

AI GRC PLAYBOOK

Principles Need Actions: Introducing India's
Financial Sector Artificial Intelligence (FSAI) Toolkit



PRINCIPLES NEED ACTIONS

Introducing India's Financial Sector Artificial Intelligence (FSAI) Toolkit

Future Shift Labs (FSL) is a strategic research and policy think tank working to build a sustainable, equitable, and digitally resilient future by advancing responsible and ethical applications of artificial intelligence. Established in 2024, the organisation conducts in-depth research and analysis on emerging AI technologies and their societal impact, advocates for inclusive AI governance, and supports strategic consulting, training, and public engagement to strengthen institutional and community capacity. With a focus on fostering collaboration among researchers, policymakers, industry leaders, and civil society, Future Shift Labs seeks to position India as a global leader in digital diplomacy, ethical AI development, and international cooperation on emerging technology challenges.

Authors

Vidhi Sharma (primary author and lead researcher), Prachi Sharma (co-author), Pranjal Dwivedi (co-author, regulatory analysis and global frameworks).

Future Shift Labs team

Sagar Vishnoi, Director, Future Shift Labs, whose belief in this project from the outset and whose continued support and encouragement throughout made it possible in every sense that matters. Bhairabi, Sugam, and Akanksha for operations and logistics. Abir Mahanta for design.

Suggested Citation

Sharma, V., Sharma, P., and Dwivedi, P. (2026). *AI GRC Playbook: From Principles to Practice: Introducing the Financial Sector Artificial Intelligence (FSAI) Toolkit*. Future Shift Labs, New Delhi.



 <https://futureshiftlabs.com/>

For publications enquiries, please contact: contact@futureshiftlabs.com

© 2026 Future Shift Labs.

This publication may be reproduced for non-commercial research, educational, or policy purposes with appropriate acknowledgement of the source

ACKNOWLEDGMENTS

AND DISCLAIMERS

ACKNOWLEDGMENTS

This Playbook is the result of months of research, consultation, and collaboration conducted between January and July 2026. It would not exist without the generosity of a great many people across India's financial sector, regulatory community, and policy ecosystem who gave their time and, more valuably, their candour.

We are grateful to the participants of the April 2026 roundtable at the Oberoi, New Delhi, including the RBI FinTech Department and representatives from technology, legal, and policy organisations whose contributions under Chatham House Rules shaped the Playbook's central argument and the architecture of the FSAI Toolkit. We are equally grateful to the practitioners at banks, NBFCs, and fintechs who agreed to one-on-one conversations during the Phase 2 consultation, including colleagues at Think360 AI, ICICI Bank, and a former Bank of Baroda practitioner and others. India's financial sector has been generous with this project, and we hope the Playbook repays that generosity.

Mandar Kagade of Black Dot Public Policy Advisors deserves particular acknowledgement. He arranged meetings that would not otherwise have happened, contributed the AI use case stories between chapters, co-authored the Mint op-ed that preceded this publication, and brought the industry and innovation perspective that prevented this document from becoming purely a compliance exercise.

DISCLAIMER

This Playbook does not constitute legal advice, regulatory guidance, or a compliance certification. Completion of the FSAI Toolkit tools does not confer regulatory approval or supervisory endorsement. The RBI Draft Guidance on Model Risk Management was in public consultation as this Playbook was finalised hence, final requirements may differ and institutions should monitor developments independently. Nothing here substitutes for qualified legal or professional advice specific to an institution's circumstances. All graphics, diagrams, and illustrations in this Playbook were generated using AI tools.

FOREWORD

I started this project because of a question I could not answer.

In late 2025, I was in a conversation with a senior compliance professional at a financial institution. We were discussing the RBI FREE-AI Framework, which had recently been published, and the growing regulatory expectation around AI governance. At some point in the conversation, she asked me something straightforward: if her institution wanted to actually implement what the framework was asking for, where would they start? in a practical way like 'what would they do on Monday morning?'.

I did not have a good answer. There were principles, frameworks, global standards from the EU, the United States, Singapore, and Canada. There was the RBI framework itself, with twenty-six recommendations. But there was no practical, India-specific, financial-sector-specific guide that translated any of this into the kind of operational tools a risk or compliance function could actually pick up and use. That gap is what this Playbook is designed to fill.

The Key Question: What to do on a Monday morning with AI Governance?

Over the course of 2025 and 2026, my team at Future Shift Labs conducted a comparative analysis of over twenty-five global, regional and sectoral governance frameworks and regulatory instruments covering AI in financial services. We studied global standards, Indian law, sector-specific regulation, maturity models, and practitioner toolkits. We held a closed practitioner roundtable in New Delhi in April 2026, bringing together senior participants from regulated financial institutions, technology organisations, legal and policy bodies, and the regulatory community. We conducted structured one-on-one interviews with compliance leads, risk officers, and data science practitioners across banks and fintechs.

What we found was consistent across every conversation, every workshop, and every piece of research. The gap in India's financial sector has moved beyond 'what responsible AI governance means' to actually 'translating that understanding into practice'. Every organisation we spoke with knew that accountability matters, that documentation is important, that risk should be proportionate to potential harm. What most of them did not have was a named accountable committee, a documented inventory of their AI systems, a vendor contract with AI-specific provisions, or a risk classification for any of the AI they were running.

Principles need actions! - That is the sole foundation of this entire project and this Playbook.

The four tools in this Playbook are collectively named the **Financial Sector Artificial Intelligence Toolkit - the FSAI Toolkit** and are designed to be used practically in an every day work scenario. The Responsible AI Adoption Assessment tells an institution where it currently stands on its AI governance journey. The Model Documentation template ensures every AI system has a baseline compliance record. The Risk Scoring tool gives any compliance or risk function a criteria-based method for classifying the governance intensity a given AI system requires. The Vendor Due Diligence Checklist provides the pre-procurement questions and minimum contract provisions that protect an institution when it deploys AI it did not build.

I have tried to write this Playbook the way I would want a practitioner to use it, which is to say efficiently and without ceremony. The context in the early chapters explains why governance matters and what the regulatory landscape looks like. The toolkit in Chapter 5 can be used independently of everything else. If you are a compliance officer with thirty minutes and a vendor contract to review, go directly to Tool 4. If you are preparing for an RBI assessment, Tool 3 gives you the risk classification your examiner will expect you to have.

The Playbook adapts the maturity framework from different Responsible AI Maturity Roadmaps like GSMA, which is one of the most practically structured maturity models we encountered in our research. The playbook draws on the RBI FREE-AI Framework as its primary regulatory anchor, specifically Recommendation 26, which identifies the need for an AI compliance toolkit for regulated entities and this Playbook is a response to that recommendation.

I am grateful to the participants of the April 2026 roundtable, who gave their time and their candour under Chatham House Rules, and to the practitioners who agreed to one-on-one conversations when a group setting was not possible. The practitioner voice in this Playbook belongs to them. The errors belong to me.

Vidhi Sharma

*Head of Responsible AI Governance,
Future Shift Labs
New Delhi, 2026*

Vidhi Sharma is Head of Responsible AI Governance at Future Shift Labs, leads the Centre of Responsible AI and is the primary researcher and author for the AI GRC Playbook. She is an economist by training, public policy strategist by practice, and AI governance specialist by conviction.

Her current work focuses on translating AI research and emerging frameworks into practical governance and global policy approaches and actions, with an emphasis on regulatory coherence, interoperability, and trust across jurisdictions.

Table of Contents

Executive Summary	9 - 12
How To Use This Playbook	13
Chapter 1: Setting the Context	14 - 19
1.1 Financial Sector at a Turning Point	14
1.2 Trust as the Foundation	14
1.3 Where India's Financial Sector Stands	15
1.4 Understanding What Is Being Governed – Three Categories of AI	16
1.5 The AI Lifecycle and Governance Across Every Stage	17
1.6 The Global Setup and What India Can Learn From It	18
1.7 What This Playbook Does	19
Use Case A: Alternative Credit Scoring & Credit Underwriting	20 - 22
Chapter 2: India's Regulatory Landscape	23 - 29
2.1 The Regulatory Context for AI	23
2.2 Mapping India's AI Regulatory Architecture	24 - 25
2.3 The RBI FREE-AI Framework: India's Sectoral Anchor for AI Governance	26 - 28
2.4 From Legal Requirements to Governance Responsibilities	28
2.5 What This Means for Regulated Entities	29
Use Case B: Fraud Detection (Transaction & Payment Fraud)	30 - 32
Chapter 3: Global Frameworks for AI Governance	33 - 40
3.1 Why Global Frameworks Matter	33 - 34
3.2 Global Approaches to AI Governance	34 - 37
A. Governing AI through Regulation and Risk Classification: The EU AI Act	34
B. Governing AI through Organisational Risk Management: The NIST AI Risk Management Framework	35
C. Governing AI through Management Systems and Audit Readiness: ISO/IEC 42001	35 - 36
D. Governing AI through Practical Implementation and Deployer Responsibility	36
E. Governing AI through Ethics and Institutional Readiness	37

3.3 International Convergence: Five Governance Signals	38
1. Clearly Assigned Accountability	38
2. Risk-Proportionate Controls	38
3. Documentation as Governance	38
4. Human Oversight for High-Impact Decisions	38
5. Continuous Monitoring Across the AI Lifecycle	39
3.4 From Global Principles to Indian Practice	40
Use Case C: AML Monitoring (Transaction Monitoring & Financial Crime Detection)	41 - 43
Chapter 4: From Principles To Practice	44 - 51
4.1 The Translation Challenge	44 - 45
4.2 Six Translations: From Regulatory Principle to Institutional Practice	45 - 51
4.2.1 From "The Board Must Approve an AI Governance Policy" to a Documented Accountability Structure	45
4.2.2 From "Maintain a Comprehensive Model Inventory" to a Governed Registry of Every AI System	46
4.2.3 From "Apply Proportionate Controls" to a Criteria-Based Risk Classification for Each System	46 - 47
4.2.4 From "Ensure Human Oversight for High-Impact Decisions" to a Documented Review Mechanism with Named Triggers	47
4.2.5 From "Validate All Models Including Third-Party Models Independently" to a Vendor Governance Framework	48
4.2.6 From "Apply Specific Controls for AI and ML Systems" to Documented Controls for Explainability, Bias, Hallucination, and Red-Teaming	49
4.3 A Seventh Translation: The Vendor Problem	50
4.4 The Role of the FSAI Toolkit	51
4.5 Sequencing and the Regulatory Alignment	51
Use Case D: Project Aurora: BIS Innovation Hub Case Study	52

Chapter 5: The Financial Sector Artificial Intelligence (FSAI) Toolkit	53 - 73
5.2 Tool 1 — Responsible AI Adoption Assessment	53 - 57
5.2.1 Part One: Awareness — Does the Institution Know What AI It Has	53 - 54
5.2.2 Part Two: The Four Levels of AI Maturity	55 - 57
5.2.3 Assessment Record	57
5.2 Tool 2 — Model Documentation	58 - 59
5.3.1 Tier A: Baseline Documentation (Every System, Regardless of Risk Tier)	58 - 59
5.3.2 Tier B: Risk-Scaled Documentation (Depth Depends on Risk Tier from Tool 3)	59
5.4 Tool 3 — Risk Scoring and Accountability	60 - 65
5.4.1 Accountability Record — Completed Before Scoring	60
5.4.2 The Eight Dimensions	61 - 63
5.4.3 Modifiers	63 - 64
5.4.4 Total Score and Risk Tier Determination	65
5.5 Tool 4 — Vendor Due Diligence and Contract Checklist	66 - 73
5.5.1 The Accountability Principle	66
5.5.2 Part A: Pre-Procurement Questions	66 - 69
5.5.3 Part B: Minimum Contract Clause Schedule	69 - 72
5.5.4 Contract Summary and Sign-Off	73
Annexures	74 - 88
Annexure A: Frameworks Reviewed for This Playbook	75 - 77
Annexure B: List of AI Governance Controls	78 - 80
Annexure C: Glossary of Key Terms	81 - 83
Annexure D: Note on Methodology	84 - 85
Annexure E: Roundtable Participants: Organisational Representation	86
Annexure F: Phase 2 Consultation: Acknowledgements	86
Annexure G: The GSMA Responsible AI Maturity Roadmap	87
Annexure H: Note on Tool 4: Vendor Due Diligence and contract checklist	88

EXECUTIVE SUMMARY

THE PROBLEM

India's financial sector is deploying artificial intelligence with growing confidence, beginning with lower-stakes applications such as voicebots, customer service chatbots, and internal process automation, and moving steadily toward more consequential functions including fraud detection, KYC verification, and customer decisioning. The governance infrastructure required to oversee this deployment has not kept pace with the rate of adoption. Regulated financial institutions broadly understand, at the level of principle, what responsible AI governance requires of them - accountability, documentation, human oversight, proportionate risk controls, and continuous monitoring. What most of them do not yet have is the operational machinery to deliver any of it: a named committee with AI governance in its terms of reference, a documented inventory of every AI system in use, a risk classification for each of those systems, or a vendor contract that includes the AI-specific provisions that regulators now expect.

The question that surfaced in every practitioner conversation we had, in every roundtable discussion and every one-on-one interview with risk officers and compliance leads across banks and fintechs, was not about principles. It was simpler and considerably harder to answer: What do you actually do on Monday morning? Which meeting do you call, which document do you create? Which vendor do you call back? Which question do you put on the board agenda?

The Reserve Bank of India's FREE-AI Framework and the Draft Guidance on Regulatory Principles for Model Risk Management, published in June 2026, have provided clear and increasingly specific answers to the question of what responsible AI governance requires. This Playbook provides the answer to what institutions do about it.

THE RESEARCH BASIS

This Playbook is grounded in a comparative analysis of twenty-five governance frameworks, regulations, and toolkits spanning global standards, Indian law, and sector-specific guidance, including the EU AI Act, the NIST AI Risk Management Framework, ISO/IEC 42001, the Singapore Model AI Governance Framework, Canada's Algorithmic Impact Assessment, UNESCO's Recommendation on the Ethics of AI, the OECD AI Principles, and the GSMA Responsible AI Maturity Roadmap, alongside the RBI FREE-AI Framework, the Digital Personal Data Protection Act 2023, the IT Act, the Consumer Protection Act, and the Bharatiya Nyaya Sanhita. The research was tested through a closed practitioner roundtable in New Delhi in April 2026, conducted under Chatham House Rules, with participants from regulated financial institutions, technology organisations, legal and policy firms, and the regulatory community. Findings were further validated through structured one-on-one interviews with risk officers, compliance leads, and practitioners across banks and fintechs across India.

THE FIVE GOVERNANCE SIGNALS

Across the twenty-five frameworks studied, five requirements appear consistently regardless of jurisdiction, legal form, or enforcement mechanism, and it is this convergence that provides the analytical foundation for the toolkit.

Someone must be clearly and formally accountable for each AI system, with accountability documented, assigned to a named committee or function, and connected through a defined escalation path to the board.

Governance intensity must be proportionate to the risk the system poses, with higher-risk systems subject to stronger controls and lighter governance applied where the potential for harm is limited.

Documentation is the governance mechanism itself, not an administrative record of governance that happens elsewhere without documentation, accountability cannot be verified, audits cannot be conducted, and incidents cannot be investigated.

Human oversight must be substantive and triggered by defined conditions for decisions that significantly affect individuals, rather than being a nominal override capability that is never used in practice.

Monitoring must be continuous across the AI lifecycle, because AI systems degrade, drift, and fail in ways that are not visible at deployment and that only ongoing surveillance can detect.

THE CENTRAL FINDING

Practitioners who participated in the research for this Playbook understood the principles of responsible AI with considerable sophistication. Few observations from the roundtable

- AI adoption is real but narrower and more fragile than commonly understood
- The governance gap is not knowledge, it is operationalisation
- Risk classification is unresolved and the resolution is in criteria, not categories
- Risk proportionate controls are important and it can vary by use case
- Vendor accountability is a contract design problem, not a governance design problem
- GenAI is a different governance problem and cannot be treated the same as classical ML
- The supervisory / examination angle is a powerful adoption lever
- The innovation-governance balance was the one thing missing from the workshop

Each of these gaps represents the distance between a principle that is understood and a practice that has not yet been built. This Playbook tries to close that distance. It takes five governance requirements that appear consistently across every major global and Indian AI governance framework and converts them into four practical tools that any regulated financial institution can begin using immediately, without waiting for new legislation, additional budget, or a dedicated AI governance function

THE FSAI TOOLKIT

The Playbook contains the FSAI Toolkit comprising four tools designed to be used in sequence for any AI system being brought under governance for the first time, and revisited annually or whenever the system or the institution's AI use changes significantly.

Tool 1, the Responsible AI Adoption Assessment, adapted from the GSMA Responsible AI Maturity Roadmap and customised for India's financial sector, allows any institution to establish honestly where it currently stands on its AI governance journey before it attempts to document or score any individual system. The assessment has two parts: a factual awareness stocktake that determines whether the institution has a working picture of its AI use across all business units and functions, and a four-level maturity ladder spanning Aware, Organising, Governing, and Embedded. The score is revealed only after all questions are answered, a design choice that reduces the tendency toward optimistic self-classification that characterises most governance self-assessments.

Tool 2, the Model Documentation template, establishes a two-tier documentation standard that distinguishes between the baseline legal and compliance requirements that apply to every AI system regardless of risk including the DPDP Act legal basis for data processing, the named accountable committee, and the system's deployment status and the risk-scaled documentation requirements that deepen proportionately as the system's risk tier increases, from a brief operational description for Tier 3 systems through to external red-teaming reports and human oversight documentation for Tier 1 systems.

Tool 3, the Risk Scoring and Accountability tool, provides eight criteria-based dimensions each scored from zero to three, plus separate modifiers for generative AI systems and agentic AI systems, producing a total score out of thirty-one that places the system in one of four risk tiers with corresponding governance intensity requirements. The tool also establishes the accountability structure for the system, assigning responsibility to a named committee or function with AI governance explicitly in its mandate rather than to a single named individual, which better reflects how banks and NBFCs actually operate.

Tool 4, the Vendor Due Diligence and Contract Checklist, addresses what practitioner conversations consistently identified as the fastest-growing and most overlooked governance gap in Indian financial services because most AI deployed in the sector is procured rather than built. The tool provides pre-procurement due diligence questions covering model documentation, bias testing results, training data disclosure, security assessment, explainability mechanisms, and audit access rights, alongside eight minimum contract clauses covering performance standards, model change notification, audit rights, incident reporting, liability allocation, DPDP compliance, explainability obligations, and AI-specific business continuity provisions.

AI IN PRACTICE - USE CASES FROM GLOBAL FINANCIAL SERVICES

Between each chapter of this Playbook, a coloured page presents a real-world AI use case from global financial services, contributed by Black Dot Public Policy Advisors. The four use cases - alternative credit scoring, fraud detection, AML monitoring, and Project Aurora illustrate how AI is already being deployed at scale in institutions including HSBC, JPMorgan, Mastercard, DBS, and HDFC Bank, and what those deployments have produced in measurable outcomes. These pages are deliberately placed at the transition between chapters and they are neither governance documents nor analysis. **They are evidence that the technology works, and that responsible adoption at scale is achievable. Governance is what makes that adoption durable.** The Playbook and the FSAI Toolkit provide the governance infrastructure. The use case pages provide the reason to build it.

BUILDING ON WHAT EXISTS

This Playbook does not emerge from a vacuum, and it does not claim to be the first serious attempt to address AI governance in financial services. Some useful work already exists, and this Playbook builds on it rather than displacing it.

Globally, the NIST AI Risk Management Framework has provided one of the most thoughtful and widely adopted approaches to organisational AI risk management, offering a flexible four-function structure that has influenced governance thinking across sectors and jurisdictions. The EU AI Act has established the most comprehensive binding regulatory framework for AI, defining risk categories and governance obligations that have become a global reference point. ISO/IEC 42001 has given institutions a management system standard for AI governance that aligns with existing certification and audit practices. The GSMA Responsible AI Maturity Roadmap, on which the maturity assessment in this Playbook is explicitly based and to which full credit is given in the references, provides one of the most practically structured approaches to organisational AI maturity assessment available.

In India, NASSCOM's Responsible AI practice has produced frameworks and toolkits that have contributed meaningfully to the conversation on responsible AI adoption in the Indian technology sector. Infosys, Microsoft, and other technology organisations have published responsible AI toolkits that address governance from the perspective of AI developers and technology providers. The RBI FREE-AI Framework itself represents a significant regulatory contribution, and the thinking behind it has shaped the architecture of this Playbook throughout.

This Playbook acknowledges all of this work with respect. The research phase reviewed twenty-five frameworks and toolkits precisely because existing work deserved to be understood before new work was produced.

DISCLAIMER

This Playbook is not a legal opinion and should not be treated as one. It is not a compliance certification, and completing the tools does not confer any form of regulatory approval or supervisory endorsement. It is not a prediction of how India's AI regulatory landscape will evolve, and institutions should continue to monitor the final requirements of the Draft Guidance on Model Risk Management and any subsequent regulatory developments. It is a governance readiness framework, designed to remain relevant across regulatory trajectories because it is grounded in governance capabilities i.e., accountability, documentation, proportionality, human oversight, and continuous monitoring that every credible regulatory framework, present and future, will require.

HOW TO USE THIS PLAYBOOK

The five tools referred to throughout this Playbook are contained in Chapter 5. They can be used independently without reading the surrounding chapters, though the context in Chapters 1 through 4 explains the design rationale behind each tool.

If you are	Start here	Then go to
A board member or independent director	Executive Summary	Chapter 2 - The Regulatory Landscape
A Chief Risk Officer or Chief Compliance Officer	Chapter 1 - Why Now	Chapters 2 through 4, then Chapter 5 Toolkit
A risk or compliance team member	Chapter 4 - From Principles to Practice	Chapter 5 Toolkit - use the tools directly
A technology or data science lead	Chapter 3 - The Global Picture	Tool 3 in Chapter 5 - Risk Scoring
An internal auditor	Chapter 4 - From Principles to Practice	Chapter 5 Toolkit, Annexure B and H
A regulator or policy professional	All chapters	Annexure A for framework mapping and B for list of controls
A policy researcher	Chapter 3 - The Global Picture	Annexure A for framework mapping, Annexure D for methodology
A legal adviser or consultant	Chapter 4 and Chapter 5 - Tool 4	Annexure H

CHAPTER 1

SETTING THE CONTEXT

1.1 Financial Sector at a Turning Point

In April 2026, India's Finance Minister convened an emergency meeting at the highest levels of the country's financial system.¹ Around the table were senior officials from the Reserve Bank of India, Indian Computer Emergency Response Team (CERT-In), National Payments Corporation of India (NPCI), and the chief executives of Scheduled Commercial Banks. The trigger was a single AI model, Anthropic's Claude Mythos – an unreleased frontier system that had demonstrated the ability to identify and exploit software vulnerabilities at a scale and speed that no human team could match. Within days of global attention turning to Mythos, India's financial authorities had mobilised. Banks were instructed to strengthen cybersecurity, undertake AI-related risk assessments, improve incident reporting, and establish real-time threat intelligence sharing across the sector.

The significance of that meeting was the revelation that AI had become consequential enough to require a ministerial-level response. And the instructions that followed, such as risk assessments, incident reporting, intelligence sharing, institutional coordination, were a combination of technology and governance, hand in hand.

Two months later, in June 2026, the Reserve Bank of India published its *Draft Guidance on Regulatory Principles for Model Risk Management*.² Where the *RBI FREE-AI Framework*³ published in 2024 had established the principles of responsible AI governance, the Draft Guidance translates those principles into specific operational requirements: a Board-approved model risk management framework, a comprehensive model inventory covering every AI system including those procured from vendors, risk-based model tiering, independent validation of third-party models, and specific controls for generative AI systems.

Now that the regulatory signal is clear and specific, the question this Playbook addresses is what regulated entities do with it.

1.2 Trust as the Foundation

Every governance framework examined for this Playbook, from the RBI FREE-AI Framework to the EU AI Act, from ISO/IEC 42001 to the GSMA Responsible AI Maturity Roadmap returns, eventually, to the same concept and that is “Trust”.

The RBI FREE-AI Framework names trust as its first foundational sutra: trust is the foundation upon which all AI deployment in financial services must rest. This framing has deep roots in Indian governance thought. The Arthashastra, Kautilya's treatise on statecraft written approximately 2,300 years ago, places trust at the centre of the relationship between institutions and the populations they serve. Kautilya's analysis begins with a premise that remains as relevant to a central bank as it was to a kingdom: the legitimacy of any institution rests not on its authority alone but on the confidence of those it governs. Institutions that lose that confidence lose, over time, their capacity to function. The principle applies with particular force to financial services, where depositors, borrowers, and customers extend trust to institutions with their savings, their creditworthiness, and their most sensitive personal information.

1. For reference, see [FM Nirmala Sitharaman meets heads of banks on AI risks following concerns over Anthropic's Mythos](#), The Economic Times, April 2026.

2. For reference, see [Draft Guidance on Regulatory Principles for Model Risk Management](#), Reserve Bank of India, 2026.

3. For reference, see [RBI Framework for Responsible and Ethical Enablement of Artificial Intelligence](#), August 2025.

The *Deeploy AI Governance Control Framework*,⁴ one of the practitioner documents reviewed in the preparation of this Playbook, captures the same insight from a different direction: the success of any AI application depends ultimately on the trust that users and other stakeholders can place in it. Governance is the mechanism through which that trust is built and sustained. It is, in this sense, institutional infrastructure is as necessary to the functioning of a financial institution in the age of AI as cybersecurity, anti-money laundering compliance, or operational resilience

AI governance allows innovation to be sustained at scale, because it creates the conditions under which customers trust the systems, regulators trust the institutions, and boards can approve the use cases.

1.3 Where India's Financial Sector Stands

India's financial sector is at an early stage of a trajectory that other financial systems have already begun to travel. The RBI's own survey of approximately 600 regulated entities, covering 90% of the sector's asset base, found that only 21% of institutions are implementing or developing AI in any form.⁵ Of those, most are using AI in low-stakes applications that have been available for several years: customer service chatbots, voicebots for transaction confirmation, internal analytics, and basic process automation. The chatbots that many institutions describe as "artificial intelligence" are, in most cases, rule-based systems operating on predefined scripts rather than systems that learn from data or generate responses. Less than 5% are experimenting with generative AI.

This picture is important for understanding what kind of governance challenge India's financial sector faces. The sector is building governance infrastructure before AI is widely deployed in the highest-stakes applications. Credit decisioning AI at scale, automated KYC rejection, fraud detection systems with account-level consequences and these are the deployments for which governance infrastructure matters most, and they have not yet arrived at most regulated entities. The window to build ahead of the curve is this present moment.

The trajectory is visible from the experience of financial systems that have moved further along it. Central banks in the United Kingdom, United States, Canada, and the European Union have integrated AI into economic forecasting, stress testing, and internal research functions.⁶

Commercial banks in those jurisdictions have deployed AI at scale in credit underwriting, fraud detection, and customer relationship management. Each of those deployments has generated governance questions about bias encoded in training data, about model drift going undetected for months, about vendor systems whose inner workings the deploying institution could not inspect and now India's sector is beginning to encounter for the first time.

The cases that have emerged from earlier deployments are instructive.

Between 2005 and 2019, approximately 26,000 parents in the **Netherlands** were wrongly accused of making fraudulent childcare benefit claims by the Tax and Customs Administration, through an automated risk-scoring algorithm that discriminated on the basis of nationality and ethnicity. Families were driven into severe financial hardship, and over 1,100 children were removed from their families. The scandal caused the resignation of the entire Dutch government in January 2021.⁷

In **Australia**, an automated welfare debt recovery programme had extensive formal documentation.⁸ The documentation did not translate into genuine oversight: critical warnings did not reach people with authority to act, and the programme caused serious harm to hundreds of thousands of people before it was shut down. An Aus\$2.4 billion in total redress and a Royal Commission followed.

4. For reference, see <https://deeploy.ai/>.

5. For reference, see *RBI Framework for Responsible and Ethical Enablement of Artificial Intelligence*, August 2025, p. 32.

6. For instance, the Bank of England's own written evidence to the UK Parliament's Treasury Committee, published 2026, confirms: "The Bank uses AI for predictive analytics, the study of non-linear interactions between variables, and analysis of larger and richer data sets, which can potentially help forecast GDP growth and bank distress." For more, see *Response to TSC inquiry report on AI in financial services*.

7. For reference, see - *Netherlands tax authorities wrongly accuse 26,000 families of fraud*, January 2021, updated May 2024.

8. For reference, see *Government to pay further \$475m to robodebt victims in class action settlement*, SBS News, September 2025.

The lesson from both cases is the same one: the Finance Minister's meeting and the RBI's Draft Guidance are signalling, from different directions, that **governance is not a consequence of having an AI problem, it is what prevents the AI problem from occurring.**

Deployment Without Governance	Governance for Risk Mitigation
Netherlands – Childcare Benefits Scandal Predictive models disproportionately targeted families based on biased historical patterns, resulting in wrongful fraud accusations and severe social harm.	Data governance and bias testing to identify discriminatory outcomes before deployment; human oversight for high-impact decisions; continuous monitoring to detect unfair outcomes over time.
Australia – Robodebt Programme Automated welfare debt recovery relied on flawed assumptions. Despite extensive documentation, governance failures meant risks were not escalated, allowing harm to continue at scale.	Clear accountability for model decisions; effective governance processes that ensure documentation informs action; independent validation, risk escalation, and ongoing monitoring throughout deployment.

Table 1: Governance Lessons from Documented AI Failures

1.4 Understanding What Is Being Governed – Three Categories of AI

Governing AI effectively requires understanding what kind of AI is present in each system, because the failure modes differ and therefore the governance controls must differ. Three categories are relevant to India's financial sector.

Classical machine learning is the category most widely deployed in Indian financial institutions today. These systems learn patterns from historical data and apply those patterns to new situations. A system that scores transactions for fraud likelihood, a tool that routes customer queries based on content classification, a model that produces internal risk reporting from operational data, these are classical machine learning systems. Their outputs are structured: a score, a flag, a category. Their failure modes are well understood and follow recognisable patterns. They can embed bias present in their training data. They degrade over time as the real world diverges from the patterns they learned - a phenomenon called model drift, which is specific to AI and which standard technology monitoring does not detect. They require documentation, validation, monitoring, and human review of adverse outputs.

Generative AI produces open-ended outputs: text, images, code, and structured content generated in response to inputs. Unlike classical systems, generative systems produce variable outputs that can be accurate or wrong, and the wrongness is often indistinguishable from the accuracy without independent verification. The specific failure mode of hallucination is the confident production of factually incorrect information and it is unique to generative systems. A classical fraud model that flags the wrong transaction produces a wrong classification. A generative AI system that tells a customer the incorrect procedure for exercising their rights under the DPDP Act, or invents product terms that do not exist in the contract, produces a confident, grammatically correct error that the customer has no obvious way to detect. The RBI Draft Guidance on Model Risk Management, Paragraph 54(2), requires specific control boundaries targeting this failure mode in generative systems deployed by regulated entities.

Agentic AI is another emerging category that takes actions rather than producing outputs for human review. A system that drafts a customer communication and queues it for a human to send is generative. A system that drafts the communication, sends it, updates the customer record, and initiates a downstream payment instruction without a human approving each step is agentic. Risk in agentic systems compounds with each autonomous step because an error early in the chain can propagate through every subsequent action. Governance of agentic systems must address the boundaries of autonomous action and the human checkpoints within and around each automated chain.

Understanding which category is present in each deployed system is the starting point for proportionate governance. A risk framework that treats a confirmatory voicebot and an autonomous account management agent as equivalent governance problems will systematically underprotect the higher-risk system while over-specifying controls for the lower-risk one.

AI Category	What it does	Sample use cases of AI in financial sector	Primary governance risk	Key governance controls
Classical Machine Learning	Learns patterns from historical data to make predictions or classifications.	Fraud detection, credit scoring, transaction monitoring	Bias, model drift, inaccurate predictions	Model validation, data governance, performance monitoring, human review
Generative AI	Creates new content such as text, code, or documents in response to prompts.	Customer support chatbot, document drafting, policy summarisation	Hallucinations, misinformation, inconsistent outputs	Prompt governance, output verification, human oversight, usage guardrails
Agentic AI	Performs tasks and takes actions with limited or no human intervention.	Automated claims processing, workflow orchestration, account servicing	Autonomous errors, cascading failures, loss of control	Human approval checkpoints, authority limits, monitoring, audit trails, kill switches

Table 2: Three Categories of AI and Their Respective Use Cases, Risks, and Controls

1.5 The AI Lifecycle and Governance Across Every Stage

Understanding AI governance also requires understanding how AI systems come into existence, operate, change, and eventually cease to operate because governance is not just a gate that is passed at deployment or a one-time check, it is a continuous process that spans every stage of a system's life. The RBI Draft Guidance on Model Risk Management makes this explicit, requiring regulated entities to implement a governance framework covering all stages of the model lifecycle.

Five stages define the AI lifecycle in practical terms, and each carries specific governance obligations.

At the **design and assessment stage**, the most important governance action is a risk classification exercise, completed before development or procurement begins. Understanding the risk tier of the proposed system before resources are committed determines the governance intensity required throughout the lifecycle. The Risk Scoring tool in Chapter 5 of this Playbook is designed for use at this stage.

At the **data and development stage**, the training data used to build or fine-tune the system must be assessed for quality, representativeness, and compliance with the DPDP Act and other global norms. Bias testing belongs at this stage, as a preventive control. When bias is detected and corrected before deployment, the cost is modest. When it is discovered after a system has been in production for six months, the cost in customer harm, reputational damage, and regulatory exposure is significantly higher.

At the **validation and approval stage**, the system is assessed by a function independent of the team that built or procured it. For Tier 1 systems, the RBI Draft Guidance, Paragraph 18, requires Board or Risk Management Committee of the Board approval before deployment. This approval is a documented event (not an informal sign-off). It requires supporting evidence such as completed documentation, validation findings, bias test results, and confirmation that the controls required for this risk tier are in place.

At the **deployment and monitoring stage**, governance becomes a continuous operational function - performance metrics are tracked against thresholds, drift is detected through monitoring rather than discovered after harm has accumulated, human review mechanisms are triggered by defined conditions, logs are maintained that allow any specific decision to be reconstructed if it is challenged. The institutions that catch model drift at month three before thousands of affected decisions have compounded are the institutions with monitoring infrastructure in place from deployment. The institutions that discover it at month twelve are the institutions that treated deployment as the endpoint of governance.

At the **review and decommissioning stage**, material changes to a system require re-validation before they are applied to live transactions. When a system is retired, an orderly process manages data, documentation, and the transition of the business processes it was supporting.

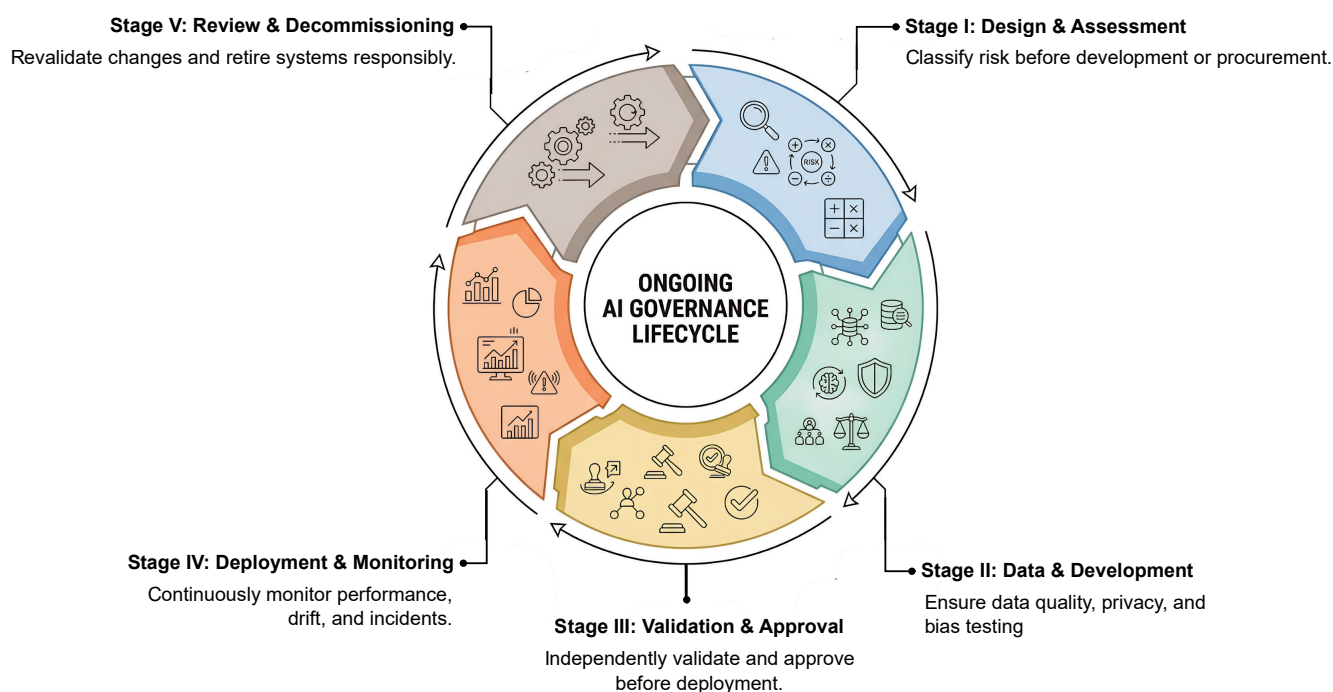


Figure 1: The Ongoing AI Governance Lifecycle

1.6 The Global Setup and What India Can Learn From It

India's regulatory development on AI governance is part of a global arc that has accelerated considerably over the past two years. The European Union's AI Act entered into force in 2024 as the first comprehensive binding regulation of AI, classifying systems by risk and imposing mandatory controls on high-risk applications. ISO/IEC 42001, the international management system standard for AI, was published and adopted by institutions in multiple jurisdictions as a governance framework. Singapore's Model AI Governance Framework and AI Verify matured into practical self-assessment and testing tools. Canada's Algorithmic Impact Assessment demonstrated how criteria-based risk scoring could operationalise proportionality without requiring legislative categories.

Chapter 3 of this Playbook examines these frameworks in detail. What is relevant here is the signal they send collectively. Despite their differences in legal form, enforcement mechanism, and institutional origin, every major framework converges on the same governance requirements: clearly assigned accountability, risk-proportionate controls, documentation as the governance mechanism itself, human oversight for high-impact decisions, and continuous monitoring across the lifecycle.

This convergence reflects the accumulated learning of jurisdictions, institutions, and researchers who have studied what happens when AI systems operate without adequate governance. The Dutch child benefit case, the Australian Robodebt programme, and numerous other documented failures all point to the same governance gaps: accountability that could not be located, documentation that did not reflect operational reality, monitoring that did not detect problems until harm had accumulated, and human oversight that was nominal rather than substantive.

India's position in this global arc is one of comparative advantage. The sector is building governance infrastructure before high-stakes AI is widely deployed, with the benefit of those earlier lessons, a maturing domestic regulatory framework, and a body of international practice to draw on. The RBI FREE-AI Framework reflects that inheritance: its seven sutras are not isolated Indian principles but a synthesis of the governance logic that global experience has validated.

India's financial sector needs a translation layer between AI principles and AI practice - that is what this Playbook attempts to provide.

1.7 What This Playbook Does

This Playbook is a direct response to RBI FREE-AI Recommendation 26, which calls for the development of an AI compliance toolkit for regulated entities. It is grounded in a comparative analysis of twenty-five governance frameworks and regulatory instruments. It has been tested through a closed practitioner roundtable in New Delhi in April 2026 and through structured one-on-one conversations with risk officers, compliance leads, and practitioners across banks and fintechs across India.

The Playbook does not introduce new governance obligations. The obligations already exist, distributed across the FREE-AI Framework, the Draft Guidance on Model Risk Management, the DPDP Act, and the other instruments examined in Chapter 2. What the Playbook provides is the operational translation of those obligations into four tools that any regulated financial institution can begin using immediately.

Chapter 2 maps the regulatory landscape including the legal and regulatory instruments that together define AI governance obligations for India's financial sector. **Chapter 3** examines the global frameworks that inform international practice and identifies the five convergence signals that underpin the toolkit's design. **Chapter 4** works through six concrete examples of the translation from governance principle to institutional practice. **Chapter 5** presents the four tools themselves.

The tools are designed to be used in sequence: an institutional maturity assessment, a model documentation template, a risk scoring instrument, and a vendor due diligence checklist. Together, they provide the governance infrastructure that the Mythos episode, the Finance Minister's instructions, and the RBI's regulatory development are all pointing toward. The institutions that build this infrastructure now, before high-stakes AI arrives at scale, will be in a materially different position from those that wait.

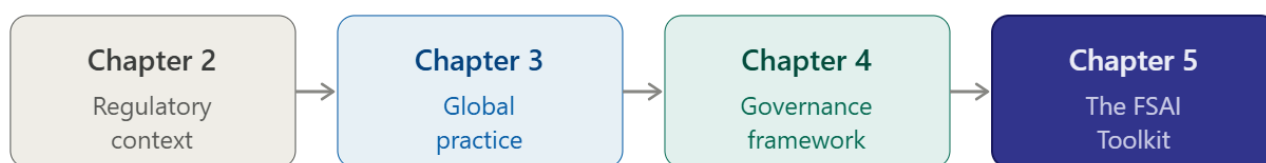


Figure 2: Structure and Chapter Overview

USE CASE A: ALTERNATIVE CREDIT SCORING & CREDIT UNDERWRITING

WHAT THE USE CASE IS AND WHY IT EXISTS

Traditional credit underwriting relies on bureau-based credit histories, excluding many “thin-file” or credit-invisible borrowers such as gig workers, young adults, MSMEs, and informal-sector participants. AI-based alternative credit scoring uses non-traditional data, including transaction histories, cash flows, utility payments, GST data, and digital behaviour, to assess creditworthiness. This enables lenders to expand credit access while improving risk assessment in underpenetrated markets.

INSTITUTIONS IMPLEMENTING IT

Institution	Type	Country	Specific implementation
UOB (with Element AI)	Bank	Singapore	Piloted Veritas Phase 1 fairness-assessment methodology specifically on a credit-risk-scoring model
HSBC	Bank	UK/Global	Participated in Veritas FEAT/Toolkit integration pilot across governance framework (credit scoring in scope)
DBS, OCBC, Standard Chartered, BNY Mellon, Singlife	Banks/insurer	Singapore/Global	Piloted integration of FEAT/Veritas fairness methodology into existing model governance
Google (Google Pay)	Fintech/BigTech	India	Applied FEAT Assessment Methodology to its fraudulent-payment detection system in the Indian market (adjacent fraud use case using the same fairness-assessment toolkit)

Swiss Re (with Accenture)	Insurer	Global	Developed and applied a fairness/transparency assessment specifically for its predictive AI-based underwriting model
CredoLab	Fintech vendor	Global (emerging markets focus)	Commercial embedded-metadata alternative-scoring tool used by lending partners
Zopa	P2P lender	UK	"Borrowing Power" tool (launched 2019) uses ML to decompose and explain a customer's credit profile
Various Indian NBFCs/fintech LSPs	NBFC/fintech	India	AI-based bank-statement analytics and GSTN-integrated MSME underwriting (RE + LSP structure under RBI's Digital Lending Directions)
Various US mortgage/consumer lenders	Banks/non-bank lenders	US	ML-based underwriting models, per CFPB Circular 2023-03 and 2026-03; CFPB's Winter 2025 Supervisory Highlights flagged models with 1,000+ variables at unnamed institutions for disparate-impact review
Apple Card / Goldman Sachs	Bank/BigTech partnership	US	Subject of CFPB enforcement (Oct 2024): \$25m fine (Apple), \$45m fine (Goldman Sachs) tied to algorithmic-decisioning failures in the Apple Card credit product

CURRENT PILOTS AND FUTURE DIRECTION

- **MAS:** moving from voluntary FEAT/Veritas guidance toward binding AI Risk Management Guidelines (consultation published November 13, 2025) - credit scoring will likely be one of the first use cases subject to mandatory fairness/governance testing rather than voluntary self-assessment.
- **RBI:** the algorithmic-fairness-audit recommendation from the original Working Group on Digital Lending remains "accepted in principle but requiring deliberation" - worth monitoring for when/if it converts into a binding directive, since India's regulatory hierarchy currently has no binding algorithmic-audit mandate for credit-scoring models, only a data-governance and disclosure regime.
- **US:** CFPB continues tightening expectations around AI explainability and adverse-action disclosures, with increasing scrutiny of complex underwriting models.
- **EU:** creditworthiness AI remains classified high-risk under the EU AI Act, meaning conformity-assessment obligations will phase in as the Act's compliance timeline matures.
- **Fintech/MSME direction:** the GSTN-integrated, real-time cash-flow-based underwriting approach used by Indian digital lenders (Input Tax Credit/Output Tax Liability analysis) represents a maturing sub-trend - moving away from static bureau/balance-sheet data toward continuously updated transactional data as the scoring substrate.

– Black Dot Public Policy Advisors

CHAPTER 2

INDIA'S REGULATORY LANDSCAPE

Over the past two decades, financial institutions have incorporated internet banking, digital payments, cloud infrastructure, and increasingly sophisticated data analytics into their operations. Each technological shift has been accompanied by new governance expectations, often through the adaptation of existing legal and regulatory frameworks rather than entirely new legislation.

AI should therefore be understood as part of a continuing evolution in India's financial sector. Although India does not yet have a dedicated AI law, financial institutions deploying AI are already subject to a range of obligations that collectively shape how these systems should be governed. This chapter examines that regulatory landscape, demonstrating how existing legal, regulatory, and supervisory instruments together provide the foundations for AI governance in India's financial sector.

First, the chapter contextualizes the adoption of AI in India's financial sector as well as the rationale for regulation. Further, it examines the applicability of RBI FREE-AI Framework, DPDP Act 2023 and Rules 2025, IT Act 2000, Consumer Protection Act 2019, Bharatiya Nyaya Sanhita 2023, and IT Intermediary Rules to BFSI. Then, it positions the Reserve Bank of India's *Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI)* as the sector's primary anchor for AI governance. The chapter concludes by translating these regulatory expectations into the practical governance responsibilities that regulated entities should understand before implementing the governance tools presented later in this Playbook.

2.1 The Regulatory Context for AI

Globally, AI is being considered increasingly for integration into core business processes, supporting functions such as credit underwriting and risk assessment, fraud detection, anti-money laundering (AML) monitoring, know-your-customer (KYC) verification, customer service through chatbots and virtual assistants, document processing, grievance redressal, and internal research.¹ AI systems supporting credit decisioning and underwriting are already deployed at scale in larger institutions,² illustrating that some of the sector's most consequential use cases are among the earliest to mature. While in India, as discussed in Chapter 1.3 (Where India's Financial Sector Stands), AI adoption across India's financial sector remains at an early stage.

However, governance cannot wait for universal adoption; it must evolve alongside deployment. In fact, early checks on risk and compliance may even help scale deployment by ensuring systems are reliable and verified at the outset.

The absence of dedicated AI legislation has sometimes led to the perception that AI systems currently operate in a *regulatory grey area*. This assumption overlooks the fact that financial services are among the most comprehensively regulated sectors in India, supported by a well-established institutional and legal architecture that has evolved over several decades.

1. For example, see The Bonadio Group – [Practical Applications of AI in Financial Services](#) (January 2026) for the Celent/Zest AI survey on lender adoption rates, and The CG AI Group – [Banking AI 2026: From Pilots to Production at Scale](#) (May 2026), for HSBC, Citi, UBS, DBS, ING deployment data and cost reduction figures.

2. See, [The intelligent bank at scale – AI leadership and accountability in global banking](#), The Asian Banker, March 2026, for JPMorgan's illustration of the scale and the structural shift in banking operating models around AI.

Beginning with the Reserve Bank of India (RBI) and the Banking Regulation Act, 1949, India's financial regulatory ecosystem has progressively expanded through the establishment of sector-specific regulators such as the Securities and Exchange Board of India (SEBI), the Insurance Regulatory and Development Authority of India (IRDAI), and the Pension Fund Regulatory and Development Authority (PFRDA). Successive waves of financial innovation and market reform have been accompanied by new regulatory frameworks governing banking, securities, insurance, payments, digital infrastructure, consumer protection, cybersecurity, anti-money laundering, and financial stability.

AI systems deployed within regulated entities operate within an established framework of sectoral regulation, data protection obligations, cybersecurity requirements, consumer protection laws, and supervisory expectations. Currently, the emergence of AI has therefore not created an entirely new regulatory challenge, but has instead required existing governance principles to be interpreted and applied to increasingly autonomous and data-driven systems.

As a result, financial institutions adopting AI are already required to navigate a range of legal and regulatory concerns. **This Playbook and its accompanying toolkits help translate these existing obligations into practical governance structures, processes, and controls that enable responsible AI adoption within India's financial sector.**

Earlier waves of technological innovation, including electronic banking, payment systems, outsourced technology services, cloud computing, cybersecurity, and model-based decision-making, were similarly governed through the evolution of existing legal principles and supervisory mechanisms rather than entirely new legal regimes. Understanding India's AI governance landscape therefore requires looking beyond the search for a single AI law. Governance obligations are distributed across multiple legal and regulatory instruments, each addressing a different dimension of responsible AI deployment, from data governance and consumer protection to cybersecurity, operational resilience, accountability, and institutional oversight. Together, these instruments form the regulatory environment within which AI systems are designed, deployed, and monitored in India's financial sector.

2.2 Mapping India's AI Regulatory Architecture

The current regulatory landscape distributes governance responsibilities across multiple legal and institutional instruments, each addressing a distinct dimension of risk. Together, these instruments create a functional governance architecture that already shapes how AI systems should be designed, deployed, and governed within regulated financial institutions.

At the centre of this architecture is the **Reserve Bank of India's Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI)**.³ As the first sector-specific framework dedicated to AI in financial services, FREE-AI articulates the Reserve Bank's expectations for the responsible adoption of AI by regulated entities. It establishes principles for governance, accountability, risk management, consumer protection, and independent assurance, while recognising that AI governance must be embedded within existing institutional risk management and compliance functions. The following section discusses in greater detail how the FREE-AI Framework serves as the primary sectoral reference point for this Playbook.

The RBI Framework, however, does not operate in isolation. Many of the governance expectations it articulates derive their legal force from existing legislation. **The Digital Personal Data Protection (DPDP) Act, 2023**⁴ and the accompanying **DPDP Rules, 2025**⁵ establish the legal foundations for responsible data governance. Together, they regulate how personal data may be collected, processed, stored, retained, and secured, while introducing obligations relating to consent, purpose limitation, security safeguards, breach reporting, and the rights of data principals. Given that many AI systems rely extensively on personal data for training, profiling, and decision-making, these obligations form a central pillar of AI governance within the financial sector.

3. For reference, see [RBI Framework for Responsible and Ethical Enablement of Artificial Intelligence](#), August 2025.

4. For reference, see [Digital Personal Data Protection \(DPDP\) Act, 2023](#), The Gazette of India, 11 August 2023.

5. For reference, see [Digital Personal Data Protection Rules, 2025](#).

Regulatory Instrument	Primary Governance Function	Application
RBI FREE-AI Framework	AI governance, risk management, accountability, oversight	Across the entire AI lifecycle
DPDP Act + DPDP Rules	Data governance, privacy, consent, security, data principal rights	Data collection, processing, model development, deployment, and monitoring
Information Technology Act, 2000 +	Cybersecurity, digital infrastructure, information security	AI systems, digital infrastructure, and operational resilience
Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026	Platform governance, user communication, grievance redressal	Customer-facing digital platforms and intermediary services
Consumer Protection Act, 2019	Consumer protection, transparency, fairness, redressal	AI-enabled products, services, and customer interactions
Bharatiya Nyaya Sanhita, 2023	Liability, fraud, criminal misuse, enforcement	Post-incident accountability and enforcement

Table 3: Regulatory Foundations of AI Governance in India's Financial Sector

Alongside data governance, institutions must continue to satisfy broader obligations relating to cybersecurity, digital resilience, and operational security. The **Information Technology Act, 2000**⁶ provides the foundational legal framework governing electronic records, computer systems, and liability arising from failures to maintain reasonable security practices. While enacted well before the emergence of contemporary AI systems, its provisions remain highly relevant wherever AI applications process sensitive information or operate within critical digital infrastructure. AI governance therefore builds upon, rather than replaces, existing expectations around cybersecurity and information security.

The updated **Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules 2026**⁷ introduce obligations relating to user communication, grievance redressal, and due diligence where financial services are delivered through digital intermediary platforms. Recent amendments further recognise “synthetically generated information” (including AI-generated and AI-modified content) and strengthen intermediary responsibilities through proactive due diligence, prompt action against unlawful AI-generated content, mandatory labelling and provenance requirements for lawful synthetic content, and the deployment of appropriate technical safeguards. Although these Rules do not regulate AI decision-making directly, they reinforce governance expectations around transparency, accountability, and responsible deployment where AI-enabled digital services interact with users.

Consumer-facing AI systems are further shaped by legislation designed to protect individuals from unfair or harmful practices. The **Consumer Protection Act, 2019**⁸ establishes safeguards against unfair trade practices, misleading representations, and deficiencies in service. As AI increasingly influences customer interactions, financial product recommendations, and automated decision-making, institutions must ensure that the deployment of AI does not compromise fairness, transparency, or consumer trust.

6. For reference, see [Information Technology Act, 2000](#), The Gazette of India, 9 June 2000.

7. For reference, see [Information Technology \(Intermediary Guidelines and Digital Media Ethics Code\) Rules, 2021](#).

8. For reference, see [Consumer Protection Act, 2019](#), The Gazette of India, 24 April 2020.

The regulatory architecture is complemented by the **Bharatiya Nyaya Sanhita, 2023**,⁹ which provides the broader framework for addressing criminal conduct such as fraud, impersonation, forgery, and deception. Although it does not regulate AI systems directly, it becomes relevant where AI-enabled systems are misused to facilitate financial offences or where automated processes contribute to unlawful conduct. In this sense, the Bharatiya Nyaya Sanhita forms part of the wider liability environment surrounding AI deployment, emphasising that institutions remain accountable for harms arising from the technologies they implement.

Viewed individually, each of these instruments addresses particular aspects of AI governance. Collectively, however, they establish a layered governance ecosystem in which different legal and regulatory mechanisms reinforce one another. Data protection law governs the responsible use of personal information; cybersecurity legislation safeguards digital infrastructure; consumer protection law protects individuals from harmful outcomes; criminal law addresses misuse and liability; and sectoral regulation provides institution-specific governance expectations. Rather than creating parallel or competing obligations, these instruments operate together to shape the governance environment within which AI systems are developed, procured, deployed, and monitored.

Compliance with AI governance cannot be achieved by reference to a single law or framework. Instead, institutions must understand how multiple regulatory obligations interact across the AI lifecycle, from data collection and model development to deployment, oversight, monitoring, incident response, and consumer redress. It is this integrated approach that the remainder of this chapter explores and that the governance tools presented later in this Playbook are designed to operationalise.

2.3 The RBI FREE-AI Framework: India's Sectoral Anchor for AI Governance

While India's broader AI governance landscape is distributed across multiple legal and regulatory instruments, the Reserve Bank of India's Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI) provides the clearest articulation of how AI governance should be understood within the financial sector. Published in August 2025, the framework represents India's first comprehensive sector-specific approach to AI governance and applies across the Reserve Bank's regulated entities, including scheduled commercial banks, non-banking financial companies (NBFCs), payment system operators, and other financial institutions within its regulatory ambit.

FREE-AI Framework

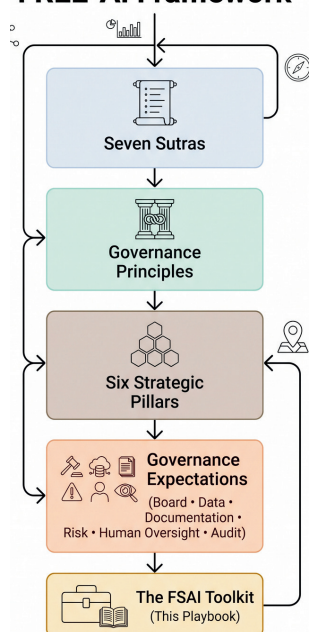


Figure 3: Structure of RBI's FREE-AI Framework

Rather than focusing exclusively on model performance or technical controls, it positions AI within broader organisational systems of accountability, risk management, consumer protection, and operational resilience

This approach is reflected in the framework's seven foundational principles, or **Sutras (Core Principles)**:

1. **Trust is the foundation**
2. **People first**
3. **Innovation over restraint**
4. **Fairness and equity**
5. **Accountability**
6. **Understandable by design**
7. **Safety, resilience, and sustainability**

Collectively, these principles establish the normative foundation for responsible AI adoption. Rather than prescribing detailed technical standards, they articulate the values that should guide institutional decision-making throughout the AI lifecycle. The emphasis is not on limiting innovation, but on ensuring that innovation is accompanied by appropriate governance, transparency, and accountability.

9. For reference, see [Bharatiya Nyaya Sanhita, 2023](#), The Gazette of India, 25 December 2023.

The framework translates these principles into practice through six strategic pillars that together provide an operational model for AI governance. The *Innovation Enablement Framework* encourages responsible experimentation through enabling infrastructure, institutional policy development, and capacity building. The *Risk Mitigation Framework* focuses on governance structures, consumer protection, and independent assurance mechanisms, recognising that effective AI deployment requires continuous oversight rather than one-time approval. Together, these pillars demonstrate how ‘innovation’ and ‘risk mitigation’ must be advanced hand in hand.

Innovation Enablement Framework:

1. **Infrastructure:** Shared data infrastructure and AI innovation sandboxes to democratize access
2. **Policy:** Clear institutional and national AI policies with proportionate compliance
3. **Capacity:** AI skills development and knowledge sharing on fairness and explainability

Risk Mitigation Framework:

4. **Governance:** Board-level oversight, reporting mechanisms, and structured AI lifecycle management
5. **Protection:** AI disclosure, grievance mechanisms, and protections for vulnerable users
6. **Independent Assurance:** Third-party audits, bias testing, and impact assessments

For regulated entities, the significance of the FREE-AI Framework lies in the governance expectations it establishes. These include board-approved AI policies that define institutional governance structures and risk appetite; formal data governance arrangements covering data quality, consent management, and security; documented lifecycle management for AI systems from development through deployment and retirement; governance processes that evaluate fairness, explainability, and potential consumer impact before implementation; transparency measures that inform customers when they are interacting with AI systems; and mechanisms for monitoring, incident reporting, and continuous review. The framework also places particular emphasis on independent assurance through activities such as audits, testing, and validation, reinforcing the principle that AI systems must remain subject to ongoing scrutiny after deployment.

Importantly, the framework recognises that *not all AI systems present the same level of risk*. Governance expectations are therefore proportionate to the potential impact of the system being deployed. High-impact applications, particularly those involving credit decisioning and loan underwriting, require more rigorous governance measures, including enhanced testing, detailed documentation, explainability, human oversight, and comprehensive audit processes. This proportional approach aligns closely with emerging international practice and reflects a broader shift towards risk-based AI governance.

Perhaps the most significant feature of the FREE-AI Framework for the purposes of this Playbook is **Recommendation 26**, which calls for the development of an AI compliance toolkit for regulated entities. The recommendation recognises that while governance principles are increasingly well established, financial institutions require practical mechanisms for translating those principles into day-to-day organisational practice. Policies alone do not operationalise governance; institutions also need structured processes, documentation templates, risk assessment methodologies, accountability mechanisms, and implementation guidance that can be integrated into existing governance, risk, and compliance (GRC) functions.

This Playbook is also designed as a direct response to recommendation 26. Rather than introducing new governance obligations, it provides practical tools to help regulated entities implement the expectations articulated within the FREE-AI Framework. The maturity assessment, model documentation template, AI risk scoring methodology, and vendor due diligence checklist presented later in this Playbook are intended to support institutions in operationalising governance expectations through existing organisational processes. In this sense, the FREE-AI Framework serves not only as the principal regulatory anchor for AI governance in India's financial sector, but also as the conceptual foundation upon which this Playbook is built.

As this Playbook was being finalised in June 2026, the Reserve Bank of India released the **Draft Guidance on Regulatory Principles for Model Risk Management**¹⁰ for public consultation, signalling the continued evolution of AI governance expectations within the financial sector. Applicable to a broad range of regulated entities and covering all models, including those based on AI and machine learning, the draft guidance introduces strengthened expectations around model governance, independent validation, lifecycle management, documentation, risk-based oversight, explainability, bias testing, human oversight, and accountability. Although under consultation at the time of writing, the guidance reinforces the central argument of this chapter: India's approach to AI governance is evolving through sector-specific regulatory frameworks that operationalise governance expectations, rather than through a standalone AI law.

2.4 From Legal Requirements to Governance Responsibilities

The preceding sections have shown that India's AI governance landscape is distributed across existing legal and regulatory instruments rather than a single AI-specific law, which may apply to the entire financial sector and others. For regulated entities, however, compliance must ultimately translate into governance practices embedded within organisational policies, processes, documentation, oversight, and day-to-day decision-making.

Rather than creating isolated obligations, these instruments collectively establish governance responsibilities across five interconnected themes that span the AI lifecycle.

Data governance and privacy: The DPDP Act and Rules require institutions to govern the collection, use, protection, retention, and processing of personal data. Because AI systems depend heavily on data, effective data governance forms the foundation of responsible AI deployment.

Cybersecurity and digital resilience: The Information Technology Act requires institutions to protect AI systems, supporting infrastructure, and sensitive information through appropriate security controls, resilience measures, and incident response capabilities.

Consumer protection and transparency: The Consumer Protection Act, together with the RBI FREE-AI Framework, reinforces that AI-enabled financial services must remain fair, transparent, and accountable, with appropriate mechanisms for explanation and customer redress.

Platform governance and grievance redressal: The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules require institutions delivering AI-enabled services through digital platforms to maintain effective user communication, grievance mechanisms, and human accountability.

Liability and accountability: The Bharatiya Nyaya Sanhita and existing regulatory obligations make clear that institutions remain responsible for the outcomes of AI systems, including those developed or procured from third parties. Responsibility cannot be delegated to technology.

Governance Theme	Primary Framework(s)	Recommended Institutional Action
Data Governance	DPDP Act and Rules	Establish consent, data quality, retention, security
Cybersecurity	IT Act	Protect systems, infrastructure, incident response
Consumer protection	Consumer Protection Act	Ensure fairness, transparency, redress
Platform governance	IT Rules	Grievance mechanisms, user communication
Accountability	Bharatiya Nyaya Sanhita	Maintain oversight, assign responsibility, investigate incidents
The RBI FREE-AI Framework acts as an overarching sectoral governance framework across each of these themes.		

Table 4: Summary of Regulatory Frameworks and their Recommendations

2.5 What This Means for Regulated Entities

As AI adoption expands across the financial sector, regulated entities should focus on operationalising governance through the following priorities:



Figure 4: Institutional Priorities for Responsible AI Governance

- **Establish clear governance and accountability.** AI governance should be integrated into existing governance, risk, and compliance (GRC) structures, with clearly assigned responsibilities, board oversight, defined governance roles, and documented accountability across the AI lifecycle.
- **Strengthen data governance and risk management.** Institutions should implement robust data governance practices, maintain comprehensive documentation, classify AI systems according to their potential impact, and apply governance controls that are proportionate to the level of risk.
- **Embed continuous monitoring and human oversight.** Governance should extend beyond deployment through ongoing monitoring for model drift, unintended outcomes, and emerging risks, supported by meaningful human review and accessible grievance mechanisms for high-impact decisions.
- **Strengthen third-party and vendor governance.** Institutions remain accountable for AI systems procured from external vendors and should conduct appropriate due diligence, understand associated risks, establish contractual safeguards, and maintain oversight throughout the vendor relationship.
- **Build evidence for audit and regulatory assurance.** Governance should be demonstrable through documentation, decision records, risk assessments, monitoring activities, and audit trails that enable regulators, boards, auditors, and assurance functions to verify responsible AI practices.

Finally, **institutions should recognise that AI governance is becoming increasingly evidence-based.** Regulators, auditors, boards, and internal assurance functions are likely to expect that governance processes exist and they can also be demonstrated through documentation, decision records, risk assessments, monitoring activities, and audit trails. In this environment, governance is measured by evidence that organisations can produce to demonstrate responsible practice.

Given this landscape, financial institutions do not need to wait for future AI-specific legislation before strengthening their governance arrangements because as established, much of the legal and regulatory foundation already exists. The challenge is to now operationalise these expectations in a manner that is practical, proportionate, and aligned with existing governance structures.

The remaining chapters of this Playbook are designed to support precisely this transition, from understanding governance obligations to implementing them in practice. Building on the regulatory foundation outlined in this chapter, the governance framework and toolkit presented in the following sections provide structured mechanisms for assessing organisational maturity, documenting AI systems, classifying AI risk, and strengthening vendor oversight. **The FSAI Toolkit, hence, aims to help regulated entities translate regulatory expectations into consistent, auditable, and day-to-day governance practices.**

USE CASE B: FRAUD DETECTION (TRANSACTION & PAYMENT FRAUD)

WHAT THE USE CASE IS AND WHY IT EXISTS

AI-powered fraud detection identifies suspicious payment, card, account takeover, and identity fraud in real time by analysing customer behaviour and transaction patterns. Unlike static rule-based systems, AI continuously adapts to evolving fraud techniques, improving detection accuracy while reducing false positives.

INSTITUTIONS IMPLEMENTING IT

Institution	Type	Country	Specific implementation
Mastercard	Card network	Global	Decision Intelligence / Decision Intelligence Pro - real-time transaction scoring across its network, generative AI (RNN + transformer) model trained on ~125-160 billion annual transactions
Visa	Card network	Global	Own network-level AI fraud detection; has invested over \$10bn in technology over five years and reports blocking \$40bn in fraudulent payments (per secondary industry source citing Visa's own figures)
PayPal	Payments/fintech	Global	AI-driven rules engine supplemented by neural-network-based anomaly detectors, injecting real-time risk scores into the authorization workflow

HSBC	Bank	UK/Global	AI-driven Dynamic Risk Assessment fraud system
DBS Bank	Bank	Singapore	AI-powered compliance/fraud system (part of its broader 350+ AI use-case portfolio)
JPMorgan Chase	Bank	US	AI-powered fraud detection in live use for 3+ years, monitoring live transactions
HDFC Bank	Bank	India	Proprietary real-time fraud-monitoring engine built in-house on own transaction data, linked to a single customer ID and governance layer, developed alongside its "Neev" GenAI platform
ICICI Bank	Bank	India	Enterprise Fraud Risk Management (EFRM) - real-time ML system analysing customer behaviour patterns and transaction anomalies
State Bank of India, Axis Bank	Banks	India	Run fraud engines and chatbots per industry comparative reporting (no independently sourced institution-specific metrics found)
Worldpay	Payments processor	Global	Uses Mastercard's Transaction Fraud Monitoring/Decision Intelligence for holistic merchant risk monitoring (per client testimonial)

CURRENT PILOTS AND FUTURE DIRECTION

- **RBI FREE-AI implementation phase:** as of the report's release (August 2025), the 26 recommendations are not yet binding regulation - RBI has proposed a phased, "self-regulation first, binding rules later" approach, meaning fraud-detection-specific AI audit requirements are still in the design/consultation phase, not enforced law.
- **BIS Project Aurora:** remains in the proof-of-concept stage using synthetic data.
- **Generative AI in fraud modelling:** Mastercard's adoption of transformer-based generative AI signals a broader industry shift toward more sophisticated fraud detection models.
- **Cross-border regulatory convergence:** RBI explicitly aligns its FREE-AI approach with the Bank of England/FCA's AI/ML discussion paper and OECD's "Regulatory Approaches to Artificial Intelligence in Finance" report - suggesting fraud-detection governance standards may converge internationally faster than credit-scoring standards have, since fraud detection carries less of a fair-lending/discrimination controversy than underwriting does.

– Black Dot Public Policy Advisors

CHAPTER 3

GLOBAL FRAMEWORKS FOR AI GOVERNANCE

The rapid adoption of artificial intelligence (AI) has prompted governments and international organisations around the world to develop frameworks for its governance. These range from legally binding regulations and international standards to risk management frameworks, technical guidance, ethical principles, and organisational maturity models. Together, they represent the most significant global effort to establish how AI systems should be governed throughout their lifecycle while enabling innovation, protecting individuals, and managing emerging risks.

Although these frameworks differ in their legal status, institutional context, and implementation mechanisms, they address many of the same governance challenges. They seek to answer common questions:



- Who should be accountable for AI systems?
- How should risk be assessed?
- What documentation should organisations maintain?
- When is human oversight required?
- How should AI systems be monitored once deployed?

Figure 5: Common Questions Across AI Governance Frameworks

Rather than viewing these frameworks as competing models, this chapter examines the different governance philosophies they represent and identifies the areas of convergence that have emerged across them. These shared governance principles form an important foundation for the governance framework and practical toolkit presented later in this Playbook.

3.1 Why Global Frameworks Matter

Why does an India-focused Playbook take into account international AI governance frameworks when regulated entities are primarily required to comply with Indian law? The answer is that AI governance is evolving globally, while AI systems, technologies, and supply chains increasingly operate across jurisdictions. Financial institutions in India may develop some AI systems internally, but they also procure software from global vendors, deploy cloud-based AI services, integrate third-party models, and increasingly participate in international financial and technology ecosystems. International frameworks also represent the most mature body of thinking currently available on AI governance. Understanding how AI governance is developing internationally is therefore becoming an important part of managing AI responsibly within India.

As discussed in Chapter 2, the RBI FREE-AI Framework incorporates governance concepts that have become increasingly prominent across global AI governance, including board-level accountability, lifecycle management, proportionate risk management, documentation, independent assurance, and human oversight. Rather than developing in isolation, India's approach is evolving alongside wider international discussions while adapting these principles to the country's regulatory architecture and institutional context.

This does not mean that Indian institutions should seek to replicate any single international framework. Legal systems, supervisory structures, organisational maturity, and institutional capacity differ significantly across jurisdictions, making direct transplantation both impractical and undesirable. Instead, the value of comparative analysis lies in understanding the governance logic that underpins different approaches, identifying areas of international convergence, and considering how these principles can be adapted to Indian financial institutions.

Accordingly, this chapter does not compare jurisdictions to determine which model India should follow. Rather, it examines how leading international frameworks approach AI governance, identifies the **common governance signals** that emerge across them, and highlights the practical lessons that inform the governance framework and operational toolkit presented in the remainder of this Playbook.

3.2 Global Approaches to AI Governance

Although governments and international organisations broadly agree that AI requires governance, they do not all govern it in the same way. Some rely on legally enforceable regulation, others emphasise organisational risk management, voluntary standards, management systems, ethical principles, or practical implementation guidance. Rather than representing competing models, these approaches illustrate different ways of addressing the same underlying governance challenge: ensuring that AI systems are deployed responsibly, transparently, and with appropriate oversight.

The frameworks discussed below are therefore **organised according to their underlying governance logic rather than their country of origin**. This makes it easier to understand the distinct contribution each framework makes to AI governance, while highlighting the complementary ideas that have shaped international practice.

A. Governing AI through Regulation and Risk Classification: The EU AI Act

The **European Union AI Act**¹ represents the most comprehensive and legally enforceable approach to AI governance currently in force. It regulates AI according to the level of risk associated with their intended use. Governance obligations increase as the potential impact of an AI system on individuals and society increases, making **risk proportionality** the organising principle of the framework.

A defining feature of the Act is that **risk is determined by how an AI system is used rather than the underlying model itself**. AI systems deployed in high-impact contexts, such as creditworthiness assessment and access to financial services, are classified as high-risk irrespective of the technology powering them.

Risk distinction is particularly relevant for financial institutions, where the same AI model may be used across both low-risk internal functions and high-impact customer-facing decisions.

The Act translates this risk-based approach into a set of legally enforceable governance obligations. High-risk systems are subject to requirements including risk management processes, technical documentation, data governance measures, human oversight, logging and traceability, conformity assessments, and post-deployment monitoring. Governance therefore extends across the entire AI lifecycle rather than focusing only on model development or deployment.

1. For reference, see <https://artificialintelligenceact.eu/>

Although the EU AI Act applies within the European Union, it has become an important global reference point for AI governance. Its structured allocation of responsibilities, emphasis on documentation, and risk-based regulatory model have influenced policymakers and organisations well beyond Europe.

For Indian financial institutions, the Act provides a useful benchmark for understanding how governance can be operationalised in high-impact sectors, even where its legal obligations do not directly apply.

B. Governing AI through Organisational Risk Management: The NIST AI Risk Management Framework

Unlike the EU AI Act, which establishes legally enforceable obligations, the **NIST AI Risk Management Framework (AI RMF)**² approaches AI governance as an organisational risk management discipline. Rather than prescribing fixed regulatory requirements or predefined categories of high-risk AI, it provides a flexible framework through which organisations can identify, assess, manage, and continuously monitor AI-related risks within their own institutional context.

The framework is organised around four interconnected functions - **Govern, Map, Measure, and Manage**, which together support governance across the AI lifecycle. These functions encourage organisations to establish governance structures and accountability mechanisms, understand the context in which AI systems operate, assess technical and societal risks, and implement ongoing monitoring and risk mitigation measures. Governance is therefore treated as an iterative process rather than a one-time compliance exercise.

A defining characteristic of the NIST framework is its contextual approach to risk. Instead of assigning risk categories in advance, organisations are expected to assess AI systems based on their intended purpose, deployment context, affected stakeholders, and potential impacts. This makes the framework highly adaptable across different sectors and use cases, while also placing greater responsibility on institutions to exercise informed judgement and maintain robust internal governance.

Although the framework is voluntary, it has become one of the world's most influential operational references for AI governance. Its emphasis on organisational responsibility, continuous improvement, and lifecycle risk management aligns closely with the needs of financial institutions, where AI systems evolve over time and governance must remain responsive to changing risks rather than relying solely on static compliance requirements.

C. Governing AI through Management Systems and Audit Readiness: ISO/IEC 42001

ISO/IEC 42001³ approaches AI governance through the lens of organisational management systems. Rather than focusing on legal compliance or risk assessment alone, it provides a structured framework for embedding AI governance into an organisation's everyday processes, responsibilities, and decision-making structures. The emphasis is on building governance that is systematic, repeatable, and capable of continuous improvement.

At the centre of the standard is leadership accountability. Organisations are expected to establish clear governance policies, define roles and responsibilities, integrate AI governance into broader organisational objectives, and ensure that appropriate resources, competencies, and oversight mechanisms are in place. Documentation is treated as a core governance requirement, supporting transparency, internal accountability, and organisational learning throughout the AI lifecycle.

The standard also places strong emphasis on governance processes rather than individual AI use cases. Organisations are expected to implement documented procedures for risk assessment, change management, internal audits, performance reviews, corrective actions, and continual improvement. In doing so, AI governance becomes part of an organisation's broader management system, similar to established standards for quality management or information security.

2. For reference, see <https://www.nist.gov/itl/ai-risk-management-framework>

3. For reference, see <https://www.iso.org/standard/42001>

Unlike regulatory frameworks, ISO/IEC 42001 derives its assurance through certification and audit rather than statutory enforcement. Organisations can demonstrate conformity through independent assessment, providing regulators, customers, and business partners with confidence that AI governance processes are consistently implemented and maintained. For Indian financial institutions, particularly those already operating within ISO-based management systems, the standard offers a practical pathway towards institutionalising AI governance and strengthening audit readiness as regulatory expectations continue to evolve.

D. Governing AI through Practical Implementation and Deployer Responsibility

While some AI governance frameworks focus primarily on regulation or organisational management systems, others concentrate on helping organisations implement responsible AI in day-to-day practice. Rather than prescribing detailed legal obligations, these approaches provide practical tools, structured assessments, and implementation guidance that enable organisations to translate governance principles into operational processes. A common feature across these frameworks is their emphasis on the responsibilities of organisations deploying AI systems, regardless of whether those systems are developed internally or procured from third parties.

Singapore Model AI Governance Framework and AI Verify⁴

Singapore's Model AI Governance Framework approaches AI governance from the perspective of responsible deployment. It recognises that many organisations adopt AI systems developed by external vendors and therefore places governance responsibility on the organisation deploying the system rather than solely on its developer. Risk is assessed through practical questions relating to decision impact, transparency, explainability, human oversight, and system robustness, making the framework highly implementation-oriented.

This practical orientation is reinforced through AI Verify, an open-source testing and assurance framework that enables organisations to assess AI systems against recognised governance principles. Rather than functioning as a regulatory certification scheme, AI Verify provides structured testing methodologies and self-assessment tools that help organisations demonstrate responsible AI practices in a transparent and consistent manner.

Canada's Algorithmic Impact Assessment⁵

Canada's Algorithmic Impact Assessment adopts a similarly practical approach but operationalises governance through structured risk scoring. Organisations complete a standardised questionnaire covering factors such as the potential impact on individuals, scale of deployment, rights implications, and degree of automation. The resulting risk score determines the level of governance required, with higher-risk systems triggering stronger expectations around documentation, human oversight, transparency, and independent review.

Rather than creating fixed categories of AI systems, the framework links governance obligations directly to the potential impact of each deployment, offering a flexible model that can be adapted across different organisational contexts.

Together, these frameworks demonstrate that effective AI governance depends not only on high-level principles but also on practical implementation mechanisms. Their focus on deployer accountability, structured risk assessment, third-party oversight, and usable governance tools is particularly relevant for Indian financial institutions, many of which deploy externally developed AI solutions while operating under diverse levels of organisational maturity.

4. For reference, see <https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf>

5. For reference, see <https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html>

E. Governing AI through Ethics and Institutional Readiness

While many AI governance frameworks focus on organisational controls and regulatory compliance, others take a broader view by emphasising the ethical principles, public institutions, and governance ecosystems required to support responsible AI adoption. Rather than prescribing operational governance processes, these frameworks establish the normative foundations and institutional conditions that enable trustworthy AI at national and societal levels.

UNESCO Recommendation on the Ethics of Artificial Intelligence and the Readiness Assessment Methodology (RAM)⁶

The UNESCO Recommendation on the Ethics of Artificial Intelligence provides the first globally agreed normative framework for AI governance. Adopted by all UNESCO Member States in 2021, it establishes a shared set of values centred on human rights, human dignity, fairness, transparency, accountability, privacy, sustainability, and meaningful human oversight. Although non-binding, it has significantly influenced international discussions on responsible AI and provides an ethical foundation for many subsequent governance initiatives.

To support implementation, UNESCO developed the **Readiness Assessment Methodology (RAM)**, which evaluates a country's preparedness for AI governance across legal, regulatory, institutional, social, economic, scientific, and technological dimensions. Rather than assessing the governance practices of individual organisations, RAM examines whether the broader governance ecosystem is equipped to support responsible AI adoption.

OECD AI Principles⁷

The OECD AI Principles similarly provide a high-level framework for trustworthy AI, promoting inclusive growth, human rights, transparency, robustness, security, and accountability. Alongside these principles, the OECD recommends that governments strengthen AI research, governance institutions, workforce capabilities, and international cooperation. Together, these recommendations position AI governance as both an organisational responsibility and a broader public policy challenge.

These frameworks demonstrate that effective AI governance depends not only on organisational controls but also on institutional readiness. As India continues to expand AI adoption across sectors, strengthening governance capacity, regulatory institutions, technical expertise, and implementation capability will be as important as developing new governance principles. This broader perspective provides important context for understanding how international governance approaches can inform India's evolving AI ecosystem.

6. For reference, see <https://www.unesco.org/ethics-ai/en/ram>

7. For reference, see <https://www.oecd.org/en/topics/sub-issues/ai-principles.html>

3.3 International Convergence: Five Governance Signals

The frameworks discussed in the previous section differ considerably in their legal status, institutional origins, and implementation mechanisms. Some rely on legislation, others on voluntary standards, organisational management systems, ethical principles, or practical implementation tools. Yet despite these differences, they demonstrate a remarkable degree of convergence in their expectations of how AI should be governed.

Rather than prescribing fundamentally different models of governance, these frameworks consistently reinforce the same organisational capabilities required to ensure that AI systems are deployed responsibly, transparently, and accountably. Across the international landscape, five governance signals emerge repeatedly.

1. Clearly Assigned Accountability

Across all major frameworks, governance begins with clearly defined organisational responsibility. AI governance cannot remain diffused across technical teams, compliance functions, and business units without clear ownership. Whether through board-level accountability, designated governance roles, or documented lines of responsibility, organisations are expected to identify who is responsible for decisions made throughout the AI lifecycle.

The precise governance structure varies across frameworks, but the underlying expectation remains consistent: accountability must be explicit, documented, and supported by appropriate oversight mechanisms.

2. Risk-Proportionate Controls

International frameworks consistently recognise that not every AI system requires the same level of governance. Instead, governance expectations should be proportionate to the potential risks associated with the system, its intended purpose, the people affected, and the consequences of failure.

Whether through the EU AI Act's risk categories, Canada's Algorithmic Impact Assessment, or NIST's contextual risk assessments, governance intensity increases alongside potential impact. High-impact applications such as credit decisioning therefore require stronger controls, more extensive documentation, and greater human oversight than lower-risk internal tools.

3. Documentation as Governance

One of the strongest areas of international convergence is the emphasis placed on documentation. Across regulatory frameworks, management standards, and organisational guidance, documentation is treated not as administrative compliance but as the primary mechanism through which governance is demonstrated.

Policies, model documentation, risk assessments, validation reports, testing records, approval processes, and monitoring logs collectively provide the evidence that governance processes exist and are functioning effectively. Without documentation, organisations cannot meaningfully demonstrate accountability, enable audits, investigate incidents, or continuously improve AI systems.

4. Human Oversight for High-Impact Decisions

No major framework supports fully autonomous AI decision-making in contexts where significant consequences may arise for individuals. Instead, organisations are expected to establish meaningful human oversight for high-impact AI systems, particularly where decisions affect access to financial services, employment, healthcare, or other important rights and opportunities.

Although frameworks differ in how they define meaningful human involvement, they consistently recognise that human judgement remains an essential safeguard for managing AI risks, addressing unexpected outcomes, and ensuring accountability.

5. Continuous Monitoring Across the AI Lifecycle

International frameworks also reject the assumption that governance ends once an AI system is deployed. AI systems evolve over time as data changes, models degrade, risks emerge, and operating environments shift. Governance is therefore treated as a continuous process rather than a one-time approval exercise. Accordingly, organisations are expected to monitor AI systems throughout their lifecycle, respond to incidents, periodically reassess risks, update controls where necessary, and maintain ongoing governance oversight as systems continue to operate.

Taken together, these five governance signals demonstrate that international AI governance is characterised less by disagreement than by convergence. While jurisdictions continue to debate the most appropriate legal and institutional mechanisms for governing AI, there is broad agreement on the organisational capabilities that responsible AI deployment requires.

For organisations seeking to strengthen AI governance, the practical question is therefore not which international framework to adopt in its entirety, but how to build governance processes that reflect these shared expectations. It is this common foundation that informs the governance framework and implementation toolkit presented in the following chapters of this Playbook.

Governance Signal	EU AI Act	NIST	ISO 42001	Singapore	Canada	UNESCO	OECD
Clearly assigned accountability	✓	✓	✓	✓	✓	✓	✓
Risk-proportionate controls	✓	✓	✓	✓	✓	✓	✓
Documentation as governance	✓	✓	✓	✓	✓	✓	✓
Human oversight	✓	✓	✓	✓	✓	✓	✓
Continuous monitoring	✓	✓	✓	✓	✓	✓	✓

Table 5: Convergence of Core Governance Signals Across Global AI Frameworks

Despite significant differences in legal form and implementation mechanisms, all major international AI governance frameworks converge on the same five organisational governance capabilities.

3.4 From Global Principles to Indian Practice

The purpose of examining international AI governance frameworks was to observe and acknowledge the fragmentation. There is no one-size-fit policy to all and that has its merits and demerits. Each framework has been developed within a different legal system, regulatory philosophy, and institutional context.

The European Union emphasises legally enforceable obligations, the United States focuses on organisational risk management, Singapore prioritises practical implementation, while international standards bodies and multilateral organisations provide management systems, ethical principles, and governance guidance. None can simply be transplanted as is into the Indian context, however, together, these approaches offer valuable lessons.

India's own approach to AI governance is already evolving through sector-specific regulation rather than a single comprehensive AI law. As discussed in the previous chapter, initiatives such as the RBI's Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI), together with existing legal and regulatory obligations, are establishing the foundations of AI governance for regulated entities. Rather than replicating any one international framework, India is progressively developing an approach that reflects its own regulatory institutions, financial system, and governance priorities.

Therefore, **the more immediate challenge for many financial institutions is not selecting the “right” governance model, but building the organisational capability to implement AI governance effectively.** Across the international frameworks examined in this chapter, there is a common assumption that organisations already possess a range of governance capabilities as if all countries and organisations have clear accountability structures or documented policies and procedures etc. In practice though, these capabilities remain uneven across countries and organisations, particularly as AI deployment continues to accelerate at different pace and scale. This is because all regions have different economic realities, political priorities and varied institutional governance architectures.

This implementation gap is again where this Playbook seeks to contribute. Rather than introducing another set of high-level governance principles, it translates internationally recognised governance expectations into practical structures, processes, and tools that are aligned with India's regulatory landscape and institutional realities. The objective is to help regulated entities operationalise the governance capabilities that those frameworks consistently expect.

Building on the regulatory context presented in Chapter 2 and the international governance lessons explored in this chapter, the next chapter moves from governance principles to governance implementation with concrete 6 case examples.

USE CASE C: AML MONITORING (TRANSACTION MONITORING & FINANCIAL CRIME DETECTION)

WHAT THE USE CASE IS AND WHY IT EXISTS

AI-driven AML monitoring helps financial institutions detect potential money laundering, terrorist financing, sanctions evasion, and related financial crimes by analysing customer behaviour, transactions, and network relationships. Compared with traditional rule-based monitoring, AI improves the detection of complex laundering patterns while reducing false positives and supporting regulatory compliance.

INSTITUTIONS IMPLEMENTING IT

Institution	Type	Country	Specific implementation
HSBC (with Google Cloud)	Bank	UK/Global	"Dynamic Risk Assessment" (DRA), built on Google Cloud's "Anti Money Laundering AI" (AML AI) product, launched 2021/announced publicly June 2023; ML models trained/fine-tuned on HSBC's own financial-crime data to score and prioritise suspicious-activity alerts
Bradesco	Bank	Brazil	Adopted Google Cloud's AML AI product (named alongside HSBC and Lunar as reporting significant value from the AI-based approach)
Lunar	Digital bank/fintech	Denmark/Nordics	Adopted Google Cloud's AML AI product

Association of Banks in Singapore members	Banking industry body	Singapore	Data analytics and ML for AML/CFT industry-wide best practice
BIS Innovation Hub (Nordic Centre)	Central-bank research	Nordic	Project Aurora - GNN-based synthetic-data proof of concept for cross-institutional laundering-network detection
Japan (national AML initiative)	Regulator/FIU-coordinated	Japan	"Proof of Concept on Machine Learning and Artificial Intelligence" - a Shared Transaction Monitoring and Screening System with AI, tested for reducing triage workload and false positives across (implicitly) multiple institutions
EU credit institutions, e-money institutions, life insurers (unnamed)	Various	European Union	Testing/piloting General Purpose AI (GPAI) for AML/CFT client and transaction profiling/clustering, per EBA's 2025 supervisory survey of Competent Authorities
SymphonyAI clients (unnamed banks)	Various	Global	Sensa Copilot - GenAI tool automating financial-crime investigation report drafting

CURRENT PILOTS AND FUTURE DIRECTION

- **EU:** EBA's 2025 Opinion signals that GPAI-based AML/CFT testing is happening now across the EU but has not yet reached widespread production status - this is a clear near-term area to watch as testing converts to deployment.
- **Japan:** Japan is exploring a shared AI-enabled transaction monitoring infrastructure across institutions.
- **Generative AI direction:** Generative AI is increasingly being used to automate investigation support and SAR drafting while keeping humans in the decision loop.
- **Adversarial dynamic:** Regulators increasingly expect institutions to address criminals' use of AI, pointing to an emerging AI-driven arms race in financial crime detection.

– Black Dot Public Policy Advisors

CHAPTER 4

FROM PRINCIPLES TO PRACTICE

4.1 The Translation Challenge

India's financial sector is navigating a period of considerable change in its relationship with AI. Across banks, non-banking financial companies, payment institutions, and fintech organisations, there is growing recognition that AI offers meaningful improvements in efficiency, risk detection, customer service, and the extension of financial access. At the same time, implementation or deployment of AI in any form, is low as-well-as low stakes with most use cases being - customer-facing chatbots, voicebots for transaction confirmation, and internal process automation.

However, the governance conversation is beginning to outpace deployment. Regulators, board members, risk committees, and compliance functions are asking, with increasing urgency, how AI should be governed once it becomes embedded in institutional operations. The question is now primarily about how to adopt AI and about whether institutions have the operational infrastructure to oversee what they deploy.

This is the challenge this chapter addresses: the gap between principle and practice. Every major governance framework relevant to India's financial sector (as discussed in chapter 3) like the RBI FREE-AI Framework, the Draft Guidance on Regulatory Principles for Model Risk Management published in June 2026, the EU Artificial Intelligence Act, the NIST AI Risk Management Framework, the GSMA Responsible AI Maturity Roadmap, and the frameworks of Singapore, Canada, UNESCO, and the OECD articulates governance requirements in the language of **common values and obligations**, which are:

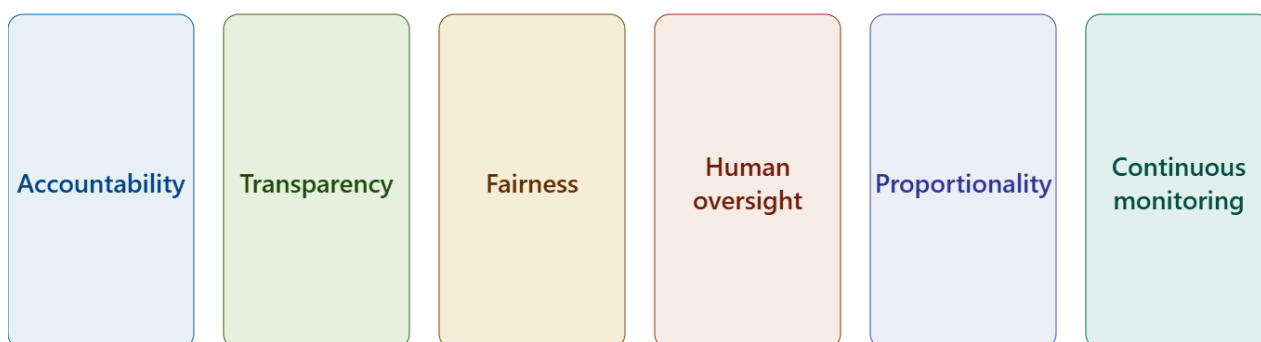


Figure 6: Common Principles across AI Governance Frameworks

The challenge today is no longer the absence of principles. It is the translation of those principles into the everyday processes, structures, and decisions through which governance is actually exercised.

This distinction may appear subtle, but from an institutional perspective it is profound. A compliance officer who reads the RBI FREE-AI Framework's Recommendation 14 that boards must approve AI governance policies with defined accountability structures, knows what is expected. What they do not necessarily have is the specific form that accountability should take, the evidence a supervisory examiner would expect to see, or the operational steps required to move from a stated principle to a documented and testable governance arrangement.

The six examples in this chapter showcase that translation. Each example takes a governance principle that appears consistently across the frameworks studied for this Playbook and bridges the gap between the regulatory statement and its operational meaning in an Indian financial institution deploying AI. The examples are drawn from the requirements of the RBI FREE-AI Framework, the Draft Guidance on Model Risk Management, and the five convergence signals identified across over twenty-five global and Indian governance frameworks reviewed for this Playbook. They are also grounded in the practitioner conversations and roundtable discussions that have informed the Playbook's design.

The four tools in Chapter 5 provide the operational forms through which these translations are implemented. This chapter explains the thinking behind them.

4.2 Six Translations: From Regulatory Principle to Institutional Practice

4.2.1 From “The Board Must Approve an AI Governance Policy” to a Documented Accountability Structure

The requirement for board-level AI governance accountability is one of the most consistently stated principles across every framework reviewed for this Playbook. The RBI FREE-AI Framework's Recommendation 14 requires board approval of an AI governance policy with a defined governance structure. The Draft Guidance on Model Risk Management, published in June 2026, advances this further: Paragraph 9 requires every regulated entity to put in place a Board-approved Model Risk Management Framework, with the Risk Management Committee of the Board providing oversight and senior management responsible for operationalisation. The principle is unambiguous and universally stated across frameworks from the EU AI Act to ISO/IEC 42001.

In practice, the picture is considerably less settled. Practitioner conversations conducted for this Playbook revealed that most regulated financial institutions either do not have an AI-specific governance policy, or have a general technology governance policy that does not explicitly address AI. Where AI risk does appear in board documentation, it is typically embedded within broader technology or operational risk reporting in a form that does not enable meaningful board oversight of specific AI systems.

The operational translation of this principle requires three specific steps.

- **First**, a risk or compliance officer must draft a written AI governance policy or, for institutions at a more advanced stage of readiness, a Model Risk Management Framework that covers at minimum the institution's risk appetite for AI use, the committee or function with explicit AI governance responsibility, and the escalation path from that committee to the board.
- **Second**, the policy must be formally approved by the board, with the date of approval recorded and the next review date stated.
- **Third**, the board must receive regular reporting on AI governance in a structured format, at a frequency and level of detail that permits meaningful oversight rather than nominal acknowledgement.

Key takeaway: A board-approved AI policy is the instrument through which the board sets the institution's risk appetite for AI, defines who is accountable, and establishes the conditions under which AI governance questions escalate to the board level. Without it, accountability for AI risk has no documented home in the institution's governance structure.

4.2.2 From “Maintain a Comprehensive Model Inventory” to a Governed Registry of Every AI System

The requirement to maintain a comprehensive inventory of all AI systems in use is among the most specifically stated obligations in India's regulatory landscape. The RBI FREE-AI Framework's Recommendation 23 requires regulated entities to maintain such an inventory, reviewed every six months and available for supervisory inspection. The Draft Guidance on Model Risk Management goes further in Paragraph 21, stating that no model may be used, relied upon, or deployed unless it is part of the model inventory. Paragraph 22 specifies the minimum content of each inventory record.

In practice, institutions systematically underestimate the number of AI systems they are operating. This is a structural reality of how AI capability arrives in institutions, it might necessarily not be a failure of diligence. Some systems are internally developed by data science or analytics teams, some are procured from vendors as standalone AI products and some arrive embedded within broader technology platforms like core banking systems, cloud services, regulatory reporting tools, customer relationship management software through product updates that introduce AI features without a separate procurement or governance decision. And some are informal like staff using generative AI tools for drafting, summarisation, research, and analysis without any institutional awareness or oversight.

An institution that asks only its technology function for its AI inventory will likely surface the first category. The complete picture requires a structured enquiry directed to each business unit and function, asking which tools produce decisions, recommendations, scores, or automated actions that affect customers or operations. The responses will surface systems across fraud, collections, KYC, customer service, analytics, and internal productivity that no single function was previously tracking.

The operational translation of this principle is a compliance exercise. A risk or compliance officer sends a structured questionnaire to each business unit and function head. The completed responses are compiled into an inventory using the Model Documentation template in Tool 2 of this Playbook, which covers the minimum fields required under Paragraph 22 of the Draft Guidance: the system's name and description, its intended use, the accountable committee, the risk tier, the deployment status, and whether it was built internally or sourced from a third party. The inventory can be reviewed every few months and updated immediately whenever a new system is deployed or a significant change is made to an existing one.

Key takeaway: A model inventory is the foundational governance document from which all other governance obligations flow. An institution that cannot account for every AI system it is operating cannot assess risk proportionately, govern vendor relationships appropriately, or demonstrate supervisory readiness. Under the Draft Guidance on Model Risk Management, no model may be used without an inventory record.

4.2.3 From “Apply Proportionate Controls” to a Criteria-Based Risk Classification for Each System

Every major governance framework establishes the principle that the intensity of AI governance should be proportionate to the risk a given system poses. The Draft Guidance on Model Risk Management, Paragraphs 17 through 20, requires every regulated entity to establish and implement a risk-based model tiering structure for all models, with the tier determining validation requirements, approval processes, monitoring cadence, and documentation intensity. The Canada Algorithmic Impact Assessment uses a structured scoring questionnaire to achieve proportionality. The EU AI Act uses legally defined use-case categories. The NIST AI Risk Management Framework leaves classification to organisational judgement within a defined framework. The convergence signal across all of these instruments is identical: governance intensity must be calibrated to risk, and a voicebot that confirms a cheque payment is not the same governance problem as a system that flags a customer account for suspension.

The challenge with proportionality as a principle is that it offers no guidance on how to classify a specific system in practice. An institution instructed to apply risk-proportionate controls to its AI is no further forward in operational terms than it was before receiving that instruction.

The operational translation requires a structured classification exercise for each AI system in the inventory.

A compliance officer and technology lead work through eight criteria for each system:

- whether the system produces outputs that directly affect individual customers;
- whether those outputs affect customers' access to financial products or services;
- the scale at which the system operates; how easily harm could be identified and corrected if the system produces a wrong output;
- whether the system intersects with a named regulatory obligation such as the RBI Fair Lending Practices Circular or the KYC Master Directions;
- whether the system can explain its outputs to a customer or auditor in plain language;
- what category of personal data the system uses; and
- whether the system was built internally or procured from a third party.

Each criterion is scored, and two additional modifiers are applied: one for systems that use any generative AI component, and one for systems that take autonomous actions without requiring human approval at each step. The resulting total score places the system in one of four risk tiers, which in turn determines the governance intensity required. This is what the Risk Scoring tool in Tool 3 of this Playbook operationalises. The classification is documented, defensible, and reproducible and it can be presented to an internal auditor or a supervisory examiner as the basis for the governance approach applied to each system.

Key takeaway: Proportionality requires a method. Without a structured classification exercise producing a documented and defensible risk tier for each system, proportionality remains a principle without operational meaning. The risk tier also determines the depth of model documentation required and the vendor contract provisions that apply.

4.2.4 From “Ensure Human Oversight for High-Impact Decisions” to a Documented Review Mechanism with Named Triggers

Human oversight for AI-driven decisions that significantly affect individuals is the principle that generates the most significant operational questions in financial sector AI governance. The RBI FREE-AI Framework identifies human oversight as a core requirement for high-risk use cases. The Draft Guidance on Model Risk Management, Paragraphs 60 through 63, is more specific: it requires regulated entities to establish human-in-the-loop or human-on-the-loop arrangements, override and kill-switch mechanisms, and periodic review of model outputs by humans. Paragraph 63 requires that human oversight arrangements, including decisions, interventions, overrides, and near-misses, are themselves reviewed and strengthened periodically.

While the principle is clear, the operational question it leaves open is **what meaningful human oversight looks like when a system is processing thousands of decisions each day, or when the business process requires near-instantaneous outputs.**

The answer varies with the type of decision, the degree of automation, and the reversibility of outcomes. Three examples from different contexts illustrate the range. A fraud detection system that flags transactions for manual review operates in a fundamentally different mode from one that automatically blocks a transaction or freezes an account. In the first case, the human is a primary decision-maker and the AI is advisory. In the second, human oversight means a named function reviews the AI's decision within a defined timeframe before any restriction that cannot be immediately reversed is applied, and that review is logged. A customer service chatbot that answers questions about account features is lower stakes, but the Draft Guidance Paragraph 59(iii) specifically requires that customers must be given the option to switch to human assistance when requested - which is a specific, documentable, and testable control. An internal analytics model that produces risk scores for use by relationship managers is advisory rather than decisional: human oversight here means that relationship managers understand the model does not make decisions for them, have received appropriate training, and operate within a process that flags anomalous outputs for review.

The operational translation for each Tier 1 and Tier 2 system in the inventory requires documentation of three specific things: the trigger condition for human involvement, defining at what point and under what circumstances a human review is required; the named function or role responsible for that review, and the timeframe within which the review must occur; and the logging mechanism that records whether the review took place, what was decided, and what action followed. Without all three elements, human oversight exists as a stated commitment but cannot be demonstrated to a supervisor or an internal auditor as an operating control.

Key takeaway: Human oversight is not just a nominal override capability that is theoretically available, in fact, it is a specific, triggered, and logged review mechanism. The question a supervisor will ask is not whether human oversight is stated in policy but whether it can be evidenced: who reviewed, when, what they found, and what happened next.

4.2.5 From “Validate All Models Including Third-Party Models Independently” to a Vendor Governance Framework

The Draft Guidance on Model Risk Management introduces one of its most significant governance requirements in Paragraph 29: all models, including those sourced from third parties, must be subject to independent validation conducted by the regulated entity. Paragraph 46 makes explicit that this obligation applies notwithstanding any validation, certification, or assurance provided by the third-party provider. The principle is that regulatory accountability cannot be transferred to a vendor, and that the institution's obligation to understand and oversee what it deploys extends to every AI system in its inventory regardless of its provenance.

This represents one of the most consequential governance requirements in the new guidance because it directly addresses a gap that practitioner conversations for this Playbook identified consistently. Most regulated financial institutions that use vendor-supplied AI have never independently validated those models. Procurement processes apply general technology due diligence rather than AI-specific governance standards. Contracts contain no provisions for documentation access, audit rights, or AI-specific incident reporting. The institution deploying the system has, in most cases, no meaningful visibility into how the system produces its outputs, what data it was trained on, or whether it was tested for bias across the populations it is now affecting.

The operational translation requires a change to the institution's procurement process for AI systems. Before any third-party AI system is deployed, the institution must have sufficient documentation to validate the model or commission its validation by an independent expert. This means that a vendor contract for any Tier 1 or Tier 2 AI system must include documentation access rights, audit provisions, performance standards, model change notification requirements, and AI-specific incident reporting obligations. Paragraph 48 of the Draft Guidance specifies each of these requirements. If a vendor is unwilling to provide documentation sufficient for independent validation, Paragraph 51 requires the institution to identify the resulting risks and put in place necessary mitigants, which may include restricting the system's use to lower-risk applications.

The vendor due diligence checklist and contract clause schedule in Tool 4 of this Playbook operationalise these requirements. The practical starting point for most institutions is to review the contracts for their most significant third-party AI systems and identify which of the required provisions are absent. Most contracts signed before 2026 will contain none of the AI-specific provisions the Draft Guidance requires. This does not make those systems ungovernable; it identifies the governance work to be done at the next contract renewal.

Key takeaway: Independent validation of vendor AI is not optional under the Draft Guidance on Model Risk Management. The institution that deploys a vendor's system without adequate documentation access, audit rights, and performance standards is not protected by the vendor's own assurances. Accountability rests with the deploying institution, and its ability to exercise that accountability depends on what its contracts provide.

4.2.6 From “Apply Specific Controls for AI and ML Systems” to Documented Controls for Explainability, Bias, Hallucination, and Red-Teaming

The Draft Guidance on Model Risk Management, Paragraph 54, introduces governance requirements specifically for AI and machine learning models that go beyond the obligations applicable to models generally. On explainability, institutions must define thresholds and ensure outputs are explainable to the degree required for the business process, with higher thresholds applied where AI is used for material decisions or has significant customer impact. Where full explainability is not achievable, enhanced controls apply: enhanced validation, mechanisms to verify outputs before use, more frequent monitoring, and restrictions on usage. On hallucination, Paragraph 54(2) requires control boundaries through system-level controls or model design features, particularly for generative AI. On bias, Paragraph 54(3) requires fairness assessments and mitigants including recalibration or redesign. On red-teaming, Paragraph 55 requires structured adversarial testing for models involving customer interaction or generative capabilities.

These four requirements are not new in the sense of being unanticipated by global frameworks such as the EU AI Act, NIST, Singapore, and UNESCO have each addressed these dimensions. **They are new in the sense of now being formal supervisory expectations for regulated financial institutions in India.**

The operational translation for each Tier 1 and Tier 2 system requires documented evidence of four things:

- **On explainability:** the institution has defined what level of explanation is required for this system’s intended use and has confirmed the system can meet that level. For a fraud detection system, this means the institution can produce, for any specific flagged transaction, an explanation of why it was flagged that would satisfy a customer complaint or a regulatory query. For a generative AI system used in customer communications, it means the system cannot produce outputs that cannot be traced to a defined and authorised source.
- **On hallucination:** the institution has tested whether the system produces confident but factually incorrect outputs, and has implemented technical controls that prevent hallucinated content from reaching customers or driving operational decisions without a verification step.
- **On bias:** a fairness assessment was conducted before deployment across relevant demographic segments, the results are documented, and a monitoring process is in place to detect emerging disparate impact.
- **On red-teaming:** an independent adversarial test was conducted before deployment, and a periodic testing schedule exists. The independence of red-teaming is important: internal teams cannot effectively red-team systems they have designed and deployed because the social dynamics of institutional life constrain the willingness to surface findings that reflect poorly on the team’s own work.

Key takeaway: The four AI-specific controls in the Draft Guidance - explainability, hallucination mitigation, bias assessment, and red-teaming are not aspirational standards imported from global frameworks. They are now supervisory expectations for regulated financial institutions in India. Each requires documented evidence and not just stated intent.

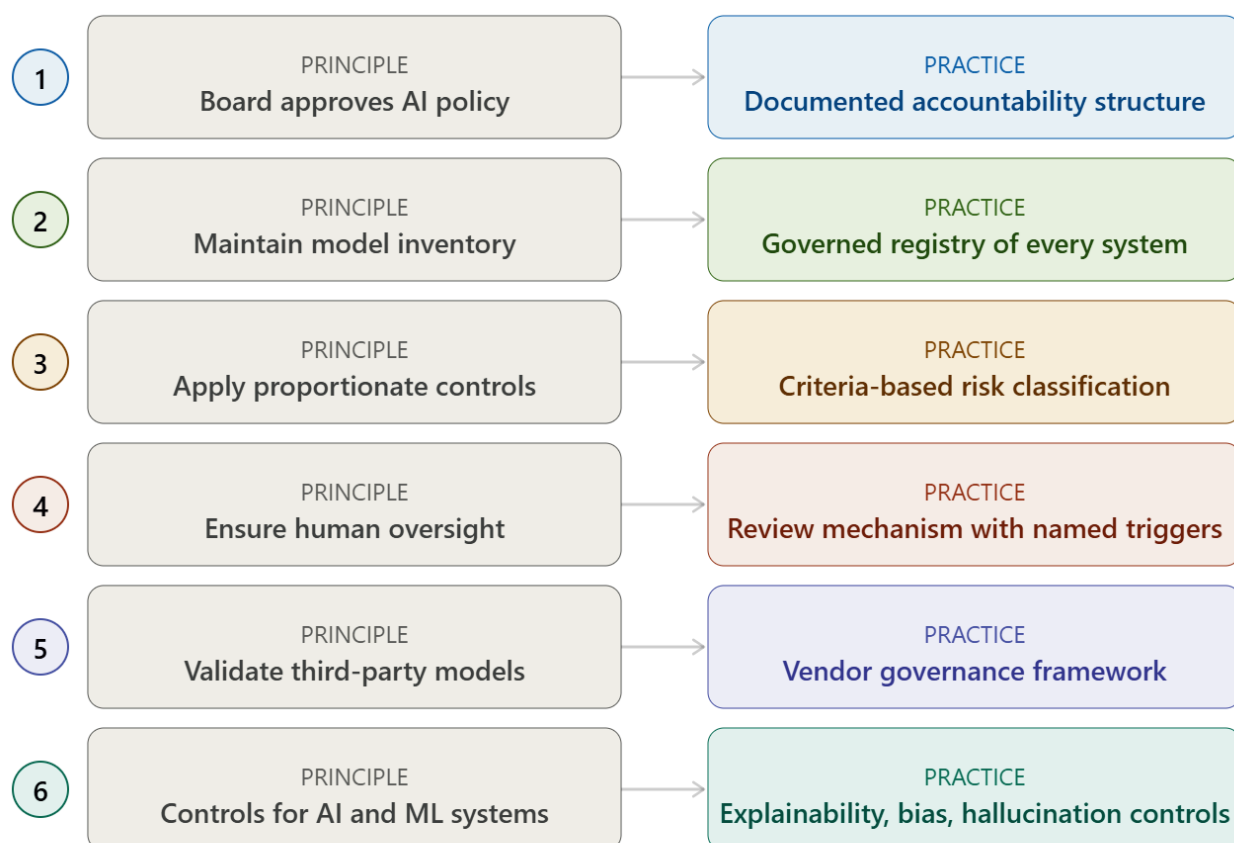


Figure 7: Translation examples of 'Principles To Practice'

4.3 A Seventh Translation: The Vendor Problem

There is a seventh translation that warrants particular attention because it is the most commonly overlooked and the one where the regulatory direction in India is now most unambiguous.

Most AI in Indian financial services is not being built by the institutions deploying it. It is being procured from fintech vendors, credit bureaus, analytics providers, and technology platforms. The regulatory position on accountability for procured AI is clear and has been consistent across both the FREE-AI Framework and the Draft Guidance. The regulated entity is accountable for the outcomes of all models it uses, regardless of whether those models were developed internally or sourced externally.

What can be managed through contract is the allocation of commercial liability between the institution and the vendor, the documentation access the institution needs to exercise its own governance obligations, and the specific AI-related provisions that govern the vendor relationship throughout the deployment period. Practitioner conversations for this Playbook found that most institutions relying on vendor-supplied AI had signed contracts containing none of these provisions. Standard technology procurement processes which are designed for software, not for probabilistic systems and that can fail silently apply to AI by default, leaving institutions without the contractual infrastructure required to govern what they have deployed.

The pre-procurement checklist and contract clause schedule in Tool 4 of this Playbook provide the minimum standard for AI vendor governance. Each clause in the schedule is grounded in a specific requirement from the Draft Guidance on Model Risk Management or the FREE-AI Framework. For institutions reviewing existing vendor relationships, the practical starting point is to identify the three to five most significant third-party AI systems in use and assess whether the current contracts include model documentation access rights, bias testing results, performance standards, model change notification obligations, audit rights, and AI-specific incident reporting provisions.

4.4 The Role of the FSAI Toolkit

The FSAI Toolkit in Chapter 5 provides the four operational forms through which the translations described in this chapter are implemented consistently and in a manner that can be demonstrated to a supervisor or internal auditor.

Tool 1, the Responsible AI Adoption Assessment, establishes where an institution currently stands in its governance journey before it attempts to govern any individual AI system. The assessment has two parts: a factual awareness stocktake that determines whether the institution has a working picture of its AI use, and a four-level maturity ladder adapted from the GSMA Responsible AI Maturity Roadmap that places the institution on a governance continuum from **Aware through Organising, Governing, and Embedded**. The assessment is designed to surface honest placement rather than aspirational self-classification.

Tool 2, the Model Documentation template, provides the baseline compliance record that every AI system must have regardless of its risk tier, and the risk-scaled documentation requirements that deepen proportionately for higher-risk systems. The baseline tier directly operationalises the inventory requirements of Paragraph 22 of the Draft Guidance. The risk-scaled tier operationalises the documentation intensity requirements of Paragraphs 24 and 57.

Tool 3, the Risk Scoring and Accountability tool, provides the criteria-based model tiering methodology required by Paragraphs 17 through 20 of the Draft Guidance. Eight dimensions plus two modifiers for generative and agentic AI produce a score out of thirty-one, which determines the risk tier and the corresponding governance intensity. The tool also establishes the accountability structure for each system, assigning responsibility to a named committee with AI governance explicitly in its mandate.

Tool 4, the Vendor Due Diligence and Contract Checklist, addresses governance of procured AI systems through pre-procurement due diligence questions and a minimum contract clause schedule. Eight clauses cover performance standards, model change notification, audit rights, incident reporting, liability allocation, DPDP compliance, explainability obligations, and AI-specific business continuity. Each clause is grounded in a specific provision of the Draft Guidance or the FREE-AI Framework.

4.5 Sequencing and the Regulatory Alignment

The tools are designed to be used in a specific sequence for any AI system being brought under governance for the first time: the maturity assessment first, at the institutional level; then baseline documentation for each system; then risk scoring; then vendor provisions for any system procured externally. The risk tier produced in Tool 3 determines the depth of documentation required in Tool 2 and the vendor provisions required in Tool 4. This sequencing is not arbitrary, it reflects the governance logic embedded in both the FREE-AI Framework and the Draft Guidance – you cannot apply proportionate controls to a system you have not inventoried, and you cannot govern a vendor relationship you have not assessed.

The RBI FREE-AI Framework and the Draft Guidance on Model Risk Management are aligned, not competing instruments. The FREE-AI Framework establishes the principles and the broad governance architecture. The Draft Guidance translates those principles into specific operational requirements across the model lifecycle. The tools in this Playbook sit at the intersection of both, providing the operational layer that enables institutions to meet the governance expectations of the first and the specific requirements of the second.

The Draft Guidance on Regulatory Principles for Model Risk Management was in public consultation as this Playbook was finalised, with comments due by July 24, 2026. Final requirements may differ from the draft. **The tools in this Playbook are designed to remain robust across the range of likely final requirements because they are grounded in the underlying governance logic of accountability, documentation, proportionality, human oversight, and continuous monitoring**, rather than in specific paragraph numbers. Institutions that build governance capability now will be better positioned to meet final requirements than those that wait for the final instrument before beginning.

USE CASE D: PROJECT AURORA: BIS INNOVATION HUB CASE STUDY

ABOUT

Project Aurora is a BIS Innovation Hub Nordic Centre proof-of-concept that evaluates whether Graph Neural Networks (GNNs), machine learning, and privacy-enhancing technologies (PETs) can improve the detection of cross-border money laundering compared with traditional rule-based AML systems. Phase 1 (May 2023) was conducted using synthetic transaction data, while Phase 2 is exploring real-world proofs of concept that could eventually lead to live pilots.

Traditional AML systems operate within institutional silos, limiting visibility into cross-border laundering networks. Since money laundering typically spans multiple institutions and jurisdictions, the BIS sought to determine whether collaborative AI models could identify suspicious networks while preserving data privacy through PETs. The project also addresses the growing cost and complexity of AML compliance, which the BIS cites as approximately US\$274 billion globally in 2022.

KEY FINDINGS

- Graph Neural Networks detected twice as many money launderers as traditional rule-based monitoring.
- GNN models outperformed other machine learning approaches, including logistic regression, isolation forests, and conventional neural networks.
- False positives were reduced by up to 85% at the individual institution level.
- The project demonstrated that combining AI with privacy-enhancing technologies enables collaborative analysis across institutions without directly exposing customer data.

WHY IT MATTERS

Project Aurora is one of the few central bank-led empirical evaluations of AI for AML, making it significantly stronger than vendor-led case studies. It demonstrates that network-based AI can materially improve cross-border financial crime detection while supporting regulatory priorities around privacy, collaboration, and supervisory oversight. The BIS also notes that technology alone is insufficient and must be supported by appropriate legal and regulatory frameworks before large-scale implementation.

CURRENT STATUS

Project Aurora remains a proof of concept based on synthetic data. Phase 2 is focused on real-world proofs of concept using collaborative analytics and privacy-enhancing technologies, with the long-term objective of progressing toward live pilot deployments

CHAPTER 5: FSAI TOOLKIT

HOW TO USE THIS

This assessment has two parts:

- Part One is a factual stocktake (do you know what AI you have).
- Part Two is a four-level maturity check (where does your institution sit in its AI governance journey).

Complete Part One first, always. If the institution cannot complete Part One confidently, that is the finding, and the action is to build the basic inventory and nothing more is needed at this stage.

Once Part One is complete, read through the four levels in Part Two and identify honestly which level's markers mostly describe the institution today.

Do not aim for the highest level you can justify, simply aim for the level that is true. **The value of this tool comes entirely from honest placement, not from a high score.**

Revisit this assessment twice a year, or whenever the institution's AI use changes meaningfully. For example, when a new AI vendor is brought on, or when AI is introduced into a new part of the business.

PART ONE: AWARENESS: DOES THE INSTITUTION KNOW WHAT AI IT HAS

Question A1: Is artificial intelligence used anywhere in this institution?

Answer Yes, No, or Not Sure.

- ☐ Yes
- ☐ No
- ☐ Not Sure

If Not Sure, that is itself meaningful information - it means there is no basic inventory.

Action point: ask each business unit and technology team to confirm whether AI is used in their area. Set a timeline - ideally of a week.

Question A2: If yes, where is it being used?

Tick all that apply, even informally:

- ☐ Customer-facing chatbot or voice assistant
- ☐ Credit or loan decisioning support
- ☐ Fraud detection or transaction monitoring
- ☐ KYC or customer onboarding
- ☐ Collections or recovery
- ☐ Internal analytics or reporting
- ☐ Internal tools like email drafting, summarisation, or coding assistance
- ☐ Something else, not listed above
- ☐ Genuinely don't know

Question A3: Was any of this AI introduced by a vendor or technology partner, without a formal internal decision to adopt AI specifically?

A CRM tool, a cloud service, or a core banking system update quietly adds as an AI feature. Answer

- ☐ Yes
- ☐ No
- ☐ Not Sure

Question A4: Has anyone in the institution: risk, compliance, technology, or otherwise ever written down a list of where AI is used?

Answer

- ☐ Yes, a current list exists.
- ☐ Yes, but it is outdated.
- ☐ No, never done.

Question A5: If a regulator asked tomorrow "where do you use AI," could someone in this institution answer within a day?

Answer

- ☐ Yes, confidently.
- ☐ Yes, but it would take longer than a day to be sure it's complete.
- ☐ No, it would be very difficult to answer with confidence.

Additional Notes

PART TWO: THE FOUR LEVELS OF AI MATURITY

This section is adapted from the GSMA Responsible AI Maturity Roadmap's four-level structure - Foundational, Evolving, Performing, Advanced, renamed here for the Indian financial sector context.

The four levels below are:

Level 1: Aware. The institution knows where AI is used. Nothing formal exists yet beyond that.

Level 2: Organising. The institution has started to put basic structure around its AI use, someone has been asked to look at it, even informally.

Level 3: Governing. Formal governance exists and is operating for at least the institution's highest-risk AI use.

Level 4: Embedded. AI governance operates as a routine part of how the institution manages risk, across all its AI use, not just the highest-risk system.

STEP 1: ASSESS AI MATURITY

Read each level description. Identify the level whose markers mostly describe the institution today. Do not aim for the highest level you can defend, aim for the level that is true.

LEVEL 1: AWARE

The institution knows it uses AI somewhere, has a rough picture of where, and at least one person has been informally asked to pay attention to it. Nothing formal exists yet.

Most of the following are true:

- We have completed Part One above and have at least a rough list of where AI is used.
- At least one person even without a formal title or mandate has been informally asked to keep an eye on AI use in the institution.
- We are aware that AI use might raise questions for risk, compliance, or customers, even if we have not yet acted on that awareness.
- No one has yet looked closely at any specific AI system to formally assess whether it is working correctly or fairly.
- We do not have a written AI policy or a named governance committee.

If this describes your institution: the next step is to turn the rough inventory from Part One into a more complete list, and to identify which one AI system matters most used, most customer-facing, or highest potential for harm. Start governance work on that one system first.

LEVEL 2: ORGANISING

The institution has started to put basic structure around its AI use. Someone has been formally asked to look at it. The highest-risk AI use has been identified even if not yet formally assessed.

Most of the following are true:

- Someone has an explicit, even if informal, responsibility for AI governance across the institution a person, a small team, or a function that AI questions are routed to.
- At least one AI system has been described in enough detail that someone could explain, in plain language, what it does and what data it uses.
- There is an emerging shared understanding of which AI use matters most which system, if it went wrong, would cause the most harm to customers or the institution.
- Basic questions have been asked about that system what could go wrong, what data does it use, who oversees it even if the answers are informal and not yet written down.
- No formal AI policy exists yet, but informal practice is beginning to take shape someone is occasionally consulted before a new AI tool is adopted.

If this describes your institution: the next step is to write down, even briefly, what has been decided informally. A one-page note on how AI decisions are made and who is responsible is worth more than a sophisticated policy document that nobody reads.

LEVEL 3: GOVERNING

Formal governance exists and is operating for at least the institution's highest-risk AI system. Some documentation is in place. There is a named home for AI governance questions.

Most of the following are true:

- A written statement of how the institution intends to manage AI risk exists it does not need to be board-approved yet, but it is in some documented form and people know it exists.
- The institution's highest-risk AI system has been assessed someone has documented what it does, what data it uses, what could go wrong, and what controls are in place.
- A named committee or function has AI governance within its mandate and is the clear destination for AI-related questions or concerns.
- If something went wrong with the highest-risk AI system, there is a documented process for who would be told, what would happen, and how it would be communicated to affected customers.
- Customers can receive some form of explanation when an AI system has made a decision that affects them, even if that explanation is not fully automated or perfect.
- Vendor AI is beginning to be looked at specifically, even if no formal vendor governance process exists yet.

If this describes your institution: the next step is to extend the governance that exists for the highest-risk system to the next two or three systems and to formalise what has been documented into a board-approved policy.

LEVEL 4: EMBEDDED

AI governance operates as a routine part of how the institution manages risk, across all its AI use, not just the highest-risk system.

Most of the following are true:

- A board-approved AI policy exists, is current, and sets out the institution's risk appetite for AI.
- A named committee has AI governance explicitly in its terms of reference and meets regularly with AI governance on the agenda.
- Every AI system in use has been documented and assessed, not just the highest-risk ones.
- External testing like red teaming, independent audit has been conducted on at least the highest-risk systems in the last 12 months.
- Vendor AI is subject to specific due diligence and contractual provisions, separate from general technology procurement.
- Customers can receive a plain-language explanation for AI-driven decisions that affect them, including adverse decisions.
- Grievances relating to AI decisions are tracked and reviewed regularly.
- The institution contributes to or engages with sector-level conversations on AI governance through IBA, NASSCOM, SRO bodies, or regulatory consultation.

Record your current level:

- **Level** ____
- **Date of assessment:** _____
- **Completed jointly by (two roles):** _____ and _____
- **Next step:** one specific action, not a list: _____
- **Next review date:** _____

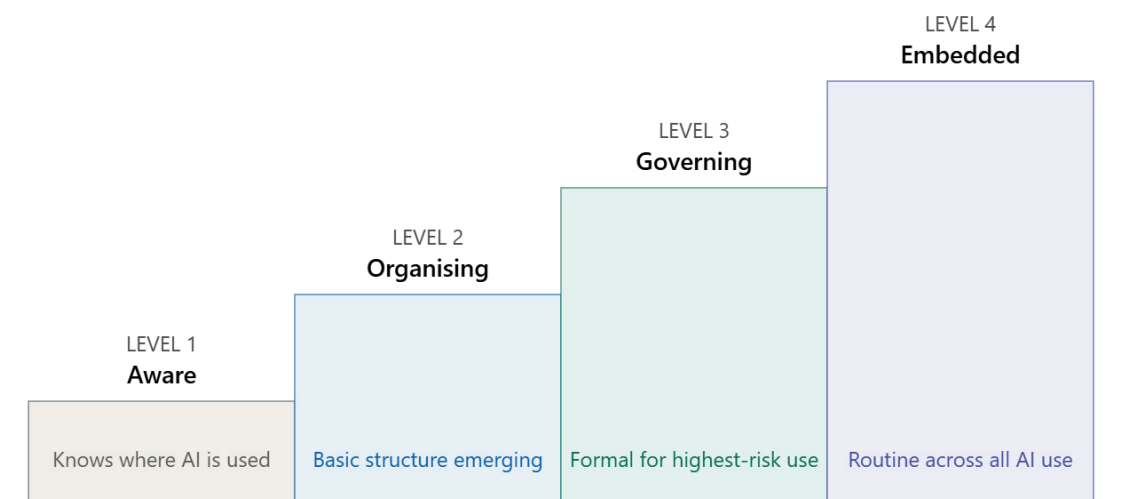


Figure 8: Four Levels of AI Maturity

STEP 2: MODEL DOCUMENTATION

To be completed for each AI system. Two tiers: a baseline that applies to every system regardless of risk, and a risk-scaled tier completed after Step 3.

TIER A: BASELINE (Every System, Regardless of Risk Tier)

These are not governance choices that scale with risk. They are baseline legal and compliance requirements. If any of these cannot be answered, the system should not be live.

System name: What do you call this system internally?

Plain-language description: What does this system do? Write this so a board member who is not technical can understand it. One or two sentences.

Legal basis for personal data processing under the DPDP Act: Which legal basis applies, consent, legitimate use, or another basis? Where is that basis documented? If you cannot answer this, the system may not be lawfully operating.

Data retention: Does the system's data retention comply with the institution's published retention schedule?

- ☐ Yes
- ☐ Not Checked
- ☐ No

Prohibited data: Does the system process any data category that is prohibited under applicable law or the institution's own policy?

- ☐ Yes (Explain)
- ☐ No
- ☐ Not Checked

Accountable committee or function: Which named committee or function has governance responsibility for this system? This is not a single person it is the standing committee, team, or function to whom AI governance questions about this system are directed.

Deployment date: When did this system go live in production?

Deployment status:

- ☐ Live
- ☐ Piloting
- ☐ Planned
- ☐ In testing
- ☐ Decommissioned

Built or procured: Was this system built by the institution, procured from a vendor, or both? If procured, what is the vendor name and contract reference?

TIER B: RISK-SCALED DOCUMENTATION (Depth Depends on Risk Tier from Step 3)

Complete this after Step 3. The depth required scales with the risk tier assigned to this system.

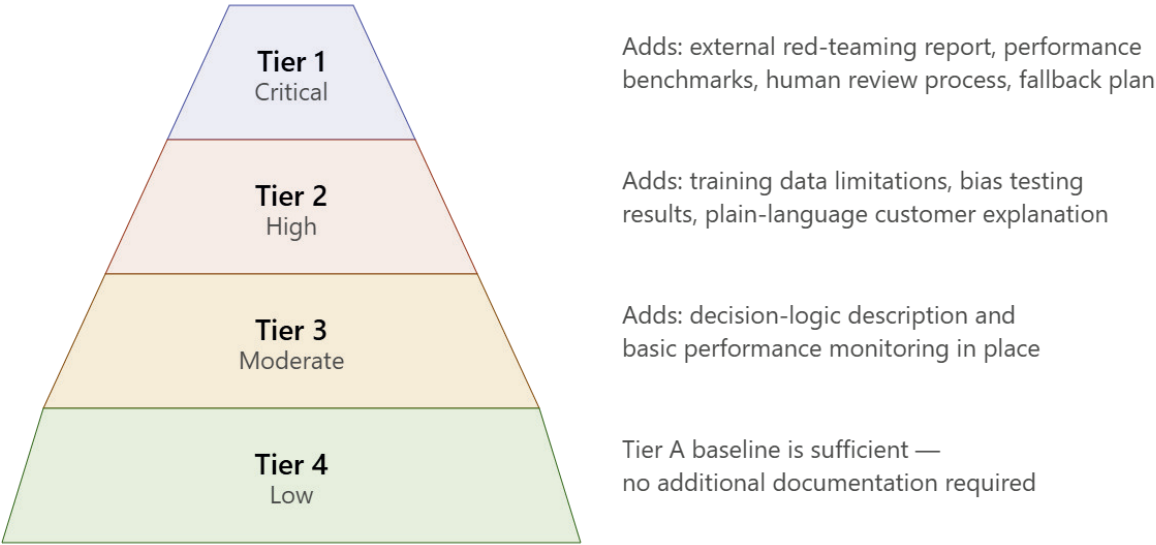


Figure 9: System Risk Tiers & Documentation Requirements

STEP 3: RISK SCORING AND ACCOUNTABILITY

One scoring sheet per AI system. Eight dimensions scored 0 to 3, plus two modifiers.

ACCOUNTABILITY: RECORD FIRST, BEFORE SCORING

Named accountable committee or function: The standing committee, team, or function with AI governance responsibility for this system. Not a single individual.

Does AI governance appear explicitly in this committee's terms of reference?

- ☐ Yes
- ☐ No — if no, this should be corrected before the system is classified as Tier 1 or Tier 2.

Meeting frequency: How often does this committee meet and review AI matters specifically?

Business sponsor: The business unit that owns the use case and requested or approved deployment.

Technical lead: The data science or technology function responsible for model performance.

Escalation contact: Who within or reporting to the committee is contacted first if something goes wrong with this system?

THE EIGHT DIMENSIONS

For each dimension, select the option that best describes this system and record the score.

D1: Customer Impact

Does this system produce outputs that directly affect individual customers?

- No — purely internal, no customer-facing output: 0
- Indirect — a human receives the output and then acts on it: 1
- Direct — the customer is affected, but a human review is possible before the action is taken: 2
- Direct and automated — the customer is affected and no mandatory human review happens before the action: 3

D1 score: ____ / 3

D2: Economic Access Impact

Does this system affect a customer's ability to obtain credit, insurance, investment, or any financial product or service?

- No economic access impact: 0
- Partial — this system is one input among several and a human makes the final decision: 1
- Significant — this system is the primary decision-support input, though a human makes the final call: 2
- Determinative — this system makes or directly drives the access decision: 3

D2 score: ____ / 3

D3: Scale of Deployment

How many customers, transactions, or decisions does this system affect each day?

- Fewer than 100 per day: 0
- 100 to 1,000 per day: 1
- 1,000 to 10,000 per day: 2
- More than 10,000 per day: 3

D3 score: ____ / 3

D4: Reversibility of Harm

If this system produces a wrong or harmful output, how easily can the harm be identified and corrected?

- Immediately reversible — the output can be corrected before any lasting effect: 0
- Reversible with effort — a customer can appeal and the decision can be revisited within a few days: 1
- Partially reversible — harm has begun to occur before it is detected: 2
- Difficult to reverse — the harm occurs before detection, or detecting it at all is unreliable: 3

D4 score: ____ / 3

D5: Existing Regulatory Obligation

Does this system's function intersect with a named RBI, SEBI, IRDAI, or statutory obligation, for example the RBI fair lending circular, KYC Master Directions, or the DPDP Act?

- No existing regulatory obligation applies to what this system does: 0
- An obligation applies tangentially: 1
- An obligation applies directly to the outputs or decisions this system produces: 2
- The system's function is central to a mandatory obligation — for example, the system produces credit rejection decisions that the RBI fair lending circular requires to be explained: 3

Name the obligation if applicable: _____

D5 score: ____ / 3

D6: Explainability

Can this system explain its output to an affected customer or to a regulator in plain language?

- A full plain-language explanation is available and accessible to affected customers: 0
- A partial explanation is available — key factors are disclosed but not in fully accessible form: 1
- A technical explanation is available internally but cannot easily be communicated to a customer or auditor: 2
- The system is not explainable — it is a black box, or producing an explanation would require prohibitive technical expertise: 3

D6 score: ____ / 3

D7: Data Sensitivity

Does this system process special category personal data under the DPDP Act, financial history, biometrics, health data, location, or political or religious affiliation?

- No special category data is used in training or operation: 0
- One category is used, at limited scale: 1
- Multiple categories are used, or one category is used at significant scale: 2
- Multiple categories including financial history or biometrics are used at scale: 3

Categories used: _____

D7 score: ____ / 3

D8: Third-Party Provenance

Was this system built internally, or procured from a vendor — and what governance visibility does the institution have over it?

- Internally built, fully documented, source code and model logic accessible to the institution: 0
- Procured, but model documentation, bias testing results, and audit access have been obtained: 1
- Procured, model documentation obtained, but limited audit access and no AI-specific contract provisions: 2
- Procured, no model documentation obtained, no audit access, no AI-specific contract provisions — black-box vendor system: 3

D8 score: ____ / 3

MODIFIERS**GenAI Modifier**

Does this system use any generative AI component that is capable of producing open-ended text, images, or other generated outputs, as opposed to returning a score, a classification, or a structured output?

Yes — add 3 points to the total / No — add 0

GenAI modifier: ____

The reason for this flat addition is that generative AI introduces failure modes that are categorically different from classical machine learning: hallucination of facts or procedures, prompt injection, and scope drift where the system starts doing things it was not designed for. A credit scoring model that produces a wrong score is a calibration problem. A GenAI system that confidently tells a customer the wrong KYC requirement is a different class of failure.

Agentic AI Modifier

Does this system take autonomous actions — not just generate output for a human to review, but actually do something: initiate a transaction, update a customer record, send a communication, or trigger a downstream process — without a human approving each step?

Yes — add 4 points to the total / No — add 0

Agentic modifier: ____

If yes: how many distinct actions can this system take in a single chain without a human checkpoint?

- ☐ One action only
- ☐ Two to three actions
- ☐ More than three actions

The agentic modifier is higher than the GenAI modifier because agentic risk compounds with each unsupervised action. A system that generates a message for a human to review and send is a GenAI system. A system that generates the message, sends it, updates the customer record, and triggers a payment instruction - all without a human approving each step is an agentic system, and the risk of each error compounding on the last is structurally different from a single-output system.

TOTAL SCORE:

D1 + D2 + D3 + D4 + D5 + D6 + D7 + D8 + GenAI modifier + Agentic modifier = ____ / 31

RISK TIER DETERMINATION

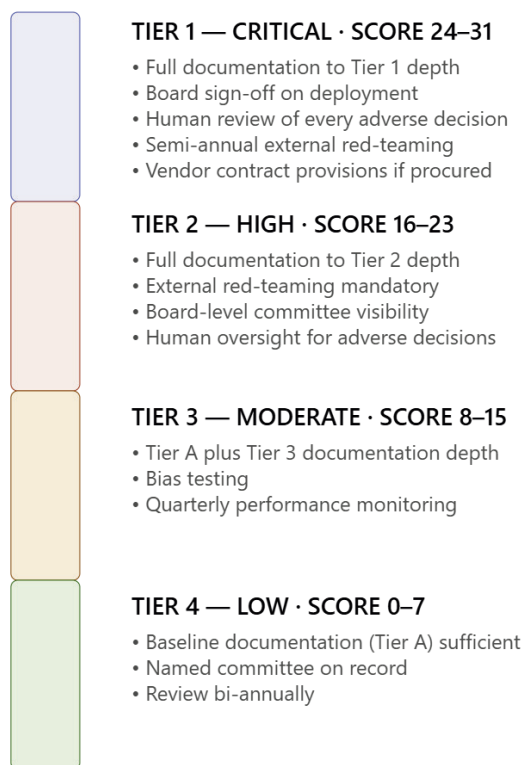


Figure 10: AI Risk Tiers: Low, Moderate, High, and Critical

Assigned tier:

- Tier ____
- Scoring date: _____
- Reviewed by (committee or function): _____

Now return to Step 2 and complete Tier B documentation at the depth corresponding to this tier.

STEP 4: VENDOR DUE DILIGENCE AND CONTRACT CHECKLIST

Complete before signing any contract for a third-party AI system.*

Two parts: pre-procurement questions and minimum contract clauses.

THE ACCOUNTABILITY PRINCIPLE

The regulated entity is always accountable to the RBI for outcomes of any AI system it deploys and no contract usually changes this. What can be contracted is the allocation of commercial liability between the institution and the vendor and this checklist provides the minimum standard for doing that.

A note on reporting: AI incident reporting should be treated as evidence of governance competence, not as an admission of liability. RBI FREE-AI Recommendation 8 states that an accommodative supervisory approach is recommended where the institution has followed appropriate safety mechanisms including incident reporting. Report early and document the response. Do not delay reporting because of concerns about how it will be perceived.

PART A: PRE-PROCUREMENT QUESTIONS

For Tier 1 and Tier 2 systems every item marked Mandatory must be obtained or the gap must be documented with an alternative assurance mechanism. For Tier 3 and Tier 4 systems seek all Recommended items and record any gaps.

Model documentation: what the system does, what it was designed for, what it must not be used for

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
- ☐ Partial
- ☐ No

Training data description: source, time period, known gaps, known biases

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
- ☐ Partial
- ☐ No

Bias testing results: which demographic or segmentation variables were tested, error rates across groups, disparate impact findings, and what remediation was applied

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
- ☐ Partial
- ☐ No

Performance benchmarks: accuracy or recall metrics against a defined test set, with a confidence interval and the methodology used

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

Regulatory compliance declaration: a written statement that the system complies with applicable Indian law including the DPDP Act, RBI guidelines, and KYC norms

Tier 1, 2, 3, 4: Mandatory

Obtained?

- ☐ Yes
☐ Partial
☐ No

Security assessment: penetration testing or security audit results, findings, and remediation status. For GenAI systems this must include prompt injection testing.

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

Explainability mechanism: how the system's outputs can be explained to affected customers and to regulators in plain language, sufficient to meet the RBI fair lending circular requirement

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

AI-specific incident response SLA: response times and escalation paths for AI-specific failures, including system uptime

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

Model update and retraining policy: when the model is retrained, what triggers a material change, and whether the institution is notified before changes go live

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

Audit and inspection rights: a contractual right to inspect model documentation, bias testing results, and performance logs. If unavailable, document the alternative assurance mechanism.

Tier 1 and 2: Mandatory, document if unavailable / Tier 3 and 4: Seek, document if unavailable

Obtained?

- ☐ Yes
☐ Alternative agreed
☐ Not available

Alternative assurance mechanism if audit access unavailable:

Foundation model disclosure: if the vendor's system is built on an open-source or third-party foundation model, name it and confirm whether commercial deployment is permitted under the licensing terms

Tier 1 and 2: Mandatory if foundation model used / Tier 3 and 4: Optional

Obtained?

- ☐ Yes
☐ Not applicable
☐ Not available

AI-specific business continuity provisions: fallback mechanisms for model drift or silent degradation, not just traditional system outage

Tier 1 and 2: Mandatory / Tier 3 and 4: Recommended

Obtained?

- ☐ Yes
☐ Partial
☐ No

Market concentration: approximate number of other Indian regulated entities using the same model. Ask and record the answer or the refusal to answer.

All tiers: Ask and record

Response:

Pre-procurement summary:**Mandatory items required for this tier:** ____**Mandatory items obtained:** ____**Outstanding mandatory items:** _____**Overall recommendation:** Proceed / Proceed with conditions / Do not proceed**Recommended by (committee or function):** _____**Date:** _____**PART B: MINIMUM CONTRACT CLAUSE SCHEDULE**

Eight clauses for Tier 1 and Tier 2 systems. For Tier 3 and Tier 4 systems seek all clauses and document any deviations with the rationale. These are starting points — have contracts reviewed by legal counsel before inclusion.

Clause 1 — Performance Standards and Accuracy Thresholds

The vendor shall maintain the model's performance at or above the benchmarks specified in the schedule attached to this agreement. Performance shall be measured by the agreed metric on a quarterly basis. The vendor shall provide the institution with assessment results within ten business days of each assessment. Where performance falls below the agreed threshold for two consecutive assessment periods, the institution may require remediation within thirty days or a written remediation plan. If remediation is not achieved, the institution may terminate without penalty.

Basis: Without this clause, performance benchmarks exist only in sales materials. The institution has no contractual basis for requiring action when the model underperforms.

Clause 2 — Notification of Material Model Changes

The Vendor shall maintain complete, accurate and auditable records relating to the Services, including applicable model documentation; model, configuration and material-change history; training-data provenance and data-governance records; performance, bias/fairness, robustness and regression-testing results; remediation records; relevant logs; and incident reports. Such records shall be retained for the longer of: (a) the period required by Applicable Law; (b) the Financial Institution's (FI)'s notified record-retention requirements, acting reasonably; and (c) seven years following creation or termination of the relevant Services.

Upon at least 30 days' prior written notice, the FI, its internal and external auditors, and its regulators or their appointed representatives may inspect, audit, and obtain copies of such records and other information reasonably necessary to assess the Vendor's compliance with this Agreement and applicable law. The FI may exercise these rights on shorter notice, or without prior notice where legally required, in connection with a material incident, a reasonable suspicion of material non-compliance, or where delay would materially impair the effectiveness of the audit.

The Vendor may protect its and third parties' confidential information, security, and intellectual property, provided that those protections do not materially impair the FI's or a competent authority's effective exercise of access, information, inspection, or audit rights.

Where direct access is impracticable, the Vendor shall provide equivalent assurance through relevant independent audit reports, certifications, targeted evidence, and reasonable follow-up access.

Basis: An institution cannot govern model changes it does not know about. RBI FREE-AI Recommendation 16 requires lifecycle governance — this clause makes it possible.

Clause 3 — Audit and Inspection Rights

The institution and its appointed auditors shall have the right, on thirty days' written notice, to inspect model documentation, training data provenance records, bias testing results and remediation records, performance logs, and incident reports for the preceding twelve months. Where full disclosure would compromise proprietary information, the parties shall agree on an independent technical audit at the vendor's cost. This right survives termination of the agreement for five years.

Basis: Most institutions currently have no contractual audit rights over vendor AI. This is the largest single vendor governance gap identified in the roundtable

Clause 4 — AI Incident Reporting — Tolerant Reporting Framework

The parties acknowledge that timely AI incident reporting is a governance control and evidence of responsible AI deployment. Reporting is not an admission of liability. The FI shall notify the Vendor expediently after becoming aware of model outputs demonstrably incorrect below the agreed threshold, bias or disparate impact detected in live deployment, prompt injection or adversarial attacks, data breaches, or any regulatory inquiry directed at the vendor relating to this system. The vendor shall co-operate and make the necessary resources available to create a root cause analysis within the period mandated by applicable law. The institution shall determine whether reporting to RBI or CERT-In is required and the vendor shall cooperate fully.

Basis: RBI FREE-AI Recommendations 8 and 22 call for a tolerant, good-faith approach to incident reporting. Institutions should not fear reporting. The tolerant reporting preamble in this clause reflects that intent.

Clause 5 — Liability Cap and Allocation

The vendor's aggregate liability in any twelve-month period shall not exceed one times the total contract value paid in the preceding twelve months, except in cases of gross negligence or wilful misconduct, breach of DPDP Act data protection obligations, or fraudulent misrepresentation in model documentation or benchmarks. Nothing in this clause limits the institution's liability to the RBI or any other regulator that accountability remains solely with the institution regardless of cause.

Basis: One time contract value is the market norm. Unlimited liability is commercially unworkable — one large bank demanded it in an RFQ and received zero vendor applications. The regulatory accountability carve-out is non-negotiable.

Clause 6 — Data Governance and DPDP Compliance

Vendor training data:

The Vendor represents and warrants that personal data used by or on behalf of the Vendor to train, develop, test, validate or improve the Model has been collected, licensed and processed in accordance with Applicable Data Protection Law and all applicable contractual restrictions.

Institution personal data:

The Vendor shall process Institution Personal Data solely on the Institution's documented instructions and exclusively for the purposes of providing the Services. The Vendor shall not use Institution Personal Data to train, fine-tune, develop, validate or improve the Model or any other product or service unless expressly authorised in writing by the Institution.

Data Principal requests:

The Vendor shall provide prompt assistance reasonably requested by the Institution to enable it to meet Data Principal requests and shall not respond directly except on documented instruction or where required by law.

Breach notification:

The Vendor shall notify the Institution without undue delay, and in any event within 24 hours, after becoming aware of an actual or suspected Personal Data Breach; it shall provide continuing updates and all information reasonably required for the Institution to meet its own notification obligations.

Deletion and return:

At termination, or earlier on instruction, the Vendor shall return or securely delete Institution Personal Data within 30 days, subject only to legally required retention and appropriately protected backups, and shall provide written certification of completion.

Basis: RBI FREE-AI Recommendation 17 requires DPDP compliance for AI systems using personal data. If the vendor's training data lacked valid consent, the institution's deployment of that model is a DPDP breach regardless of the institution's own data practices.

Clause 7 — Explainability Obligations

For Tier 1 or Tier 2 systems, the vendor shall provide an explainability mechanism that enables the institution to give affected customers a plain-language explanation of any adverse decision, including credit rejection, service denial, or fraud flag. The explanation must accurately reflect the model's actual decision logic, not a generic template. The vendor shall update the mechanism within thirty days of any material model change and shall provide the institution with the explainability methodology in sufficient detail for the institution to meet its obligations under applicable law.

The Vendor shall design, operate, and maintain the Model and associated services so that the Institution can provide each affected customer with a clear, accurate, and plain-language explanation of any adverse decision (including but not limited to credit rejection, service denial, or fraud flag) that is materially influenced by the Model.

Where the Model is not inherently interpretable, the Vendor shall implement and maintain one or more post-hoc explainability techniques (such as SHAP, LIME, counterfactual explanations, or other industry-accepted methods) that:

- (a) operate at the level of individual decisions;
- (b) identify the principal factors or features that materially contributed to the decision;
- (c) are validated for stability, fidelity, and consistency with the Model's behaviour across relevant customer segments; and
- (d) produce outputs that the Institution can reasonably translate into specific reasons for adverse action notices and other required communications.

The Vendor shall:

- (i) make available all necessary technical resources, documentation, and support to enable the Institution to meet its obligations under Applicable Law relating to explanations of automated decisions, fairness, and non-discrimination;
- (ii) provide the Institution with a detailed description of the explainability methodology, including the techniques used, their assumptions and limitations, the features and data elements exposed for explanation, and the mapping between those features and the reasons provided to customers;
- (iii) ensure that explanations accurately reflect the Model's decision logic to the extent technically feasible, and do not constitute generic or templated reasons that are inconsistent with the actual factors considered for the individual decision; and update the explainability mechanism, documentation and associated outputs within thirty (30) days of any material Model change (including but not limited to changes to model architecture, training data population, feature set, or decision thresholds) and promptly notify the Institution of any known limitations or degradation in explanation quality.

Basis: Illustratively, The RBI fair lending circular, over ten years old, already requires a reason for credit rejection. Technology does not change this obligation. A credit AI system that cannot explain a rejection is non-compliant with existing regulation today

Clause 8 — AI-Specific Business Continuity

The Vendor shall implement and maintain automated monitoring for AI-specific failure modes, including but not limited to model drift, input or output distribution shift, concept drift, calibration degradation, and silent model degradation, distinct from traditional infrastructure or application outage monitoring. The Vendor shall define, document, and agree with the Institution a set of model performance and stability metrics, thresholds, and evaluation windows (the “Model Monitoring Framework”).

The Vendor shall:

- (a) continuously evaluate the Model against the Model Monitoring Framework using production or representative data;
- (b) promptly alert the Institution when monitored metrics indicate degradation beyond agreed thresholds, even if the system remains technically operational; and
- (c) provide sufficient detail in such alerts to enable the Institution to assess business impact, including the nature of the degradation, affected segments or use cases, and recommended mitigation steps.

The Institution shall share with the Vendor, in a timely manner, relevant business-outcome data and feedback (such as approval rates, loss rates, fraud confirmation rates, complaint volumes, or override rates) necessary for the Vendor to calibrate and validate the Model Monitoring Framework and to distinguish between model-related degradation and changes in business conditions or policy.

Fallback Mechanisms for AI-Specific Failures:

The Vendor shall design, document, and maintain fallback mechanisms specifically for AI-related failure scenarios, including significant model drift, distribution shift, unexpected bias emergence, and silent degradation in model performance.

Business Continuity and Testing:

The Vendor shall incorporate AI-specific failure scenarios into its business continuity and incident management procedures for the System. Upon the Institution's request, the Vendor shall participate in business continuity and disaster recovery drills or tabletop exercises relating to this System, including scenarios involving model drift, distribution shift, and silent degradation, up to two (2) times per calendar year at no additional charge. The Vendor shall designate appropriate technical and operational personnel to participate in such drills and shall cooperate in developing realistic scenarios, runbooks, and post-exercise remediation plans.

Coordination of Detection and Response:

The Parties acknowledge that either Party may be the first to observe indicators of model performance degradation. Each Party shall promptly notify the other upon becoming aware of any material or suspected model degradation, AI-specific failure, or significant change in business outcomes that may be linked to the Model. The Parties shall cooperate in good faith to investigate, attribute root causes, and implement remediation or fallback measures as appropriate.

Basis: RBI FREE-AI Recommendation 21 requires institutions to extend their BCP frameworks to cover AI-specific failures. A vendor system can be fully operational in a technical sense while silently producing unreliable outputs due to model drift. The institution cannot manage this risk without vendor cooperation.

Contract summary

Clause 1 Performance Standards:

Included / Not included / Modified _____

Clause 2 Model Change Notification:

Included / Not included / Modified _____

Clause 3 Audit Rights:

Included / Not included / Modified _____

Clause 4 Incident Reporting:

Included / Not included / Modified _____

Clause 5 Liability Cap:

Included / Not included / Modified _____

Clause 6 DPDP Compliance:

Included / Not included / Modified _____

Clause 7 Explainability:

Included / Not included / Modified _____

Clause 8 AI Business Continuity:

Included / Not included / Modified _____

Any clause not included or materially modified — rationale and compensating control:

Reviewed by legal counsel:

☐ Yes

☐ No

Approved by accountable committee:

☐ Yes

☐ No

Date: _____

*This tool is for governance planning purposes only and does not constitute legal advice; all contract terms should be reviewed and finalised by qualified legal counsel before execution.

Tool 4 was co-developed with Black Dot Policy Advisors

 **BlackDot.**

ANNEXURES

ANNEXURE A: FRAMEWORKS REVIEWED FOR THIS PLAYBOOK

S.No	Framework / Instrument	Category	Issuing Body	Key Highlights
1	EU AI Act	International Framework	European Union	Risk-tier governance, mandatory documentation, external oversight
2	NIST AI RMF	International Framework	NIST (US)	Contextual and iterative AI risk management
3	ISO/IEC 42001	International Framework	ISO and IEC	AI management system with audit readiness
4	Singapore Model AI Governance + AI Verify	International Framework	Singapore Government	Practical checklists and deployer responsibility
5	Canada Algorithmic Impact Assessment	International Framework	Government of Canada	Risk score based proportional governance
6	RBI FREE-AI Framework	National Framework	Reserve Bank of India	Board oversight, lifecycle management, audits
7	UNESCO Ethics of AI + RAM	International Framework	UNESCO	Global ethical principles, readiness assessment
8	Infosys Responsible AI Toolkit	Toolkit	Infosys	Enterprise AI governance practices
9	NASSCOM Responsible AI Toolkit	Toolkit	NASSCOM IndiaAI	Awareness and guidance

10	Microsoft Responsible AI Toolbox	Toolkit	Microsoft	Technical fairness and explainability tools
11	OECD AI in Asia's Financial Sector	Research	OECD	Regional adoption and risks
12	OECD AI Principles	International Framework	OECD	Responsible AI principles, accountability
13	Information Technology Act 2000	Indian Law	Government of India	Digital liability and cyber obligations
14	IT Intermediary Rules 2026	Indian Law	Government of India	Platform governance, due diligence
15	Bharatiya Nyaya Sanhita 2023	Indian Law	Government of India	Criminal liability for harm and fraud
16	DPDP Act 2023	Indian Law	Government of India	Consent, purpose limitation, data rights
17	DPDP Rules 2025	Indian Law	Government of India	Operational data governance compliance
18	Consumer Protection Act 2019	Indian Law	Government of India	Unfair trade practices, consumer protection
19	UNESCO Readiness Assessment Methodology	International Framework	UNESCO	Five-dimension country readiness assessment
20	UNESCO Ethical Impact Assessment	International Framework	UNESCO	System-level ethics assessment for procurement
21	GSMA RAI Maturity Roadmap	Toolkit	GSMA	Four-level organisational maturity model

22	Economic Survey 2025-26	Research	Government of India, MoF	AI adoption data, international central bank use cases
23	IMF Financial Stability Report Oct 2025	Research	IMF	Systemic risk, herding behaviour from shared AI models
24	RBI Draft Guidance on Model Risk Management	National Framework	Reserve Bank of India	Model inventory, risk tiering, validation, GenAI controls
25	Deeploy AI Governance Control Framework	Toolkit	Deeploy	Practitioner AI governance controls

Table 6: List of Reviewed Frameworks

ANNEXURE B: LIST OF AI GOVERNANCE CONTROLS

This is a reference list of potential controls, not a mandatory checklist that every institution must fully implement. Not every control here will apply to every AI system, and no institution is expected to adopt all fifty at once. Which controls are relevant, and how many, depends on the risk tier assigned to each system in FSAI Tool 3 - a Tier 4 system may need only a handful of preventive and oversight controls, while a Tier 1 system will draw on most or all four categories.

Use this list as a menu to select from when building out the controls appropriate to a specific system's risk tier, not as a compliance requirement in itself. This annexure exists to give risk and compliance teams the fuller catalogue of options available when a more tailored or advanced control environment is being designed, particularly for Tier 1 and Tier 2 systems where the FSAI Toolkit's baseline requirements may need to be supplemented.

PREVENTIVE CONTROLS

1. Risk classification completed using FSAI Tool 3 before development or procurement begins
2. Bias testing on training data before deployment demographics tested, error rates documented, remediation applied
3. External red teaming by an independent party before deployment mandatory for Tier 1 and 2
4. Design review by risk and compliance function before deployment
5. Formal approval gate named committee sign-off with supporting evidence before go-live
6. DPDP Act compliance assessment for training data and operational data before deployment
7. Data quality assessment confirming training data is fit for the intended purpose
8. Scope definition documented what the system can and cannot do, and what triggers human escalation
9. Hallucination controls configured and tested before deployment (generative AI systems)
10. Toxicity and content guardrails configured before deployment (generative AI systems)
11. Prompt injection testing completed before deployment (generative AI systems)
12. Deterministic override rules defined for regulated processes KYC, credit rejection, regulatory disclosures that the AI layer cannot override
13. Pre-procurement due diligence completed using FSAI Tool 4 Part A before any vendor contract is signed
14. Vendor contract includes all required clauses from FSAI Tool 4 Part B before deployment of third-party AI
15. Foundation model licensing terms confirmed for commercial deployment before use

DETECTIVE CONTROLS

16. Continuous performance monitoring with defined thresholds and named responsible function
17. Model drift detection periodic assessment of whether model accuracy is degrading as real-world data diverges from training data
18. Post-deployment bias monitoring on a defined frequency with documented results
19. Logging of all inputs and outputs sufficient to reconstruct any specific decision if challenged
20. Guardrails in CI/CD pipeline detecting hallucination, toxicity, and scope drift before outputs reach customers (generative AI systems)
21. Scope drift detection mechanism to identify when a generative AI system begins performing functions outside its authorised scope
22. Customer grievance tracking specifically for AI-driven decisions, volume, nature, resolution, and whether the system was modified in response
23. Model change notification from vendor before any material retraining or architecture change (third-party systems)
24. Vendor performance monitoring independent of vendor's own reporting (third-party systems)
25. Anomaly detection on model outputs, flagging unusual patterns in outputs that may indicate degradation or adversarial manipulation

CORRECTIVE CONTROLS

26. Documented AI incident response process distinct from general IT incident management, with named responsible function
27. AI incident definition documented, what constitutes an incident for this specific system
28. Root cause analysis mandatory before redeployment following any AI incident
29. Customer remediation process for individuals adversely affected by an AI failure
30. Rollback procedure documented and tested, how the system is taken offline or reverted to a previous version
31. Regulatory incident reporting to RBI within required timeframes
32. Incident reporting to CERT-In where there is a cybersecurity dimension
33. Tolerant reporting: incident reporting treated as governance competence, not liability admission, per FREE-AI Recommendations 8 and 22
34. Lessons from incidents fed back into preventive controls for the same system and for other AI systems
35. Post-incident re-assessment of risk tier where the incident reveals new risk dimensions
36. Vendor incident reporting obligation: vendor notifies institution within four hours of AI-specific failure per FSAI Tool 4 Clause 4

OVERSIGHT CONTROLS

- 37. Named accountable committee with AI governance explicitly in its terms of reference
- 38. Escalation path from system-level alert to named committee to board, with defined trigger conditions at each step
- 39. Human review mechanism triggered for adverse decisions, named function, timeframe, and logging requirement documented
- 40. Kill-switch or override mechanism, documented ability to suspend or override the AI system without disrupting the underlying business process
- 41. Board or Risk Management Committee of the Board receiving structured AI governance reporting at defined frequency
- 42. Board or RMCB approval for Tier 1 system deployment, documented decision with supporting evidence per Draft Guidance Paragraph 18
- 43. Internal audit of AI governance on an annual cycle, scope, findings, and remediation documented
- 44. Independent validation of all models including third-party models before deployment per Draft Guidance Paragraph 29
- 45. Periodic validation, revalidation triggered by material model changes, significant incidents, or defined time interval
- 46. Audit and inspection rights over vendor systems, contractual right to inspect documentation, logs, and test results per FSAI Tool 4 Clause 3
- 47. Human oversight periodic review, structured review of human override decisions, near-misses, and anomalies to identify patterns per Draft Guidance Paragraph 63
- 48. BCP testing for AI-specific failures, confirmed the institution can continue operations if a critical AI system fails or degrades silently
- 49. Vendor BCP participation, vendor participates in institution's BCP drills for AI systems per FSAI Tool 4 Clause 8
- 50. Annual independent assurance or certification of AI governance programme (Level 4 institutions)

ANNEXURE C: GLOSSARY OF KEY TERMS

Agentic AI

An AI system that takes autonomous actions beyond generating output for human review — initiating transactions, updating records, sending communications, or triggering downstream processes without a human approving each step. Risk in agentic systems compounds with each autonomous step in the chain.

AI Incident

For the purposes of the FSAI Toolkit, an AI incident is any event in which an AI system produces outputs that cause or risk causing harm to customers, the institution, or third parties, or that constitute a breach of applicable legal or regulatory obligations. This includes: model outputs demonstrably incorrect above a defined threshold; bias or disparate impact detected in live deployment; prompt injection or adversarial attacks on a generative AI system; data breaches affecting AI training data or operational inputs; and autonomous actions taken by an agentic system outside its defined scope.

Accountable Committee

The named standing committee or function with AI governance explicitly in its terms of reference, to whom AI governance questions, escalations, and incident reports for a specific system are directed. Not an individual's name — a named institutional body or function, such as a Model Risk Committee or Technology Risk Committee.

Classical Machine Learning

AI systems that learn patterns from historical data and apply those patterns to new situations, producing structured outputs such as a score, a classification, or a recommendation. Failure modes include model drift, bias encoding, and calibration errors.

Deployer

The regulated entity that deploys an AI system, regardless of whether it built the system. Bears full regulatory accountability for the system's outcomes under the RBI FREE-AI Framework and the Draft Guidance on Model Risk Management.

Developer

The entity that built, trained, or fine-tuned the AI system. May be the same as the deployer where the system was built internally, or different where the system was procured from a third-party vendor.

DPDP Act

The Digital Personal Data Protection Act 2023, India's primary data protection legislation, governing how personal data may be collected, processed, stored, and used, and establishing the rights of data principals.

Explainability

The ability of an AI system to provide an accurate, plain-language explanation of a specific output to an affected customer, an internal auditor, or a regulator. Distinct from general technical documentation — explainability refers to the capacity to explain why a specific decision was made for a specific individual.

FREE-AI Framework

The Reserve Bank of India's Framework for Responsible and Ethical Enablement of Artificial Intelligence, published in 2024. India's primary sector-specific AI governance instrument for regulated financial entities. Built on seven foundational sutras and six strategic pillars, with 26 recommendations including Recommendation 26, which calls for an AI compliance toolkit for regulated entities.

Generative AI

AI systems capable of producing open-ended outputs, text, images, code, or structured content in response to inputs, as distinct from classical systems that return a score or classification. Specific failure modes include hallucination, prompt injection, and scope drift.

Grievance

For the purposes of the model inventory, a complaint or concern raised by a customer, a staff member, or a third party that specifically relates to an AI-driven decision or output affecting them.

Hallucination

The production by a generative AI system of outputs that are factually incorrect, internally inconsistent, or fabricated, presented with apparent confidence. Particularly concerning when a generative AI system hallucinates regulatory requirements, product terms, or KYC procedures.

Human Oversight

A substantive, informed review by a named function or role against defined criteria, triggered by specific conditions defined in advance. Not a nominal override capability that is theoretically available but never used in practice.

Model Drift

The gradual degradation in an AI system's accuracy or reliability as the real-world data it encounters in production diverges from the data it was trained on. A system experiencing model drift may remain technically operational while producing increasingly unreliable outputs, a failure mode not detected by standard technology monitoring.

Model Documentation

The set of records describing what an AI system does, what data it uses, what its known limitations are, who is accountable for it, and what controls govern it. Both the baseline Tier A record and the risk-scaled Tier B record in FSAI Tool 2 constitute model documentation.

Prompt Injection

An attack on a generative AI system where adversarial inputs embedded in user prompts or external data sources attempt to override the system's instructions or safety guardrails, causing the system to produce outputs outside its authorised scope.

Red Teaming

A structured adversarial testing exercise conducted by an independent party, designed to identify failure modes, biases, prompt injection vulnerabilities, hallucination risks, and other risks not visible through standard testing. For generative AI systems, includes testing for scope drift and adversarial manipulation.

Risk Tier

The classification assigned to an AI system using FSAI Tool 3 (Tier 1 Critical, Tier 2 High, Tier 3 Moderate, or Tier 4 Low), based on the total score across eight dimensions plus applicable modifiers. The risk tier determines the governance intensity required, including the depth of documentation, validation requirements, oversight mechanisms, and vendor contract provisions.

Scope Drift

The phenomenon in which a generative AI system begins performing functions it was not designed or authorised to perform, for example, a customer service chatbot that begins providing product advice or regulatory guidance outside its defined scope.

Tolerant Reporting

The principle, stated in RBI FREE-AI Recommendations 8 and 22, that AI incident reporting is evidence of governance competence rather than an admission of liability, and that an accommodative supervisory approach is recommended where the regulated entity has followed appropriate safety mechanisms including timely incident reporting.

A note on the sources for these definitions

The definitions in this glossary are derived from three sources, applied in order of priority.

Where a term is defined or described in the RBI FREE-AI Framework, that language is the primary basis. Where a term is addressed in the RBI Draft Guidance on Regulatory Principles for Model Risk Management, that instrument's language is used and the relevant paragraph is cited. Where neither instrument provides a working definition, because the term reflects a governance design choice rather than a regulatory category, the definition is derived from the FSAI Toolkit's own design logic, which itself is grounded in the practitioner conversations and roundtable discussions that shaped this Playbook.

For example, the definition of Accountable Committee reflects a deliberate design choice in FSAI Tool 3: the toolkit assigns accountability to a named committee or function rather than a single named individual, because practitioner conversations for this Playbook established consistently that individual accountability structures are not how banks and NBFCs actually operate for major risk categories. That design choice is not drawn from a specific regulatory provision, it is a translation of governance intent into institutional reality. The definition explains this reasoning.

The definitions of Hallucination, Prompt Injection, and Scope Drift are drawn from the FREE-AI Framework's Chapter 2 discussion of generative AI failure modes and from Draft Guidance Paragraph 54, which introduces specific regulatory requirements for each of these risks in deployed AI systems.

The definition of Tolerant Reporting is drawn directly from FREE-AI Recommendations 8 and 22. The definition of Risk Tier is drawn from Draft Guidance Paragraphs 17 through 20. The definition of Deployer reflects the accountability principle stated in FREE-AI Recommendation 22 and Draft Guidance Paragraph 8.

This Playbook is not a legal document and the glossary does not constitute legal definitions. Regulated entities should refer to the primary instruments for definitive regulatory language and should seek qualified legal advice where the interpretation of specific terms has compliance implications.

ANNEXURE D: NOTE ON METHODOLOGY

Research Basis

This Playbook is grounded in two phases of primary research conducted between January 2026 and June 2026, alongside a comparative analysis of twenty-five governance frameworks, regulations, and toolkits. The full list of frameworks reviewed is in Annexure G.

The Phase 1 comparative analysis examined how over twenty-five global and Indian governance instruments approach accountability, risk proportionality, documentation, human oversight, and continuous monitoring. These five themes correspond to the five convergence signals described in Chapter 3 and form the analytical foundation of the FSAI Toolkit.

The comparative analysis was tested through a closed practitioner roundtable held at the Oberoi, New Delhi, on 16 April 2026, with eight senior participants from regulated financial institutions, technology organisations, legal and policy firms, and the regulatory community. All contributions were made under Chatham House Rules. Participating organisations are listed in Annexure I.

Eight findings from the roundtable directly shaped the Playbook's design — most significantly, that the governance gap in India's financial sector is in operationalisation rather than knowledge, that risk classification should be criteria-based rather than category-based, that red-teaming is the non-negotiable minimum control for high-risk systems, and that vendor accountability is the most urgent unresolved governance challenge. These findings are reflected throughout Chapters 3 and 4 and in the design of all four FSAI Toolkit tools.

PHASE 2: WHY THE FORMAT CHANGED

The original Phase 2 design called for a second closed roundtable with operational practitioners inside regulated financial institutions. Following the Phase 1 roundtable and outreach for Phase 2, the format was changed to structured one-on-one interviews with risk officers, compliance leads, and practitioners across banks and fintechs.

Three factors drove this decision. The RBI's own survey data showing only 21% adoption meant that AI governance is not yet a live operational conversation for most institutions, making a group format unsuitable the discussion would have been dominated by larger, more sophisticated participants while the institutions the Playbook most needs to serve remained silent. Reluctance to participate in a named group setting was also higher among operational practitioners at regulated entities, where current governance gaps might become visible to peers and regulators simultaneously. And the specific purpose of Phase 2 stress-testing the four tools against what institutions actually have in place requires the candour that individual interviews produce and group settings systematically suppress.

Conducting interviews across dispersed institutions is more operationally demanding than a single group session. The format change was made specifically to protect the quality of the research. The Playbook's publication timeline was not affected.

Phase 2 interviews were conducted between April and June 2026. Organisations that participated are acknowledged in Annexure J.

WHAT THE TWO PHASES PRODUCED

The comparative framework analysis produced the five convergence signals and the regulatory mapping in Annexures B and C. The Phase 1 roundtable produced the eight practitioner findings that shaped the Playbook's architecture including the central finding that the governance gap is in operationalisation, not knowledge, which became the argument of Chapter 4.

The Phase 2 interviews produced two things. First, the calibration of Tools 2, 3, and 4 against the reality of what Indian financial institutions currently have in place, the risk tier cutoffs, the documentation depth requirements, and the vendor checklist provisions were all refined against practitioner feedback on what is achievable. Second, and more significantly, Phase 2 produced FSAI Tool 1 itself. Practitioner conversations made clear that institutions could not meaningfully use a risk scoring tool or a vendor checklist without first establishing a basic factual picture of where they are on their AI governance journey. The Responsible AI Adoption Assessment did not exist in the Phase 1 design. It emerged directly from Phase 2, when it became apparent that many institutions needed to answer the question of whether they even knew what AI they had before they could begin to govern any specific system. Tool 1 is therefore the most practitioner-driven element of the FSAI Toolkit.

All roundtable and interview contributions are anonymised in accordance with Chatham House Rules. Any errors of interpretation are the authors' own.

ANNEXURE E: ROUNDTABLE PARTICIPANTS ORGANISATIONAL REPRESENTATION

The April 2026 roundtable was conducted under Chatham House Rules. Contributions are not attributed to individuals. The following organisations were represented:

1. Reserve Bank of India: FinTech Department
2. Dell Technologies: BFSI Practice
3. Black Dot Public Policy Advisors
4. Salesforce: Security, Compliance and Customer Trust
5. Open Network for Digital Commerce (ONDC)
6. IKIGAI Law
7. India Fintech Foundation
8. Future Shift Labs (Facilitating)

The roundtable was moderated by Vidhi Sharma, Head of Responsible AI, Future Shift Labs, and Pranjal Dwivedi, Visiting Fellow, Future Shift Labs.

ANNEXURE F: PHASE 2 CONSULTATION

ACKNOWLEDGEMENTS

The Phase 2 structured one-on-one interviews that validated and refined the FSAI Toolkit were conducted with risk officers, compliance leads, and practitioners across India's banking and fintech sector. Conversations were conducted under Chatham House Rules. The Playbook acknowledges the following organisations for their participation in Phase 2 consultation:

- Think360 AI
- ICICI Bank
- Former Bank of Baroda, practitioner interview
- Standard Chartered
- A few other participants remain unnamed.

The Playbook is also grateful to all individuals who participated in structured conversations during the research phase and whose insights shaped the design of the FSAI Toolkit, even where those conversations were informal and are not formally acknowledged here.

ANNEXURE G: THE GSMA RESPONSIBLE AI MATURITY ROADMAP

STRUCTURAL BASIS FOR FSAI TOOL 1

FSAI Tool 1, the Responsible AI Adoption Assessment, adapts the structure of the GSMA Responsible AI Maturity Roadmap (2024). The GSMA Roadmap is one of the most practically structured organisational AI maturity models available, built around four maturity levels and five governance dimensions with twenty sub-dimensions, each with examples of evidence and artefacts. Future Shift Labs acknowledges the GSMA Roadmap as the structural basis for FSAI Tool 1 and is grateful to GSMA for making this framework available. The GSMA Roadmap is available at gsma.com.

The original GSMA structure:

Four maturity levels: Foundational, Evolving, Performing, Advanced.

Five dimensions: Vision, Operating Model, Technical Controls, Third-Party Ecosystem, Change Management and Communications.

The FSAI Tool 1 adaptation:

The four levels are renamed for the Indian financial sector context: Aware, Organising, Governing, Embedded. The level descriptors are rewritten for the specific governance landscape of RBI-regulated entities. The five dimensions are mapped to Indian financial sector governance realities: Vision becomes Board Commitment and Risk Appetite; Operating Model becomes Governance Structure; Technical Controls becomes Controls Across the Lifecycle; Third-Party Ecosystem becomes Vendor and Procurement Governance; and Change Management and Communications becomes Workforce Readiness and Customer Communication.

The hidden-score design, revealing the maturity level only after all questions are answered, is an addition by our team, introduced to address the specific tendency toward optimistic self-classification observed in practitioner conversations for this Playbook.

ANNEXURE H: NOTE ON TOOL 4

VENDOR DUE DILIGENCE AND CONTRACT CHECKLIST

Tool 4 was developed to help institutions ask the right questions and understand the right categories of contractual protection when procuring AI from third-party vendors. Its reasoning draws on the RBI FREE-AI Framework's accountability principles as well as international third-party risk management practice, including the EBA Guidelines on Outsourcing Arrangements and the MAS Outsourcing and Third-Party Risk Management guidance, both of which establish similar principles around audit rights, proportional diligence, and vendor accountability that informed this checklist's structure.

Tool 4 is a reasoning and readiness aid, not a legal instrument. It is designed to help institutions identify what to ask a vendor and what categories of protection to seek, so that conversations with legal counsel start from an informed position rather than a blank page. The specific clause language provided is illustrative and starting-point material only. Actual contract drafting, negotiation, liability apportionment, and enforceability all require qualified legal counsel familiar with the specific transaction, vendor, and institution. This Playbook and its authors do not accept responsibility for reliance on Tool 4's content in place of legal advice.

.....

This Playbook is available for digital download at futureshiftlabs.com.



Visit Our Website
www.futureshiftlabs.com

FUTURE SHIFT LABS
Skymark One, 13TH Floor, Sector 98, Noida

