RESPONSIBLE AI
#AIForAll

Approach Document for India:
Part 2 - Operationalizing Principles for Responsible AI

AUGUST 2021

RESPONSIBLE AI
#AIFORALL

Approach Document For India:
Part 2 - Operationalizing Principles For Responsible AI

AUGUST 2021

Acknowledgements

In writing this report, Rohit Satish and Preeti Syal from NITI Aayog have made valuable contributions.

We are pleased to have collaborated with the World Economic Forum Centre for the Fourth Industrial Revolution as the Knowledge partner in developing the Responsible AI for All approach document. The valuable contributions of Ms. Arunima Sarkar from World Economic Forum are acknowledged. Legal inputs by Shardul Amarchand Mangaldass are also acknowledged.

We are also grateful for the support and contributions of several experts from India and globally including Prof Amit Sethi, Prof Balaraman Ravindran, Google India, Mr John Havens and Srichandra (IEEE), Prof Mayank Vatsa, Dr P Anandan and Dr Rahul Panicker from Wadhwani Institute for Artificial Intelligence, Dr. Rohini Srivathsa (Microsoft), Tanay Mahindru (NITI Aayog) and Vidhi Center for Legal Policy. Valuable inputs were also provided by various Ministries/ Departments of the Government of India and regulatory institutions, namely MeitY, DST, DBT, Office of the PSA, MoHUA, RBI and NHA.

Anna Roy
Sr. Advisor,
NITI Aayog

Foreword

Artificial Intelligence has seen significant growth in India in the past few years. India with its robust startup ecosystem with AI powered innovation has the highest AI skill penetration rate in the world. The job market for AI has also shown promising growth with "AI Specialist" being among the top job roles in India in 2020. Various Government entities have also leveraged AI powered innovations in offering efficient services and enabling transparent governance.

This is just the beginning and the momentum needs to be sustained. The National Strategy for Artificial Intelligence underlines the importance of a trusted ecosystem for accelerated adoption of the technology. This is particularly relevant for India as *'AI for All'* is the core of the national strategy and the well documented diversity, digital divide, scale and lack of awareness provides a fertile ground for the risks of AI to amplify. The importance of ensuring responsible use of technology was echoed by the Hon'ble Prime Minister himself at the Davos Summit of World Economic Forum in January, 2021. In this regard, an approach document on 'Principles of Responsible AI', based on wide ranging consultations, was released in February 2021. The document identified seven principles derived from the tenets of the Indian Constitution which provide a guiding framework for various stakeholders in leveraging AI.

While identifying the principles is an essential starting point, operationalizing them is the next important step. Ensuring that AI systems adhere to the principles requires a multi-disciplinary approach and a behavioral shift in organizational processes and practices. The multi-faceted role of the Government as a policy maker, regulator, and procurer makes it an important stakeholder in the operationalization of the principles. However, it is also important to note that Government interventions alone are not sufficient and it is important for the entire ecosystem to play its role in ensuring to put in place a trusted AI ecosystem.

This document identifies the various mechanisms needed for operationalizing these seven principles. It is a culmination of a series of interviews with experts and AI practitioners over the past year. This follows a working document that was placed for public consultation last year. It outlines the specific role for the Government and recommends a multi-disciplinary advisory body to guide the various activities. It is extremely important that any measures taken to regulate the technology must be proportional to the risk and must be balanced to encourage innovation. The document also recommends measures for the private sector, research and academia to build an institutional capacity to evaluate the risks and undertake actions to appropriately address them.

We hope the country and the AI community at large will join and support us in this effort to create a responsible AI ecosystem and unleash the enormous potential of AI in the society.

Dr. Rajiv Kumar

Contents

> "
> Making AI more sensitive to the full scope of human thought is no simple task. The solutions are likely to require insights derived from fields beyond computer science, which means programmers will have to learn to collaborate more often with experts in other domains.
>
> *-Fei Fei Li, Computer Scientist*
> "

Whether we understand it or not, AI is ever-pervasive, rendering a new meaning to the words 'automatic', 'intelligence' and 'machines'. For India, the era of AI holds promise beyond economic growth – the promise and potential of solving some of the country's most difficult social and societal challenges. As outlined and identified by National Strategy for AI in 2018, we are already beginning to see vast impact of AI across healthcare, agriculture, education and entertainment. During COVID-19 AI image recognition solutions and ML-based resource allocation disaster platforms greatly enhanced state's capability to deliver services bridging gaps of limited access, resources, healthcare delivery and knowledge.

While the potential of AI to solve complex problems and societal issues is beyond misgiving, the risks and challenges of leveraging AI have emerged in parallel, requiring dealing with trust issues towards enabling adoption of AI at-scale. Besides the usual large data set biases that gets perpetuated, the 'black box' nature of certain types of AI compounds the problems. The inherent nature of AI systems lacking transparency does not credit or help build user-trust, making it more difficult. Most recently, accentuating digital divide and denying access to healthcare by their very nature, AI-powered applications draw their fair share of societal dislike and unless there are measures in place to address these, we will continue to see a rise in the skepticism. Not to forget the use of AI for malicious intent (fake news, deep fakes etc.) to create misinformation is already beginning to accumulate as negative externality pitted against the benefits to society.

The National Strategy for AI, while laying down its vision for implementation of AI, addressed these issues by emphasizing the need to foster Responsible use of AI. Taking that vision forward, a roadmap for the Responsible use of AI in the country is key to bringing the benefits of *'AI to All'*, i.e. inclusive and fair use of AI. In Part-1 of the Responsible AI paper released in February 2021, the various systems and societal considerations of AI systems have been studied and the principles for Responsible AI have been outlined.

While the overarching Responsible AI principles will guide the overall design, development and deployment of AI in the country, operationalizing these principles by the ecosystem is essential to realize the results. This paper –Part 2 of the strategy – lays that groundwork. A delicate balance guides the adoption of these principles in the AI ecosystem in India, with a focus on maximising the benefits of AI for all, while minimising AI-related risks. The paper notes that this paradigm of promoting risk-minimised AI rests on two key concepts: calibration, in that regulatory and policy interventions designed for realising the principles must be calibrated to the uses and the risk-profile of AI systems, and continuous assessment, in that these principles are ingrained into an AI system's lifecycle.

This paper identifies a series of actions that the ecosystem must adopt to drive responsible AI. These actions are divided among three stakeholders; governments, the private sector and research institutions. Among these stakeholders, the actions are further divided into areas, with each area identifying a series of related measures for implementing the AI principles. These are:

- ***For the government*** – designing ideal regulatory and policy interventions, creating awareness, accessibility and capacity building, and facilitating precise procurement strategies.
- ***For the private sector and research institutions*** – incentivising ethics by design, creating frameworks for compliance with relevant AI standards and guidelines, and the promotion of Responsible AI practices in research.

In the context of regulation, the paper recommends a risk-based mechanism for regulating AI in India. Regulation must be proportional to the likelihood of harm that can be occasioned by an AI system; greater the risk of harm, greater the regulatory scrutiny attracted by the relevant AI system. In order to determine the risk posed by AI systems, the paper proposes the adoption of specific policy interventions, such as sandboxing and controlled deployments. Further, in instances where the perceived risk of harm is low, governments may prefer regulatory forbearance and allow market players to lead with self-regulation. Sectoral regulators may however, continue to oversee AI-related developments in their field to avoid conflicting guidelines in the future.

Presently, policy and regulation-building on AI is being explored by various limbs of government. It is important however, to augment the capacities of such bodies, and ensure cohesive policymaking on AI. In light of this, the paper proposes the setting up of an independent, multi-disciplinary advisory body at the apex-level, whose remit covers the entire digital sector. This proposed Council for Ethics and Technology (CET) will aide sectoral regulators in formulating appropriate AI policies, and serve as a think-tank for creating quality research products around issues related to AI. The CET will be also responsible for devising model guidelines or ethics review mechanisms that will evaluate the efficacy of AI systems.

In addition to proposing these government-driven measures, the paper notes that the delivery of ethical AI will also be influenced by the private sector. In light of this, the recommendations include mandating responsible AI practices for any public-sector procurement of AI systems and in the adoption of high-risk AI. The private sector is also encouraged to devise unique ways to ensure cost-effective compliance with AI standards, with the paper recommending the assignment of relevant roles to specific personnel and the leveraging of open tools and materials to achieve the same.

Lastly, the paper identifies high-quality research as a priority in aiding the implementation of the AI principles, including through government-formulated guidance on measuring the impact made by AI research initiatives. At the same time, the paper recognises that responsible AI principles should be a critical consideration for the research itself.

Author of "Branding India"
JNU (MA in International Relations)

Amitabh Kant
CEO, NITI Aayog

Principles for Responsible AI– A Background

1.1. Pursuant to the recommendations of the National Strategy for Artificial Intelligence (NSAI)[1], NITI Aayog in 2021 released an approach document on the Principles for Responsible Artificial Intelligence. The document based on widespread consultations with experts across research, law, non-profit, civil society, private sector and the government had studied various ethical ramifications for the development and use of Artificial Intelligence (AI) across two levels (Refer Box 1):

    a. impact on various stakeholders (eg: users, individuals/organisations impacted by AI's decision, auditors, etc) of a specific AI system; and

    b. broader impact on the society (eg: impact of automation on jobs, social discord due to malicious use).

1.2. The document also benchmarks the technology and legislative approaches for responsible AI and identifies seven principles to drive convergence across various stakeholders in the development of the AI ecosystem in India.

Box 1: Considerations for Responsible AI

Consideration	Description	Implications
Understanding the AI system's functioning for safe and reliable deployment	While accuracy gives a reasonable view into how a system performs, understanding decision making process is important to ensure safe and reliable deployment	The system could pick spurious correlations, in the underlying data, leading to good accuracy in test datasets but significant errors in deployment
Post-deployment–can the relevant stakeholders of the AI system understand why a specific decision was made?	With 'Deep Learning' systems have become opaque, leading to the 'black box' phenomenon; Simple linear models, offer interpretable solutions but their accuracy is usually lower than deep learning models;	Leads to: • A lack of trust by users, discouraging adoption • Difficulty in audit for compliance and liability • Difficulty in debugging/ maintaining/verifying and improving performance • Inability to comply with specific sectoral regulations

---

1 National Strategy on Artificial Intelligence released by NITI Aayog in 2018

Consistency across stakeholders	Different types of cognitive biases have been identified and tend to be 'unfair' for certain groups (across religion, race, caste, gender, genetic diversity); Since AI systems are designed and trained by humans, based on examples from real-world data, human bias could be introduced into the decision-making process;	Large scale deployment of AI, leads to a large number of high-frequency decisions, amplifying the impact of unfair bias. Leads to lack of trust and disruption of social order
Incorrect decisions leading to exclusion from access to services or benefits	There are a variety of means of assessing or evaluating the performance of an AI system (accuracy, precision, recall, sensitivity, etc.); In some cases, despite a high accuracy a system may fail in other measures;	May lead to exclusion of citizens from services guaranteed by the state
Accountability of AI decisions	Decisions by AI systems are influenced by a complex network of decisions at different stages of its lifecycle. Deployment environment also influences self-learning AI Assigning accountability for harm from a specific decision is a challenge	Lack of consequences reduces incentive for responsible action Difficulty in grievance redressal ↘ VAGUENESS IN DECISION MAKING PROCESS
Privacy risks	AI is highly reliant on data for training, including information that may be personal and/or sensitive (PII), giving rise to: Risk that entities may use personal data without the explicit consent of concerned persons; Possible to discern potentially sensitive information from the outputs of the system	Infringement of Right to Privacy
Security risks	AI systems are susceptible to attack such as manipulation of data being used to train the AI, manipulation of system to respond incorrectly to specific inputs, etc; Given some AI systems are 'black boxes', the issue is amplified	Real-world deployments may lead to malfunctioning and potentially impact the fundamental rights if underlying AI models are manipulated; Risk to IP protection due to potential of 'model steal' attacks

Societal considerations	
Consideration	**Recommendations**
Impact on jobs	Track changes in job profiles, both nationally and internationally Identify policies to harness upcoming job profiles through skilling and education and safeguard interests of citizens in those roles Have a long term strategy to harvest the potential of AI to create additional job roles
Malicious use of AI: psychological profiling and false propaganda	Advance research efforts towards flagging of malicious content in local languages

1.3. The Supreme Court of India has, in various instances, benchmarked prevailing morality in India with the principle of Constitutional morality[2]. The Principles for Responsible AI in India (Refer Box 2) thus flow from the Constitution of India and all laws enacted thereunder and are also compatible with the principles identified by international bodies such as the Global Partnership on Artificial Intelligence (GPAI).

Box 2: Principles for Responsible AI

Principle	Description
Principle of Safety and Reliability	AI should be deployed reliably as intended and sufficient safeguards must be placed to ensure the safety of relevant stakeholders
Principle of Equality	AI systems must treat individuals under the same circumstances relevant to the decision equally
Principle of Inclusivity and Non-discrimination	AI systems should not deny opportunity to a qualified person on the basis of their identity. It should not deepen the harmful historic and social divisions based on religion, race, caste, sex, descent, place of birth or residence in matters of education, employment, access to public spaces, etc. It should also strive to ensure that an unfair exclusion of services or benefits does not happen.
Principle of Privacy and Security	AI should maintain privacy and security of data - of individuals or entities that is used for training the system. Access should be provided only to those authorized with sufficient safeguards
Principle of Transparency	The design and functioning of the AI system should be recorded and made available for external scrutiny and audit to the extent possible to ensure the deployment is fair, honest, impartial and guarantees accountability
Principle of Accountability	All stakeholders involved in the design, development and deployment of the AI system must be responsible for their actions
Principle of protection and reinforcement of positive human values	AI should promote positive human values and not disturb in any way social harmony in community relationships

Operationalizing Principles – An Evolving Landscape

1.4 The principles are based on current understanding and AI landscape and must evolve with innovation and technology advances and with a greater understanding of the impact of AI. Identifying Principles is the essential first step, that needs to be complemented by the mechanisms required for adherence to these principles towards ensuring a responsible AI ecosystem. Adherence to the Principles may require new institutional mechanisms, certain changes in processes and operations of various entities involved, and requisite governance frameworks. This document identifies mechanisms for enforcement of the Principles of Responsible AI, broad governance structures and policies for the creation of a responsible AI ecosystem in India.

---

2   https://main.sci.gov.in/supremecourt/2016/14961/14961_2016_Judgement_06-Sep-2018.pdf

02

Responsible AI and India

Significance of AI for India

2.1. The NSAI advocates for responsible use of AI and the approach document on Principles for Responsible AI identifies a core set of principles to guide the various stakeholders of the AI ecosystem. This chapter outlines the various considerations needed to ensure practice and operationalization of these principles. The institutional framework required to guide the responsible AI lifecycle across public sector, private sector and research institutions and the policies to enable responsible AI, are further explored in the subsequent chapters.

2.2. Several studies have quantified the economic impact of AI for the Indian economy[3]. The NSAI also identifies potential social benefits especially in sectors like health, education, agriculture, viz. increased access to quality health facilities, inclusive financial growth for large sections of the population that have historically been excluded, real-time and customized advisory to farmers, and building smart and efficient cities and infrastructure. **AI has also been recommended by the Indian Judiciary in various instances to uphold the fundamental rights of citizens and improve efficiency** (examples in Box 3)

Box 3: Use of Artificial Intelligence to uphold rights and improve efficiency

The Supreme Court of India and various High Courts have recommended the use of AI as a tool to meet the objectives of various laws and improve efficiency:

Location of Missing Persons

- *Sri C. Shiva S/O Chikka Chowdappa vs The State of Karnataka (2006):* The Karnataka High Court discussed the use of AI based facial recognition software to help Bangalore City Police identify and locate missing persons.

Child Protection

- *In re Prajwala (2018):* Certain social media companies highlighted, before the Supreme Court, the possibility of using AI for proactive detection of content amounting to Child Sexual Imagery.

Trade Name Protection

- *Tata Sky Limited vs. National Internet Exchange of India (2019):* The Delhi High Court suggested that AI be used to prevent identical or deceptively similar domain names to be registered.

---

3   Rewire for Growth: Accelerating India's Economic Growth with AI, Accenture (2018)

> **Efficiency in the judicial process**
>
> • In April 2021, the Supreme Court launched its AI portal SUPACE (Supreme Court Portal for Assistance in Courts Efficiency) to leverage machine learning (ML) to aid scrutiny of cases and address existing bottlenecks.[4]

2.3. Building a robust AI ecosystem is also crucial for India as it seeks to establish itself as a hub for AI development.[5] The Stanford AI Index Report (2021) shows that India has the highest AI skill penetration rate in the world.[6] According to a recent NASSCOM report, data and AI have the potential to add USD 450-500 billion to India's economy by 2025.[7] AI also has a significant presence in the startup ecosystem, with 44% of deep-tech startups in India leveraging AI technology.[8] The job-market for AI is also showing promising growth, with 'AI Specialist' being the #2 among emerging job-roles in India in 2020.[9] The export of software services contributed USD 128.6 billion in 2019-20, registering a growth of 9.1 per cent.[10] Robust and reliable frameworks serve to increase confidence in AI-powered products and services from India.

The need to adopt AI responsibly

2.4. At the same time there are documented risks relating to this technology as outlined in the approach document on the Principles for Responsible AI (2021). India has one of the highest smartphones user bases in the world, providing a large platform for applications to scale.[11] The diversity, scale, digital divide, lack of awareness and inequality serves a fertile ground for the negative effects of AI to amplify. Creating a trusted AI ecosystem is important to realise both the economic and social potential of AI.

2.5. Addressing the risks needs a consistent approach and clarity on acceptable behaviour of AI systems under various situations and across use cases. AI also depends on data and therefore is enabled by high quality data availability, robust data protection and sharing protocols. Guidelines and frameworks therefore need to be evolved with advances in technology and increase in use cases.

2.6. The approach for operationalizing the Principles in India needs to therefore strike a balance between creating the necessary guardrails and enabling research and innovation to flourish. The goal must be to maximize the benefits of AI for the citizens, businesses and research and minimize the risks. There is extensive literature on how well-calibrated guidelines and frameworks on ethics can provide clarity, improve trust and define expectations, thus promoting research and innovation.[12,13] The operationalization of the

---

4   https://webcast.gov.in/scindia/6apr2021.html

5   Artificial Intelligence Market Forecasts | Omdia; DC FutureScape: Worldwide IT Industry 2018 Predictions

6   Artificial Intelligence Index Report 2021, Stanford University HAI

7   https://nasscom.in/knowledge-center/publications/unlocking-value-data-and-ai-india-opportunity

8   https://nasscom.in/knowledge-center/publications/indias-deeptech-start-ups-next-big-opportunity

9   LinkedIn: 2020 Emerging Jobs Report India

10 https://rbi.org.in/scripts/BS_PressReleaseDisplay.aspx?prid=51278

11 https://icea.org.in/wp-content/uploads/2020/07/Contribution-of-Smartphones-to-Digital-Governance-in-In-dia-09072020.pdf

12 Economic Survey (2019-20)

13 https://www.eiu.com/n/staying-ahead-of-the-curve-the-business-case-for-responsible-ai/

Principles for Responsible AI in India must not only look at the regulatory aspects of the technology but also consider enabling policies for responsible innovations.

2.7. Part 1 of the responsible AI series studied the various considerations for responsible AI under systems and societal considerations. Systems considerations identify the various aspects that need to be examined for the use of individual AI systems. Societal considerations identify broader potential ramifications arising from the interaction of AI systems with the society. The responsible AI ecosystem must be calibrated to address both these considerations.

2.8. The growth of AI has been relatively recent and its adoption in India is at a nascent stages. Understanding the societal risk requires an ongoing monitoring and study of the influence of AI systems in India and around the world in an institutional manner. While issues such as the impact on jobs or malicious use of AI may not be sector specific, certain sectors may see a greater impact than others. It is therefore important to create a multi-disciplinary institution for research, enabling private sector, legal, social and policy thinking on empowering effective interfacing with relevant Ministries and the States.

## Tearing down barriers – promoting adoption of Responsible AI

2.9. In addition to studying the risks to the society, there is also a need to remove barriers for responsible AI and for the advocacy of responsible AI systems and the benefit it offers. Lack of trust in technology and AI systems has inhibited their adoption in various sectors. The limited digital literacy and skewed digital footprint inhibits creation and adoption of large-scale responsible AI systems. The NSAI identifies a key role for India to serve as a leader in AI for social good and solve for challenges in the developing and emerging economies. It is therefore important that such challenges are represented and considered in international dialogue on AI. A mechanism for this has been recommended in Chapter 3.

2.10. Various organizations are involved in the research and development of AI systems and the risks of the technology depends on the specific context for which it is used and the environment it is deployed. It is therefore **infeasible to identify prescriptive one-size-fits-all guidelines to ensure adherence to the Principles**. Instead, the focus must therefore be on instituting governance mechanisms that would enable the creation of reliable, predictable and trustworthy applications. imp.

2.11. Chapters 3 and 4 identify such governance mechanisms across the Government, the private sector and research institutions. It is important that these mechanisms start with stakeholder awareness and education on both capabilities of AI and the risks. There must also be an institutional mechanism to consider multi-disciplinary perspectives and address AI-related risks. **The responsible AI considerations cannot be a one-time activity and must be embedded into the lifecycle of the AI system**. In addition, thinking through the various considerations requires a wide-ranging perspective and should ideally involve a cross-disciplinary representation. Institutional capacity of regulatory systems must be augmented to enable creation of standards,

guidelines and benchmarks for individual use-cases or specific technologies based on the social, economic, political, and cultural realities of the nation, while maintaining an international outlook.

2.12. NSAI recommends that the Government must drive adoption of AI systems especially in the social sector. AI has also seen a sharp growth in private sector and research outputs in the recent years. It is important for the entire ecosystem to play its role in ensuring a trustworthy AI ecosystem. In this regard, the subsequent chapters identify actions for the Government, private sector and research institutions.

03
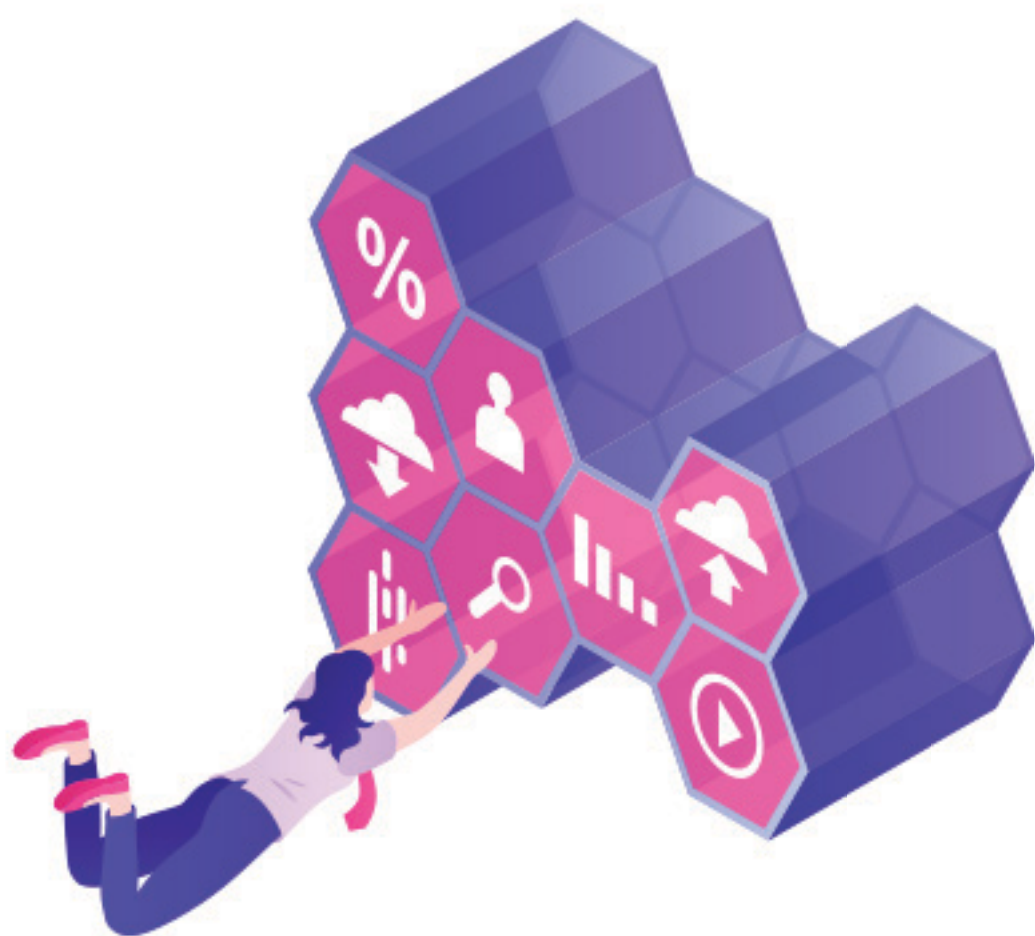
Role of Government

Role of Government

3.1. The NSAI (2018) argued that while the private sector has a significant stake in the development of AI in India, it is the role of the Government to drive adoption of AI in social sectors. The adoption is primarily aimed at achieving various goals such as overcoming access barriers, increased and efficient access to government schemes and services, and enabling high quality skill-based services at the all levels of the Government and inclusive growth. Due to the sheer scale of Government programs and initiatives, ensuring an institutional mechanism for procurement of AI systems to follow responsible AI principles goes a long way in improving trust in the technology and improve acceptance of AI systems by the public.

3.2. This Chapter looks at the broad areas for Government intervention and identifies an institutional mechanism to support the implementation

Areas for Government intervention

3.3. As discussed in Part 1 of the responsible AI series, various legislations and regulations already influence development and use of AI systems. The diversity of the country and limited digital literacy of the population makes it important for the Government to undertake enabling measures to empower various innovators across private sector, research and academia to adhere to responsible AI principles for AI based innovations. In this regard, the interventions by the Government must strengthen the following pillars of a responsible AI ecosystem

    a. Regulatory interventions towards creating a trusted AI ecosystem

    b. Policy interventions to enable a responsible AI adoption

    c. Awareness and capacity building on responsible AI in the public sector

    d. Facilitate alignment of procurement mechanisms with responsible AI principles

Area 1: Regulatory Interventions

3.4. The approach document *Principles for Responsible AI*[14] notes that various considerations and risks with AI systems already find an expression in the Constitution of India and existing laws. Specific rules and regulations may need to be augmented to include the AI/ML-specific risks. In addition, the growth of AI has been relatively recent and approaches to govern AI systems are still evolving in most parts of the world. India has also seen AI-specific regulatory interventions and, in certain cases, existing regulations define the expectations from AI systems.

3.5. There is also an enabling role that regulations may play to boost the adoption of AI. The NSAI identified the lack of ethical regulations as being a key barrier for AI adoption. For instance, clarity around doctor-patient confidentiality, the informed consent process, explainability standards and liability framework are a few of the areas in which the Government may consider enabling AI innovators in the digital healthcare industry.[15]

3.6. Approaches to regulate AI systems must aim to protect individual rights while promoting innovation. A one-size-fits-all approach to AI regulation, by design, is not feasible as the risks depend on the given use case and context in which it is deployed. An evolving, risk-based approach is needed to encourage innovation and safeguard the consumer and citizen interests. Various bodies around the world are exploring regulatory mechanisms on similar lines (see Box 4).

Box 4: Global approaches to AI regulation

On 21 April 2021, the European Commission published its proposal for a Regulation on Artificial Intelligence. The regulation follows a risk-based approach, differentiating between uses of AI that create (i) an unacceptable risk, (ii) a high risk, and (iii) low or minimal risk.16 Whether an AI system is classified as high-risk depends on the intended purpose of the system and on the severity of the possible harm and the probability of its occurrence.[17]

The U.S. Food and Drug Administration (FDA) issued the "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" from the Center for Devices and Radiological Health's Digital Health Center of Excellence.[18] The paper leveraged practices from the International Medical Device Regulators Forum's risk categorization principles.

Australia's Artificial Intelligence (AI)[19] Ethics Framework examines the probability of risk, together with the consequence via suggestive frameworks. When a risk has both a high probability of occurring and more negative outcomes, the consequences become more severe. (details in Appendix 2).

3.7. A risk-based regulatory mechanism is recommended for India. The principle underlying this approach is this: the greater the potential for harm, the more stringent the requirements and the more far-reaching the extent of regulatory intervention. In cases where the AI system has the potential to violate the

---

14   NITI Aayog (2021). Approach Document for India. Part 1- Principles for Responsible AI

15   CIS report- AI in Healthcare

16   https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

17   https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence

18   https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

19   https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework

fundamental rights, or it is highly likely to cause harm or a negative impact, the Government should consider increased scrutiny and mandate responsible AI practices.

3.8. When assessing the potential for harm, the sociotechnical system as a whole must be considered. All components of an algorithmic application, including the people involved, from the design phase through to its implementation in an application environment and any evaluation and adjustment measures, should be assessed.[20] The assessment of risk must also take both the direct and indirect impact of the system into consideration and may be done through policy instruments like sandboxing.

3.9. In low-risk AI applications where the risks are low, there must be an effort to minimize the regulatory burden. Self-regulation and awareness campaigns may offer the best approach for responsible AI practices for such use cases. Supporting structures to enable accountability, transparency and grievance redressal may be required for self-regulation to be effective.

3.10. In areas where the risks are not clear, regulatory mechanisms may be developed through policy sandboxes and controlled deployments where market reactions and impact could be closely monitored.

3.11. Standards offer a flexible and evolving approach to promote innovation and industry participation for AI. Areas of interest have been identified and relevant standards are being developed by various international standards organizations. For example, IEEE P2089 establishes a framework for developing digital services for children.[21] The National Digital Health Mission (NDHM) proposes the use of the FHIR (Fast Health Interoperability Resources) standard for interoperability and data exchange[22].

3.12. Standards are maintained by experts and go through a transparent due process that is recognized internationally. For India, the standards and benchmarks may be identified based primarily on the prevalent social, economic, political and cultural factors. International standards may be leveraged when the goals are common.

3.13. Regulatory mechanisms have historically not kept pace with innovation. Until the necessary guidelines are in place, the principles for responsible AI may serve as a guide and, where feasible, the development of AI systems may be done in collaboration with multi-disciplinary stakeholders to ensure adherence.

3.14. India has an extensive and robust system of sectoral regulators that oversee various activities, products and services. These regulators already have elaborate mechanisms to regulate and govern innovations in their domain, with some releasing rules and guidelines for AI applications (Box 5).[23,24] Extant regulation may continue to oversee AI-led innovations in domains under their purview for the time being. This would also avoid the risk of conflicting or confusing guidelines and reduce compliance overhead.

20 https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf;jsessionid=-0F3AEDD276064F891DC87DBC08CB473A.1_cid334?__blob=publicationFile&v=2

21 https://standards.ieee.org/project/2089.html

22 https://nha.gov.in/home/emr_faq

23 https://ndhm.gov.in/documents/ndhm_strategy_overview

24 https://www.sebi.gov.in/legal/circulars/may-2019/reporting-for-artificial-intelligence-ai-and-machine-learning-ml-applications-and-systems-offered-and-used-by-mutual-funds_42932.html

3.15. Legislative interventions may be needed as the use-cases of AI in regulated or high-risk areas mature and may be considered at a relevant stage. For example, the electronic evidence is currently governed by the Indian Evidence Act – Ss. 65A and 65B specifically. However, the increasing use of biometrics, or algorithms in predictive policing is not deemed to be "electronic evidence" within these provisions and may require amendments or bespoke legislation. The draft PDP bill has provisions for an "AI sandbox" with the intention of incentivising innovation in a regulatory lenient environment, before putting it to public use.

Box 5: Regulations impacting AI systems

The Securities and Exchange Board of India (SEBI) has issued a circular on reporting requirements for AI/ML applications and systems.

The National Digital Health Mission strategy identifies a key role of the mission to *"keep a check on the reliability of AI systems by laying down guidelines and standards"*[25] and has created a sandbox to allow products to be tested in a contained environment and evaluate consumer and market reactions to it.[26]

The Personal Data Protection Bill, 2019 has provisions to regulate personal and sensitive data and proposes to establish a Data Protection Authority to prevent misuse of personal data

The *Code of Civil Procedure, 1908* requires a judge to pronounce his judgement after stating the reasons for his finding on each issue. Similarly, administrative authorities and tribunals are required to give '*sufficiently clear and explicit reasons*' in support of the orders made by them, to inspire confidence in their adjudicatory processes.[27] It is likely that the automation of judicial and quasi-judicial functions under Indian law would need to be accompanied by reason-giving and require AI to be explainable.

Area 2: Policy Making

3.16. While Government alone cannot ensure effective operationalizing of the Principles for Responsible AI, it needs to play the lead role. In this regard, its envisaged actions can be categorized under following headings:

  i.   Manage and update the Principles for Responsible AI in India

  ii. Research into technical, legal, policy and social aspects of responsible AI in India

  iii. Enable access to data, responsible AI tools and techniques

  iv. Develop India's (and other emerging economies') perspectives on responsible AI

I. Manage and update the principles for responsible management of AI in India

3.17. NITI Aayog's approach paper on Responsible AIforAll introduced seven Principles by studying various AI use cases in India and around the world. The paper acknowledges that the growing number of use cases requires the principles to adapt and reflect the latest capabilities, risks, policies and legal environment. Some emerging considerations include impact of model training on the environment, the impact on trade and the security implications of AI.

---

25   National Health Authority (July 2020). National Digital Health Mission Strategy Overview

26   National Digital Health Mission (2020). NDHM Sandbox- Enabling Framework

27   The Siemens Engineering and Manufacturing Co. of India Ltd. v. Union of India, AIR 1976 SC 1785.

3.18. In this regard, there is a need for a custodian of responsible AI principles. The custodian shall monitor the responsible AI environment, update the Principles and identify mechanisms to translate them to practice on an ongoing basis. A mechanism for this is identified later in this chapter.

II. Research into technical, legal, policy and social aspects of Responsible AI

3.19. The NSAI highlights the need to incentivise research for harnessing the benefits of AI. As the adoption of AI increases, it is also important to dedicate research efforts towards ensuring AI is beneficial to society. Such research must cover a broad spectrum across social, policy, legal and technology aspects of AI systems and their interaction with individuals and society. Relying on private initiative for areas relating to responsible AI may not be sufficient and national governments as well as international collaboration should be proactive in initiating, funding and supporting such research projects.

3.20. Social research must be aimed at understanding the interaction of AI systems with the local and marginalised communities. This includes understanding how different communities are impacted by the deployment of AI technologies for the delivery of benefits and services, and if benefits are reaching the population as intended, ramifications of risks and considerations such as discrimination, inclusivity, privacy, etc on local and marginalised communities, and identify any other concerns, both in the short term and long term, shaped by the introduction of Artificial Intelligence. This research is further expected to inform the responsible AI principles, guide policies and inform technology research and innovation.
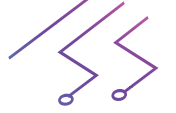
3.21. Policy and empirical research is needed to adapt policies towards AI and technology-driven economies, maximise the benefits and minimise the adverse effects. The approach paper on Principles for Responsible AI identified the need to track changes in the job environment both locally and internationally.[28] While the economic potential of AI is well documented, various studies also warn that AI could create wealth concentration and inequality, and displace less-skilled job roles.[29,30] Education and skilling programs to build human capacity, incentives to encourage reskilling, social safety measures to guard against the malicious use of AI, growth and management of the gig economy, leveraging global AI supply chains, design and relevance of universal basic income are some potential research areas that could inform both short term and long term policy decisions.

3.22. In addition to economic impact, policy research could also be dedicated towards accelerated adoption of responsible AI in India. Policies such as responsible data sharing to enable machine learning, responsible development and deployment framework for AI systems, streamlining public procurement to enable innovative solutions to be procured and scaled, and incentivising research and innovation must also be considered for research. The research outcomes could then inform approaches of the relevant regulators.

---

28 NITI Aayog, (2021). Approach document for India. Part 1- Principles for Responsible AI

29 https://blog.irvingwb.com/blog/2015/04/the-rise-of-the-digital-capital-economy.html

30 https://www.usnews.com/news/articles/2016-10-11/1-in-3-workers-employed-in-gig-economy-but-not-all-by-choice

3.23. The use of AI systems for consequential decision making also raises legal concerns that warrant research. Policies on data ownership involving physical safety, informed consent, confidentiality, and security would be beneficial for identifying liabilities. Identifying high risk use cases, liability and accountability frameworks, IP related considerations for AI innovations, privacy and security considerations with advances in AI across sectors, evolution of law and legal frameworks to account for AI capabilities are potential areas of research.

3.24. The NSAI advocated for the use of technology itself to solve for concerns raised by AI. The various challenges identified through social, policy and legal research could feed into technology research. The sources of demographic data in India have skews that are well documented.[31,32,33] Building robust and reliable ML models with limited data is an upcoming field of research and may be considered in Indian context. Chatbots are increasingly being used in India across sectors to improve user-experience and enhance productivity.[34] According to a Google report, 90% of internet users in India prefer to use vernacular language for searching and other tasks but Indian language content on the internet is abysmally low.[35,36] NLP tools, translational services, multi-lingual datasets could enable inclusive development of the AI ecosystem and accelerate adoption of AI systems.

3.25. The research on responsible AI is, by design, multi-disciplinary. Research in one domain feeds into another. For example, social, legal and policy research must be aware of technology's capabilities and technology research must be informed by the social, policy and legal context. The Government may, therefore, support research in Responsible AI and incentivize cross-disciplinary research. The Government may, either directly or indirectly, support research on responsible AI in the Indian context across technology, legal, policy and social aspects by prioritizing funding opportunities and fellowship programs.

3.26. Research areas that are rewarding for the private sector (such as identification of false and mis-information) may be identified. This will facilitate co-investment and enable leveraging private sector efficiency and international experience and facilitate conversion of research into impact on the ground.

3.27. Responsible AI has gathered attention around the world and there is an increasing recognition for international collaboration. The GPAI has a working group on responsible AI. International alliances and partnerships may be leveraged to facilitate the exchange of multidisciplinary talent, data, and consolidation of research efforts, especially in areas of social good.

---

31 https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/05/GSMA-The-Mobile-Gender-Gap-Report-2020.pdf

32 https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2020/05/GSMA-The-Mobile-Gender-Gap-Report-2020.pdf

33 Sambasivan, N. et al. "Re-imagining Algorithmic Fairness in India and Beyond." ArXiv abs/2101.09995 (2021): n. pag.

34 https://www.livemint.com/companies/news/oracle-sees-uptick-in-adoption-of-ai-enabled-chatbots-in-india-11593937775357.html

35 https://www.thinkwithgoogle.com/intl/en-apac/marketing-strategies/search/year-in-search-2020-india/

36 https://w3techs.com/technologies/overview/content_language

3.28. Academic conferences offer a variety of benefits to researchers, including networking, learning new techniques, recognition for their work. Conferences on responsible AI may be incentivised to be hosted in the country so that challenges and approaches around the world can be studied and motivate indigenous research.

III. Enable access to data, responsible AI tools and techniques

3.29. India has a rich and diverse portfolio of AI efforts by the private sector at various stages of revenue and funding.[37] The Government may play an enabling role by promoting awareness, access and affordability of responsible AI knowledge materials, tools and technologies.

3.30. In this regard, hackathons, workshops, open challenge mechanisms may be used to develop tools and mechanisms that encourage adherence to Principles. Such activities may also be leveraged to introduce responsible AI practices to the community. Existing responsible AI practices may also be compiled and made available to the community. The Government may initiate this by documenting responsible AI practices in the public sector AI deployments.

3.31. Ensuring that AI systems are inclusive and non-discriminatory is important, especially in high-risk use cases. This requires availability of high quality and representative datasets. The digital divide in India makes it difficult to ensure sufficient coverage.[38,39] Lack of reliable proxies also make it difficult to evaluate AI models for fairness.[40] The Government, in its activities, generates a large amount of data across the socio-economic spectrum of the country. But the data is currently not available at the unit level and is published as summary statistics. There is also a lack of consistent adherence to meta-data and data standard.

3.32. The Government may work towards identifying mechanisms to make India-specific and application specific data available for AI/ML research and innovation. To enable the data to be used for machine learning, the data quality considerations may need to go beyond data cleaning and resolve concerns such as data source reliability, missing data, duplicate data, correlated variables, and outliers[41].

3.33. It is important that any policy on access to data balance the competing interest of privacy preservation and harnessing datasets for model fairness and innovation. Data may need to undergo privacy preserving transformations to reduce the sensitivity of data shared. The Government could enable better AI models by supporting such efforts and creating data sharing policies to safeguard citizen interests and promote development of reliable AI models.

---

37 NASSCOM Startup pulse survey 2 | Indian tech startups. On the road to recovery, Nov 2020

38 https://scroll.in/article/824882/missioncashless-few-use-mobiles-fewer-know-what-internet-is-in-adivasi-belts-of-madhya-pradesh

39 https://ceda.ashoka.edu.in/picture-this-how-bad-is-indias-digital-divide/

40 Sambasivan, N. et al. "Re-imagining Algorithmic Fairness in India and Beyond." ArXiv abs/2101.09995 (2021): n. pag.

41 Gudivada, V., Apon, A., & Ding, J. (2017). Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations.

IV. Develop India's perspectives on responsible AI and inform the global point of view

3.34. The perspectives on the ethics of AI are mostly dominated by western concerns and philosophies [42,43]. <mark>As the adoption of AI matures in India and research on social and policy ramifications develops,</mark> the perspectives on responsible AI in India is expected to evolve. In addition, since India shares its socio-economic context with several emerging economies, such perspectives could represent the concerns of 40% of the world (NSAI, 2018).

3.35. These learnings may be shared at international forums to inform the global strategy on responsible AI. The Government may facilitate dialogues on this, through focused research studies and publications. In addition to providing local perspectives, the NSAI recommended leveraging international partnerships(including research partnerships) to solve various challenges for social good through research partnerships. The Government may also identify facilitation mechanisms for such partnerships such as cross border data sharing and the creation of dedicated funds for such collaborations.

Area 3: Awareness & Capacity Building

3.36. The NSAI (2018) highlighted the need for awareness and capacity building within the government. These measures must include responsible AI practices. Government may curate awareness initiatives on AI not only to provide perspectives on the capabilities but also highlight the weaknesses of AI systems and the need for responsible AI practices. Academic experts, industry bodies and independent organizations may be leveraged for needs assessment, development of training curriculum and conduct training programs for public sector officials. The content of the awareness campaigns may also depend on the needs and the role of the stakeholder (Figure 1). The objectives of these programs may include:

- Raising awareness of capabilities of AI in order to ensure that the <mark>expectations from AI are practical and the supporting factors for the success of AI initiatives are well understood</mark>
- Underlining the need for responsible AI for promoting investment into responsible AI practices
- Showcasing industry practices for responsible AI, including governance systems, tools and processes
- Identifying and facilitating the availability of datasets, policy measures and other instruments needed to enable responsible AI in India
- Reducing information asymmetry, trust issues and apprehensions of AI systems and develop skills to identify and think through ethical problems
- Staying abreast of global developments on responsible AI

2.37. In collaboration with concerned stakeholders, course contents may be developed on technology as a whole, enabling factors for adoption and associated risks. This may be made an integral part of all training programmes of different streams of government services at all levels.

---

42   https://iep.utm.edu/ethic-ai/

43   https://www.pri.org/stories/2018-02-16/what-can-ai-learn-non-western-philosophies

3.38. In addition to knowledge and information dissemination, awareness programs may also include case studies, research projects, proofs-of-concept or multi-party consultations in relevant sectors and publicizing examples emanating from an India. States with successful AI deployments may be encouraged to host other states for knowledge transfer. Case studies pertaining to the public sector's adoption of responsible AI may also be documented to create a repository and knowledge base for responsible adoption to scale.

3.39. The stakeholders involved in using the AI technology must be made aware of specific capabilities of the system and the standard operating protocols. It is important that they are also aware of limitations so that human interventions are made at the right time. For example, in technologies such as facial recognition, it is helpful to understand the innate bias that may be exhibited by even the most sophisticated algorithms.[44,45] A sustained awareness program may be needed to gradually shift the behaviour of various stakeholders involved towards effectively use the AI system.

Decision Maker	Procurer/ Influencer	Implementing Agency	User	Impacted Stakeholder
How AI/ML works	How AI/ML works	Standards, guidelines, best practices	Capabilities of a specific AI technology	Awareness of rights
Need for ethical thinking	Identify and anticipate ethical problems	Tools and techniques for responsible AI	Awareness of its limitations and safe usage protocols	Awareness of role, capabilities, limitations of AI
Best practices in procurement	Ability to reason on potential solutions	Grievance redressal mechanisms, SOPs, etc		Awareness of grievance redressal mechanisms
	Ability to communicate ways of addressing the problems			

Fig. 1: Examples of potential topics for various roles. Depending on the needs of the use case and role of the stakeholder, training needs may be different

3.40. It is also important to reach the intended beneficiaries, especially of public sector AI deployment, in sensitive or high-risk cases and other impacted stakeholders of the AI system to understand how the system is perceived, understand issues and gaps in implementation. This will also help facilitate targeted awareness campaigns. These campaigns must ensure stakeholders are aware of their rights and grievance redressal mechanisms. The existing state and local bodies along with regional social organisations may facilitate such programs with necessary support from the relevant Ministry. The strengths and shortcomings of such campaigns must be monitored and evaluated and a mechanism to support this is identified later in the chapter (see Advisory Body needed to guide the various interventions)

---

44   https://indiaai.gov.in/article/webinar-wind-up-mitigating-bias-in-facial-recognition-systems

45   https://jolt.law.harvard.edu/digest/why-racial-bias-is-prevalent-in-facial-recognition-technology

3.41. The effectiveness of awareness campaigns must be reviewed for their strengths, challenges and efficacy in improving the understanding and trust among stakeholders. The learnings from the review must be used to make appropriate corrections in the strategy.

Area 4: Procurement

3.42. Despite its emergence as a crucial element of good governance, the public procurement system in India continues to suffer from several weaknesses. Mired in inefficient monitoring processes, limited accountability and governance, limited awareness, and organizational culture, **initiatives like having model documents have greatly eased the procurement process**, especially in the infrastructure sector. The Government e-Marketplace (GEM) portal has further helped in enhancing the transparency in the procurement system, thereby, establishing groundwork for trust mechanisms.

3.43. However, public procurement for an emerging technology like AI is no mean task and one needs to be cautious against further complicating the process by adding more regulatory layers which would be counter-productive. Emphasis must be given on laying down specific indicators, their measurement techniques, tools and sandboxes through which, based on sectoral use case, AI systems may be adjudged for their trustworthiness. It should also be kept in mind that the process of procurement should not lead to administrative delays or simply exist as a mechanism to issue clearances but must be setup to guide responsible procurement of AI at project level.

3.44. Initiatives like evolving model procurement mechanisms and documents need to be pursued proactively to guide the overall process of procurement and ensure that the interventions are transparent and unambiguous. The issue of liabilities if AI is used in violation of the principles must also be addressed in procurement documents.

3.45. Depending on use case and deployment specifics of the proposed AI (or emerging technology) project, an institutional mechanism, similar to expert advisory committees that are constituted for complex projects, may be formulated to ensure that proposed projects are designed, developed and operated in adherence to the responsible AI principles. The composition of this body may include experts relevant for the use case- such as computer science, data science and machine learning experts, domain experts, legal experts, social science experts etc.

Advisory Body needed to guide the various interventions

Facilitating operationalization of a trusted responsible AI ecosystem

3.46. The Government already has an extensive machinery dedicated to the four areas of interventions mentioned in this Chapter. India is at a relatively nascent state of AI maturity and creating parallel structures for these tasks may not be necessary. However, the capacity of extant Government mechanisms must be augmented to take responsible AI considerations into their purview. The unpredictable nature of AI growth and emerging areas of impact (ex: impact on ecology and environment) requires an evolving mechanism for

frameworks, guidelines and benchmarks and liaison with regional, industry and global best practices.

3.47. In this regard, an advisory body with multi-disciplinary expertise is proposed to strengthen and advice the existing Government machinery, driving convergence across sectors and States. The body should endeavor to provide overall guidance and uniformity in approach while at the same time avoid unnecessary barriers and centralization.

3.48. An advisory body at the apex level should be set up as an independent, multi-disciplinary and highly participatory entity and provide a forum for all stakeholders to have a representation. This will enable accounting for the advances in the field and incorporate perspectives of various stakeholders of the AI ecosystem. This could co-exist with the sectoral instruments that can continue to oversee systems involving AI within their regulatory regime.

3.49. **The remit of such a body may go beyond AI and cover the entire digital space with a focus on key sector specific use-cases**. This is important as AI exists in an ecosystem of other emerging and established technologies. In addition, risks are being identified in other emerging technologies as well. For example, internet-of-things (IoT) applications are being considered for critical scenarios like crisis warnings and public safety, with systems needing to ensure reliability and integrity to be effective[46]. Augmented and Virtual Reality (AR/VR) devices must consider ethical implication of data collection, location tracking, privacy, etc.[47]

3.50. Further, the proposed expert advisory body must be an independent technology wheelhouse advising relevant Government agencies. It should be autonomous to work with individual regulators and Ministries to help draft legislations for AI powered innovations wherever the need arises.

Box 4: Approaches from around the world

The approach for oversight of AI around the world has primarily been through institutionalisation of an independent advisory body to inform governance.

The Centre for Data Ethics and Innovation (CDEI) in the United Kingdom has been established as an advisory body to provide the Government with access to independent, impartial and expert advice on the ethical and innovative deployment of data and artificial intelligence.[48]
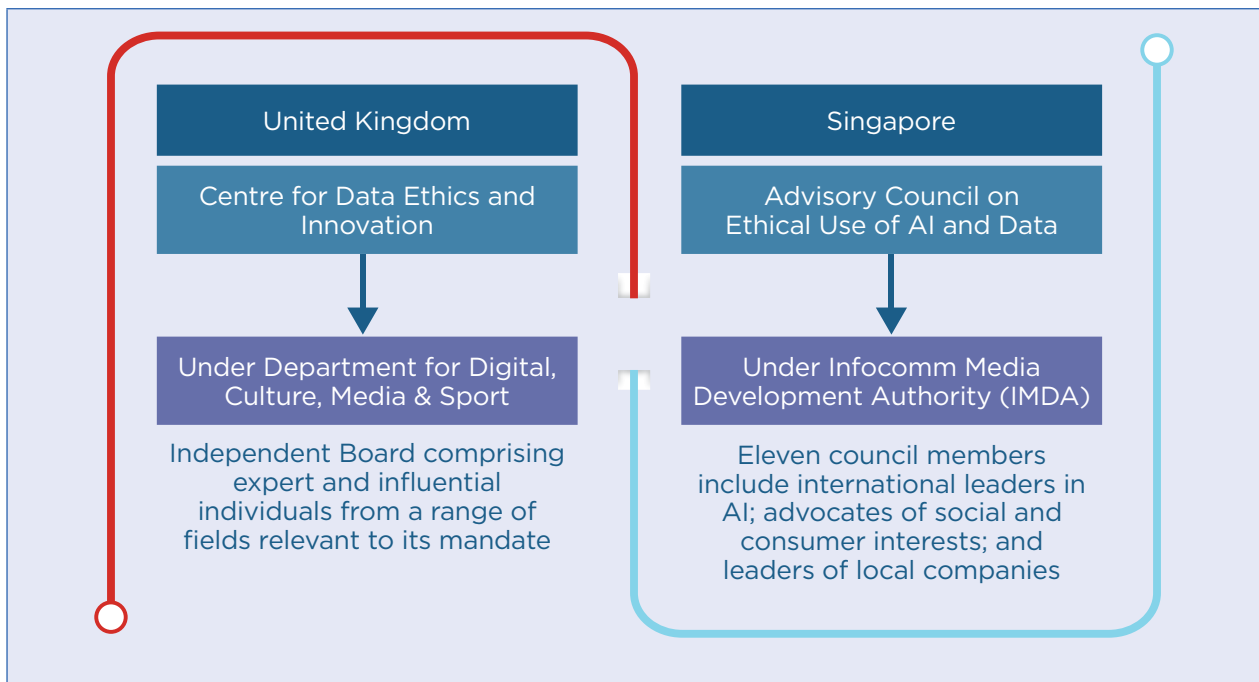
Singapore's Advisory Council on Ethical Use of AI and Data has been set up to advise and work on the responsible development and deployment of AI.[49]

---

46 Digital India Action Group- Whitepaper. Internet of Things (IoT) for Effective Disaster Management

47 Nishith Desai Associates (September 2019). Augmented, Virtual and Mixed Reality– A Reflective Future. Strategic, Legal, Tax and Ethical Issues

48 https://www.gov.uk/government/publications/framework-agreement-between-the-department-for-digital-cul-ture-media-sport-and-the-centre-for-data-ethics-and-innovation/framework-agreement-between-the-depart-ment-for-digital-culture-media-sport-and-the-centre-for-data-ethics-and-innovation

49 https://www.imda.gov.sg/news-and-events/Media-Room/Media-Releases/2018/composition-of-the-advisory-council-on-the-ethical-use-of-ai-and-data

Fig. 2: Institutional advisory body to guide responsible AI ecosystem in United Kingdom and Singapore

3.51. Keeping in mind what is envisioned as an independent and empowered think tank interfacing across various ministries and state departments, **a Council for Ethics and Technology (CET)** is proposed for India.

3.52. Given the mandate to enable preparedness for AI and emerging technologies along with driving innovations in a responsible manner, it is recommended that CET have the following composition:

a. Computer Science and AI experts,

b. Legal experts,

c. Relevant sectoral experts,

d. Civil societies,

e. Humanities and Social Science experts

f. Private sector and industry representatives

g. Environmental expert

h. National Security expert

i. Cybersecurity expert

j. Representatives from standard setting bodies

with the option of coopting of additional experts as and when the need arises.

3.53. The formulation of CET must take into cognizance the sectoral regulators' roles and be complementary to and in conjunction with the same to ensure CET isn't just another layer of unnecessary supervision hampering innovation. Since CET's mandate is envisioned to be multi-faceted, reducing bureaucratic hurdles while guiding the implementing hands of sectoral regulators via

ethical and unbiased implementation will be a delicate balance that the advisory body is envisioned to withhold.

3.54. In order to ensure effective functioning, the CET may consider forming sub-groups for emerging technologies of interest and evaluate ethical considerations arising from their usage. In addition, sectoral sub-groups could also be considered on similar lines.

3.55. The CET may also function as a knowledge hub on policy matters by publishing policy papers and promoting any such activities towards realizing the benefits of AI while managing its risks. It may monitor and coordinate policy approaches across sectoral regulators to avoid duplication of effort, and prevent the enactment or operationalization of inconsistent policies. It may serve to support existing authorities with identifying use cases and defining policies, benchmarks and relevant rules and guidelines. It may also support the policy initiatives specified in Chapter 3 and 4 and advise various Ministries and States towards protecting individual interests and enabling responsible AI research and innovation.

3.56. States have varying degrees of AI adoption and responsible AI strategies and roadmap must reflect the relative AI maturity of the state. States such as Telangana and Tamil Nadu have identified policies for responsible AI. Telangana AI framework recognises the need for governance and has identified a working committee with multidisciplinary expertise to develop guidelines for AI use cases.[50] The Government of Tamil Nadu released 'Safe and Ethical Artificial Intelligence Policy' that identifies a framework for evaluation of AI systems before roll out.[51] In order to ensure that state specific considerations are addressed, the CET may leverage learnings from individual states and develop guidelines for the constitution and mandate of State specific committees. While it is crucial that individual states identify policy actions depending on the regional needs and relative maturity of the AI ecosystem, it is also important to have convergence and consistency in AI policies to enable innovation to scale across the country and also prevent exploitation of policy gaps in certain states. In this regard, the CET may play the role of fostering "cooperative federalism" between center and the states.

3.57. The CET may be also be tasked with driving convergence across stakeholders, and leverage the experience of existing initiatives around ethics and technology, like the ones undertaken by SEBI, Indian Council of Medical Research and NDHM, creating model guidelines as well as ethics review mechanisms that other Ministries, States and private organizations may build upon.

---

50 https://it.telangana.gov.in/wp-content/uploads/2020/07/Govt-of-Telangana-Artificial-Intelligence-Frame-work-2020.pdf

51 https://tnega.tn.gov.in/assets/pdf/TN_Policy_for_Safe_and_Ethical_AI.pdf

Center for Ethics and Technology			
Subgroups for individual emerging technologies		Subgroups for individual sectors	
Regulatory	Policy	Procurement	Awareness
Work with sectoral regulators to drive convergence and identify risk based regulatory interventions Collaborate with standards agencies to identify relevant standards and benchmarks for Indian context	Work with MeitY, NITI Aayog, MoE, MHRD on policy interventions towards societal goals Work with individual Ministries on relevant sector specific policies	Create model documents for RFP, SOPs for EAC by leveraging learnings from individual AI system procurement in India and around the world	Create a capacity building program on ethics and technology in collaboration with DoPT, MeitY and NITI Aayog

Fig. 3: Conceptual framework for operations of the CET. The CET shall co-exist and collaborate with the existing Government instruments
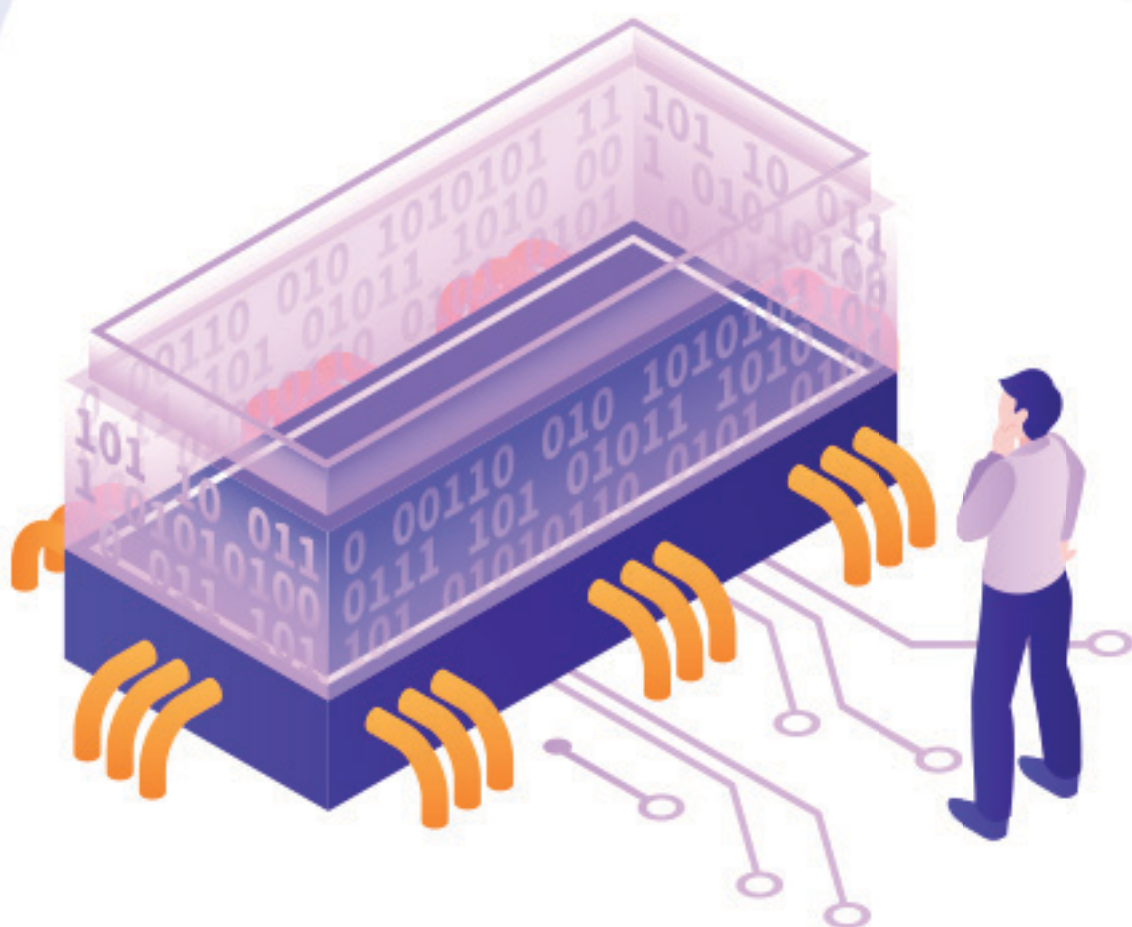
Actions for the Private Sector and Research Institutions

This Chapter explores the institutional mechanisms for operationalizing the principles of responsible AI across private sector and research institutions.

4.1 Private Sector

4.1.1. India has a vibrant private sector ecosystem of AI, with over 950 startups focused on AI. The number of startups has seen significant growth recently, with a 5-year CAGR of 45-50% in 2020.[52] During the COVID-19 pandemic, over 40% of deep-tech solutions for COVID leveraged AI.[53] A report by NASSCOM suggests that data and AI have the potential to add $ 450-500 billion to India's economy by 2025.[54]

4.1.2. There is also an increasing market demand for responsible AI practices. A survey by NASSCOM shows that trust is essential for enterprise-wide adoption of AI. 88% of the respondents identified the need to address AI ethics-related concerns in their risk management framework.[55]

4.1.3. Globally, both monetary and non-monetary benefits of responsible AI have been acknowledged, with responsible AI is being seen as a competitive advantage. Improved data privacy and security practices increases the trust in an organisation and boosts the availability of data. Inclusive and non-discriminatory practices of an AI model, allows user-profiles across a wider demography to be served efficiently. Interpretable AI helps identify use cases and improves product quality.[56]

4.1.4. As mentioned in the previous chapter, it is also important for organizations to prioritize and commit to responsible AI practices. Awareness on the need for responsible AI and associated risks of non-adherence is important to drive commitment towards good organisational practices. Industry-led and collaborative workshops, conferences and knowledge sharing seminars may

---

52 NASSCOM Startup Report 2020

53 NASSCOM Startup pulse survey 2 | Indian tech startups. On the road to recovery, Nov 2020

54 https://nasscom.in/knowledge-center/publications/unlocking-value-data-and-ai-india-opportunity

55 https://nasscom.in/knowledge-center/publications/can-enterprise-intelligence-be-created-artificially-survey-in-dian

56 "Why addressing ethical questions in AI will benefit organizations", Capgemini Research Institute; "Staying ahead of the curve The business case for responsible AI", a report by The Economist Intelligence Unit

be leveraged to raise awareness about the risks and best practices. Industry bodies and Government may facilitate the creation of open materials, tools and technologies, sharing of such tools with the ecosystem.

4.1.5. Additionally, standards and guidelines may provide a general direction for responsible AI behaviour. Community engagement should also be considered for absorption of best practices and sensitization of risks. Internal ethics boards, self-assessment guides and external audits could be leveraged as mechanisms for private sector enforcement. A few examples of toolkits for responsible AI in the private sector are provided in Appendix-1.

4.1.6. Thinking through ethical considerations requires a multi-disciplinary and multi-stakeholder perspective. Till the time guidelines, standards and benchmarks are in place, the private sector may be encouraged to use responsible AI principles as a starting point and collaborate with multi-disciplinary stakeholders (social sector experts, legal experts, representatives of end users who may be impacted, etc.) and relevant organisations (civil society, research institutions, etc.) to effectively identify and address the risks.

Incentivizing and enabling ethics-by-design

4.1.7. **Mandate responsible AI practices in Government Procurement.** The NSAI noted that the Government should play a major role in the procurement of AI systems. Most of the AI systems currently used by the Government in its projects and initiatives have been developed in collaboration with the private sector. By mandating the institutionalization of responsible AI practices in public sector procurement, the Government could create a demand for such practices and boost the adoption of ethics-by-design practices in the country. The NSAI also recommends that the Government guide AI innovations through Moonshot Challenges. Support to such challenges could be conditioned on to the participating entity adopting responsible AI practices.

4.1.8. **Government may mandate responsible AI practices for high-risk AI use cases.** The identification of high-risk use cases may be done by the CET in consultation with the sectoral authorities. This will also create an ecosystem of trust and enable export of Indian AI innovations to the global market

Compliance mechanisms

4.1.9. Compliance with responsible AI standards and guidelines has sometimes raised concerns in terms of increasing cost and creating a barrier to entry for start-ups. However, start-ups around the world have found unique ways to manage such costs. Some of the practices adopted by start-ups include,

a. Assigning accountability for responsible AI to a member of the leadership team;

b. Leveraging online courses, workshops, open materials so the entire team is aware of the risks and develops the skill to ask the right questions;

c. Leveraging open tools and techniques to ensure adherence

4.2 Research and Educational Institutions

4.2.1. The impact of AI research on society either in the present or in the future has gained significant attention around the world. AI institutes around the world have identified an institutional mechanisms for ensuring that research is conducted in a responsible manner. Such mechanisms start with the introduction of ethical reasoning in the curriculum.

4.2.2. The development of curriculum and the best mechanisms to deliver responsible AI courses must be explored. Foundational courses in AI are already being introduced in the secondary and senior secondary curricula in India.[57] The ethical aspects of AI may be introduced in these courses so that need for responsible AI is recognised at a nascent stage. Graduate and post-graduate programs on AI may include a further training on the subject so that the skills needed to identify and anticipate ethical issues are developed and the students are trained to identify effective ways of addressing them. Such courses may be included in the model curriculum and should not be limited to just the technical aspects but must also explore social considerations, including such considerations that vigorously debate the creation of these technologies. In this regard, both standalone courses and embedded modules in computer science programs are being explored around the world[58].

4.2.3. Institutions offering engineering degrees along with social sciences, philosophy, humanities studies are limited in India. In 2019, the All India Council for Technical Education (AICTE) issued approval for engineering colleges to provide courses in humanities and the arts.[59] This move could boost cross disciplinary courses in the engineering curriculum. Individual institutions could be incentivised to document their approaches and learnings for others to leverage. A common cross-disciplinary curriculum on responsible AI may also be provided through SWAYAM and NPTEL online courses to make them accessible in universities where relevant multi-disciplinary faculty is not be available. Cross-university collaboration and guest lectures may also be considered to augment pedagogy in such universities.
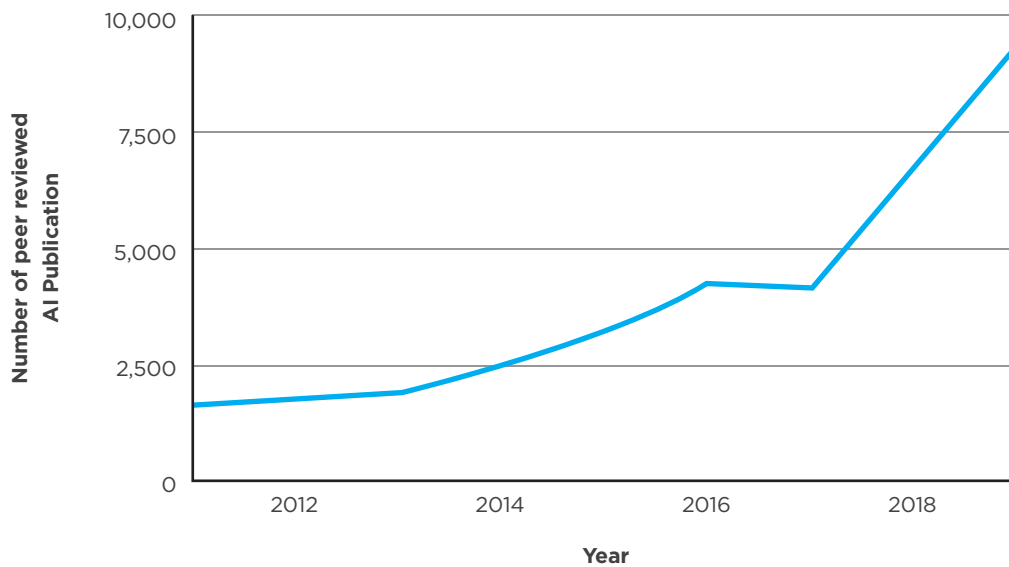
Responsible AI practices in research

4.2.4. Research on AI in India has shown steady growth in the past decade. The number of peer-reviewed AI publications has grown by over six-fold in the last decade. In the last five years, the number of publications in arXiv, the online repository of electronic preprints and post-prints, has grown almost five-fold. It is important for us to now start thinking about responsible AI practices in research.

---

57 https://ncert.nic.in/pdf/syllabus/CSHSS.pdf; http://cbseacademic.nic.in/web_material/CurriculumMain22/SrSec/Computer_Science_SrSec_2021-22.pdf;
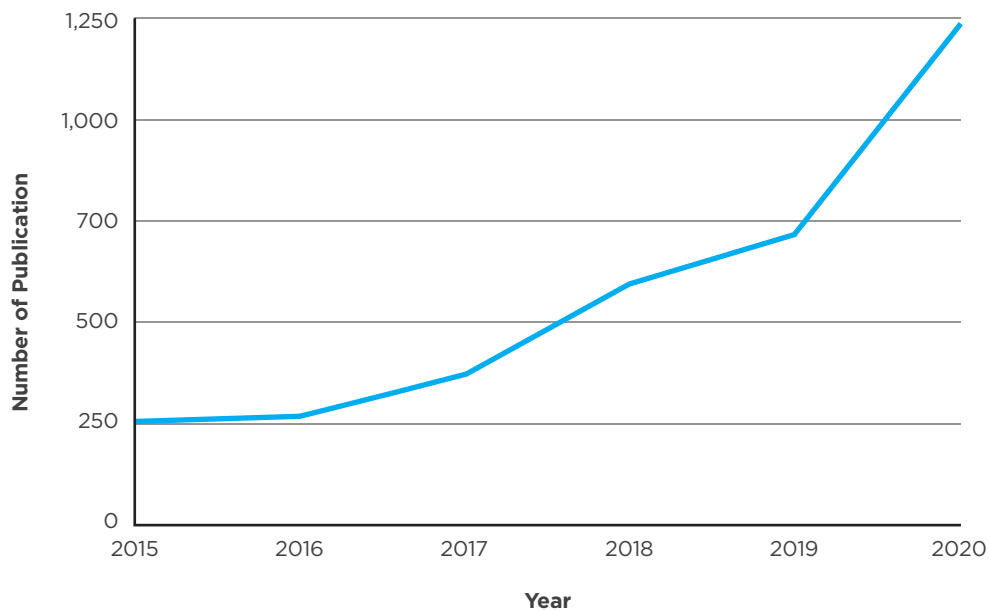
58 https://dl.acm.org/doi/10.1145/3330794

59 https://www.hindustantimes.com/education/students-can-now-pursue-humanities-alongside-engineering-degrees/story-qZO0r9Qe8LyrwB6gYb3daO.html

Fig. 4: Peer-reviewed AI publications from India 2010- 2019

(Source: Elsevier/ Scopus | Data from Stanford AI Index 2021)

Fig. 5: AI-related publications on arXiv from India 2015- 2020

(Source: arXiv | Data from Stanford AI Index 2021)

4.2.5. The ethical guidelines and enforcement structures for research in India are mostly limited to clinical and biomedical research. These structures include the creation of an ethics committee in research institutions. NSAI (2018) had highlighted the need for ICTAI and CORE to include ethics councils to ensure and institutionalize responsible practices. In Universities around the world Institutional Review Boards (IRB) play the role of ensuring that research follows ethical principles. Some research institutions in India already include a review board and could be augmented to review AI research. The current peer-review mechanism within institutes may also be reinforced with reviewers across humanities and social sciences. In Institutions where relevant skills are not available, cross-university collaborations may be considered.

4.2.6. In 2020, the Conference on Neural Information Processing Systems (NeurIPS) mandated all paper submissions to include *"a statement of the potential broader impact of their work, including its ethical aspects and future societal consequences"*[60], a move that started a debate in the research community. While this could incentivise AI researchers to improve their understanding of the broader consequences of their research and improve cross-disciplinary collaboration, concerns have also been raised on the complexity of, and the lack thereof, clear mechanisms for determining the impact of AI solutions. In 2021, NeurIPS released ethics guidelines to assist the researchers and include a provision for reviewers to flag submissions for ethics review.

4.2.7. The practice of including the impact of research and innovation is also practised by certain state funding agencies.[61] The Government may monitor the effectiveness of such approaches and consider requiring a statement of impact in all Government AI research funding and AI fellowship opportunities. It may be useful to formulate guidance on reliably evaluating the impact of research and the expertise of CET may be leveraged in this regard. A platform can be provided to enable stakeholder consultations centering around the issues relating to responsible AI in research, best practices, the identification of new areas for research for promoting responsible AI, etc.

---

60 https://neurips.cc/Conferences/2020/CallForPapers

61 Prunkl, C.E.A., Ashurst, C., Anderljung, M. et al. Institutionalizing ethics in AI through broader impact requirements. Nat Mach Intell 3, 104–110 (2021)

Conclusion

Drawn up across two distinct documents, the strategy for Responsible AI consolidates several best practices to ensure that AI solutions are socially conscious and travel beyond the digital divide. The strategy builds upon the pervasive approach of AI for All, first discussed in the NSAI, to bring under its ambit, an accountable and utility-maximising approach to deploying AI solutions. The essence of AI for All includes within itself the maxims, Good AI for All, and AI for Good, which the strategy for Responsible AI sets to work on.

In Part I of the strategy, the focus was on acknowledging the risks and considerations that require addressing in the pursuit of responsible AI. To respond to these challenges, several guiding principles were recognised as a means to navigate these considerations and to set the narrative on accountable, transparent and beneficial AI. These principles seek to strengthen the Indian AI ecosystem's commitment to privacy, security, equality, inclusivity and non-discrimination, accountability, transparency, and safety.

This paper – Part II of the strategy – sheds light on the manner in which these principles can be operationalised and enforced within the AI ecosystem. The interventions described, and requiring the attention of the government, the private sector and research institutions, are set to bring about a paradigm shift in AI-related policymaking, moving governance practices from risk-agnosticism to a risk-based approach regulation. The paper's timely contributions in this regard are critical: AI must be subject to such scrutiny that befits the risk it undertakes; innovations should flourish, while the likelihood of harm should be minimised.

The mechanisms outlined in this paper seek to achieve this balance between innovation and responsibility. Sandboxes and controlled deployments will control for malicious AI at an early stage. Standards and benchmarks evolved in cognisance of Indian socio-economic and cultural factors will be more responsive to uniquely Indian challenges, such as adherence to the rights outlined in the Constitution of India or addressing the extant digital divide in parts of the country. Research on these subjects can achieve dynamic decision-making for novel challenges in AI, ensuring that forthcoming risks or considerations are not met with laggard policy responses.

A significant task entailed in bringing about responsible AI involves bridging sectoral and regional gaps to drive a coordinated response to challenges arising out of AI. A

multidisciplinary apex level advisory body like the proposed CET is poised to resolve for this concern, and possesses tremendous potential for good. With time, a robust and expert CET will not only unlock uniform appropriate and necessary standards for harnessing and governing AI solutions in India, but its research capabilities may inform discourse on development of AI at a global level.

It is also important to inculcate attitudes promoting responsible AI among private sector players and academia, given the crucial positions they hold in the overall ecosystem. By recommending mandatory adherence to the principles for high-risk AI and AI procured by the government, this paper seeks to narrow the margin for error or malice among AI used to perform sensitive functions while ensuring that innovation and utility is encouraged. Similarly, by recommending that government-funded research incorporate tools of impact assessment, this paper commits to enhancing the welfare-capacity of AI solutions.

The takeaways contained in this paper respond to the current challenges faced by at-scale adoption of AI systems in India and lay down the first steps to be taken in adequately addressing these challenges, especially when India is rapidly establishing itself as a hub of AI innovation. Implementing these measures and adopting an enabling framework for implementing responsible AI principles will contribute meaningfully towards unlocking AI for All.
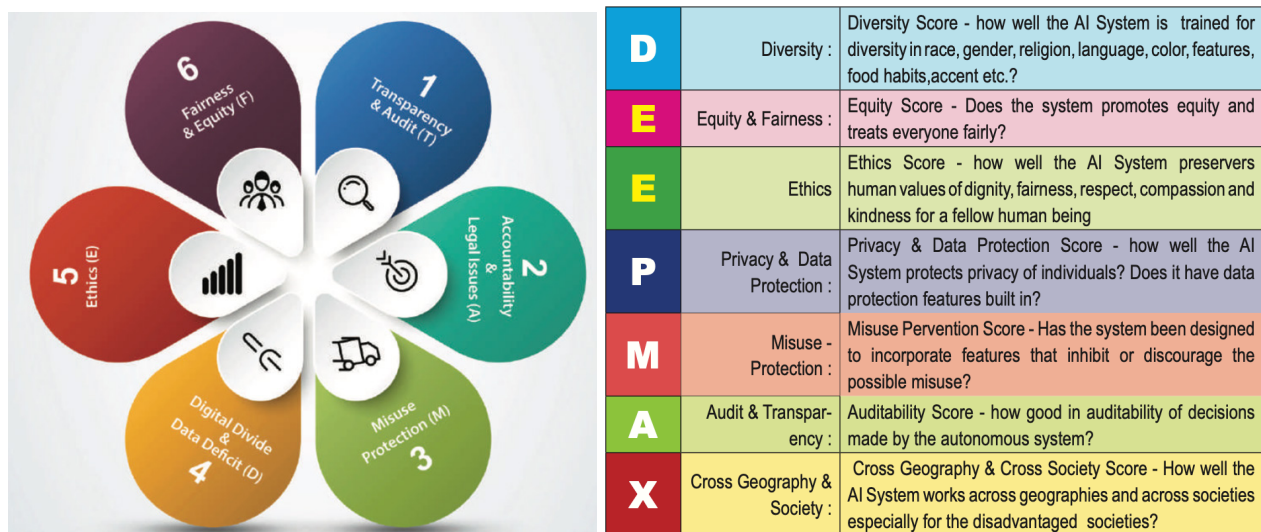
Example responsible AI frameworks to evaluate AI systems and identify governance mechanisms

DEEP-MAX Scorecard, Government of Tamil Nadu

The Tamil Nadu Government issued a "Safe and Ethical Artificial Intelligence Policy" in 2020 to guide implementation and deployment of AI systems in the state. The policy identifies a six-Dimensional **TAM-DEF** Framework along with **DEEP-MAX** Scorecard for evaluating all AI Systems before public roll out

D	Diversity :	Diversity Score - how well the AI System is trained for diversity in race, gender, religion, language, color, features, food habits,accent etc.?
E	Equity & Fairness :	Equity Score - Does the system promotes equity and treats everyone fairly?
E	Ethics	Ethics Score - how well the AI System preservers human values of dignity, fairness, respect, compassion and kindness for a fellow human being
P	Privacy & Data Protection :	Privacy & Data Protection Score - how well the AI System protects privacy of individuals? Does it have data protection features built in?
M	Misuse - Protection :	Misuse Pervention Score - Has the system been designed to incorporate features that inhibit or discourage the possible misuse?
A	Audit & Transpar-ency :	Auditability Score - how good in auditability of decisions made by the autonomous system?
X	Cross Geography & Society :	Cross Geography & Cross Society Score - How well the AI System works across geographies and across societies especially for the disadvantaged societies?

Fig. 6: TAM-DEF framework and DEEPMAX scorecard
for evaluating AI systems before roll-out

IEEE

IEEE P7000™, 'IEEE Standards Project for Model Process for Addressing Ethical Concerns During System Design'[62] has been developed to provide engineers and technologists with an implementable process aligning innovation management processes, IT system design approaches, and software engineering methods to minimize ethical risk for their organizations, stakeholders and end-users. There are

---

62   https://ethicsinaction.ieee.org/p7000/

four pathways by which these standards can be assimilated by stakeholders in their design and development journey:

1. ***AI and Ethics in Design*** are ten courses aimed at creators of ethically aligned design.

2. ***AI Ethics Glossary*** features more than two hundred pages of terms that help provide a common understanding and terminology of AI ethics to multidisciplinary teams.

3. ***The Open Community for Ethics in Autonomous and Intelligent Systems (OCEANIS)*** is a global forum for discussion, debate, and collaboration for organizations interested in the development and use of standards to further the creation of autonomous and intelligent systems.

4. ***The Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)*** has the goal to create specifications for certification and marking processes that advance transparency, accountability, and reduction in algorithmic bias in autonomous and intelligent systems.

World Economic Forum

The World Economic Forum has developed an Oversight Toolkit for Boards of Directors[63]. The ethics module[64] outlines five tools to help a board of directors oversee the setting of ethics standards and the establishment of an ethics board.

1. ***The AI ethics principles development tool*** helps boards of directors and AI ethics boards develop an AI ethics code.

2. ***AI Ethics Board Goals and Guidance tool*** provide questions to consider before establishing an AI ethics board.

3. ***AI Ethics Board Member Selection tool*** for selecting the members of the AI ethics board suggests requirements to consider when appointing members to the AI ethics board.

4. ***AI Ethics Code Assessment tool*** - Assessing the draft AI ethics code provides questions to help directors evaluate the draft code presented by the AI ethics board.

5. ***Implementation, Monitoring and Enforcement tool*** - Assessing implementation, monitoring and enforcement of the AI ethics code include questions to help boards evaluate whether they are receiving the information they require to carry out their oversight responsibilities and whether the management team of the AI ethics board is effectively carrying out these responsibilities.

Chatbots RESET[65]:

A framework for governing responsible use of conversational AI in healthcare by bringing together chatbot developers, chatbot platforms, the medical community, civil society, academia and healthcare regulators.

---

63 https://spark.adobe.com/page/RsXNkZANwMLEf/

64 https://wef-ai.s3.amazonaws.com/WEF_Empowering-AI-Leadership_Ethics.pdf

65 https://www.weforum.org/reports/chatbots-reset-a-framework-for-governing-responsible-use-of-conversation-al-ai-in-healthcare

The framework consists of two parts:

1. A set of principles selected by the multistakeholder community to govern the use of chatbots in healthcare. The principles have been drawn from AI ethics principles and healthcare ethics principles and interpreted specifically for the use of chatbots in healthcare applications.

2. Actions that stakeholders can take to operationalize the principles in various stages of the use of chatbots in healthcare.

The framework has been developed with three types of stakeholders in mind: Developers, providers and regulators and provides recommendations for actions to be performed during three operationalization stages: 1. Develop 2. Deploy 3. Scale.

Because of the different types of risk levels involved in the use of different types of chatbots, the operationalization actions of the framework are not equally applicable across the spectrum of chatbots. To address this diversity of risk levels, the framework includes a preliminary classification of Chatbots into four types (Types I, II, III, or IV) based on the severity of the healthcare condition and the significance of the information provided by the chatbots to healthcare decisions.
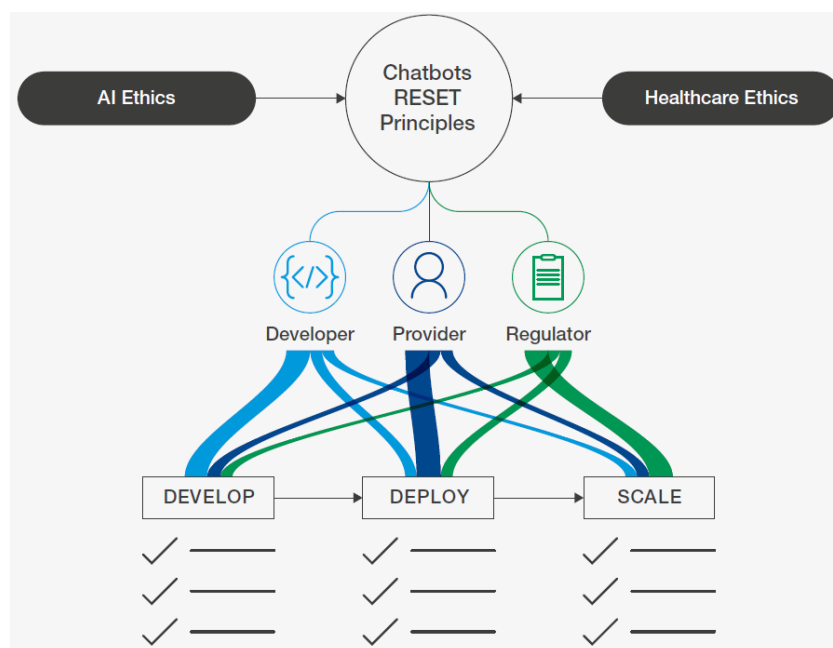
Fig. 7: World Economic Forum: Chatbots RESET[66]

---

66   http://frwww3.weforum.org/docs/WEF_Governance_of_Chatbots_in_Healthcare_2020.pdf

Global Practices towards a risk-based approach to regulating AI

There have been various international bodies that have proposed guidelines and frameworks using a risk-based approach to govern AI for varying applications across sectoral use cases.

Germany

The German Data Ethics Commission recommends adopting a risk-adapted regulatory approach to algorithmic systems (shown below).
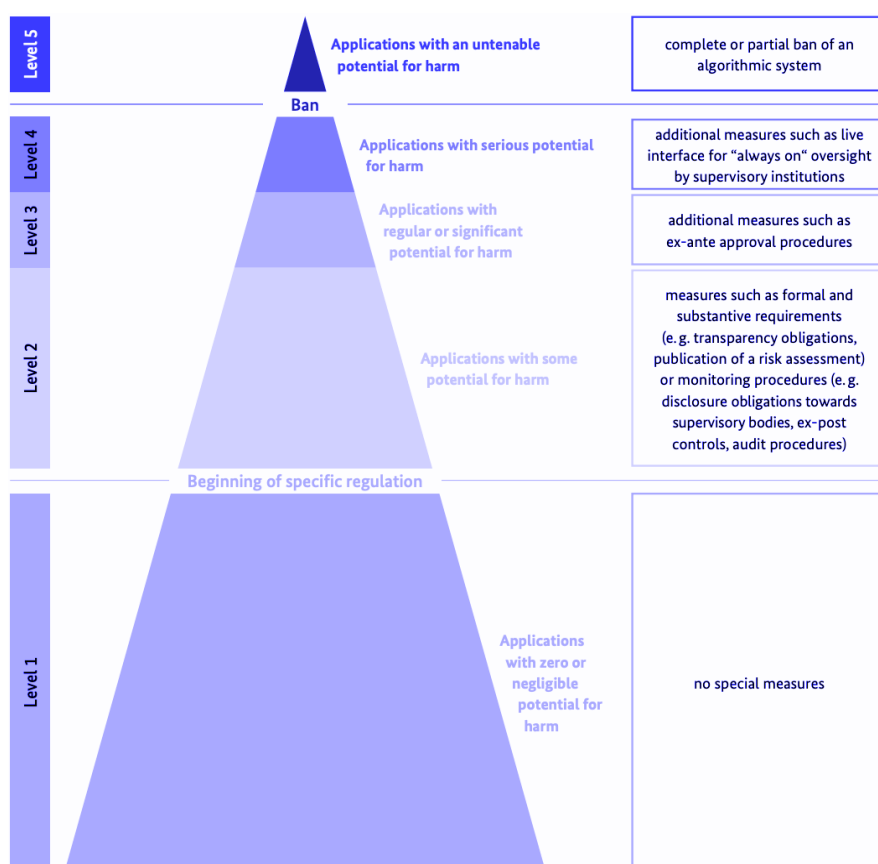

Fig. 8: Criticality pyramid and risk-adapted regulatory system for the use of algorithmic systems

The principle underlying this approach should be as follows: greater the potential for harm, more stringent the requirements and the more far-reaching the extent of regulatory intervention. When assessing this potential for harm, the sociotechnical system as a whole must be considered, or in other words all the components of an algorithmic application, including the people and data involved, from the development phase right through to its implementation in an application environment and any evaluation and adjustment measures.[67]

European Union

The European Commission in its white paper titled Artificial Intelligence - A European approach to excellence and trust, recommends that a given AI application should generally be considered high-risk in light of what is at stake, considering whether both the sector and the intended use involve significant risks, in particular from the viewpoint of protection of safety, consumer rights and the fundamental rights.[68]

On 21 April 2021, the European Commission published its proposal for a Regulation on Artificial Intelligence. The regulation follows a risk-based approach, differentiating between uses of AI that create (i) unacceptable risk, (ii) high risk, and (iii) low or minimal risk.[69] Whether an AI system is classified as high-risk depends on its intended purpose of the system and on the severity of the possible harm and the probability of its occurrence. The proposal provides that high-risk AI systems need to respect a set of specifically designed requirements and lays down a ban on a limited set of uses of AI that contravene European Union values or violate fundamental rights. Under the proposed regulation, other uses of AI systems are only subject to minimal transparency requirements.[70]
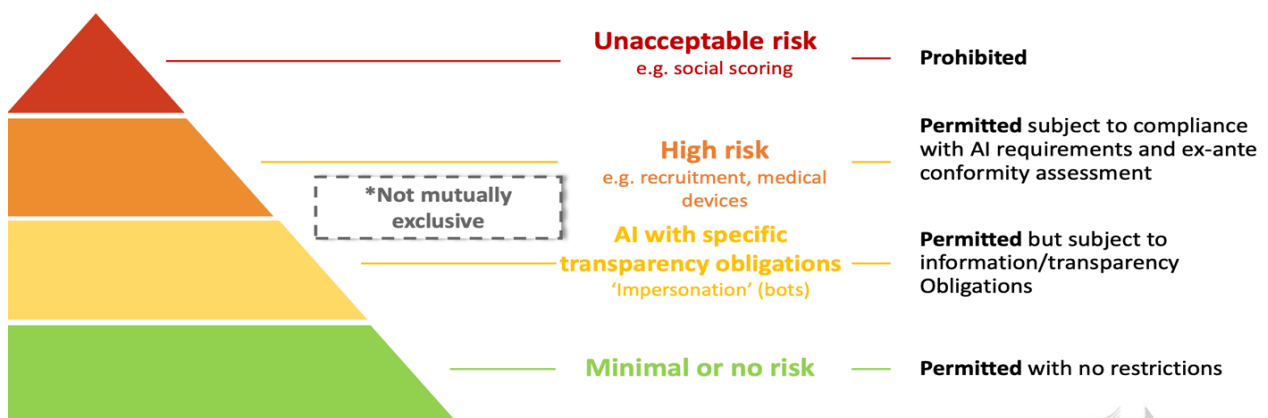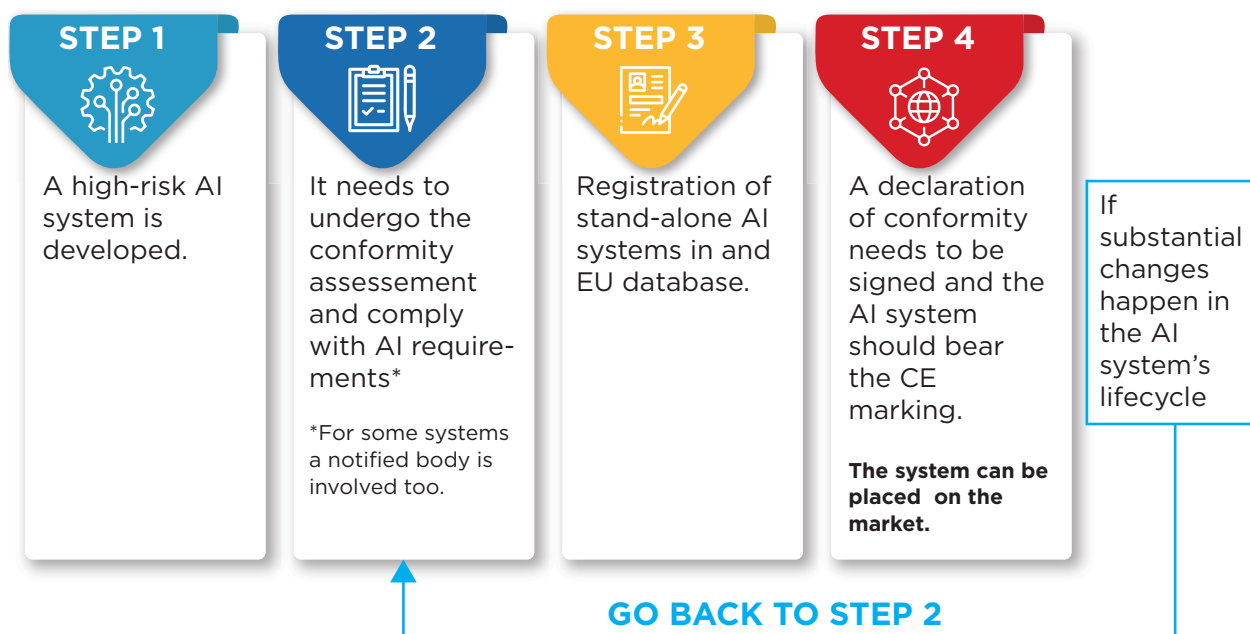

Fig. 9: Risk based approach for AI regulations

67 https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf;jsessionid=-0F3AEDD276064F891DC87DBC08CB473A.1_cid334?__blob=publicationFile&v=2

68 https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

69 https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

70 https://digital-strategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificial-intelligence

Fig. 10: Practice for providers of high-risk AI systems

United States of America

The U.S. Food and Drug Administration (FDA) issued the "Artificial Intelligence/ Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" from the Center for Devices and Radiological Health's Digital Health Center of Excellence.[71] The paper leveraged practices from the International Medical Device Regulators Forum's risk categorization principles.

Australia

Australia's ethics framework for AI[72] examines the probability of risk, -along with the consequences of such risks - via suggestive frameworks like the one shown in the table below. When a risk has both a high probability of occurring and carries with it, the possibility of an increased number of negative outcomes, the consequences become more severe.

Likelihood of risk	Consequence				
	Insignificant risk	**Minor risk**	**Moderate risk**	**Major risk**	**Critical risk**
Rare	Low	Low	Moderate	High	High
Unlikely	Low	Moderate	Moderate	High	Extreme
Possible	Low	Moderate	High	High	Extreme
Likely	Moderate	High	High	Extreme	Extreme
Almost certain	Moderate	High	High	Extreme	Extreme

Fig. 11: Example Risk Assessment Framework for AI Systems[73]

---

71 https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learn-ing-software-medical-device

72 https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-eth-ics-framework

73 https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf