation of such AI is all safe, reliable, all the while being robust.

The rapid expansion of AI in India displays greatly the need for a strong security and reliability measure, particularly in sectors like digital governance, healthcare, finance, and defence[44]. As India undergoes the foreseeable digital shift, providing for dependability of AI systems is crucial to guard against cyber threats, misinformation, and data breaches. To expand, platforms like Aadhaar and UPI[45] handle a great amount of sensitive personal data, making their security a top priority. Similarly, the Ayushman Bharat Digital Mission[46] aims to enhance healthcare accessibility through a digital health system[47]. In the financial sector, AI helps banks streamline operations but also exposes them to highly sophisticated cyber threats. Securing AI-driven financial services is vital to maintaining consumer trust and economic stability. Additionally, AI applications in national security, such as those developed by the DRDO, require strong protections to prevent potential security breaches which could hamper national safety[48]. The following table provides for the present risks in the AI security landscape in India along with the viable mitigation measures:

AI Security Risks in India

Risk	Description	Mitigation Measures
Data Breaches & Privacy Concerns	Large-scale data systems like Aadhaar, UPI, and the Ayushman Bharat Digital Mission store sensitive information, making them prime targets for	• Implement robust encryption (both in transit and at rest) • Adopt strong multi-factor authentication • Enforce strict access controls and role-based permissions

[43] Ai TA, "Goals of Artificial Intelligence" (*Applied AI Blog*, December 17, 2024) <https://www.appliedaicourse.com/blog/goals-of-ai/>

[44] Dham V, "Opinion: India Should Focus on Investing in AI across Cyber Security, Border Security" *The Economic Times* (August 6, 2024) <https://economictimes.indiatimes.com/tech/artificial-intelligence/et-opinion-india-should-focus-on-investing-in-ai-across-cyber-security-border-security-and education/articleshow/112308431.cms>

[45] "Transforming Governance with AI and DPI" (*Drishti IAS*) <https://www.drishtiias.com/daily-updates/daily-news-editorials/transforming-governance-with-ai-and-dpi>

[46] "Ayushman Bharat Digital Mission Marks a Transformative Three-Year Journey towards Enabling Digital Health | Ministry of Health and Family Welfare | GOI" <https://www.mohfw.gov.in/?q=pressrelease-87>

[47] India R, "Bridging the AI Security Gap: Challenges for Indian Enterprises" (*Risk Management Association of India*, January 28, 2025) <https://rmaindia.org/bridging-the-ai-security-gap-challenges-for-indian-enterprises/>

[48] Staff Report and Staff Report, "India Introduces Trustworthy AI Framework for Defence Reliability and Security | OEM Update |" (*OEM Update Original Equipment Manufacturer - Industrial Manufacturing*, October 19, 2024) <https://www.oemupdate.com/technology/india-introduces-an-ai-framework-for-defence-reliability-and-security/>.

	unauthorized access and misuse.	• Comply with emerging data protection legislations
Insecure AI Model Development & Deployment	Poorly secured AI development pipelines can lead to infiltration or tampering with models (e.g., injection of malicious code or biased training data).	• Ensure secure development life cycle (DevSecOps) • Adopt code-signing and checksums for model integrity • Regularly audit datasets for quality and bias • Use containerization and sandboxing for development and testing environments
Adversarial Attacks & Model Manipulation	Attackers can exploit vulnerabilities in AI models (e.g., adversarial examples in facial recognition or voice authentication) to manipulate outputs, resulting in incorrect predictions or unauthorized access.	• Conduct regular adversarial testing (Red Team exercises) • Deploy robust anomaly detection systems for inputs • Use defensive distillation and adversarial training • Maintain continuous monitoring of AI inputs for suspicious patterns
Cyber Threats to AI-Driven Financial Systems	AI-powered banking and payment services can be targeted for sophisticated cyberattacks that exploit real-time decision-making processes (e.g., credit scoring, fraud detection) to commit financial fraud at scale.	• Employ advanced threat intelligence and real-time monitoring • Deploy fraud detection systems with continuous machine learning updates • Regularly patch and update software components • Establish sector-wide incident response protocols and share threat intelligence
Misinformation & Deepfake Proliferation	AI-based tools can create highly realistic deepfakes and spread misinformation, potentially undermining trust in digital governance, influencing elections, and causing social unrest	• Develop and deploy deepfake detection algorithms • Strengthen content moderation policies on digital platforms • Educate the public about identifying manipulated content. • Encourage cross-industry collaboration for misinformation tracking and reporting.
Regulatory & Compliance Gaps	India's AI regulatory framework is still evolving. Gaps in policy and enforcement can lead to inconsistent security practices, especially when	• Adopt standardized guidelines (e.g., NITI Aayog's AI strategy, MeitY guidelines) • Align with global best practices like ISO 27001 and GDPR-style data protection • Strengthen compliance monitoring across government and industry

	stakeholders lack clarity on data protection and governance requirements.	• Encourage collaboration between policymakers, academia, and private sector
Shortage of Skilled AI Security Professionals	Rapid AI adoption outpaces the availability of trained personnel in AI risk assessment, cybersecurity, and data governance, leading to a skills gap in effectively safeguarding AI systems.	• Invest in targeted skill-building programs, specialized certifications • Foster public-private partnerships for workforce development • Incorporate AI security modules in higher education curricula • Leverage global collaboration and knowledge transfer programs.
Supply Chain Vulnerabilities	Dependence on external technology and hardware (e.g., IoT devices, third-party APIs) can introduce hidden backdoors or compromised components in AI systems, threatening critical infrastructure.	• Implement zero-trust architecture for data exchange • Conduct regular supply chain audits and vendor risk assessments • Enforce strict procurement standards with cybersecurity requirements • Employ blockchain/secure ledgers for transparent tracking of supply chain components
National Security & Defense System Vulnerabilities	AI applications in national security—such as those under DRDO—if compromised, could jeopardize intelligence, surveillance, and defense systems critical for the country's strategic interests.	• Use air-gapped or highly secure networks for critical AI systems • Mandate robust encryption for sensitive defense data • Engage in constant vulnerability assessments and threat intelligence sharing with strategic partners • Develop secure hardware solutions and cryptographic modules for defense applications.

IV. Ethical and Social Considerations

7. Privacy and Data Security in AI

The newly released report by the Ministry of Electronics and Information Technology on AI Governance Guidelines Development outlines avenues for AI-driven growth as well as risks and challenges[49], which

[49] Ministry of Electronics and Information Technology, "Report on AI Governance Guidelines Development" (2023) <https://indiaai.s3.ap-south-1.amazonaws.com/docs/subcommittee-report-dec26.pdf>

creates a need for government mechanisms that ensure development of AI systems. This paper focuses on issues of Privacy and Security, accentuating the necessity of AI compliance to data protection laws (the Digital Personal Data Protection Act, 2023 in India) and safeguarding user privacy as a whole while promoting innovation.

A. Overview of AI systems' privacy and security challenges

India has emphasised digital technology and growth more so than anything else, with its "Digital India" mission enabling access to better services for education, health care and agriculture and helps ensure transparency and accountability.[50] Given the rapid technological evolution and adoption of digital tools, there are understandable concerns surrounding privacy, especially when taking data protection and AI into account. The results produced by machine learning systems, which scan huge datasets to extract insights, might inherently be biased by the information they use. As such, artificial intelligence reflects the prejudices and biases inherent in its training data, thus perpetuating systemic biases that could persist even after its algorithms are constructed with positive intentions, while still representing a risk for privacy infringements in cases of improper use.[51]

An example is the American facial recognition company, Clearview AI, which scraped photos from social media platforms such as Instagram, Facebook and YouTube without consent to develop facial recognition software and identify people with a very high accuracy.[52] This technology was then deployed to law enforcement agencies. The company also had the ability to manipulate the results that the police see.[53] But the lack of regulation allowed anyone to access Clearview's database if they could afford it.[54] The Dutch government used a wealth fraud detection system called "Systeem Risico Indicatie" (SyRI) but a Dutch court invalidated it in 2020, stating that it collected too much data and did not provide clear purposes for collecting such data.[55]

Thus it is important to create regulations that direct AI systems and models in their usage and prevent invasive practices. The Supreme Court of India, through its landmark ruling in 2017, formally recognized privacy as a Fundamental Right.[56] While this legal acknowledgment is crucial to the protection of personal data from abuse, the concern remains that the artificial intelligence further amplifies already existing biases - especially when historical disparities are inherent in the dataset. Without such transparency and accountability, these risks become increasingly unidentifiable and unmitigated. The regulatory framework

[50] "6 Years of Digital India | MyGov.In" <https://www.mygov.in/campaigns/digitalindia/>
[51] Mabel V. Paul, 'Technical, Legal and Ethical Opportunities and Challenges of Governing Artificial Intelligence in India' (2023) 5 Indian JL & Legal Rsch 1
[52] Hart R, "ClearView AI—Controversial Facial Recognition Firm—Fined $33 Million for 'Illegal Database'" *Forbes* (September 4, 2024) <https://www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database/>
[53] Hill K, "The Secretive Company That Might End Privacy as We Know It" *The New York Times* (November 2, 2021) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html>
[54] Isra Ahmed, "ACLU v. Clearview Ai, Inc.," (*Digital Commons@DePaul*) <https://via.library.depaul.edu/jatip/vol33/iss1/4/>.
[55] Borgesius FZ, "Digital Welfare Fraud Detection and the Dutch SyRI Judgment" *IAPP* (February 7, 2024) <https://iapp.org/news/a/digital-welfare-fraud-detection-and-the-dutch-syri-judgment>
[56] Supreme Court Observer, "Fundamental Right to Privacy - Supreme Court Observer" (*Supreme Court Observer*, October 28, 2024) <https://www.scobserver.in/cases/puttaswamy-v-union-of-india-fundamental-right-to-privacy-case-background/>

of India is also evolving gradually, and there would be components that would remain sorely deficient across different domains.

B. Aligning AI Governance with the DPDP Act, 2023

As previously stated, AI technology relies on large datasets to function effectively, which raises critical concerns about the privacy of the user and their data. Effective alignment of AI governance with the Digital Personal Data Protection Act, 2023 requires on preventing misuse and addressing potential vulnerabilities in AI.

As per DPDP act's principles of Purpose Limitation (given in Section 4)[57] which states that personal data can only be processed for a specific, clear and lawful purpose, AI systems must collect and process only the necessary data to reduce risks of misuse and repurposing, also known as data minimization.[58]

To address security concerns, AI systems must adopt secure-by-design principles, as mentioned in the report, integrating encryption, access controls and real time threat detection to ensure safety and quick action in case of data breaches. Moreover, the implementation of Privacy-Enhancing Technologies (PETs)[59], which operate according to the principle of "data protection-by-design", can assist in complying with data minimization and providing robust anonymisation and pseudonymisation. PETs can help reduce the risk to individuals, while enabling further analysis of personal data without a controller necessarily sharing it, or a processor having access to it.[60]

As part of the DPDP Act, Data Fiduciaries are required to specify the purpose for which the Data Principal has voluntarily provided their personal data and the Data Fiduciaries can thus process that data for the specified purpose only.[61] In the context of AI, this may translate into displaying model cards[62], which describe the machine learning tool, including but not limited to the data it used, its intended features and even the limitations.

8. Ethics and Human-Centric Values in AI

Needless to say, AI is reshaping the world in unprecedented ways, offering transformative opportunities across industries. From enabling accurate medical diagnoses to optimizing supply chains and fostering

---

[57] Parliament, "THE DIGITAL PERSONAL DATA PROTECTION ACT, 2023" (2023) <https://prsindia.org/files/bills_acts/bills_parliament/2023/Digital_Personal_Data_Protection_Act,_2023.pdf>

[58] Home K in I|, "Decoding Digital Personal Data Protection Act, 2023" (*KPMG*, December 19, 2024) <https://kpmg.com/in/en/insights/2023/08/decoding-digital-personal-data-protection-act-2023.html>

[59] Oecd, "Emerging Privacy-Enhancing Technologies" (2023) <https://doi.org/10.1787/bf121be4-en>

[60] European Union Agency for Cybersecurity (ENISA), "Draft Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance" (2022) <https://ico.org.uk/media/about-the-ico/consultations/4021464/chapter-5-anonymisation-pets.pdf

[61] Parliament, "THE DIGITAL PERSONAL DATA PROTECTION ACT, 2023" (2023) s 7 <https://prsindia.org/files/bills_acts/bills_parliament/2023/Digital_Personal_Data_Protection_Act,_2023.pdf>

[62] "Responsible AI: The Role of Data and Model Cards" (*Datatonic*) <https://datatonic.com/insights/responsible-ai-data-model-cards/>

personalized user experiences,from facilitating healthcare diagnoses to enabling human connections through social media and creating labour efficiencies through automated tasks the rapid rise in AI has created many opportunities globally. However, these rapid changes also raise profound ethical concerns.These arise from the potential AI systems have to embed biases, contribute to climate degradation, threaten human rights and more.AI systems at times encounter ethical dilemmas which require nuanced understanding and judgement , which it lacks. Such problems associated with AI have already begun to compound on top of existing inequalities, resulting in further harm to already marginalised groups. Several keys areas of the ethical dilemmas presented by AI can be categorised as :

- Bias and Discrimination , by reinforcing the social inequalities.
- Lack of Autonomy and Responsibility in case of any decision that resulted in harm.
- Privacy concerns regarding the vast amount of personal data being collected and processed by AI.
- Transparency and Fairness and explainability of actions.

And at the centre of these challenges , is the need to ensure that these systems uphold the human values as fairness , autonomy , transparency etc. No doubt that the AI technology brings major benefits in many areas, but without the ethical guardrails, it risks reproducing real world biases and discrimination, fueling divisions and threatening fundamental human rights and freedoms. Among this human oversight is the bridge that connects AI's technical potential with the broader mission and values, ensuring that AI-driven innovations do not come at the expense of fairness, accountability, and trust. It acts as a critical safeguard in this context ensuring that the outputs and decisions align with the ethical principles.

A. Human Oversight in Artificial Intelligence (AI)

Human oversight refers to the involvement of human actors in developing, deploying and using AI systems to ensure that they respect human dignity, autonomy and values. Human oversight can take different forms and degrees , depending on the context and purpose of the AI system . For Example , human oversight can mean human-in-the-loop (HITL)[63],where a human can intervene and modify the outcome of an AI system ; human-on-the-loop (HOTL), where humans can monitor and stop an AI system or human-in-command(HIC)where humans have the ultimate authority and responsibility over an AI system.[64]

So today when AI is developing at a rapid pace , human oversight is of paramount importance to ensure ethical decision making , maintaining accountability , mitigating risks , biases and ensuring transparency which may arise due to the automated systems.

All generative artificial intelligence (AI)[65] systems are fallible.Machine Learning (ML) finds patterns in huge data sets and produces results based on those patterns. The AI system needs no human direction to produce

[63] Xiao-Li Meng, "Data Science and Engineering with Human in the Loop, behind the Loop, and above the Loop" (2023) 5 Harvard Data Science Review <https://doi.org/10.1162/99608f92.68a012eb>.
[64] Marco Repetto, "How Human Oversight and Transparency Can Ensure Trustworthy AI in the EU | CERTX" <https://certx.com/ai/how-human-oversight-and-transparency-can-ensure-trustworthy-ai-in-the-eu/>.
[65] Kim Martineau, "What Is Generative AI?" (*IBM Research*, September 1, 2024) <https://research.ibm.com/blog/what-is-generative-AI>.

outputs. And, if lacking human direction, and relying solely on big data and ML, the AI system will produce results that are misleading, biased, or simply wrong.

B. Ensuring Ethical AI through Human Oversight and Governance

It is vital, for specific AI systems built for specific purposes, that organisations rely upon the knowledge of experts. In the legal sector, for example, people with an in-depth understanding of law should continuously monitor and improve the performance of AI systems, inserting their specialist knowledge to build much improved AI solutions.

Human oversight should start with the inputting of data. Experts should carefully curate the data used to train systems and then constantly evaluate that data, looking for any outliers or anomalies and correcting sources to maintain high quality. They can minimise inaccuracies and bias through bias-reducing procedures and algorithmic detection tools.An example of this is seen in predictive policing tools[66]. In the United States, datasets used to train predictive policing AI systems often contained historical biases. By including experts who understand social contexts, the data was cleaned to reduce over-policing in marginalized areas.In India the government could implement a national framework requiring the inclusion of domain experts, such as sociologists and criminologists, in AI development for law enforcement to ensure the training data is free from historical biases. This could help prevent over-surveillance of vulnerable communities

Data quality assessments can also help with monitoring. The assessments highlight missing values, outliers, and general issues within data. The results allow humans to clean data sets and handle missing data. Perhaps most importantly, data assessments ensure that the data is representative and reflects the real-world scenario in which the AI operates.For instance, in healthcare, AI systems used for diagnosing diseases must be trained on diverse datasets. A study by researchers found that skin cancer detection algorithms performed poorly on darker skin tones because the datasets primarily consisted of images of lighter skin[67]. This highlights the importance of diversity and quality in training data. In India, given its wide demographic and genetic diversity, healthcare AI systems should be mandated to include data from all regions and communities, ensuring their efficacy across the population.

Organisations are responsible for the results of the AI and should always consider the real-world impact. That's why AI systems should also apply human oversight through auditing. At the highest level, AI systems can perform model performance evaluation , developing a series of metrics – accuracy, precision, speed, relevance, so on – and judging outputs against the metrics. Shortcomings should become clear through the process.[68]An example is the auditing of credit-scoring algorithms used by financial institutions. Experts have

---

[66] Bilel Benbouzid, "To Predict and to Manage. Predictive Policing in the United States" (2019) 6 Big Data & Society 205395171986170 <https://doi.org/10.1177/2053951719861703>.

[67] "Towards Fairness in AI for Melanoma Detection: Systemic Review and Recommendations" <https://arxiv.org/html/2411.12846v1>.

[68] LexisNexis, "Generative AI: The Importance of Human Oversight in the Law" <https://www.lexisnexis.co.uk/insights/generative-ai-the-importance-of-human-oversight-in-the-law/index.html>.

flagged instances where these systems discriminated against certain demographic groups. By incorporating fairness metrics and regular audits, organisations have worked to mitigate these biases[69].In India, the Reserve Bank of India could introduce similar guidelines for AI-driven credit systems to ensure fairness and inclusivity, protecting citizens from discriminatory financial practices.

AI systems should judge outputs in real-world scenarios, inviting feedback from the people using the system.For example, in the legal domain where AI tools are used for contract analysis, it has been observed that the common clauses and flagging discrepancies are efficiently identified,but, lawyers have observed that often these tools miss jurisdiction-specific nuances and sometimes even interpret legal terminology incorrectly. For this ,incorporating feedback from practicing lawyers, these systems have been refined to handle complex legal language and contextual variations more effectively and efficiently .Similarly, in India, a feedback mechanism where suggestions for improving the applicability of the AI tools to the Indian laws and jurisdiction could be given by the legal professionals could be established.

So, the seemingly ever-increasing use of artificial intelligence, AI, Human oversight has been much stressed and discussed as a safeguarding measure to ensure human centrism in AI deployment. The notion of 'human centric' AI does not imply a given regulatory strategy. The normative content of 'human centrism' is primarily of an ethical quality but can nevertheless be operationalised to provide legal guidance on more specific issues. Human centricity, as it has come to be (broadly) understood, does not only reflect that human needs are to be met by new technologies, but also incorporates the aim to safeguard individual rights and increase human well-being. Human centricity in AI is therefore a concept that places human-beings at the centre of any reflection about AI, its development, features and use.

Hence ,we can conclude that the rapid proliferation of artificial intelligence undoubtedly has unlocked huge potential across the industries ,the significant ethical dilemmas it poses cannot be ignored.The best AI systems are accountable systems, with individuals providing human oversight and accountability, minimising risks and amplifying the benefits.

As AI continues to grow and evolve , by ensuring robust human oversight principles as fairness, trust , accountability , fairness can be upheld along with technological advancements.AI technology brings major benefits in many areas, but without the ethical guardrails, it risks reproducing real world biases and discrimination, fueling divisions and threatening fundamental human rights and freedoms.

UNESCO recommends four core values[70] which should be the foundations for AI systems that work for the good of humanity, individuals, societies and the environment. These are:

- Respect , protection and promotion of human rights and fundamental freedoms and human dignity.
- Living in peaceful , just and interconnected societies

[69] Ștefan Spiridon, "AI Bias in Credit & Loan Processing: Is AI Biased When Assessing Credit Worthiness?" (*AI Bias in Credit & Loan Processing*, December 19, 2024) <https://www.itmagination.com/blog/credit-loan-processing-ai-biased-when-assessing-credit-worthiness>.

[70] "UNESCO's Recommendation on the Ethics of Artificial Intelligence" (*www.unesco.org*) <https://unesdoc.unesco.org/ark:/48223/pf0000385082.locale=en>.

- Environment and ecosystem flourishing
- Ensuring diversity and inclusiveness.

Along with that to operationalize human oversight effectively, following are some recommendations:

- Establishing a clear regulatory framework mandating oversight mechanisms for possible high-risk AI systems and embedding frameworks like HITL , HOTL , HIC .
- Invest in training programs to equip experts with the skills to manage AI-related ethical and technical challenges.
- Promoting transparency through mandatory disclosures on AI system processes, decision-making, and accountability structures.
- Encouraging collaborations between governments, academia, and the private sector to develop ethical standards for AI governance.
- Regularly audit AI systems for fairness, accuracy, and alignment with ethical principles.

V. Broader Governance Approaches

9. Perusal of Global Approaches to AI Regulations

European Union

The EU AI Act, which came into effect in August 2024, is a dedicated AI law in the European Union. The EU AI Act is a comprehensive regulatory framework which categorizes risk into different types to cater to specific risks in different applications. The risk-based approach helps to manage complex AI technologies, and its impact varies widely in society.[71]

Risk Classification

The EU Act classifies AI systems according to different levels of risks:

1. Unacceptable Risk: AI systems used for social scoring and those AI systems which uses deceptive or manipulative tactics that cam influence a person's behaviour or their will in a way that can cause harm falls under this category. The AI systems under this category are prohibited.

2. High Risk: AI systems under this category have the most detailed compliance obligations under the EU AI Act. They are further divided into two categories-

    i) AI systems used as a safety component of a product

[71] IBM. 'What is the EU AI Act?' (IBM, 2024) <https://www.ibm.com/think/topics/eu-ai-act> accessed 23 January 2025.

ii) AI systems deployed in 8 specific areas, including (but not limited to), employment, education, law enforcement, administration of justice, essential private and public services, and migration.

3. Limited Risk: Those AI systems that directly interact with the public at large falls under this category. These AI systems include, Chatbots like- ChatGPT and DeepSeek, Deepfake software, emotion recognition systems and biometric categorization systems. These systems are obligated to report/disclose to the user that the content generated is artificial or manipulated.

4. Low/Minimal Risk: Any AI system not under any of the above categories is of low/minimal risk.[72]

This risk-based approach increases safety and accountability by prohibiting AI systems that pose unacceptable risks and also by making sure that the developers implement rigorous oversight mechanisms. This approach also requires transparency which helps build trust among the users.

Integration with Data Protection

The EU AI Act is in coherence with the General Data Protection Regulation or GDPR which ensures that with the rapid evolution of AI technologies, data privacy is not compromised. By taking into consideration both data protection and AI governance, the EU has created a regulatory environment that addresses multiple aspects of AI use. The compliance of data protection laws also increases the trust of the public in AI technologies.[73]

Enforcement

The EU AI Act is legally binding in nature and establishes clear consequences for non-compliance. Penalties range from €35 million or up to 7% of a company's total worldwide annual turnover for not following the prohibited AI practices, to €7.5 million or up to 1% of a company's total worldwide annual turnover for supply of misleading or incorrect information. This potential for heavy fines/penalties makes sure that the developers follow the regulations set by the EU.[74]

United States Of America (USA)

The AI regulations in the US rely heavily on existing laws and follow sector-specific guidelines instead of a

[72] White & Case LLP. 'AI Watch: Global Regulatory Tracker - European Union' (White & Case LLP, 2023) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-european-union> accessed 23 January 2025.
[73] European Commission. 'Regulatory Framework for AI' (European Commission, 2023) <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 23 January 2025.
[74] White & Case LLP. 'Long Awaited EU AI Act Becomes Law After Publication in the EU's Official Journal' (White & Case LLP, 2024) <https://www.whitecase.com/insight-alert/long-awaited-eu-ai-act-becomes-law-after-publication-eus-official-journal> accessed 23 January 2025.

unified/centralized framework. The sector-specific approach fosters rapid innovation, as companies can adapt quickly to the changes in the regulations rather waiting for specific federal laws. The US uses the existing laws to govern AI applications, for example, guidelines for the Medical AI given by the Food and Drug Administration (FDA), or the Federal Trade Commission which oversees consumer protection. Although the existing laws provide some degree of oversight, a dedicated unified regulatory framework is still needed to overcome inconsistencies in oversight across different sectors.[75]

Risk Classification

As of now, the US has no comprehensive legislation which regulates AI. Also, in the existing frameworks and laws, the US generally fails to classify AI according to risks, except a few sectors which follow a risk-based approach, such as, FDA's framework which takes into consideration the potential impact of AI tech on patient safety.[76]

Integration with Data Protection

Although the US does not have a federal data protection law, it does have sector specific regulations like HIPAA (Health Insurance Portability and Accountability Act) for healthcare data privacy and COPPA (Children's Online Privacy Protection Act) for children's privacy. These sector specific approaches can lead to vulnerabilities in data protection towards AI as lack of a comprehensive law/regulation means that a lot of areas may not have required coverage/protection against data misuse where AI may be applicable.[77]

Enforcement

As there are no comprehensive laws in place, most of the AI guidelines are non-binding in nature with enforcement limited to only some sectors. For instance, The Federal Trade Commission (FTC) has taken an active step against Rite Aid over facial recognition misuse, resulting in a 5 year ban and strict compliance requirements. The enforcement also varies from state to state, as according to The Colorado AI Act, the State Attorney General has enforcement authority. Also, in California, many bills, for example- Senate Bill 942, imposes penalties on violations.[78]

[75] Tech Policy Press. 'The Coming Year of AI Regulation in the States' (Tech Policy Press, 2025) <https://www.techpolicy.press/the-coming-year-of-ai-regulation-in-the-states/> accessed 23 January 2025.

[76] White & Case LLP. 'AI Watch: Global Regulatory Tracker - United States' (White & Case LLP, 2025) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states> accessed 23 January 2025.

[77] Software Improvement Group. 'AI Legislation in the US: A 2025 Overview' (Software Improvement Group, 2025) <https://www.softwareimprovementgroup.com/us-ai-legislation-overview/> accessed 23 January 2025.

[78] The White House. 'Removing Barriers to American Leadership in Artificial Intelligence' (The White House, 2025) <https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/> accessed 23 January 2025.

Japan

Japan follows a hybrid model of regulatory framework which combines ethical principles with sector-specific guidelines. This framework strives to promote human-centric AI which allows for adaptability and flexibility in governance. It also addresses societal concerns while promoting innovation. The government believes that strict regulations could hinder investment and technological advancement in AI. By the way of industry collaboration, Japan creates an environment conducive to innovation.[79]

Risk Classification

In Japan, AI systems are not classified according to risks in the relevant guidelines. However, a risk-based approach is necessary according to some government officials. The lack of a rigid classification system exposes the people to threats while also helping the regulators adapt to rapidly evolving AI technology.[80]

Integration with Data Protection

Japan has a comprehensive data protection act known as APPI (Act on the Protection of Personal Information) which aligns with the GDPR principles and ensures data privacy. The APPI helps maintain the trust of the people by providing a framework that protects personal data in AI systems.[81]

Enforcement

Japan has chosen voluntary compliance together with industry collaboration as its AI regulatory system instead of legally binding regulations. The method allows adaptable governance control by adjusting to technological advances. At present, companies face no legal obligation to comply with ethical guidelines because Japan uses voluntary compliance instead of legal enforcement.[82]

Australia

Australia is devoid of any comprehensive legal framework and is in the developing stage of a risk-based framework. Australia has published many guidelines such as The AI Ethics Principle in 2019, The Voluntary AI Safety Standard and also the Interim Response, but none of these covers AI regulation in depth as much as a specific statute or regulation would. Australia has also been focusing on sector specific guidelines, for instance Australian Securities and Investments Commission (ASIC) has issues guidelines on use of AI in

[79] CSIS. 'Japan's Approach to AI Regulation and Its Impact: 2023 G7 Presidency' (CSIS, 2023) <https://www.csis.org/analysis/japans-approach-ai-regulation-and-its-impact-2023-g7-presidency> accessed 23 January 2025.

[80] White & Case LLP. 'AI Watch: Global Regulatory Tracker - Japan' (White & Case LLP, 2023) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-japan> accessed 23 January 2025.

[81] Clifford Chance. 'Understanding the New AI Operator Guidelines in Japan' (Clifford Chance, 2024) <https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2024/06/understanding-the-new-ai-operator-guidelines-in-japan.html> accessed 23 January 2025.

[82] Diligent. 'Japan AI Regulations' (Diligent, 2023) <https://www.diligent.com/resources/blog/japan-ai-regulations> accessed 23 January 2025.

services related to finance and Therapeutic Goods Administration (TGA) which regulates AI-based medical devices and software.[83]

Risk Classification

The Australian Government in the Interim Response stated that they would adopt a risk-based framework. The Proposals Paper further elaborated on this framework. Two categories were identified in the Proposals Papers:

Category 1: The measurement of risk for known and foreseeable AI system applications falls under this category.

Category 2: It focuses on predictive risk assessment of unknown AI systems and their developing hazards.[84]

Integration with Data Protection

Australia has their own data protection act known as the Privacy Act, 1988. This act provides a basic framework for data protection but is less comprehensive than GDPR. The latest amendments of the Privacy Act sought to achieve better privacy protection for personal information while providing clear insights about automated processes.[85]

Enforcement

The regulatory framework governing AI within Australia, much like Japan's, operates on voluntary principles because the country lacks sufficient enforcement systems. The government plans to amend its guidelines through time by adopting best practices from other jurisdictions like the EU and Japan. The need for stricter enforcement mechanisms increases with the fast-growing landscape to ensure compliance with ethical standards and protect people from potential harm.[86]

Lessons From Global Approaches

What's Working: Best Practices :

1. Risk-based Regulation (EU)

   The EU's AI Act applies four risk levels (unacceptable, high, limited, and minimal risk) to guide regulatory concentration on critical applications in healthcare along with criminal justice operations.

[83] Minister for Industry and Science. 'Albanese Government Acts to Make AI Safer' (Minister for Industry and Science, 2024) <https://www.minister.industry.gov.au/ministers/husic/media-releases/albanese-government-acts-make-ai-safer> accessed 23 January 2025.

[84] White & Case LLP. 'AI Watch: Global Regulatory Tracker - Australia' (White & Case LLP, 2023) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-australia> accessed 23 January 2025

[85] Digital Transformation Agency. 'Australia's Artificial Intelligence Ethics Principles' (Digital Transformation Agency, 2024) <https://architecture.digital.gov.au/australias-artificial-intelligence-ethics-principles> accessed 23 January 2025.

[86] Technology's Legal Edge. 'Shaping the Future: Australia's Approach to AI Regulation' (Technology's Legal Edge, 2024) <https://www.technologyslegaledge.com/2024/09/shaping-the-future-australias-approach-to-ai-regulation/> accessed 23 January 2025.

Resource distribution becomes efficient through this model which allows risk managers to direct their efforts toward the most critical safety hazards. Through this approach high safety and ethical standards in high-risk fields have been effectively established. The system has promoted transparency as well as accountability for critical infrastructure AI systems.[87]

Why It Works: The EU targets its regulatory focus on high-risk applications which enables it to resist heavy-handed control of low-risk AI systems and create favourable conditions for consumer app and entertainment innovation.

2. Sector-Specific Guidelines (US)

The US relies on sector-specific regulations, such as FDA guidelines for medical AI and FTC oversight for consumer protection. Different industries can utilize this approach to achieve flexibility along with adaptability. AI innovation has progressed rapidly within US territory especially within the sophisticated technology environment of Silicon Valley. For example, AI-driven healthcare solutions have flourished under FDA guidelines.[88]

Why It Works: Each industry sector receives specialized regulations to handle unique risks and opportunities which promote innovation and enables oversight of new solutions.

3. Ethical Principles and Collaboration (Japan)

The Social Principles of Human-Centric AI in Japan put equal emphasis on fairness and transparency along with accountability. The government promotes collaboration between industry, academics and regulatory bodies. Japanese industries use AI technology in manufacturing and robotics applications without compromising public support for these technological solutions. For example, modern manufacturing uses AI-powered robots that enhance operational efficiency because these systems show minimal ethical concerns.

Why It Works: Japan achieves innovative technological development alongside public trust through its combined effort in establishing ethical AI development culture and collaboration frameworks.[89]

What's Not Working: Gaps and Challenges

1. Rigid Regulations (EU)

Startups and SMEs face problems with EU regulatory restrictions as while their strict framework delivers both safety and ethical standards, teams have criticized it for being too complex. Communication costs along with implementation complexity inhibit innovation in targeted sectors.

---

[87] Schellman. 'What Next and What Now After the EU AI Act' (Schellman, 2023) <https://www.schellman.com/blog/cybersecurity/what-next-and-what-now-after-the-eu-ai-act> accessed 23 January 2025

[88] ISACA. 'The EU AI Act: Adoption Through a Risk Management Framework' (ISACA, 2023) <https://www.isaca.org/resources/news-and-trends/industry-news/2023/the-eu-ai-act-adoption-through-a-risk-management-framework> accessed 23 January 2025.

[89] RAND Corporation. 'The EU AI Act: A Comparative Analysis' (RAND Corporation, 2023) <https://www.rand.org/pubs/research_reports/RRA3243-3.html> accessed 23 January 2025.

For example, small businesses find it challenging to fulfil strict standard requirements that apply to high-risk artificial intelligence systems.[90]

Why It's a Problem: By imposing excessive rules on innovation the AI market experiences both slowed developments and reduced competition among smaller startups.

2. Lack of a Unified Framework (US)

The United States does not have a single federal AI law therefore it implements various sector-specific regulations in addition to non-binding guidelines. The current regulatory oversight lacks consistency as it allows for different monitoring throughout the various sectors which generates spaces where AI applications operate between sectors with the lack of transparency. For example, AI systems which are used for hiring procedures have sparked criticism regarding both their biased nature and their unclear functioning.[91]

Why It's a Problem: The absence of a unified framework creates uncertainty for businesses and risks insufficient oversight in high-risk areas.

3. Voluntary Ethical Guidelines (Australia)

The AI Ethics Framework of Australia operates as a voluntary program without mandatory controls. The framework only achieves modest success in making people aware of ethical matters but it lacks power to avoid unethical behaviour or hold people responsible. For example, public service AI systems face negative feedback due to their bias and lack of transparency.[92]

Why It's a Problem: Systemic problems of bias and discrimination require more than voluntary guidelines to solve them in high-stakes points of application.

Recommendations

1. Adopt a Risk-Based Regulatory Framework

India needs to implement an AI regulatory system comparable to the EU AI Act that fits specific socio-economic demands of its domain. A risk-based classification system should divide AI systems into four risk groups beginning with unacceptable followed by high and limited and minimal risk so that strict requirements can be applied to high-risk applications like healthcare, criminal justice and critical infrastructure.[93]

---

[90] Thoropass. 'EU AI Act' (Thoropass, 2023) <https://thoropass.com/blog/compliance/eu-ai-act/> accessed 23 January 2025.
[91] HeyData. 'SMEs in the AI Era: The Impact of EU AI Act' (HeyData, 2023) <https://heydata.eu/en/magazine/sm-es-in-the-ai-era-the-impact-of-eu-ai-act> accessed 23 January 2025.
[92] White & Case LLP. 'AI Watch: Global Regulatory Tracker - Australia' (White & Case LLP, 2023) <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-australia> accessed 23 January 2025.
[93] Carnegie Endowment for International Peace. 'India's Advance on AI Regulation' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en&center=india> accessed 23 January 2025.

Rationale:
Through its risk-based model the EU directs its regulatory focus toward high-risk areas but enables free innovation in low-risk fields. India could establish an ordered approach to allocate resources effectively through which it would focus on crucial risks keeping startup companies and SMEs free from excessive regulation. Also, Indian regulators face resource constraints which would benefit from implementing a risk-based system to distribute their enforcement capabilities.

2. Strengthen Data Protection and Privacy Laws

India needs to establish AI governance procedures that exactly follow procedures outlined in the Digital Personal Data Protection Act (DPDP) 2023. This includes mandating transparency in data usage, ensuring informed consent, and requiring impact assessments for AI systems that process personal data.[94]

Rationale:
Robust data protection legislation must be integrated into AI regulation standards because this approach exists in both the GDPR from the EU and the APPI from Japan. The DPDP Act stands as a base to protect data privacy yet its enforcement capabilities should be expanded to manage AI-specific problems like biased algorithms and incorrect data handling procedures. As India lacks effective enforcement powers the government should enhance its regulatory skills through technology-based monitoring systems to improve compliance across the board.

3. Establish a Dedicated AI Regulatory Body

India needs a special AI regulatory organization such as an AI Governance Authority to supervise the development alongside deployment and continuous assessment of AI systems. The body should combine members from technology fields with legal experts and ethical specialists as well as industrial representatives[95].

Rationale:
Specialized oversight becomes necessary as also seen in the EU AI Act's centralized approach and Japan's collaborative model. The newly created authority would create clear and uniform AI regulations by solving regulatory fragmentation in different departments while requiring particular expenditure and specialized competency.

[94] Carnegie Endowment for International Peace. 'India's AI Strategy: Balancing Risk and Opportunity' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/posts/2024/02/indias-ai-strategy-balancing-risk-and-opportunity?lang=en> accessed 23 January 2025.
[95] Carnegie Endowment for International Peace. 'India's Advance on AI Regulation' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en&center=india> accessed 23 January 2025.

4. Encourage Industry Self-Regulation and Collaboration

The Indian government should support the development of industry-wide self-regulatory initiatives consisting of AI system conduct codes and certification frameworks. The initiative must be supported by government oversight for the purpose of maintaining accountability.[96]

Rationale:
Industrial participation through the US's sector-specific guidelines and Japan's collaborative model shows that involved industries create successful regulation. The Indian startup economy and advancing AI sector would benefit from adaptable rules that support innovation. However, self-regulation needs government monitoring to stop misuse while maintaining public trust.

5. Invest in Regulatory Capacity and Public Awareness

India needs to allocate money towards developing regulatory capabilities through official training in AI technology standards and ethical principles alongside legal frameworks. The country needs to embark on an awareness initiative that instructs residents on the advantages as well as dangers AI presents.[97]

Rationale:
Australia along with other nations has been spending money to establish regulatory expertise which helps dealing with new technologies. The regulatory agencies in India struggle to monitor AI systems due to their deficiency in technical competence. Building trust with the public depends on raising public awareness which ensures accountability. Constructing organizational strength and public understanding demands continuous financial support because it stands essential to achieve effective governance.

6. Foster International Collaboration
India needs to join global efforts to standardize AI regulations through its participation in the Global Partnership on AI (GPAI). The country should unite its efforts with the EU, the US and Japan to exchange learning about AI while solving problems that affect neighbouring borders.[98]

[96] Carnegie Endowment for International Peace. 'India's AI Strategy: Balancing Risk and Opportunity' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/posts/2024/02/indias-ai-strategy-balancing-risk-and-opportunity?lang=en> accessed 23 January 2025.
[97] Carnegie Endowment for International Peace. 'India's Advance on AI Regulation' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en&center=india> accessed 23 January 2025.
[98] Carnegie Endowment for International Peace. 'India's AI Strategy: Balancing Risk and Opportunity' (Carnegie Endowment for International Peace, 2024) <https://carnegieendowment.org/posts/2024/02/indias-ai-strategy-balancing-risk-and-opportnity?lang=en> accessed 23 January 2025.

10. AI as an Enabler to Advance SDGs

TABLE: Role of AI in Advancing SDGs

Development Focus	SDG Objectives	Role of AI
Healthcare Delivery	SDG 3 focuses on ensuring healthy lives and promoting well-being for all.	From aiding in medical diagnosis and drug discovery to personalising treatment plans and automating repetitive tasks, AI presents a significant opportunity to improve healthcare delivery. AI-powered tools can analyse medical images for faster disease detection, provide virtual consultations in remote areas, and monitor chronic illnesses remotely. Cognitive robotics can merge data from preoperative medical records with live operational data, aiding physicians in refining their instrument precision during procedures. These advancements not only improve surgical outcomes but also foster trust in the application of AI across various surgical specialties. A study by Accenture suggests that AI can improve healthcare productivity by up to 40 percent, leading to better access to quality healthcare for all.
Education	SDG 4 aims to ensure inclusive and equitable quality education for all.	AI-powered tutoring systems can personalise learning experiences, catering to individual student needs and paces. AI can analyse student performance data to identify learning gaps and provide targeted interventions. Additionally, AI-powered translation tools can facilitate knowledge sharing across language barriers, promoting inclusive education.

Source[99]

[99] Bhowmick S, 'Ai as a Catalyst for Sustainable Development' (orfonline.org, 1 June 2024) <https://www.orfonline.org/expert-speak/ai-as-a-catalyst-for-sustainable-development> accessed 18 January 2025

The most active geographical regions/ countries around the world focusing on AI4SDG research are the United States, Western Europe, China, Japan, Australia, and India (Figure 2). Areas on the east coast of United States (New York, Washington, Hawaii), Canada (Montreal, Ottawa, Toronto, etc.), United Kingdom, Norway, Sweden, France, Germany, Italy, India (New Delhi, Bangalore), Singapore, China (Beijing, Shanghai), Hong Kong, Japan, and Australia (Sydney, Melbourn) show the highest density in the map and would have the greatest number of research publications on AI4SDG. On the other hand, South American countries Mexico, Brazil, Argentina, Africa and parts of Middle East had lower number of publications. In addition to this, the number of publications in different regions were also found to vary by different SDGs. For examples, it was observed that the African region has most of its publications on SDG 2 (Zero Hunger) and SDG 3 (Good Health and Wellbeing). In India and China, the larger percentage of publications were related to SDG 3, SDG 7, and SDG 11. Whereas, the USA, UK, Europe, and Japan had publications relating to all six top-performing SDGs. Among the South American countries, Brazil and Columbia had the largest number of publications. These were mostly related to SDG 3 and SDG 7, whereas other SDGs did not have many publications related to AI4SDG from this region. It is observed that research giant countries have been focusing more on SDGs relating to infrastructure, environment, education, and health and the smaller and underprivileged countries target specifically SDGs relating to health and hunger reflecting their societal needs.

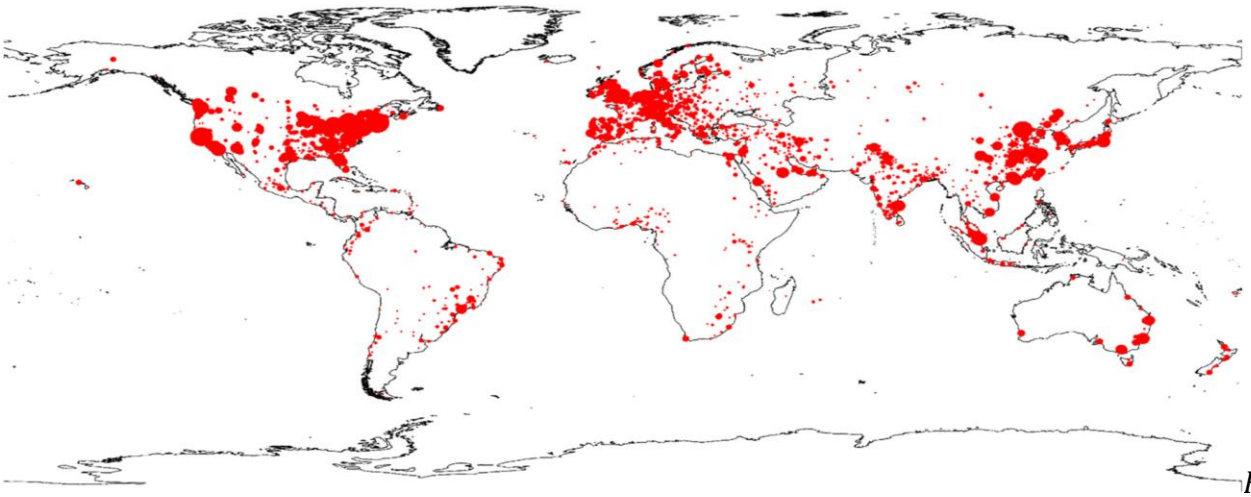

Figure: Regional distribution of research on AI4SDG.[100]

AI and Inclusivity

Across the world, businesses are clamouring to adopt the latest AI technologies, and they're willing to invest significantly. But the benefits of AI can extend beyond large enterprises and make a considerable difference to small businesses too if adopted responsibly.[101] According to the World Health Organization, more than one billion people worldwide have disabilities. The field of disability studies defines disability through a social lens; people are disabled to the extent that society creates accessibility barriers. AI

[100] Singh A and others, 'Artificial Intelligence for Sustainable Development Goals: Bibliometric Patterns and Concept Evolution Trajectories' (2023) 32 Sustainable Development 724

[101] Galea-Pace S, 'Inclusive Innovation: Why Ai Isn't Just for Big Businesses' (Interface, 19 December 2024) <https://interface.media/blog/2025/01/03/inclusive-innovation-why-ai-isnt-just-for-big-businesses/> accessed 18 January 2025

technologies offer the possibility of removing many accessibility barriers; for example, computer vision might help people who are blind better sense the visual world, speech recognition and translation technologies might offer real-time captioning for people who are hard of hearing, and new robotic systems might augment the capabilities of people with limited mobility.

The inclusivity of AI systems refers to whether they are effective for diverse user populations. Issues regarding a lack of gender and racial diversity in training data are increasingly discussed; however, inclusivity issues with respect to disability are not yet a topic of discourse, though such issues are pervasive. These inclusivity issues threaten to lock people with disabilities out of interacting with the next generation of computing technologies.[102] The creation of a network of experts and resources for AI and inclusion could help to address the "unmet need of assistive products crucial to achieve the Sustainable Development Goals, to provide Universal Health Coverage, and to implement the UN Convention on the Rights of Persons with Disabilities", that India has ratified.[103]

To make the changes needed for a more inclusive AI that centres equity, the field must first find agreement on foundational premises regarding inclusion. These are four guiding principles for ethical engagement grounded in best practices:[104]

1.	All participation is a form of labour that should be recognized.
2	Stakeholder engagement must address inherent power asymmetries.
3.	Inclusion and participation can be integrated across all stages of the development lifecycle.
4.	Inclusion and participation must be integrated into the application of other responsible AI principles

Recommendations to Make AI More Accessible and Inclusive:

1. **Through AI-Powered Accessibility Tools**: These include Speech Recognition. Natural Language Processing (NLP), Adaptive Interfaces. These AI-driven solutions not only improve accessibility but also empower individuals with disabilities to engage more fully in educational, professional, and social activities, fostering greater inclusivity.

AI TOOL	USE & IMPACT

[102] Alley, 'Ai and Accessibility' (Communications of the ACM, 1 September 2023) <https://cacm.acm.org/opinion/ai-and-accessibility/#R4> accessed 23 January 2025

[103] 'Assistive Technology' (World Health Organization) <https://www.who.int/news-room/fact-sheets/detail/assistive-technology> accessed 23 January 2025

[104]Park T, 'Making AI Inclusive: 4 Guiding Principles for Ethical Engagement' (Partnership on AI, 9 December 2023) <https://partnershiponai.org/paper/making-ai-inclusive-4-guiding-principles-for-ethical-engagement/> accessed 18 January 2025

Speech Recognition	AI-powered speech recognition tools like Google's Live Transcribe or Microsoft's Azure Cognitive Services can convert spoken language into written text in real-time, helping individuals with hearing impairments participate fully in conversations and public events.
Natural Language Processing (NLP)	NLP technologies can facilitate communication for individuals with speech disabilities by interpreting their input and translating it into understandable text or speech. This can be particularly beneficial for those who rely on alternative communication methods.
Adaptive Interfaces	AI can create adaptive user interfaces that adjust to the needs of users with different abilities. For instance, AI systems can learn a user's preferences and behaviors over time, providing personalized experiences that make technology more intuitive and easier to navigate.

2. **Through Promoting Inclusivity and Equal Opportunities**: AI has the potential to level the playing field by providing equal access to educational resources and opportunities. However, to fully realize this potential, it's essential to address systemic biases that may exist within AI systems and ensure that these technologies are developed and deployed in ways that promote inclusivity and equity.
    a. Addressing the bias in AI: One of the critical challenges in making AI inclusive is addressing the biases that can be embedded in AI systems. These biases often arise from the data used to train AI models, which may reflect historical inequalities or systemic discrimination. If not carefully managed, AI systems can perpetuate or even exacerbate these biases.To create more inclusive AI systems, it's essential to use diverse data sets that represent a wide range of perspectives and experiences. This can help reduce the risk of bias and ensure that AI tools are fair and equitable. AI developers must prioritize ethical considerations in their work, including the potential impact of AI on marginalized communities.
    b. Equal Access to AI Tools: To promote equal opportunities, it's also crucial to ensure that AI tools are accessible to all, regardless of socioeconomic status, location, or background. This can be achieved through initiatives that make AI technologies and education more affordable and available to underserved communities.Open-source initiatives allow individuals and organizations with limited resources to leverage AI for their needs. Moreover, Governments and nonprofits can play a vital role in promoting access to AI by funding educational programs, providing scholarships, and supporting community-based technology initiatives.

3. **Through Comprehensive AI Education Programs:** To prepare individuals for the AI-driven future, educational institutions and organizations must offer robust AI education and training programs. These programs should aim to equip learners with the knowledge and skills needed to understand, develop, and

utilize AI tools responsibly. Courses on AI and machine learning can provide students with a strong foundation in the technical aspects of AI, including programming, algorithm development, and data analysis. This knowledge is critical for those looking to enter AI-related fields.[105] AI tools are useful in teaching to both teachers and students. For students, AI tools can help with coding, mathematical reasoning, improve writing and presentation of graphs and slides, explore topics being covered in the course, clarify doubts, provide practice problems, provide a personalised learning experience, and summarise and synthesize literature. For teachers, AI tools can help with tasks such as generating course plans, syllabi, and course policies, facilitating grading, and providing assistance to students outside class. AI tools can provide a personalised learning experience, help with the exploration of a topic, provide help with problem-solving, programming, and data analysis, and improve writing. Emphasising the importance of originality, proper citation, and maintaining academic integrity is crucial. Educators and institutions should clearly communicate guidelines on the ethical use of AI tools and thoughtfully integrate them into the learning process; these can serve as supplements rather than replacements for traditional teaching and research methods..[106]

Additionally, The Artificial Intelligence and the Futures of Learning project of UNESCO builds on the Recommendation on the Ethics of Artificial Intelligence adopted at the 41st session of the UNESCO General Conference in 2019 and follows up on the recommendations of the UNESCO global report Reimagining our futures together: a new social contract for education, launched in November 2021. It is implemented within the framework of the Beijing Consensus on Artificial Intelligence and Education and against the backdrop of the UNESCO Strategy on technological innovation in education (2021-2025).[107]

VI. Implementation and Policy Challenges

11. Operationalizing AI Governance Principles

Operationalisation of AI governance principles is the process of translating high level abstract AI governance principle into actionable guidelines that require a multifaceted approach. The focus should shift from “what” to “how”, this means rather than knowing what the principles are to how to effectively implement it.

[105] Kmetz R, ‘Making AI Accessible and Inclusive’ (Medium, 16 August 2024) <https://ryankmetz.medium.com/making-ai-accessible-and-inclusive-32ef2c279d47> accessed 18 January 2025

[106] (Emerging AI tools for Education and research) <https://iisc.ac.in/wp-content/uploads/2024/03/Report-of-Committee-on-AI-Tools-for-Education-and-Research.pdf> accessed 23 January 2025

[107] ‘Artificial Intelligence and the Futures of Learning’ (UNESCO.org) <https://www.unesco.org/en/digital-education/ai-future-learning> accessed 23 January 2025

Why do we need proper Operationalisation?

The need for proper operationalisation of AI governance principle especially in India stems for various factors, as India has one of the highest smartphone user bases in the world,[108]providing a platform for applications to scale, the diversity, scale, digital divide and lack of awareness among the populace makes it an ample breeding ground for the negative effects of AI, shown through the fact that second largest user base of Chatgpt in the world is India[109], considering these factors operationalisation of AI governance principle is the need of the hours to ensure that the AI principles adopted is effective and enforceable in the real world. AI also depends on data and therefore is enabled by high quality data availability, robust data protection and sharing protocols. The approach for operationalizing the Principles in India needs to therefore strike a balance between creating the necessary guardrails while enabling research and innovation. The goal must be to maximize the benefits of AI for the citizens, businesses and research and minimizing the risks.

One thing to keep in mind while trying to operationalise AI principle is prioritisation, it is a need to when adopting guidelines and regulatory framework for adoption of AI principle that they should be characterised based on the risk factor they pose. That is characterising them as "*high risk AI systems*" under Chapter 3 Section 1 of the EU artificial intelligence act[110].[1]. High-risk AI systems are those that have the potential to cause significant harm or negatively impact fundamental rights. The classification of an AI system as high-risk depends on its intended purpose and the probability and severity of potential harm. High-risk AI systems are not exclusive to the European Union (EU), though the EU has been at the forefront of establishing regulatory frameworks for AI based on risk.

How do different jurisdictions view this?

EU Approach: The European Commission's proposed AI regulation differentiates between AI uses that create (i) unacceptable risk, (ii) high risk, and (iii) low or minimal risk. High-risk systems in the EU are subject to specific requirements and are banned in a limited number of cases where they contravene EU values or violate fundamental rights. The EU's AI Act also addresses how responsibility and liability for demonstrating compliance with AI regulatory principles should be allocated to various actors in the AI lifecycle. The EU also envisions a substantial role for standards bodies in drafting technical standards to support key technical areas covered by the Act. The U.S. Food and Drug Administration (FDA) has issued an action plan for AI/ML-based software as a medical device, leveraging risk categorization principles from the International Medical Device Regulators Forum.[111] Australia's AI ethics framework examines the probability of risk along with the consequences of risk using a framework. When a risk has a high

[108]IBEF, 'India's smartphone market becomes second largest globally by unit volume | IBEF' (*India Brand Equity Foundation*, 8 November 2024) <www.ibef.org/news/india-s-smartphone-market-becomes-second-largest-globally-by-unit-volume>

[109] Duarte F, 'Number of ChatGPT Users (Jan 2025)' (*Exploding Topics*, 3 December 2024) <https://explodingtopics.com/blog/chatgpt-users>

[110] EU Artificial Intelligence Act, Law No 1689/2024, 12 July 2024 <https://artificialintelligenceact.eu/chapter/3/>

[111] FDA, *Proposed Regulatory Framework for Modfiications to AI/ML based software as a Medical Device* <http://fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf>

probability of occurring and has more negative outcomes, the consequences are considered more severe.[112]German Data Ethics Commission recommends a risk-adapted regulatory approach to algorithmic systems.[113]

Across different approaches, the assessment of potential harm should consider the socio-technical system as a whole, including people and data involved throughout the AI lifecycle. It's also important to consider both the direct and indirect impacts of the system. High-risk AI applications include:

AI systems involved in sensitive uses, such as the denial of credit, employment, education, or healthcare services. Surveillance systems and other AI systems that pose risks to personal freedoms, privacy, and human rights. AI systems that create a risk of significant physical or emotional harm. Government use of predictive algorithms for fraud prevention in welfare. AI applications in human resources (HR), such as software for hiring or promoting employees.

Such practices should be adopted to ensure a more precise operationalization of AI principles. By implementing risk-tiered regulations, policymakers can ensure that AI used in critical sectors like healthcare, finance, law enforcement, and employment screening is subjected to rigorous testing and ethical compliance before deployment. Meanwhile, lower-risk AI applications, such as AI-driven chatbots or recommendation algorithms, can operate with lighter regulatory burdens while still adhering to overarching governance principles.

The Need for a Centralized Regulatory Body for AI Governance

Establishing a specific regulatory body is one of the major recommendations that can be made to effectively operationalise AI governance principles. As Sam Altman, CEO of OpenAI said a specialised regulatory body to address AI concerns and licensing needs is necessary[114]. As governance regulation will go a long way in making sure AI is effectively operationalised in the country. Presently the policy and regulation building on AI is being conducted by the various wings of the government[115] but there is a need for a singular apex authority on AI to augment the efforts taken by various ministries into one uniform code. This multidisciplinary advisory body that covers the entire digital sector. In their responsible AI paper the Niti Ayog and other authors recommend the establishment on of the "*Council for Ethics and Technology"* which may be made responsible for such create and regulation of AI throughout the country[116]. The council will

[112]Australia's AI Ethics Principles' (*Home page | Department of Industry Science and Resources*) <www.industry.gov.au/publications/australias-artificial-intelligence-ethics-principles/australias-ai-ethics-principles>
[113]Report by the German data ethics Commission on data handling and the use of algorithmic Systems | Clifford Chance' (*Clifford Chance*) <www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2019/11/report-by-the-german-data-ethics-commission-on-data-handling-and.html>
[114] Kang C, 'OpenAI's Sam Altman Urges A.I. Regulation in Senate Hearing (Published 2023)' (*The New York Times*, 16 May 2023) <www.nytimes.com/2023/05/16/technology/openai-altman-artificial-intelligence-regulation.html>

[115]'RBI sets up panel to develop a framework on ethical use of AI in financial sector' (*The Hindu*) <www.thehindu.com/business/rbi-sets-up-panel-to-develop-a-framework-on-ethical-use-of-ai-in-financial-sector/article69029678.ece#:~:text=The%20Reserve%20Bank%20of%20India,with%20the%20monetary%20policy%20announcement.>
[116]

function similarly to a thinktank, conducting research on the technical, legal, social and policy aspects of responsible AI in India. The council may also be responsible for creating model guideline and ethics review mechanisms for evaluating AI systems.

The Challenge of Translating AI Principles into Practical Guidelines

Another major problem that needs to be addressed when creating a plan to operationalise AI principle is how to translate it effectively into a workable guideline. The AI governance principles commonly have one feature, that there exists substantial difference between the high-level principle and the practical guideline needed to create a responsible AI framework.[117] The primary reason for this is the gap between the principle and the practical guidelines as there is lack of alignment between the two. The absence of well-defined methods for translation from high level principle to low and mid-level norms and operational requirements underscore the complexity in such conversion rendering the principles mostly useless in certain sectors.[118] For example, take the principle "AI should be fair." Fair is a term that is commonly used but when fit into the context of practical application of principles it is too less, while making such a principle is necessary it should be classed together with operational requirement like there should be regular bias assessments through existing technologies such as the *equity evaluation corpus*.[119][120]

Addressing the Oversight of AI Deployers in Governance Frameworks

Another major problem that blocks the proper operationalisation of AI governance Principles is that existing guidelines and AI principles focus on the end user and those affected by the AI, but not on the deployer of the AI. [121]It is necessary for creating a functioning system of AI governance that can be effectively operationalised. As a result, organisations must create guidelines with the organisational perspective in mind. By making such a shift, guidelines can be made more practical while also including the objective of the organisation in mind. Focusing on decision making framework and organisational frameworks, organisations can proactively fight unintended consequences and on part of the AI that may appear in the future rather than focusing solely on the end user concerns. but it is necessary to understand rather than abandoning the end user rights this will help to empower the organisation with the necessary frameworks and guidelines to uphold those rights from within their AI initiatives.

---

[117] C. Sanderson et al., "AI Ethics Principles in Practice: Perspectives of Designers and Developers," in IEEE Transactions on Technology and Society, vol. 4, no. 2, pp. 171-187, June 2023, doi: 10.1109/TTS.2023.3257303.

[118] Morley, J., Floridi, L., Kinsey, L. *et al.* From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. *Sci Eng Ethics* **26**, 2141–2168 (2020). https://doi.org/10.1007/s11948-019-00165-5

[119] Kiritchenko S and Mohammad SM, 'Examining Gender and Race Bias in Two Hundred Sentiment Analysis Systems' National Research Council Canada <https://saifmohammad.com/WebDocs/EEC/ethics-StarSem-final_with_appendix.pdf>

[120] Akbarighatar, P. Operationalizing responsible AI principles through responsible AI capabilities. *AI Ethics* (2024). https://doi.org/10.1007/s43681-024-00524-4

[121] ibid

Fostering a Culture of Responsibility for Ethical AI -H3

Another key feature that can help effectively to operationalise AI governance principles is the creation within the organisation a culture of responsibility. It is necessary as a responsible culture fosters increased awareness and sensitivity to potential ethical implications, societal implications and risks associated with AI systems among all stakeholders within the organisation. When responsibility is deeply ingrained within the organizational culture, ethical concerns change from being a mere afterthought but are integrated into every stage of the AI lifecycle.[122] It leads to reduced risks and enhanced trust and reputation for the organisation. But to create such a culture there needs to be a top-down approach, with the leadership setting the tone and demonstrating clear commitment to ethical AI principles. One suggestion to help create such a culture is establishing an ethical review process for example Microsoft's *Aether Committee,* an AI advisory committee that helps shape organisational policies and research priorities to AI and operationalisation of AI governance principle.[123]

Combatting Superficial AI Governance Principles

One of the major problems facing the AI governance ecosystem around the world is the toothless principles published by companies. Some companies publish AI governance principle as a mere publicity stunt to ensure false sense of security, but they are focusing on short term profits and such principle are not genuinely integrated into the decision-making process[124].For example, Clearview AI, one of the largest private companies in the United States, was recently fined a staggering thirty million euros by the Dutch Data Protection Authority for using AI to develop facial recognition software that illegally scraped images from the web and stored them in an unauthorized database. This highlights that, regardless of their public statements and commitments, companies are prone to violating governance principles unless they are held accountable to reasonable and enforceable standards[125]. So, to combat the toothless principle dilemma the nation needs to enact robust regulations and standards that hold the companies legally responsible to integrate the AI governance principle into their decision-making process, this includes defining specific requirements for fairness, transparency and accountability and data privacy. Another crucial part of ensuring that RAI is adhered to is to raise public awareness and empower the consumers to demand a RAI practice from companies can become a crucial part of the drive to operationalisation of AI governance principles.[126]

Organizations need to cultivate specific responsible AI capabilities to effectively implement responsible AI principles. These capabilities should address both AI-specific and end-to-end considerations throughout

---

[122] *Report on the Core Principles and Opportunities for Responsible and Trustworthy AI* (Ref: PS22477, Innovate UK) <https://iuk-business-connect.org.uk/wp-content/uploads/2023/10/responsible-trustworthy-ai-report.pdf>

[123] *Putting principles into practice: How we approach responsible AI at Microsoft* (Micrososft) <www.microsoft.com/cms/api/am/binary/RE4pKH5>

[124] 'Why Are We Failing at the Ethics of AI?' (*Carnegie Council for Ethics in International Affairs | Home*) <www.carnegiecouncil.org/media/article/why-are-we-failing-at-the-ethics-of-ai>

[125] Hart R, 'Clearview AI—Controversial Facial Recognition Firm—Fined $33 Million For 'Illegal Database'' (*Forbes*, 3 September 2024) <www.forbes.com/sites/roberthart/2024/09/03/clearview-ai-controversial-facial-recognition-firm-fined-33-million-for-illegal-database/>

[126] Ibid 29

the AI lifecycle. Understandable AI models are needed and organizations need to ensure that their AI systems are transparent and understandable, allowing stakeholders to comprehend how decisions are made. This involves using explainable AI techniques and providing clear documentation about the AI system's functionality bias remediation[127].Addressing bias in AI systems is crucial for ensuring fairness and equity. This requires implementing processes for identifying, mitigating, and monitoring bias throughout the AI lifecycle. Tools like Fairlearn,[128] Interpret.m[129], AIF360, and AIX360 can be leveraged for this purpose. Responsiveness is alosa must in this new age of AI governance, The AI landscape is constantly evolving, so organizations must be responsive and adaptable. This involves staying informed about emerging trends, best practices, and regulatory developments. The government's role in updating and managing AI principles is crucial in this regard.

The Need for Sector-Specific AI Governance Guidelines

Sector specific guideline is the most crucial for proper operationalisation in AI governance principle as there are vast difference in the needs of different sectors, and to address these differences[130]Operationalisation should be made based on the sector they are in while adopting the overarching principles, we should adapt and refine existing overarching AI principles to align with the specific need of that sector. For example the American Food and Drug Administration body released its Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device report emphasized the need for industry specific AI governance Guidelines. There should also be the creation of sector specific governance structures that will go a long way to oversee the implementation and enforcement of Responsible AI, this could include ethic review boards or regulatory frameworks tailored to specific sectors. Engaging with stakeholders and conducting a detailed analysis of the sector will help in identifying key AI applications and the unique ethical considerations and risks that use of AI might cause in that sector.[131]

12. Legal Framework and AI

The Artificial Intelligence market size in India is projected to reach 8.30 Billion USD in 2025 wih an annual growth rate on 27.86%[132]. This implies that the level of AI technology deployed in Indian industries such

[127]Newman JC, *Decision Points in AI Governance* (CENTER FOR LONG-TERM CYBERSECURITY) <https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf>

[128] Bird S and others, *Fairlearn: A toolkit for assessing and improving fairness in AI* (Microsoft) <www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf>

[129] Bird S and others, *Fairlearn: A toolkit for assessing and improving fairness in AI* (Microsoft) <www.microsoft.com/en-us/research/uploads/prod/2020/05/Fairlearn_WhitePaper-2020-09-22.pdf>

[130] G, *Recommendations for Regulating AI* (Google) <https://ai.google/static/documents/recommendations-for-regulating-ai.pdf>

[131] https://cltc.berkeley.edu/wp-content/uploads/2020/05/Decision_Points_AI_Governance.pdf

[132] Statista, "Artificial Intelligence - India | Statista Market Forecast" (*Statista*) <https://www.statista.com/outlook/tmo/artificial-intelligence/india> accessed January 22, 2025.

as healthcare, technology, the workforce, and education will also be on a rise, and thus, forcing the government to take active steps towards regulating AI[133].

The Indian Legal Landscape falls short of legal provisions that exclusively govern the regulation of AI[134]. Legal developments to address the growing need for AI-centred regulations and Laws are fast-paced yet nascent and thus, necessitating the ad-hoc recourse of interpreting existing principles and adhering to guidelines and frameworks released by the government. In the status quo, India's AI mission is guided by NITI Aayog's foundational National Strategy for Artificial Intelligence[135], Principles for Responsible AI, and operationalizing principles for responsible AI[136]. Amongst the prevailing laws, the IT Act, of 2000, the Competition Act, the Tort Law, the Copyright Act, the Consumer Protection Act, of 2019, and the Digital Personal Data Protection Act, of 2023 assume relevance from the perspective of AI regulations.

In the status quo, existing statutory provisions are suitably interpreted and applied to deal with issues arising from AI involvement in various fields:

Harm Caused by the AI Algorithm	**Applicable statutory provisions**
Usage of copyrighted material by generative AI when consent is not taken from the author/ owner	The Copyright Act, 1957
Usage of personal data in AI training when consent is not taken from individuals.	Information Technology Act, 2000 Digital Personal Data Protection Act, 2023
Unauthorized impersonation using AI-generated deepfakes	Bharatiya Nyaya Sanhita, 2023 Information Technology Act, 2000
Depiction of a child in sexually explicit videos generated by AI	Bharatiya Nyaya Sanhita, 2023 Information Technology Act, 2000 Protection of Children from Sexual Offences Act, 2012
Discrimination in hiring decisions using AI Recruitment tools	Equal Remuneration Act, 1976, Specific provisions based on grounds of discrimination such as - Scheduled Castes and Scheduled Tribes (Prevention of

[133] "AI Regulation in India Current State and Future Perspectives" (*Tech and Sourcing at Morganlewis*, January 26, 2024) <https://www.morganlewis.com/blogs/sourcingatmorganlewis/2024/01/ai-regulation-in-india-current-state-and-future-perspectives> accessed January 22, 2025

[134] Priya Singh, "'No Regulations for Artificial Intelligence in India': IT Minister Ashwini Vaishnaw," *Business Today*, April 06, 2023, https://www.businesstoday.in/technology/news/story/no-regulations-for-artificial-intelligence-in-india-it-minister-ashwini-vaishnaw-376298-2023-04-06

[135] NITI Aayog, *National Strategy for Artificial Intelligence*, 2018, https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

[136] NITI Aayog, Responsible AI: Approach document for India, 2021, https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf ; NITI Aayog, Responsible AI: Adopting the Framework – A Use Case Approach on Facial Recognition Technology, 2022, https://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdf

	Atrocities) Act, 1989, Transgender Persons (Protection of Rights) Act, 2019 et cetera.

The insufficiency of the prevailing legislations is primarily felt while addressing the harm caused while deploying the AI. This could look like the AI tool using copyrighted content to generate results or the tool aiding in creating sexually explicit videos featuring children. These are fundamentally new risks and prevailing laws are not equipped to deal with them. Looking at it from the perspective of different laws individually,

I. Criminal Law

The present criminal law system in India ascertains liability by inferring mens rea or the intention to have committed the act. This emerges as the most significant legal issue in establishing AI criminal liability esspecially with the emergence of Automatic AI which can take decisions without human direction. Thus, this severely restricts the scope of what the prevailing legislations can deal with when AI enter the picture.. Example, Indian courts can never ascertain intention of AI, they would thus fail to categorize AI actions in "Sections 100" and "106" of BNS of culpable homicide and negligence, respectively. The thinking that underpins these statutes presupposes a power of reason that AI is constitutionally incapable of, and there is no clear answer as to who, if anyone, should be held criminally liable[137].

II. Intellectual Property Rights Law

The intersection of AI and IPR presents two primary lacunae tha the present laws fail to address adequately - Copyrights over AI-generated content and AI using copyrighted content.

The current Copyright Act requires human authorship to be eligible for a copyright which means that outputs of generative AI, however unique, cannot be protected throught copyright. Simultaneously, since the present Copyright Act does not recognize AI as a person, the algorithm is also not legally capable of committing copyright infringements.

Suggestions

1. Defining Artificial Intelligence

One of the most crucial inclusions to any framework aiming to govern the functioning of AI must be to define what is perceived as AI. Currently there is no universally accepted definition of AI[138]. However, with the rapid expansion of AI involvement across fields, not defining what the law perceives as Artifical Intelligence will give raise to loopholes that maybe misused.

[137] "AI and Intent Data: How Do They Contribute to Full-Blown Intent Intelligence?", <http://www.algolia.com/blog/ai/ai-and-intent-data-how-do-they-contribute-to-full-blown-intent-intelligence/> (Accessed on January 15, 2025)

[138] "IP and Frontier Technologies" <https://www.wipo.int/about-ip/en/frontier_technologies/ai_and_ip.html> accessed January 22, 2025

2. Liability for harm cause by AI

As a generic approach across fields, liability in harm caused by AI can be ascertained either upon the owner or the user. Clear definitions of who qualifies as a 'owner' and 'user' is also essential to avoid loopholes which may be misused.

The ownership of the algorithm maybe granted to the one who either

i) was directly involved in developing the algorithm, or

ii) commissioned the development of the algorithm to a third party.

And the deployer or user of the AI maybe any person making use of the algorithm to generate results, but themselves have no influence over the development of the technology in any manner whatsoever.

Ascertaining liability for harm caused by AI must be on a case to case basis - in the case of AI operating upon user instructions, the first step could be to determine whether the harm resulting in the output generated could have been reasonably foreseen based on directions provided by the person deploying it. If yes, liability for the harm caused could be completely that of the deployer of the algorithm and if not, the liability will be of the owner who will be deemed to have not taken sufficient care while developing the algorithm. In the context of IPR, provisions must be added/ amended to accommodate that AI is capable of committing infringements and liability in that regard can also be determined through the same method mentioned above.

3. Granting IPR Rights to AI-Generated Content

With specific reference to Intellectual Propery Right for AI generated content, firstly, two tests could be adopted to decide if a work can be granted copyrights determining the level of AI deployment in the content created.

First, The human-AI ratio of work done to generate the end results will be used to subsequently determine if the user of the algorithm is eligible for a copyright protection for the work. The EU Report on Challenges to IPR Framework[139] suggests an approach based on the involvement of the user in the content creation process. According to which, If the role of the system user is so constricted that he cannot exercise free choices at any stage of the creative process, the user is a passive player who will not qualify as author of the ensuing production. This leaves the owner of the AI System with authorship of the AI assisted work.

Second, the aspect of the work for which AI has been used - if the creative part or the descriptive part of the work is algorithm generated. If creative angle to the work in introduced through the deployment of AI, the same is unlikely to result in a copyright for the author. The present EU IPR Framework provides a similar approach in this regard. The framework recommends the division of mahine-aided production into distinct phases of the creative process - (i) Conception - involving creating and elaborating the design/plan

[139] European Commission: Directorate-General for Communications Networks, Content and Technology, Hartmann, C., Allan, J., Hugenholtz, P., Quintais, J. et al., *Trends and developments in artificial intelligence – Challenges to the intellectual property rights framework – Final report*, Publications Office of the European Union, 2020, **https://data.europa.eu/doi/10.2759/683128**

of work, (ii) Execution - converting the plan into drafts of the final output and (iii) Redaction where the drafts are processed and reworked to deliver the final output. Here, production of output by an AI system could qualify as an work protected under the EU Copyright law on condition that a human being initiated and conceived the work and subsequently redacted the AI-assisted output in a creative manner. Essentially, mere human intervention at the conception and redaction stages could suffice for copyright protection.

Combating AI-Induced Bias

Understanding What is Bias in Artificial Intelligence

Bias is commonly understood as prejudice and this understanding extends to A.I. systems. Bias in AI can be defined as the systems' systematic predilection or prejudice, resulting in unequal treatment or inaccurate outcomes for specific persons or groups.[140] It can lead to discriminating results based on race, caste, sex, social standing, etc.

Bias is a long-standing issue for AI algorithms, in part because they are frequently trained on data sets that itself is biased or not entirely representative of the people they serve, and in part because they are created by humans, who have their own intrinsic biases.[141] An experiment conducted by Bloomberg used the generative AI model Stable Diffusion to generate thousands of images about job titles and crime. It was found, "*The world according to Stable Diffusion is run by White male CEOs. Women are rarely doctors, lawyers or judges. Men with dark skin commit crimes, while women with dark skin flip burgers.*"[142]

According to Shivangi Narayan, a researcher who has studied predictive policing in Delhi, "*It is going to directly affect the people living on the fringes - the Dalits, the Muslims, the trans people. It will exacerbate bias and discrimination against them*,"[143] Further, Siva Mathiyazhagan, an assistant professor at the University of Pennsylvania, reported to Thomson Reuters Foundation, "*If you ask a chatbot the names of 20 Indian doctors and professors, the suggestions are generally Hindu dominant-caste surnames - just one example of how unequal representations in data lead to caste-biased outcomes of generative AI systems*".[144]

At present, companies such as Microsoft and Amazon take the help of AI models to aid in recruitment and hiring.[145] However, if the datasets these AI systems rely on are biased, it could lead to unfair hiring practices and discriminatory outcomes. This not only would it stall the progress of the development of AI and

---

[140] Singh, S. (2025) *Understanding bias in artificial intelligence: Challenges, impacts, and mitigation strategies - E&ICT Academy, IIT Kanpur*, *E&ICT Academy, IIT Kanpur*. Available at: https://eicta.iitk.ac.in/knowledge-hub/artificial-intelligence/understanding-bias-in-artificial-intelligence-challenges-impacts-and-mitigation-strategies/ (Accessed: 18 January 2025).

[141] *Rise of AI puts spotlight on bias in algorithms - WSJ* (no date) *Wall Street Journal*. Available at: https://www.wsj.com/articles/rise-of-ai-puts-spotlight-on-bias-in-algorithms-26ee6cc9 (Accessed: 18 January 2025).

[142] *Humans are biased. Generative AI is even worse* (no date) *Bloomberg.com*. Available at: https://www.bloomberg.com/graphics/2023-generative-ai-bias/ (Accessed: 18 January 2025).

[143] https://www.thehindu.com/sci-tech/technology/racist-sexist-casteist-is-ai-bad-news-for-india/article67294037.ece

[144] Id.

[145] Writingsofrach (2023) *Microsoft, Amazon among the companies shaping AI-enabled hiring policy*, *CNBC*. Available at: https://www.cnbc.com/2023/10/11/microsoft-amazon-among-the-companies-shaping-ai-enabled-hiring-policy.html (Accessed: 18 January 2025).

undermine public trust in AI.

Global Framework

The EU AI Act have established certain legal requirements to promote fairness in AI systems. Article 10(2)(f)[146] and 10(2)(g)[147] provide that Data sets must be examined for potential biases that could negatively impact health, safety, fundamental rights, or lead to prohibited discrimination, especially when the AI system's outputs influence future inputs. Appropriate measures must be implemented to detect, prevent, and mitigate any biases identified.

From a bare reading of the text, it is clear that the Act urges the developers to be more cautious against any biases that the model might have. To further mitigate the biases, the Act requires member states to create at least one AI regulatory sandbox.[148] These sandboxes are controlled environments where AI systems can be developed, tested, and validated before being released to the market. This enables authorities to oversee AI systems and guarantee that their output is fair and transparent as required.

Way Forward

The best way of removing biases in an AI system is at the stage of training the model. While training, it must be ensured by the developers to avoid group stereotypes. Researchers can force the model to ignore attributes like race, class, age, and gender. According to one researcher at the MIT-IBM Watson AI Lab, "*It's not the algorithm that's to blame, it's the data*".[149] This method was tested by IBM and was found by the researchers, "*If you just remove them, it turns out you drastically improve on fairness. The beauty of this technique, for foundation models especially, is you can avoid retraining the model*."[150] If applied to India, AI models, being forced to ignore caste and religious barriers, will eliminate stereotypes by evaluating individuals based on their unique qualities rather than grouping them. The government therefore must introduce a policies prohibiting the stereotyping of individuals based on the groups.

After the development of the AI, the Government must create a regulatory sandbox similar to the one present in the EU Act. Within this sandbox, the system should undergo rigorous testing, including red-teaming exercises whereby experts consciously try to find and take advantage of AI systems' flaws, restrictions, and vulnerabilities to strengthen these models. Red teaming makes it possible to depict actual scenarios in a regulatory sandbox where AI can lead to unjust and discriminatory outcomes. This technique is essential for finding systemic flaws that conventional testing methods could miss.

Bias detection and mitigation strategies must be integrated in the process of development of AI models. AI has the potential to develop into a technology that benefits all people and promotes a more just and equitable society if these ethical issues are given top priority.

---

[146] EU Artificial Intelligence Act, Article 10(2)(f).
[147] EU Artificial Intelligence Act, Article 10(2)(g).
[148] EU Artificial Intelligence Act, Article 57.
[149] Martineau, K. (2022) *Debugging Foundation models for Bias*, *IBM Research*. Available at: https://research.ibm.com/blog/debugging-AI-bias (Accessed: 18 January 2025).
[150] Id.

13. Conclusion

As AI continues to evolve and integrate into every aspect of our lives, India must take a proactive and balanced approach to its governance. While AI offers incredible opportunities for economic growth, efficiency, and innovation, it also brings challenges like privacy risks, biases, and security threats. Without a strong yet flexible regulatory framework, the risks could outweigh the benefits.

The MeitY report provides a good foundation for AI governance, emphasizing accountability, transparency, safety, and fairness. Ensuring that AI systems operate without bias, remain transparent in their decision-making, and function reliably in real-world conditions is key to building trust and promoting responsible AI use. However, regulations cannot be static; they must evolve alongside technology to keep pace with AI's rapid advancements. For AI governance to be effective, it needs a collaborative effort—not just from the government, but also from tech developers, businesses, regulators, and society at large. A lifecycle approach—where AI systems are monitored from development to deployment—combined with automated oversight mechanisms will help enforce regulations without stifling innovation. Recognizing the interconnected role of stakeholders, from data providers to end users, is equally crucial in creating a regulatory framework that works for everyone.

Moving forward, India must strike the right balance—creating an environment where AI can thrive while ensuring ethical safeguards are in place. By embracing a structured yet adaptable governance model, India can harness AI's potential for national progress without compromising public interest. Though the challenges are complex, with collective effort and a forward-thinking approach, the country can lead the way in responsible AI governance, ensuring that technology serves the people, not the other way around.