

Responsible Artificial Intelligence and Banking: Navigating the FREE-AI Framework

Pramod C Mane is with National Institute of Bank Management, Pune

(Email: pramod.mane@nibmindia.org)

Abstract

The Reserve Bank of India's (RBI's) recently released FREE-AI framework serves as a strategic blueprint for the responsible and ethical adoption of Artificial Intelligence (AI) in the nation's banking and finance sectors. This framework encourages safe innovation through shared infrastructure and institutional capacity building while moderating risk through governance and consumer protections. However, we believe that its effectiveness is contingent on overcoming significant implementation hurdles, mainly the absence of standardized measures for fairness, robustness of models, and clear third-party responsibility. This commentary explores the strategic vision of the FREE-AI framework, exploring the collaboration required across the financial ecosystem and critically assessing the challenges of implementation.

Introduction

Artificial intelligence (AI) seems to be capturing the imagination and concern of all, while we have some idea about the shape of things to come. Lots of speculations are also going on. These speculations encompass both the euphoria and concerns of ordinary people. Hence, the role of the regulator is of paramount importance in this uncertain world. It is how the recently proposed FREE-AI framework of the RBI aims to facilitate the responsible and ethical implementation of AI, specifically in the banking and financial sector.

The FREE-AI framework is intended to promote safe use of AI in Indian banking, while building trust, and balancing innovation versus risks. It is comprised of two halves; one half enables innovation by building shared infrastructure, providing guidance/ agile policy and building skills and institutional capacity to enable firms to responsibly adopt AI. The second half addresses risk through a governance of AI decisions, protection for customers and data, and through continuous assurance to testing and oversight of AI decisions. The two halves hope to accelerate safe, trustworthy use of AI in finance combined.

Artificial Intelligence (AI), we commonly refer it as the ability of machines to replicate human-like intelligence (including self-learning, making complex decisions, and even identifying subtle patterns). In banking and finance, AI is not simply a technological curiosity, it is a significant departure from how we carry out banking. On the one hand, banking is poised to gain greatly from AI in the form of automated efficiencies and more advanced fraud detection during banking transactions, as well as more customized client services, among other things. On the other hand, the implementation of AI does raise important issues. For instance, algorithmic bias, transparency, and increased cybersecurity and data privacy risks. This is the reason; regulators worldwide are attempting to balance meaningful innovation with accountability and fairness.

The conversation on AI is moving from *how to police AI* to *how to use it responsibly*. In this context, the RBI's FREE-AI framework offers a significant strategic vision. While welcoming this shift, this commentary raises some issues on the implementation front.

Regulating AI to Responsible AI

Who holds the steering wheel when algorithms start deciding who gets a loan or how markets move? And what happens when efficiency comes at the cost of fairness or stability? These concerns explain why AI regulation has come into the picture. At its core, regulation refers to the legal and supervisory guardrails that ensure AI is deployed safely, transparently, and accountably in sectors like banking and finance (Crisanto et al. 2024). Across the world, countries are taking different approaches to regulating AI.

In the finance domain, the earliest discussions regarding AI regulation suggested the need for rigid legal regulations in order to mitigate risks associated with lending, trading, and compliance (EC 2021). But soon afterwards, the rigidity of laws turned out to be insufficient: they could fall behind new developments in AI practice, they would restrict innovation (and therefore improvement), and they would also have difficulty dealing with contextual harms, for example, discrimination, which is prominent with biased credit scoring, and lack of transparency regarding robo-advisors (Jobin et al., 2019). This led to a pivot toward “Responsible AI” (RAI) understood as the design and deployment of AI systems that are ethical, explainable, human-centric, and socially-valued.

Responsible AI (RAI): Why it matters

AI has moved from small, rules-based tools to smart systems that generate text, code, and decisions at scale. Banks now depend on external models, cloud compute, and fast-changing software that they do not fully control. This raises hard questions: Who is accountable when a model shifts or a vendor updates a system? How do we check fairness, explainability, and robustness in models that are not fully transparent? Meanwhile, supervisors are embedding AI into prudential and conduct standards, and courts along with data-protection laws now require clear, intelligible reasons for high-stakes decisions. Together, these forces pushed RAI from an ethical ideal to an operational requirement.

In the financial sector, RAI has gained prominence as institutions seek to harness the benefits of innovation while upholding public trust and meeting regulatory standards that safeguard fairness and accountability. Specifically, banks and finance institutions are now expected to pay attention not only to cybersecurity and privacy, but also to issues of self-governance and how AI affects their workforce. RAI is often described through four main pillars: building systems that allow people and machines to work together, putting in place strong governance, ensuring technical reliability and transparency, and aligning AI use with the firm’s values and accountability. Even small lapses, such as a hidden bias in data, can quickly grow into wider risks when scaled across financial operations (Belelieu et al., 2025).

FREE-AI Framework: Sutras and Recommendations

In recent years, academia and the financial services industry have been working to identify what RAI means. According to Goellner et al. (2024), “Responsible AI is human-centered and ensures users’ trust through ethical ways of decision making. The decision-making must be fair, accountable, not biased, with good intentions, non-discriminating, and consistent with societal laws and norms. Responsible AI ensures, that automated decisions are explainable to users while always preserving users’ privacy through a secure implementation.” They outlined five pillars of RAI: Human-centered, Trustworthy, Ethical, Explainable, Privacy-preserving and Secure AI.

The FREE-AI framework for responsible AI draws its strength from integrating all of these core pillars. This detailed and comprehensive framework includes seven sutras and twenty recommendations for responsible incorporation of AI in India's banking and financial sector.

FREE-AI framework consider trust as the foundational pillar in AI adoption. It requires AI systems to be fully explainable and accountable to those they impact. In order to adopt AI, it must serve as a tool, which augments human decision-making, prioritising citizen welfare always and in critical situations defer to human judgment. Moreover, innovation must be purpose-driven, ensuring outcomes are not only efficient and secure but also rigorously fair, non-discriminatory, and environmentally sustainable to maintain long-term societal legitimacy.

The RBI's FREE-AI approach to AI is comprehensive. It goes beyond a list of individual rules. Instead, it organizes its twenty recommendations into six strategic, interwoven pillars.

The framework’s core ambition is clear and that to create a robust and responsible ecosystem from the ground up, and this starts with Foundational Infrastructure & Public Goods. This is a vital move to build a shared, high-quality base of data and tools. This base democratizes AI access. It ensures even smaller institutions can build trustworthy models. This foundational effort must be enabled by Regulatory Enablers & Adaptive Governance. This means the RBI must implement flexible, dynamic policies, such as a graded liability framework. These policies encourage innovation. At the same time, they maintain systemic stability and quickly adapt to new risks.

To translate these external guidelines into safe practice, Internal Governance & Risk Management acts as the crucial corporate safeguard. It requires banks and FinTechs to embed strong, trustworthy policies. These policies control operational, ethical, and regulatory risks. They protect both their customers and their operations. For this whole system to function, Capacity Building & Knowledge Sharing is non-negotiable. A skilled workforce and shared experience are essential to responsibly harness AI's benefits. It also helps avoid duplicating efforts.

Critically, the framework anchors AI adoption to a public purpose through Financial Inclusion & Social Impact. It ensures the technology actively closes access gaps. It also builds confidence in formal finance, rather than deepening existing inequalities. Ultimately, Consumer Protection & Ethical Design serves as the bedrock. It transforms abstract concepts like transparency and accountability into tangible user trust and market stability. This is essential for sustained adoption.

Beyond Checklists: A Thematic Reading of FREE-AI

RBI's Free-AI framework, is not merely a regulator's or a bank's rulebook, it is a blueprint for collaboration across the entire financial ecosystem. To make such principles as trust, fairness, and innovation a reality in everyday life, all players in the spectrum, from fintech and universities to civil society and government institutions have a role to play: banks need to deliver transparency and accountability, fintech and incubators should promote inclusive and responsible AI innovations, regulators must align policies and develop capacity, and investors should fund projects in line with societal values. The framework's objectives, such as financial inclusion, cybersecurity, and explainable AI, can happen only if all these actors work together for swapping best practices, testing innovations in sandboxes, and monitoring risks repeatedly. Without this collaboration, the principles will stay on paper, and the chance to build safe, fair, and scalable AI in finance will slip away.

It is critical to go over the questions, including what FREE-AI's recommendations are meant to accomplish, who in the ecosystem are key players in putting them into practice, what their expected roles are, and what obstacles exist in putting them into practice.

Table 1 outlines the expected roles of key stakeholders in putting the framework guidelines into action. Additionally, it identifies the primary implementation challenges that these players are likely to encounter.

Theme	Crucial Players and Their Expected Role	Key Implementation Challenges
1. Foundational Infrastructure & Public Goods	<p>Governments / Government Agencies</p> <ul style="list-style-type: none"> • Fund and spearhead high-quality financial data infrastructure. • Integrate infrastructure with national platforms (e.g., AI Kosh). <p>Technology Companies / FinTechs</p> <ul style="list-style-type: none"> • Develop and offer Indigenous AI Models as a public good. • Participate in the AI Innovation Sandbox for testing. <p>Regulators</p>	<p>Data Fragmentation & Quality: Difficulty harmonizing siloed, legacy data from different institutions for a central infrastructure.</p> <p>Funding & Sustainability: Ensuring sustained public/private funding for development and long-term maintenance of the infrastructure.</p> <p>Cybersecurity at Scale: Securing a massive, interconnected DPI from sophisticated, AI-powered cyber threats.</p>

	<ul style="list-style-type: none"> • Collaborate on funding and incentive establishment. • Oversee ethical governance of public data. 	
2. Regulatory Enablers & Adaptive Governance	<p>Regulators</p> <ul style="list-style-type: none"> • Issue comprehensive AI Policy and consolidated Guidance. • Constitute and oversee the AI Standing Committee. • Administer the graded AI Liability Framework. <p>Governments / Government Agencies</p> <ul style="list-style-type: none"> • Participate in the AI Standing Committee. • Align existing legal frameworks with new policy. <p>Academic Research Institutions.</p> <ul style="list-style-type: none"> • Contribute expert advice to policy development. 	<p>Pace of Innovation: Regulation moves slower than rapidly evolving AI/GenAI technology, leading to constant policy obsolescence.</p> <p>Model Opacity (Black Box): Regulating and auditing proprietary models whose decision-making process is fundamentally opaque or difficult to interpret.</p> <p>Talent Gap: A shortage of internal AI-fluent talent within the regulatory body to effectively assess complex technological risks.</p>
3. Internal Governance & Risk Management	<p>Banks (Regulated Entities - REs)</p> <ul style="list-style-type: none"> • Establish and enforce the Board-Approved AI Policy. • Implement robust AI Model and Data Governance. • Execute Red Teaming and upgrade cybersecurity. <p>Technology Companies / FinTechs</p> <ul style="list-style-type: none"> • Design auditable AI systems with built-in safety. 	<p>Legacy Infrastructure:</p> <p>Integrating advanced AI systems with outdated, rigid core banking legacy systems.</p> <p>Third-Party Risk:</p> <p>Over-reliance on a few external AI vendors, making compliance and liability tracking complex.</p> <p>Cost of Compliance:</p> <p>High implementation cost for rigorous model validation, red teaming, and ongoing monitoring for smaller REs.</p>

	<ul style="list-style-type: none"> • Integrate AI products into institutional risk frameworks. <p>Regulators</p> <ul style="list-style-type: none"> • Mandate and audit compliance with governance frameworks. • Issue guidance on risk mitigation (e.g., red teaming). 	
4. Capacity Building & Knowledge Sharing	<p>Academic Research Institutions</p> <ul style="list-style-type: none"> • Develop and conduct training/upskilling for the workforce. • Support regulators' research and capacity development. <p>Regulators</p> <ul style="list-style-type: none"> • Invest in institutional capacity (e.g., AI institute). • Facilitate industry-wide best practice sharing. <p>Banks (REs)</p> <ul style="list-style-type: none"> • Develop Board/C-suite AI governance competence. <p>Industry Associations</p> <ul style="list-style-type: none"> • Support smaller entities with policy templates. 	<p>Talent Retention: Difficult for banks and regulators to compete with large tech salaries to retain top AI and data science talent.</p> <p>Curriculum Lag: Academic programs struggle to update curricula fast enough to match the industry's real-time AI tool and governance needs.</p> <p>Resistance to Sharing: Institutions may be reluctant to share sensitive AI use cases or "lessons learned" due to competitive concerns.</p>
5. Financial Inclusion & Social Impact	<p>Governments / Government Agencies</p> <ul style="list-style-type: none"> • Establish framework to integrate AI with DPI for scaled services. • Direct incentives toward inclusion initiatives. 	<p>Bias Amplification:</p> <p>Risk that AI models, trained on historically biased data, replicate or worsen financial exclusion for underserved groups.</p> <p>Digital Divide:</p>

	<p>Banks (REs) & FinTechs</p> <ul style="list-style-type: none"> • Drive AI innovation for financial inclusion • Apply for recognition/rewards for social impact AI. <p>RBI / Regulators</p> <ul style="list-style-type: none"> • Provide regulatory flexibility for inclusion-focused AI. 	<p>Ensuring equitable access for low-bandwidth, low-digital-literacy users in rural and remote areas.</p> <p>Measuring Impact:</p> <p>Difficulty in defining common, quantitative metrics to measure social impact and fairness for reward programs.</p>
6. Consumer Protection & Ethical Design	<p>Banks (Regulated Entities - REs)</p> <ul style="list-style-type: none"> • Establish consumer protection framework (transparency, recourse). • Invest in consumer education campaigns. <p>RBI / Regulators</p> <ul style="list-style-type: none"> • Enforce compliance with consumer protection frameworks. <p>Civil Society</p> <ul style="list-style-type: none"> • Surface user risks and potential harms. 	<p>Explainability vs. Simplicity:</p> <p>The inherent difficulty in making legally accurate AI decisions simple enough for the average consumer to understand.</p> <p>Fragmented Grievance:</p> <p>Customer grievance systems can become slow and confusing when multiple third-party vendors are involved in a single AI service.</p> <p>Adversarial Attacks:</p> <p>New cybersecurity risks where attackers poison or manipulate AI models, leading to unfair customer decisions (e.g., biased credit scores).</p>

Table 1. Mapping Ecosystem Roles and Challenges for FREE-AI Implementation (Source: Authors created)

Why AI in Finance Needs More Than Just Rules

When we try to put the AI framework into practice, we run into some major practical issues that make it hard to apply consistently. One of the primary issues is the interpretation of the guidelines. Different actors in the industry, the banking executives, developers, auditors, as well as fintech companies can define different terms (“high-risk”, “human-in-the-loop”, or “autonomous AI”, etc.) in completely different ways. For example, a particular bank might view a chatbot as a simple, low-risk tool for interacting with clients, while another bank may view it as a high-risk component that needs intense tracking. This discrepancy means that banks end up adopting different levels of tracking as well as divulging information in different ways.

The second issue is that there is no clear, standardized way to implement key requirements like fairness, robustness, or explainability. Everyone agrees these things are important, but how do you actually measure them? Should fairness be about equal outcomes across different groups? or should it focus on reducing negative impacts? What level of model drift is serious enough to pause a system? Without clear answers, banks end up making their own rules as they go, which makes it nearly impossible for auditors to compare practices fairly or for regulators to get consistent, reliable information.

RIA refers to using AI with clear and defined rules so it is explainable, fair, private, and secure. That only works when data/ML teams, business experts (credit, fraud, AML), and risk/governance work together to set checks like bias tests, reason codes for decisions, consent and data controls, and ongoing monitoring. Without this mix, models may look good in tests but fail in the real, world causing compliance problems, extra costs, and damage to trust.

Another issue is how this framework fits with other existing rules, like those for IT outsourcing, cybersecurity, and data protection. Let us take an example, if a bank uses an AI service from an outside vendor, then in that case, who is responsible for testing it for vulnerabilities or reporting problems; the bank or the vendor? And where do the stakeholders draw the line between data privacy records and AI model documentation? These gray areas then possibly lead to important tasks being overlooked, unnecessary duplication of effort, or even attempts to exploit regulatory loopholes. Next, working with outside companies like cloud services and data firms adds another layer of complexity. Each one has its own rules and ways of doing things. Without clear, standard agreements, it becomes very unclear who is responsible if something goes wrong. Accountability simply vanishes.

Let it be frank; some controls we want are not hard to achieve today. For instance, deep models would not be fully explainable; GenAI would not be hallucination-free; bias cannot be zero; provenance and vendor transparency are patchy.

In conclusion, we believe that the FREE-AI framework is a strong start and it successfully connecting AI technology with ethical rules, pointing toward a more open and fair financial system. Even so, its real success depends on turning these big ideas into practical steps for everyone involved. Although the road ahead is difficult, this framework offers a hopeful foundation for building AI that the public can trust.

Acknowledgement

The authors gratefully acknowledge Prof Partha Ray (Director, NIBM, Pune) for their invaluable feedback and guidance on this commentary.

References

Belelieu, A, D Propson, D Parker (2025). Artificial Intelligence in Financial Services. White Paper, World Economic Forum.

Crisanto, J C, J Ehrentraud, M Fabian (2024). Artificial intelligence in financial services: Regulatory considerations and challenges. Bank for International Settlements (BIS) Financial Stability Institute Occasional Paper.

EC (2021). Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act); COM (2021) 206 final; European Commission: Brussels, Belgium, 2021. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>

Jobin, A, M Ienca, E Vayena (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>

Goellner, S, M Tropmann-Frick, B Brumen (2024). Responsible artificial intelligence: A structured literature review. arXiv preprint arXiv:2403.06910.

RBI, (2025). FREE-AI Committee Report. Available at: <https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/FREEAIR130820250A24FF2D4578453F824C72ED9F5D5851.PDF>. Access on: September 23, 2025.