aapti institute

Downstreaming Responsible Artificial Intelligence in India: A White Paper

This research was conducted by **Aapti Institute**, a public research institute that works on the intersection of technology and society. Aapti examines the ways in which people interact and negotiate with technology both offline and online.

Author: Vaishnavi Patil
Project Guidance: Kunal Raj Barua

Report design: Meher Rajpal | **Cover illustration:** Monjira Sen

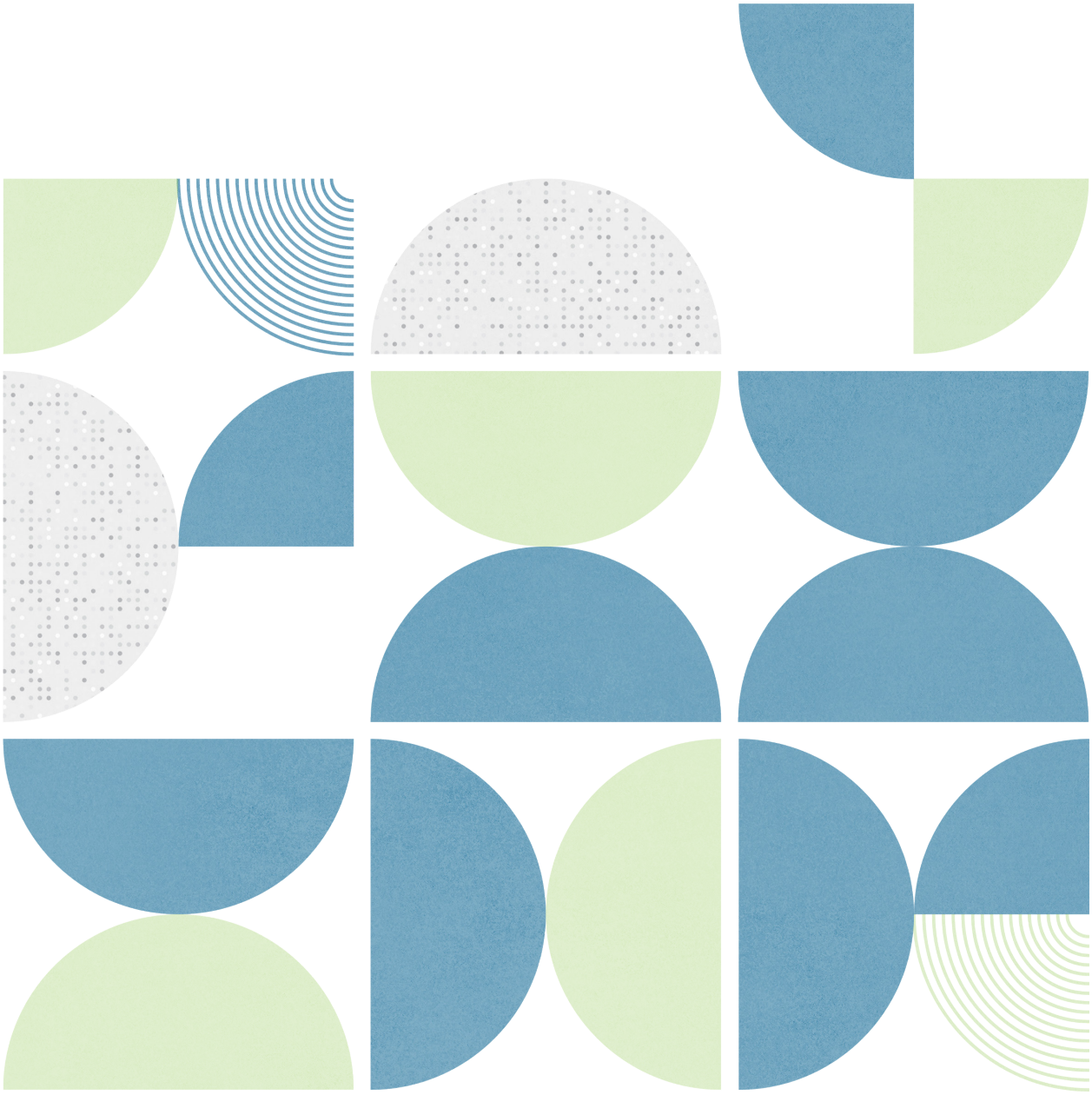
ACKNOWLEDGEMENTS

In addition to contributions from the wider Aapti team, this report also draws upon the expertise of numerous academic and industry experts, practitioners.

NAME	DESIGNATION	ORGANISATION
Varun Agrawal	Director of AI Safety Fellowships	Impact Academy
Leslie D'Monte	Author and Writer	The Mint
Sidharth Deb	Associate Director of Public Policy	The Quantum Hub
Anwesha Sen	Assistant Programme Manager	The Takshashila Institution
Lahar Appaiah	Counsel	IBM India/South Asia
Venkatesh Hariharan	India Representative	Open Invention Network
Vikram Vincent	Founder and CTO	FreedomThought Labs
Apar Gupta	Advocate and Founder Director	Internet Freedom Foundation
Minesh Mathew	ML Scientist	Wadhwani AI

We are grateful for their input during interviews and feedback sessions.

Downstreaming Responsible Artificial Intelligence in India: A White Paper



aapti institute

Table of Contents

About the White Paper

This white paper explores the principles, frameworks, and strategies required to advance Responsible AI (RAI) in India. It highlights the importance of balancing innovation with regulation, addressing societal concerns, and fostering collaboration among policymakers, industry leaders, academia, and civil society. By prioritising fairness, transparency and accountability, privacy, and safety and security, India has the opportunity to establish itself as a global leader in Responsible AI, setting benchmarks for ethical technology development in the Global South while ensuring equitable benefits for all its citizens.

The study was conducted over a period of three months, undertaking a mixed-methods approach that combined desk research and expert interviews.

1. **Desk research:** The examination of the literature and resources was two-pronged:
    a. The research aimed at generating a landscape review of AI systems across the world, and what role responsible AI played in that ecosystem. The landscape review focused on the Indian AI ecosystem, particularly with the tech startup industry in mind. This helped develop a broad understanding of opportunities towards RAI, the benefits of RAI, challenges in implementing RAI, and existing efforts and themes towards widespread RAI adoption.
    b. The desk research also explored the many ways, technologies, techniques, processes and best practices towards downstreaming RAI principles into real-world AI product development. This effort allowed for a study into how developers themselves can introduce and implement RAI into their workflows. Thus, the desk research granted the study both a high- and ground-level perspective into the RAI discourse.

2. **Expert interviews:** The study necessitated a cross-disciplinary investigation into AI systems in India, and RAI considerations and processes that are or could be embedded within them. Therefore, 17 semi-structured conversations were conducted with AI experts — including developers, industry leaders, legal and policy experts, researchers, and AI journalists. The interviews illuminated the challenges and opportunities for RAI implementation in India, particularly in relation to resources, knowledge, tools, processes, and regulation of AI initiatives in India. They also helped identify the core stakeholder groups that could enable widespread RAI adoption in India.

The research was triangulated through the processes of inductive and axial coding. From the desk research related to RAI practices emerged an RAI Primer for Developers in India. The landscape analysis and expert interviews came together to form the main arguments in this white paper.

CHAPTER I

Responsible AI — A Landscape Analysis

TABLE OF CONTENTS

CHAPTER I

Responsible AI — A Landscape Analysis

1. Artificial Intelligence: Transforming the world

The artificial intelligence (AI) revolution is here. Over the last decade, especially with the growing use of deep learning technology in general artificial intelligence use-cases, AI's vast potential has been recognised as truly transformational. According to a PwC report, AI is expected to increase global GDP by 14%, contributing US$15.7 trillion to the global economy by 2030. A survey conducted across industries in different sectors by The Economist Intelligence Unit (The EIU) uncovered that 94% of global executives described AI as crucial to solving their organisational challenges. These respondents believed that AI would boost growth (90%), productivity (86%), innovation (84%), and job creation (69%) over the next five years in their country and industry stating that AI provides improved user experience, increases employee satisfaction and workflows, generates faster and more data-driven decision-making, and increases overall revenues.

With the many recognised benefits of AI, both AI development and AI adoption has seen a global surge. A McKinsey survey claims, AI adoption in respondent organisations in 2023 increased to 72% as compared to the 50% mark it remained at for the preceding 6 years. The growing accessibility of AI has shaped new opportunities for even stakeholders in the Global South who are exploring new avenues to solve entrenched challenges of a difficult landscape. From boosting agricultural productivity to enabling easier diagnoses of diseases, from making educational services accessible to creating environmental models — AI has been the technology at the forefront of mitigating global challenges across domains. It is critical that we acknowledge the tremendous impact of AI systems, while ensuring that system design is holistic and emerging risks are being addressed dynamically.

Risks of AI

Increased use can often translate into increased risks. Aapti Institute's previous research on the impact of AI on Human Rights, and the implications of Generative AI on the human rights of women and non-binary genders highlighted various areas where AI systems could present harms. Beyond systemic risks like job displacement, environmental degradation, dis- or mis-information, and global inequality, some of the major threats posed by AI are as follows:

1. **Blackbox decision-making:** AI systems frequently function as "black boxes," rendering judgments that are opaque to users, making it near-impossible for them to comprehend or question its logic.
2. **Loss of control:** Humans may struggle to keep an eye on or influence important decisions as AI systems grow more independent, leading to a loss of accountability within AI initiatives.
3. **Discrimination:** AI can reinforce or magnify biases in training data, producing unfair or unjust results and discrimination against individuals or groups.
4. **Privacy violations:** AI systems often depend on vast amounts of personal data, raising the risk of abuse or breaches of sensitive personal information.
5. **Dangerous or harmful outputs:** AI systems could generate harmful or dangerous content, including violent instructions that cause unintended harm.
6. **Cybersecurity threats:** AI can be exploited by malicious actors to develop more sophisticated cyberattacks, compromising systems and data security.

Such risks are predominantly threatening the growth of the AI landscape. In particular, the use of AI in sensitive sectors like healthcare and finance has prompted a fierce discussion around the ethical use, deployment, and governance of this technology. Collective efforts like "Pause Giant AI Experiments" respond to growing concerns regarding unrestrained and unethical AI development. Naturally, this environment of founded and unfounded fears has stoked the fears of policymakers, leading to new regulations across the globe.

Regulatory landscape of AI

In addition to technology-agnostic data protection laws like the General Data Protection Rules (GDPR) in the EU and the Personal Data Protection Act (PDPA) in Singapore, countries have now begun to align AI development to broad frameworks and laws that encourage ethical design, development, and deployment of AI products. Around 37 laws related to AI were passed by legislative bodies in 127 countries in 2022 alone. A leading example is the EU AI Act. It attempts to classify AI systems according to the levels of risk they pose, and to tailor their regulatory obligations in proportion to the same. Canada has also published the Artificial Intelligence and Data Act (AIDA) to focus on responsible AI, user protection and fair competition. As opposed to such binding regulations, Singapore's Model AI Governance Framework takes a more self-regulatory approach. Similarly, the USA, through the American Artificial Intelligence Initiative, emphasises voluntary commitments by encouraging a principle-based approach that promotes industrial best practices and risk mitigation tools. India's regulatory approach is expounded further in section 3 below.

While regulation is essential, it must not stifle innovation. Overly restrictive regulations, by increasing compliance costs or creating uncertainty, can shut down a growing industry with immense transformative potential. Many countries realise the need to foster a landscape where developers can create AI solutions creatively and responsibly by providing clear guidelines, incentives for ethical practices, and support for research and development. For instance, Spain, through its National Strategy for Artificial Intelligence, is developing a trustworthy AI certification, i.e. a national seal to certify the security and quality of AI technology. The country is also encouraging dialogue between developers and policymakers by leading a pilot for a regulatory sandbox.

While the role of regulation and policy is critical and should aim for balance, other measures that involve key ecosystem actors are necessary when building such systems.

2. Responsible AI: Mitigating AI risks

Regulation is just one facet of addressing and discussing AI risks. There are several fears related to AI that require a broader discussion with other stakeholders. It is widely accepted that AI is going to revolutionise how we live. However, we do not know when this change will arrive and how it will impact the world. The Stanford Encyclopedia of Philosophy lists surveillance, job displacement, human-robot interactions, automated systems, and manipulation of behaviour as some of the existential apprehensions that AI inspires.

Alternatively, skeptics are less enthusiastic about the sustained benefits of AI systems. For instance, a survey conducted by Boston Consulting Group (BCG) and Massachusetts Institute of Technology (MIT) found that 90% of company respondents had made investments in AI but only two out of five of those firms reported any business gains in the last three years. Even if data is available everywhere, the practices of using it sustainably and efficiently without violating privacy regulations is still growing as a discourse. Thus, there is a dire need to invest in practices that can promote long-term, sustained AI development that addresses both current and emergent risks to ensure widespread AI adoption.

Responsible AI (RAI) represents this comprehensive toolkit of practices, processes and frameworks, providing myriad ways of embedding ethics into AI design, development and deployment. RAI attempts at maximising the positive benefits of AI while minimising its risks of negative consequences, allowing developers to navigate the complex matrix of financial, social and legal obligations. The aim is to create safe, trustworthy and ethical AI that is accessible to all.

What is Responsible AI?

There is no single dominant definition of RAI. Industry leaders across the globe have developed their own sets of definitions, principles, frameworks and practices. For instance, IBM defines Responsible AI as "Responsible artificial intelligence (AI) is a set of principles that help guide the design, development, deployment and use of AI—building trust in AI solutions that have the potential to empower organisations and their stakeholders. Responsible AI involves the consideration of a broader societal impact of AI systems and the measures required to align these technologies with stakeholder values, legal standards and ethical principles. Responsible AI aims to embed such ethical principles into AI applications and workflows to mitigate risks and negative outcomes associated with the use of AI, while maximising positive outcomes.

The Case for RAI

While there is no universally accepted framework defining RAI and how it should be implemented, some early adopters have already begun their journey towards safe, trustworthy and ethical AI. This white paper has consolidated these frameworks and tools below to strengthen the understanding around embedding RAI . However, many of them still minimally implement RAI, instead of attempting to gain the full long-term benefits of RAI. In reality, RAI is the key ingredient in scaling AI initiatives. A 2019 Accenture report uncovered that a firm incorporating RAI is 1.7 times more likely to scale. However, RAI cannot be defined only through a top-down approach - it is meaningful only if there is sufficient buy-in from firms and technical practitioners themselves.

FOR BUSINESS

Competitive advantage	90% executive respondents in the EIU survey consider long-term benefits and cost savings outweigh initial investment in RAI, enhancing product quality and competitiveness.
Talent acquisition & retention	Unethical practices discourage diversity of talent, compromising quality and competitive advantage. RAI mitigates staff attrition costs and boost productivity.
Enhanced market reach	Geographies and demographics offer differing expectations, norms and regulations. RAI helps foresee such challenges and access these markets.
Better risk management	Upfront investment in RAI reduces downstream risks and their costs like business performance, reputational damage, sunk costs, lost sales, cancelled contracts, non-compliance fees, etc.
Broader revenue streams	Product differentiation due to RAI helps increase market share. 2019 Ethisphere analysis shows World's Most Ethical Companies outperform Large Cap Index companies over 5 years by 14.4% and over 3 years by 10.5%.
Procurement advantages	In bidding processes, RAI gives procurement advantages (VCs, public grants, government projects, etc.). >90% of company respondents in the EIU survey include ethical consideration in these processes.
Increased pricing power	RAI can increase pricing power through better branding and reputation – which leads to price premiums of 26% on average. 2015 Nielsen survey showed 66% consumers are willing to pay more for ethical goods.

FOR THE ECOSYSTEM

Increased public trust	RAI reduces the risk of public scandal and distrust by prioritising users, leading to strong brand development, engagement and public relations.
Increased collaboration	Multistakeholder collaborations that enhance RAI allow tech companies to become thought leaders.
Sustainable investing	Conscious investors have already begun to advocate for ethical practices before deciding to invest.
Preparedness for compliance	When the GDPR became law, only 31% businesses believed they were ready. RAI helps prepare for such upcoming regulations.

FOR THE USERS

Increased public trust	Security and privacy are the biggest obstacles to AI adoption, especially in heavily regulated sectors. RAI addresses them.
Increased collaboration	RAI compels technology companies to perform well across broad user profiles, thereby boosting product value.
User Engagement	Capgemini analysis suggests 70% of consumers expect ethics in AI services, while Salesforce suggests 95% of customers are more loyal to industries they trust.

Frameworks of RAI

Organisations, including private firms, non-profit organisations, multilaterals and government bodies, have attempted at addressing RAI through their own frameworks, proposing principles and practices that align with them. Each carries their own benefits, particularly in terms of localising RAI practices to their use-cases, domains, and geographies. Some of them are listed as follows:

RELEVANT FRAMEWORKS	DESCRIPTION
Google's AI principles	Emphasises AI's ethical use by prioritising social benefit, safety, and fairness while explicitly avoiding harmful applications, such as weapons or surveillance.
Microsoft's responsible AI standard v2	Focuses on integrating responsibility-by-design principles into AI lifecycle stages, emphasizing fairness, inclusivity, transparency, and privacy.
NIST's trustworthy and responsible AI	Defines trustworthiness through principles like explainability, reliability, and resilience, tailored to application contexts and risk management.
OECD's AI principles	Advocates for human-centric AI promoting inclusivity, accountability, and sustainable economic growth with a global cooperation perspective.
GPAI's scaling responsible AI solutions	Speaks on practical collaboration to scale AI responsibly, addressing global challenges such as ethics, bias, and inclusivity.
Facebook's (Meta) five pillars of responsible AI	Focuses on transparency, fairness, privacy, robustness, and empowering users, particularly through user control and safety.
UNI Global Union's top 10 principles for ethical artificial intelligence	Stresses labor rights and social justice, ensuring AI respects workers' rights, privacy, and human dignity.

Internet Society's Artificial intelligence and machine learning	Highlights the need for user-centric AI that respects digital rights, promoting openness, interoperability, and decentralised solutions.
UNESCO's Ethics of artificial intelligence	Advocates for global ethical norms ensuring AI benefits humanity, with a focus on diversity, sustainability, and protecting human rights.
European Commission: AI High-Level Group's Ethics guidelines for trustworthy AI	Defines trustworthy AI through legality, ethical robustness, and technical reliability, promoting human autonomy and fairness.

> **The consolidation of these frameworks provides a larger view at the landscape and tries to provide an objective landscape and analysis for how Responsible AI could be viewed.**

Tools of RAI

There has been a steady increase in the number of AI-related projects on GitHub — from a mere 845 in 2011 to ~1.8 million in 2023, according to the 2024 Stanford AI Index report, with 59.3% rise being added in 2023 alone. Congruent with this interest, new research and tools have emerged in the field of RAI as well. The World Economic Forum has also launched the AI Governance Alliance to accelerate RAI adoption. In general, open sourcing of tools and techniques has increased both the development and adoption of RAI.

The Defense Advanced Research Projects Agency (DARPA)'s launched the Explainable AI (XAI) Programme to develop AI systems that can transparently and clearly explain their choices, therefore increasing user trust and confidence. They also launched other initiatives like Guaranteeing AI Robustness against Deception (GARD), Understanding Group Biases (UGB), and Machine Common Sense (MCS). In a similar vein, SHapley

Additive exPlanations (SHAP) employs cooperative game theory to explain how each feature contributes to a model's output, helping identify model biases while Local Interpretable Model-Agnostic Explanations (LIME) enables developers to create interpretable explanations for black-box models, aiding users understand individual predictions.

To ensure that AI systems produce fair results across a variety of industries, IBM's AI Fairness 360 Toolkit provides a broad range of algorithms to assess and reduce bias in datasets and models. With the use of Microsoft's Fairlearn, developers may evaluate differences in model results and rectify unfair advantages or disadvantages for particular demographic groups. TensorFlow's Fairness Indicators is also a valuable tool, offering visualisations to evaluate the fairness of models across demographics. Another fairness auditing tool, Aequitas, created by the University of Chicago, helps identify how machine learning models affect various demographic groups while promoting moral decision-making in predictive algorithms.

Privacy and security are also critical in RAI. Technologies like TensorFlow Privacy allows developers to incorporate differential privacy techniques into machine learning models, protecting sensitive data from exploitation. Google's What-If Tool creates an interactive interface that tests various model behaviours, granting the developers ways to assess hazards. IBM's Adversarial Robustness Toolbox contributes to the security of AI models by making them resistant to adversarial attacks. Furthermore, to promote accountability and transparency, Google's Model Cards employ frameworks to record model performance. By guaranteeing that AI systems are equitable, open, and safe, these and other open-source tools aid in the implementation of Responsible AI.

3. A Focus on India

Displaying the highest AI skill penetration rate in the world, India is emerging as a global hub for AI innovation and application. The country's AI ecosystem has witnessed exponential growth, driven by a combination of a vast talent pool, government initiatives like FutureSkills PRIME and IndiaAI, a burgeoning startup culture, and a robust technology services sector. The Department for Promotion of Industry and Internal Trade (DPIIT) recognises 140,000 startups in the country. Moreover, a National Association of Software and Service Companies (NASSCOM) report suggests that 44% of Indian deep-tech startups leverage AI technology. India has developed its technology services sector, with the export of software services increasing by 2.8% during 2023-24 to USD 190.7 billion. Unsurprisingly, NASSCOM estimates that AI could contribute significantly to India's GDP, adding USD 450-500 billion by 2025.

There is enthusiastic interest in the job market for AI as well. 'AI Specialist' is ranked second among emerging job roles in India in 2020. Interestingly, the Indian environment has also been deemed more AI aware. In a BCG survey with 21,000 consumer respondents from 21 countries, 93% of Indian respondents were aware of ChatGPT and 45% had used it (or other AI service). This represents the highest figures in both cases as compared to other countries, even mature AI markets like the USA and the UK. With both the market players and users primed for AI development, we now see AI technology across industries like healthcare, e-governance, education, financial services, agriculture, and manufacturing. For instance, during the COVID-19 pandemic, over 40% of deep-tech solutions for COVID leveraged AI.

Responsible AI in India

Despite India's trajectory towards developing a vibrant AI ecosystem, the unique challenges that plague this country threaten this eventuality. The digital divide between populations could lead to only a few benefiting from AI development, exacerbating existing inequalities. Scarcity of annotated and high-quality datasets coupled with the country's multilingual social fabric makes it difficult to build AI systems to scale. The lack of collective efforts to enhance trust in

AI systems could erode the public's confidence in AI initiatives. Moreover, Indian developers struggle to gain a competitive advantage over large-scale AI services provided by global technology giants.

Such concerns can only be addressed through the lens of RAI. AI systems that are fair, transparent, accountable and inclusive can realise India's full AI potential by prioritising ethics and equity in AI design and deployment. RAI can bridge the digital divide and ensure that underserved communities also benefit from AI advancements. For example, designing with regional languages and local contexts in mind can bolster last-mile access to AI solutions in healthcare and education. By promoting explainable AI, accountability mechanisms and governance frameworks, stakeholders can build trust among users.

To prime India for responsible AI development, the government of India tasked NITI Aayog with creating the National Strategy for Artificial Intelligence (NSAI) in 2018. The strategy proposes a pro-innovation and welfare-based approach that emphasises on risk management while striving to make the technology accessible to all. This light-touch regulatory framework resonates well with the developing economy of India and hopes to form cross-disciplinary flows with existing industries. Moreover, India enacted the Digital Personal Data Protection (DPDP) Act in 2023 to robustly address privacy concerns in AI systems.

Industry stakeholders and academic institutions are also actively researching and implementing RAI practices. Research hubs like CeRAI and ARTPARK are advocating for ethical AI use and developing guidelines for responsible AI deployment. According to a NASSCOM report, businesses across key industries are seeking to mature their RAI practices, particularly banking and financial services (38%), telecommunications (30%) and healthcare (21%). The report makes apparent that businesses that have higher AI maturity in India also exhibit higher RAI maturity. Indian AI initiatives seem to be steadily progressing towards widespread RAI adoption, with 30% of the respondents reported having mature RAI practices, another 30% having initiated formal steps towards RAI, another 30% demonstrating basic RAI awareness, and only 10% having no interaction with RAI at all.

Indian Startups: A centrepiece in the ecosystem

Despite this intellectual progress, the actual ability to govern and implement responsible AI frameworks is unevenly distributed. Larger companies have larger teams with deeper experience that are capable of aligning product development, deployment and dissemination cycles to principles of responsibility quickly and effectively. NASSCOM's report corroborates this, discovering that large enterprises are 2.3 times more likely than startups to report matured RAI practices and policies.

And yet, startups are the backbone of India's tech industry. With over 31,000 tech start-ups, India's tech startup ecosystem ranks third globally. With around 70% of these startups investing in AI, India's private AI ecosystem has seen significant growth. Indian developers working within this set-up struggle to gain a competitive advantage over large-scale AI services provided by global technology giants. This gap not only creates an inefficient market but also externalities like safety and privacy harms that affect all consumers. This erosion of trust can lead to stringent, shut-down regulations that stifle innovation.

RAI's prescribed proactive and principle-oriented approach allows for robust solutions that address such inequality between technology stakeholders. Through innovative models and open-source technology, it can make AI solutions affordable and scalable for small businesses and local communities, democratising access to AI and empowering individuals and businesses across socio-economic strata.

To achieve this reality, buy-in from developers in the Indian startup ecosystem is essential. Engaging directly with the cadre of Indian developers will ensure bottom-up strategies that remain grounded in reality and are viable and feasible to implement by all. Therefore, this study focuses on engaging with Indian developers and startups. The research specifically illuminated the gaps in the Indian AI startup ecosystem that need to be bridged for broader RAI implementation. They also investigate the opportunities towards building the RAI discourse in India. As such, they interact with diverse stakeholders across industries and fields of expertise to gain a holistic understanding of the potential of embedding and downstreaming RAI practices in the Indian AI startup ecosystem.

CHAPTER II

Analysed insights from expert conversations

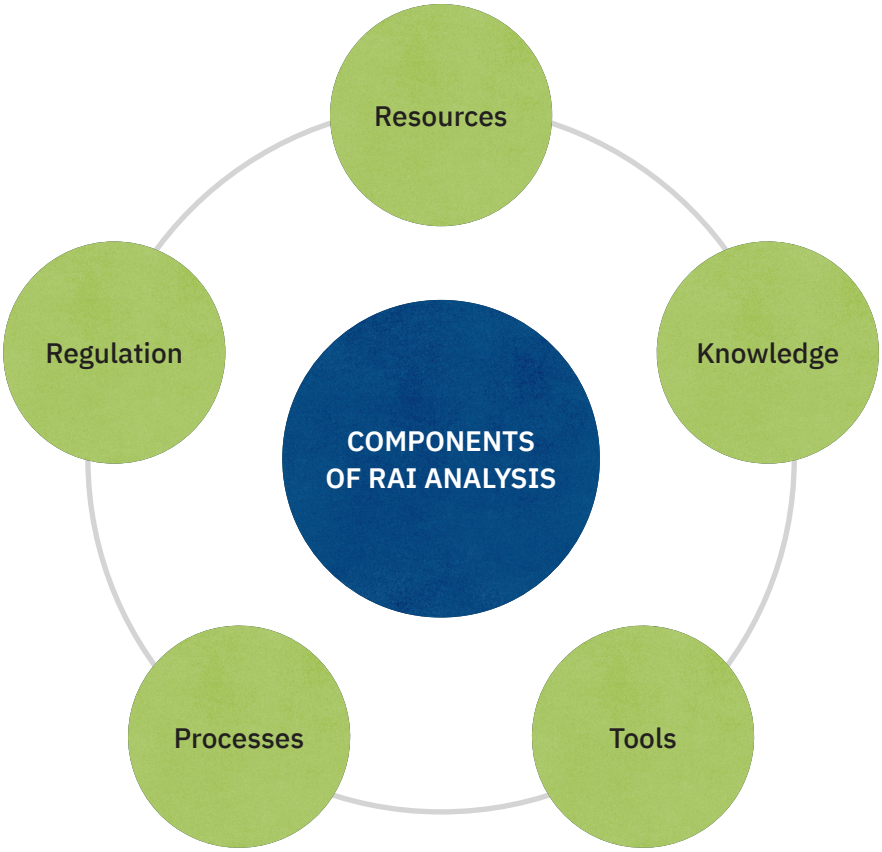
TABLE OF CONTENTS

CHAPTER II

Analysed insights from expert conversations

Framework for analysis

The expert interviews uncovered challenges and opportunities that fell under the five components of analysis — **resources, knowledge, tools, processes, and regulation.**



The interviews also delineated the major stakeholder groups within this ecosystem, noted as follows:

Developers	Create, train, and optimise AI systems; ensure ethical design and deployment practices
Startup management	Strategically integrate RAI principles into operations, driving adoption and accountability
Industry leaders	Set benchmarks for ethical AI use and influence industry-wide standards
Government	Regulate AI to protect public interest and enforce accountability
Research community	Advance AI ethics through studies, innovations, and critique of current practices
Domain experts and community leaders	Provide contextual insights to ensure RAI aligns with societal needs
Venture capitalists	Fund AI initiatives, pushing for responsible innovation and ROI
Standard-setting bodies	Develop and enforce ethical AI guidelines to promote safe and fair practices
Users	Drive adoption by valuing trust, usability, and fairness in AI systems
Citizens	Influence policies and hold stakeholders accountable through advocacy and societal feedback

These stakeholder groups will allow for grounding the analysis in real-world practice.

1. Resources

Stakeholder groups involved:
Developers, Startup management, Venture capitalists

CHALLENGES

A large obstacle faced by the Indian startup ecosystem in the widespread adoption of RAI practices are material constraints embedded within the system. With increasing innovation in AI, there is an uptick in competition amongst AI practitioners in India leading to a prioritisation of output specific goals. This creates an environment where embedding RAI is considered low on the list of priorities, resulting in limited resources being allocated to it. Insights from developers suggest that meeting KPIs is considered most important, and none of the existing KPIs take RAI adoption into consideration.

- **Financial constraints:** Developers might perceive RAI as an expensive venture with low returns. Investments into audits, bias mitigation strategies, transparent and human-centric design, and explainability tools are, thus, avoided.
- **Time pressures:** The race to bring products to market often leads startups to prioritise short-term functionality and profitability over long-term usability and ethical considerations. This leads to a culture of conceptualising ethics as a retroactive afterthought.
- **Talent shortage:** There is a global shortage of RAI experts who can guide AI product development in a financially sound and time-sensitive manner. A policy expert pointed out that the lack of relevant talent led to smaller companies and startups to ignore vital safety and privacy guardrails, often leading to their eventual failure to compete with larger players with well-equipped and well-trained teams.
- **Lack of high-quality data:** Data is at the core of several RAI processes. In NASSCOM's State of Responsible AI in India report, businesses deem lack of high-quality data as one of the biggest challenges when it comes to RAI adoption. While some sectors have no dearth of data (for e.g., healthcare and finance), the available data is often fragmented and unstructured. Inadequate labelling and inconsistencies in data

formats are additional obstacles to using data for AI systems in India. Furthermore, well-annotated and feature-rich local datasets, necessary for upholding RAI principles, are few and far between.

OPPORTUNITIES

Resource constraints in the Indian AI industry are a reflection of the growing pains of a developing economy. By acknowledging such limitations, pathways towards responsible and sustainable development of the AI industry should be chalked out.

- **Increasing Return on Investment:** The unwillingness of investing in RAI can be combatted by emphasising on the market incentives of adopting RAI. RAI should be reimagined as a smart cost-saving technology that delivers maximisation of profits by ensuring savings, enhancing brand reputation, and mitigating risks. Such conversations of integrating RAI for long-term return on investment can eventually lead to building a robust talent infrastructure in India for developing RAI expertise.
- **Expanding market reach:** An educator and AI developer pointed out that while it is important to adhere to local and contextual needs, an aspirational AI startup will be willing to look forward to expanding their product's market reach to international markets. For which, it becomes imperative for them to preemptively address ethical concerns that are current points of contention in the global AI ecosystem.
- **Building trust:** The interviews emphasised that AI stakeholders in India should be more aware of using RAI practices to build user trust. An industry leader posited that B2B startups have to build trust with clients and B2C startups have to build trust with end-users. In both cases, RAI processes are crucial. Creating trustworthy AI is intrinsic to broader branding of an AI product — an investment that can be considered worthy even by management and investors.
- **Future-proofing:** Some developers look at RAI processes and governance as 'future-proofing' their investments and efforts into building a product. Scaling a product entails embedding robustness across the AI value chain, and RAI principles are

the primary way of achieving this. Studies have pointed out how preventive and proactive risk management is a lot more cost-effective than retroactive harm mitigation.

- **Bridging data silos:** Data is integral to RAI, so there is a dire need for contextually relevant and high-quality data in India. Developers interviewed hailed Bhashini AI, a multilingual translation model, as an exemplar for Indian efforts towards developing a robust local data infrastructure. The Ministry of Electronics and Information Technology (MeitY)'s plan to establish an IndiaAI Innovation Centre will also help develop indigenous LLMs and Indian domain-specific foundational models. In addition to developing home-grown systems, global algorithms also need to be retrained on Indian datasets to be applicable in this ecosystem.

2. Knowledge

Stakeholder groups involved:
All

CHALLENGES

The lack of experts is just the tip of the iceberg in RAI awareness, training, and education disparities. RAI discourse in India is plagued with information asymmetry — i.e. unequal distribution of knowledge of expertise that creates power imbalances and inefficiencies. It is a systemic problem that affects all stakeholders involved in the AI ecosystem.

- **Educational gap:** Democratic dissemination of RAI concepts and practices would entail changes in the pedagogy of AI development. Developers unanimously communicated how they wished they were taught about RAI in their academic career itself.
- **Lack of on-the-job training:** Interviewees mentioned how Indian AI startups neither train their employees on RAI processes nor restructure their existing practices in line with RAI principles. A developer criticised how his company did not provide any RAI training despite mentioning RAI in their policies, leading to significant skill gaps, especially amongst those who were newly onboarded.

- **Lack of awareness of benefits of RAI:** Some expert interviewees went beyond capacity building and education to suggest that the main obstacle to RAI adoption is the lack of motivation among AI developers. This can be attributed to unawareness on how RAI can provide long-term benefits to AI startups. Startups, thus, often prioritise rapid development and scaling, leaving little room for making use of even the publicly available and time-sensitive ethical tools.
- **Lack of awareness amongst other stakeholders:** The entire brunt of implementing RAI should not fall squarely on the shoulders of just the developers, as interviewers stated that most decision-making is made top-down. Like developers, even upper and middle management can also be unaware of the commercial and social benefits of incorporating RAI practices in their products. Additionally, a global leader suggested that venture capitalists in India are less knowledgeable than their global counterparts of the long-term risk of ignoring ethical AI development. A legal expert in his interview claimed that government stakeholders controlling the AI narrative in India are also in dire need of RAI capacity building. Moreover, lack of awareness about RAI among users and citizens can fail to inspire a sustained and holistic effort towards RAI adoption, as it has in more mature AI ecosystems.

OPPORTUNITIES

There is a need for a systemic push towards greater and inclusive RAI awareness across all stakeholders in the AI ecosystem. These activities would ensure that all stakeholders, from developers to policymakers, are equipped with the necessary understanding to integrate ethical principles into AI systems. Moreover, these efforts towards knowledge-sharing should be inclusive and accessible to all to ensure developers and startups of all levels can derive maximum benefit.

- **RAI curricula:** Equipping students with RAI skills is congruent to preparing them for jobs of the future. Interviewed experts focused on the need for Indian RAI education to drive localised efforts to address local AI challenges responsibly and sustainably. A RAI educator mentioned how many of their students prefer leaving the country to work for global

technology companies. They hypothesised that ensuring public and private funding for AI education will drive innovation and nurture future RAI talent in India. An industry leader pointed out that holistic RAI education will not limit itself to technical practices, but also include compliance requirements, sectoral specificities, documentation practices, and data governance.

- **Enabling environments:** Institutions like incubators and research centres have been leaders in driving India's startup growth. These same institutions, thus, have the capacity for providing a healthy space to explore RAI without resource limitations. A few interviewees mentioned that mechanisms like grants, mentorship programs, fellowships and workshops can grant new startups and aspiring AI entrepreneurs the necessary support to take their first step towards RAI. Such opportunities can allow even mid-career startups to align their product lifecycle to RAI principles.
- **Industry-relevant capacity building:** A legal expert, in his interview, shared that the organisation is the main lever for encouraging responsibility in AI practices and processes. Thus, the organisation becomes the site where RAI can be explored as well as implemented. Such targeted vocational learning can encourage AI developers to develop new RAI competencies and keep pace with ongoing trends, iterate on existing products to match new user needs and address emergent risks. Several interviewees argued that the training should not be limited to just developers. All important stakeholders, especially those that are in charge of decision-making, governance and funding, should be willing to participate in broader RAI awareness campaigns and programs.
- **Cultural significance:** Motivation to build a responsible AI practice comes from the collective responsibility of stakeholders. A policy expert referred to the book *Crossing the Chasm* by Geoffrey A. Moore, to suggest that there is a big chasm between early adopters and the mainstream market. A heightened attempt at practicality is critical to make this jump. Far-reaching and intentional conversations around the application of RAI within the Indian context will allow for the diffusion of AI ethics.

3. Tools

Stakeholder groups involved:
Developers, startup management, industry leaders

CHALLENGES

Several open-source as well as proprietary tools are already available to developers. However, many interviewees stated that existing tools may not always be applicable to particular contexts. India, being a diverse country with different sets of resources and differing context, is a complex matrix of AI practices and processes. With different sectors harbouring different expectations, output-based tools are not translatable across domains.

- **Lack of access to relevant tools:** The Indian AI startup ecosystem continues to play catch-up with their global counterparts. Interviews suggest that Indian developers struggle to gain easy and free access to relevant RAI tools. One developer mentioned that one member of their team gained access to an international RAI conference however, could not circulate pictures from the conference with the rest of the team for legal reasons.
- **Inapplicability of tools:** Some developers stated that the frameworks were too abstract to be devolved into their practice. Such instructions were deemed difficult to adopt in their existing workflows. Sectoral differences further make it difficult to adapt tools to specific use-cases.
- **Difference in context:** A policy expert mentioned that the dominant frameworks are made by global technology giants, for regulatory environments that are unlike India. Indian developers, especially in the early stage, cannot afford to direct resources to meet standards that do not translate to their problem context. Infrastructure gaps to utilise existing AI tools make it difficult to implement large-scale RAI processes.
- **Socioeconomic and linguistic barriers:** Indian developers have the additional responsibility of engaging ethically with a culturally-nuanced multilingual population who are themselves resource constrained. NLP models especially struggle in this environment. Therefore, inviting principles like inclusivity, fairness, and transparency into Indian AI products are particularly challenging.

OPPORTUNITIES

India is a complex geography that resonates better to decentralised solutions that remain contextually relevant. There is no 'one-size-fits-all' solution to the problem of RAI in India. However, this is an opportunity to amplify self-regulatory practices. This could encourage AI startups to actively identify, assess and manage risks that their product might face or create.

- **Principle based approach:** Some experts stated that output-based standard-setting is less effective than a process- or input-based approach. In light of this, principles can be applicable and universal across the value chain. Focusing on embedding principles within the general processes of an AI product's lifecycle will create adaptive frameworks that can be widely applicable even in specific solutions. This approach will also allow for a better mapping of existing and potential risks, thereby creating frameworks that seek to reduce harms and manage risks with deliberation and intention.
- **Small-scale best practices and tools:** While large-scale RAI processes might be difficult to integrate, certain best practices can be helpful in ensuring AI development remains responsible to its context. The expert interviews delineated the following as crucial:
  - **Purpose limitation:** Purpose limitation is not just useful in upholding privacy requirements underlined in the DPDP Act but also in defining scope. A developer pointed out how limiting datasets to a purpose also allows developers to assess and address anomalies and vulnerabilities more easily.
  - **Human-in-the-loop within systems:** Many interviewees corroborated on the benefits of contending with the 'black box' nature of AI by ensuring humans were in control of interventions that applied principles and tools of RAI into the system.
  - **Data governance:** An AI consultant emphasised that any implementation of responsibility in AI, being a data-hungry system, should begin with data governance. An industry leader suggested that localising and customising data governance frameworks, inspired by global standards, should be a priority for startups from an early stage itself.

- **Inclusivity and bias mitigation:** For AI products to be successful in India, they need to be tailored to its unique cultural landscape. Particular attention to inclusion and bias mitigation will allow for a quality product that can be accessible to all. A legal expert mentioned how AI products should not only respond and interact well with the physical and digital infrastructure of India but also the social and cultural superstructure. For instance, community-driven data collection is more representative of India's cultural matrix, as put forth by the MeitY's Expert Committee Report on Non-Personal Data Governance Framework. Additionally, iterative mechanisms of transparency and accountability allow for greater and more meaningful interactions with users.

4. Processes

Stakeholder groups involved:
Developers, startup management, industry leaders, government, research community, domain experts and community leaders, users, citizens

CHALLENGES

Ethical and safe AI necessitates cross-disciplinary collaboration across its different processes. However, expert interviews highlight that the Indian AI ecosystem is suffering from siloisation. There is a lack of coordination, collaboration, and conversation between different stakeholders that interact or participate in the AI ecosystem.

- **Control over the value chain:** Developers mentioned that they did not have control over upstream AI processes like data collection. Without collaboration among actors across the value chain, organisations generate unnecessary risks or unwelcome harms, while also running the risk of duplicating efforts and wasting resources - both up and downstream.
- **Between policy and technology:** There is a lack of dialogue between policymakers and technology, leading to policies that do not speak to the real-world process, needs, expectations, priorities, and goals of other stakeholders in the Indian AI ecosystem.

- **Between experts and technology:** Many interviewees claimed that the company policies and governance structures are often made by people who are not in charge of technology development. This creates a significant gap that eventually leads to a product that is not responsive to the needs of the users. Without inputs from ethicists, social scientists, domain experts or community leaders, AI products could fail in the real world.

OPPORTUNITIES

The interviews unanimously signalled the need for inspiring collective action to encourage widespread RAI implementation. Collective action will not only strengthen the core foundations upon which AI is built in India, but also prioritise inclusion and equity within the system. A journalist mentioned how there should be stronger links between stakeholders like academic and domain experts, ethicists, technological experts, venture capitalists, community leaders, and industry associations. This needs to span roles, sectors, geographies, and domains.

- **Transparency and knowledge sharing:** There needs to be increased collaboration between internal teams that are dedicated towards creating an AI product. Transparency in regards to processes and governance, in addition to consistent knowledge sharing, will create an environment that is conducive to implementation of RAI.
- **Interdisciplinary collaboration:** Domain experts, academic experts, ethicists, social scientists and community leaders are valuable voices to be included not just in the initial stages but throughout the product development process. In particular, compliance experts can guide baseline RAI considerations, ensuring sector-specific requirements are met.
- **Platforms for dialogue and exchange:** Conferences, community of practitioners (CoPs), knowledge portals and forums, and workshops are some of the ways that interviewees mentioned networks could be formed between early adopters or potential adopters of RAI in India. Academic discourse and technological breakthroughs should be frequently deliberated upon to ensure emergent risks are recognised at the onset.

Such platforms will also empower smaller startups to catch up to standards set by dominant technological players.

5. Regulation

Stakeholder groups involved:
Developers, startup management, industry leaders, standard-setting bodies, government

CHALLENGES

Several interviewees noted that AI's speed is unmatched by global regulation's attempts. India's regulatory ecosystem is lagging behind, especially considering the country's enormous AI potential. While the DPDP Act is a step in the right direction, more proactive regulatory systems could support and encourage responsible AI innovation and adoption.

- **Inadequate compliance structures:** All interviews commented on the importance of regulatory bodies in curbing bad actors but to also be a guiding force towards a standardisation of RAI best practices for developers to imbibe. Nevertheless, the current compliance landscape in India faces its own set of challenges. A 2024 Carnegie paper points out the following limitations of Indian state capacity in regulating AI: lack of technical expertise, failure to issue clear and timely regulatory guidance, lack of investigative powers, ineffective or inconsistent enforcement, and lack of grievance redressal mechanisms. One developer, while conveying his experience with developing an AI product for the government, mentioned the need for inbuilt mechanisms like checklists and assessments that could have helped ensure a safer rollout of the product.
- **Inadequate external regulatory bodies:** In addition to limitations in the legislative capacity, the Indian ecosystem also suffers from a lack of bodies that could provide positive reinforcement to adopt RAI through certifications and standard-setting - limiting pathways towards voluntary commitments to RAI.

OPPORTUNITIES

India should prioritise AI ecosystem building by taking a pro-innovation approach. This would bolster the growth of the AI industry and possible ancillary industries.

- **Self regulation:** Institutions like NITI Aayog have stated that hard regulation like the EU AI Act may not translate as well in India. To ensure inclusive innovation, self-regulation is seen as the ideal way forward. NITI Aayog suggests a graded risk-based approach that focuses on harm reduction. Non-governmental organisations like auditory bodies, self-regulatory organisations (SROs), industry associations, startup hubs and advocacy groups have the potential of leading this conversation, as advocated by many of the experts interviewed. Risk assessments, checklists, audits and evaluations were brought up across interviews as the preferred methods of incentivising RAI in India. Similarly, many interviewees spoke about the benefits of introducing certifications and publicly displayed seals, ideally granted by reputed external bodies to reward investment in RAI best practices. It is believed that it will help the Indian AI ecosystem mature to the same levels set internationally. A journalist who was interviewed brought up ESG as an exemplar to follow in terms of integrating RAI into mainstream AI practices.
- **Binding regulation:** Some interviewees were of the opinion that self regulation and market incentives are not enough. They believe that governments should play a greater role in penalising bad AI practices and actors. A legal expert did not see it as contradictory to implementing self regulation, but posited maintaining SROs as a difficult venture and inferior to legislation. One policy expert preferred updating existing sectoral policies to comprehend the risks of integrating AI into those domains over introducing new legislation around AI. Nonetheless, interviewees agreed that punitive regulation is incompatible with an ecosystem as immature as India's. The industry needs to flourish and be guided by the expectations set through the government. Respecting plurality and diversity of the stakeholders within the AI ecosystem in India also needs to be at the forefront of this discussion. An industry leader gave the example of the 'Singapore model' that provided regulation that encouraged innovation by developing a shared repository of resources that is accessible to all practitioners.

CHAPTER III

Interventions and Recommendations

TABLE OF CONTENTS

CHAPTER III

Interventions and Recommendations

1. Intervention: An RAI primer for developers

The landscape study and expert interviews articulated that developers are key stakeholders to engage with. It is necessary to provide developers with the right resources and tools with which they can start thinking about different ways of incorporating RAI into their practice. As such, an intervention was designed that could empower developers to begin their journey in RAI implementation. This intervention took the form of a primer <link to the primer> that describes best practices relevant to Indian startups.

Principles of the primer

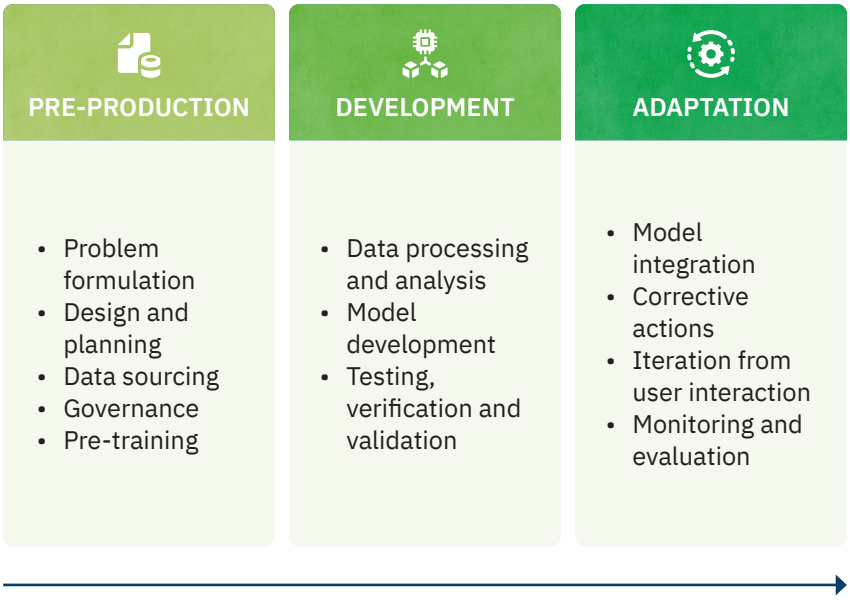
Existing frameworks and literature have provided different understandings of the principles that are foundational to RAI. The primer condenses the literature into four principles overall:

1. **Fairness:** Fairness refers to the principle that allows an AI system to operate equitably — preventing bias, discrimination, inequality, and injustice.
2. **Transparency and Accountability:** While transparency gives one the ability to interpret AI systems, accountability allows stakeholders to intervene in its design and correct automated decisions.
3. **Privacy:** Privacy refers to the safeguarding of personal data and respecting an individual's right to control their personal information when AI systems are designed, developed and deployed.
4. **Safety and Security:** While safety tries to calibrate AI systems towards mitigating broad physical, digital, ethical, and psychological risks, security protects them from internal or external threats.

The best practices are in line with the principles they uphold, divided further using the value-chain approach.

THE VALUE CHAIN APPROACH

The primer is structured along the lines of the value chain: three stages that determine the AI system lifecycle. The methods of RAI intervention are nestled within these three stages.



2. Recommendations for the ecosystem

Stakeholders other than the developers should also be engaged to address the problem of imbibing responsibility into AI systems holistically. Keeping in mind the ecosystem constraints in India, the following recommendations could be the first step towards RAI adoption and maturity.

Shared repository of knowledge and resources

Stakeholder groups involved:
Developers, startup management, industry leaders, research community

Addresses challenges of:
Resources, Processes, Knowledge, Tools

Lack of resources is cited as one of the most intrinsic barriers to the adoption of RAI practices. It has become increasingly important to make RAI tools accessible to all. A shared repository of resources, including tools, knowledge, frameworks, and practices, should be made available to developers of all experience levels, democratising RAI knowledge across the ecosystem. While open-source libraries, checklists, and templates can help plug RAI directly into existing workflows, domain-specific case studies will tailor practices to their industry standards. The most useful consolidation of such resources could also form the basis of a possible standardisation of best practices. As contributions pour in from different sources, the repository will reflect the changing landscape, dynamically matching the need of the hour. Moreover, localised insights from diverse geographies will help ensure that developers can also tackle legal, cultural, and contextual nuances.

Community of practitioners

Stakeholder groups involved:
Developers, startup management, industry leaders, standard-setting bodies, research community

Addresses challenges of:
Resources, Processes, Knowledge, Tools

Collective problem-solving and maintaining peer accountability will allow for democratic diffusion of RAI. Early adopters, by sharing their successes and failures from diverse projects, can help other smaller players avoid common pitfalls and replicate effective strategies. Cross-disciplinary inputs will lead to holistic and contextual solutions to RAI challenges, accelerating innovation. Peer reviews, audits and assessments within the community would help identify risks and potential solutions and opportunities in AI systems, fostering collective accountability. By coming together to

understand and tackle compliance challenges, the entire developer ecosystem could meet and negotiate with existing regulations more effectively. Such a unified community approach will empower developers to influence industry and regulatory standards. Moreover, a collaboration between practitioners will be most effective at addressing emergent risks — integrating advancements while ensuring sustained relevance and impact.

Collaboration between actors

Stakeholder groups involved:
Developers, startup management, industry leaders, standard-setting bodies, research community, government

Addresses challenges of:
Knowledge, Processes, Regulation

Bridging the gap between compliance standards and real-world AI practices has become the key concern in the developer ecosystem. There is a need for a more involved conversation between practitioners, policymakers, and community members to make sure that regulatory oversight shepherds AI towards ethical practices without stifling innovation. Developers can act in advisory and advocacy capacities to educate policymakers about technical capabilities and limitations of AI, enabling realistic and informed regulations. In turn, policymakers can standardize best practices to build public trust, reduce uncertainties, and encourage wider adoption of safe and responsibly built AI. The community can provide iterative feedback loops that can embed contextuality and responsibility within AI systems. Sandboxing is one of the techniques that can spark these interactions, allowing for the co-creation of ethical as well as innovative solutions. Such joint efforts will make sure that potential risks are mitigated proactively, effectively, and efficiently. Open and transparent communication between these stakeholders will, thus, lead to healthier and more equitable practices and outcomes.

Better handholding and training

Stakeholder groups involved:
Developers, startup management, industry leaders, research community, domain experts and community leaders

Addresses challenges of:
Resources, Processes, Knowledge, Tools

There also needs to be greater awareness amongst practitioners about the benefits of RAI. Industry leaders, pioneers, and early adopters of RAI practices should take up the challenge of diffusing responsibility to developers of all experience levels. Other stakeholders within academia, policy and research should also collaborate in ensuring that the knowledge gap among practitioners is bridged. The existing resources can be demystified and communicated clearly to practitioners through workshops, conferences, and mentorship programs. Within organisations as well, personnel should be trained to think about ways to embed responsibility in AI systems. It is also recommended to bring in community and subject experts to coach developers on the ethical particularities of the AI system's operating domain. Such handholding should be iterative — maintaining continuous engagement through feedback loops, audits, monitoring, and refresher courses.

Certifications and external standards

Stakeholder groups involved:
Developers, startup management, industry leaders, venture capitalists, standard-setting bodies, research community

Addresses challenges of:
Knowledge, Regulation

Currently, the use of AI across industries has generated a lot of fear and mistrust among the general public. A way to combat such fears is to advocate and advertise the use of responsible AI practices publicly. Clear external guidelines and certifications, verified independently by respected bodies, could be one way to accomplish this. By adhering to such standards, practitioners would not only secure a long-term plan for product development but also be rewarded with a loyal user base that responds

positively to certified ethical considerations. Moreover, venture capital can also learn to respond positively to such certifications, creating a market mechanism that encourages RAI best practices. By exhibiting a commitment to responsibility, companies can also expand their market reach internationally. Being at the forefront of cross-border trustworthiness and industry consensus would promise competitive advantage and broader avenues for growth.

Greater Public Awareness

Stakeholder groups involved:
All

Addresses challenges of:
Knowledge, Processes

The public anxiety around AI should not be considered a hindrance but an opportunity towards a cultural shift that establishes RAI practices as the norm. A well-informed public who is aware of the potential and limitations of AI can foster a robust demand for ethical AI systems. Stakeholders within the AI ecosystem, thus, need to encourage broader discussions, using accessible forms of media, to reach out to diverse audiences. Transparent communication, such as reporting on operations that uphold a company's commitment to RAI, is the first step towards building public trust. Similarly, to combat unfounded fear mongering, interactive tools and platforms can be used to allow the public a peek into AI decision-making processes. RAI specialists and leaders can leverage blogs, videos, talks, and social media posts to have a far reaching impact.

3. The Primer

In line with the recommendations provided above, the study also produced a primer for developers in India. The primer compiles some of the best practices, tools and frameworks to make RAI knowledge accessible to Indian developers who often find themselves constrained by resources.

It seeks to align its RAI recommendations with existing workflows. It serves as a structured yet modular step-by-step guide on RAI adoption that can assist future training, workshops, and mentorship programmes. While detailed, it is flexible,

allowing the developer to pick and choose the recommendations that make most sense to their processes — making it ideal for peer audits, reviews and accountability practices. It is also useful for non-practitioners like the public and policymakers to gain insights into RAI practices. By being open to potential iterations, the primer hopes to be a living document for ongoing learning and skill development that matches the fast-evolving AI industry.

As such, the primer acts as a catalyst for knowledge sharing, industry alignment and public trust, addressing critical challenges like lack of resources, fragmented processes, and regulatory uncertainty.

Concluding Remarks

As we develop AI's disruptive potential, it is the duty of all stakeholders to guarantee its ethical and inclusive use. As the designers of this transformation, developers play a particularly important role among them. Their involvement is essential to the goal of Responsible AI since their choices at every level—from design to development to deployment—shape how AI interacts with society.

Developers are in the best position to proactively address possible risks, harms, or unintended effects because they are often the first to recognise them. Their knowledge is crucial for integrating moral values into the technical underpinnings of AI systems. We can produce solutions that represent a wide range of perspectives and foresee emergent societal issues by bringing developers closer to interdisciplinary discussions with ethicists, researchers, policymakers, and users.

To realise the vision of Responsible AI, we must place a high priority on transparent communication, thorough documentation, and proactive teamwork. Developers, in particular, play a pivotal role in maintaining transparency by sharing methodologies, documenting decisions, and engaging openly with scrutiny from peers and communities. Their active involvement improves the general robustness of AI systems and strengthens trust and integrity.

The path to responsible AI is ongoing and collaborative. Let's give developers the freedom to lead the way in establishing moral guidelines and making justice and diversity a priority in their work. By working together, we may establish and develop ethical technology that fosters trust, advances social progress, and leaves a constructive legacy for next generations.

aapti institute

Aapti is a public research institute that works at the intersection of technology and society. Aapti examines the ways in which people interact and negotiate with technology both offline and online.

contact@aapti.in | www.aapti.in