DSCI
PROMOTING DATA PROTECTION

Brief on

RBI'S DRAFT GUIDANCE ON REGULATORY PRINCIPLES FOR MODEL RISK MANAGEMENT, 2026

DSCI Brief on RBI's Draft Guidance on Regulatory Principles for Model Risk Management

I. Background

The Reserve Bank of India has released its draft Guidance on Regulatory Principles for Model Risk Management, 2026 for public comment, inviting public feedback and responses due by **July 24, 2026.** The Data Security Council of India (DSCI) welcomes the draft, noting that while its guidance spans all financial quantitative and statistical models, its most critical provisions target artificial intelligence and machine learning (AI-ML). The takeaway for regulated entities is clear: the era of informal AI adoption is over.

II. Who This Applies To?

The guidance reaches every corner of the regulated financial ecosystem. Commercial banks (including foreign banks), small finance banks, payments banks, and regional rural banks are all covered, as are urban and rural co-operatives, NBFCs across all four layers (Base, Middle, Upper, and Top), all-India financial institutions like NABARD, EXIM Bank, NaBFID, NHB, and SIDBI, plus asset reconstruction companies and credit information companies.

Crucially, the obligations apply regardless of whether models are built in-house, licensed from a vendor, or delivered via API. There is no carve-out for third-party systems - the regulated entity owns the risk, always.

III. How the RBI Defines a "Model"?

The scope of this definition is purposefully inclusive. The draft defines a model as any system that includes algorithms, analytics, interfaces, and decision-based rules - taking inputs, applying processing logic, and producing outputs that materially affect business or operational decisions.

The draft explicitly illustrates this: a spreadsheet-based loan pricing calculator that derives lending rates from borrower inputs and applies interest rate grids qualifies as a model. This broad framing means entities cannot exempt legacy tools or low visibility decisioning systems from compliance on the grounds that they are not formally recognized as AI.

IV. The Core AI-Specific Requirements

1. *Governance and board accountability.* Every regulated entity must establish a Board-approved Model Risk Management Framework (MRMF) covering all AI/ML models. The board is not merely informed - it is responsible for approving the entity's risk appetite for model risk, reviewing the framework periodically, and approving tiering policies. The Risk Management Committee of the Board (RMCB) must review high-risk models before deployment and maintain enhanced oversight of all AI models and third-party models, regardless of their assigned risk tier. This is a structural elevation of AI governance from a technology matter to a board-level accountability.

2. *Risk-based model tiering.* Entities must classify every model by risk level, factoring in materiality (financial impact, consumer exposure), complexity (explainability challenges, use of unstructured data), and specifically for AI: the

degree of autonomy the model exercises in decision-making. Tiering determines validation intensity, approval authority, monitoring scope, and business continuity planning. Higher-risk models require RMCB approval to deploy; lower-risk ones may be delegated. Critically, the guidance prevents entities from gaming the system: a low-complexity score cannot disproportionately reduce the overall risk tier of a highly material model.

3. *Explainability as a hard requirement*. Regulated entities must define explainability thresholds for every AI model and ensure outputs are interpretable to the degree the business process demands. Where a model materially drives decisions or significantly affects customers, higher explainability standards apply. Where full explainability is not technically achievable as is commonly the case with large neural networks - the entity must compensate with enhanced validation, output verification mechanisms, continuous monitoring, and usage restrictions. This effectively forces a design trade-off conversation: if a model cannot be explained, it must be constrained.

4. *Hallucination controls for generative AI.* The guidance specifically addresses generative AI models, requiring system-level controls or model design features to mitigate hallucination risk, particularly where outputs drive customer interaction or downstream decision-making. This is a significant operational requirement for any entity deploying LLM-based chatbots, document processing tools, or automated advisory systems.

5. *Bias and fairness assessment.* Entities must actively identify bias and discriminatory outputs - particularly in use cases involving differential treatment of customer segments. Fairness assessments are mandated, with required remediation including recalibration or redesign. For complex models, the guidance recommends constraining model complexity (regularization) and limiting feature selection as risk controls, not just post-hoc audits.

6. *Human oversight and kill-switch requirements*. Every AI deployment must include robust human oversight arrangements. This encompasses human-in-the-loop or human-on-the-loop mechanisms, override and suspension capabilities, and explicit kill-switch arrangements that allow immediate deactivation. Personnel performing oversight must possess genuine understanding of model functioning, not just operational familiarity. The guidance also flags automation bias and decision fatigue as risks that the oversight design must address.

7. *Customer disclosure obligations.* When AI or ML systems interact with customers or drive decisions affecting them, entities must disclose that the customer is interacting with an AI-based system, including its limitations. Customers must also be given the option to switch to human assistance on request. This converts what many entities currently treat as an internal governance question into a consumer-facing obligation.

8. *Third-party AI vendor requirements.* The guidance is unsparing on third-party models. Entities must conduct independent validation of all vendor models regardless of any certification or assurance the vendor provides. Contracts must include access to sufficient technical documentation to enable validation, audit rights for both the entity and RBI, and continuity and exit arrangements. Where a vendor does not disclose adequate information about an AI model, the

entity must identify the resulting risks and implement mitigants including restricting usage. For material third-party AI dependencies, entities must additionally assess supply chain concentration risk and the impact of provider-driven model updates.

V. Areas Open for Public Comment

As this is a draft guidance released for public consultation, DSCI members that include regulated entities, NBFCs, technology firms, and other stakeholders have an opportunity to shape the final framework before it is enforced. Several areas in the draft appear to invite substantive feedback, given their breadth, technical complexity, or operational ambiguity.

- Scope of the model definition
- Proportionality across regulated entity types
- Explainability thresholds and technical feasibility.
- Third-party and vendor model validation
- Generative AI and hallucination controls
- Human oversight design and automation bias
- Customer disclosure requirements
- Kill-switch and deactivation practicalities
- Timeline and transition arrangements

VI. Interaction with existing RBI directions

The guidance notes that in cases of inconsistency, existing RBI Directions will prevail.

VII. Open for Feedback

We encourage our members and stakeholders to seek clarification on how this guidance interacts with existing frameworks on IT risk, outsourcing, and cybersecurity, and whether any harmonization or consolidation is intended over time. Additionally, members are also encouraged to write their suggestions, inputs and comments to policy@dsci.in for a consolidated industrial response.

The public comment period closes **July 24, 2026.**