Comments to the Draft Guidance on Regulatory Principles for Model Risk Management, dated 24 June 2026

On 24 June 2026, the Reserve Bank of India (RBI) released the **"Draft Guidance on Regulatory Principles for Model Risk Management, 2026"** (hereafter "**Draft Guidance**")

The Draft Guidance proposes broad regulatory expectations for model risk management across the model lifecycle. It intends to strengthen governance, oversight, risk management and controls of Regulated Entities (REs) using models (including AI / ML models).

In this response we present our comments to the Draft Guidance, through six recommendations

1. Including a formal cross-functional governance and escalation mechanism could further strengthen the governance framework. Specifically, we recommend that the Draft Guidance:

   a) establish a cross-team committee or similar forum where different teams jointly determine model fitness, effectiveness of the proposed mitigants, escalation mechanisms for resolution, and other aspects of model deployment and behaviour;

   b) offer indicative principles for ensuring the independence of validators, which account for expertise, organisational standing and incentives for independent validators and

   c) In line with the recommendations of the FREE-AI Committee, ground the Model Risk Management Framework (MRMF) in a distinct Board-approved AI policy that establishes a common institutional position on responsible AI adoption before individual models enter the development and validation pipeline.

2. We propose strengthening customer protection rights, specifically by complementing explainability requirements with rights such as the right to information, contestability, grievance redressal, right of compensation founded on a clear delineation of liability.

3. We recommend that REs establish an AI incident management framework comprising standardized taxonomies of potential model harms and risks, reporting and resolution protocols, and mechanisms to systematically identify, respond to model-related harms.

4. We propose acknowledging the practical limitations in independently validating proprietary third-party models and provide guidance on complementary governance measures that REs could adopt. Specifically, we recommend the Draft Guidance to consider alternative assurance mechanisms from third-party model providers and enhanced post-deployment controls.

5. We recommend the Draft Guidance to consider providing more granular details on identification of concentration risk, documentation of concentration risk using AI inventory and exit preparedness.

6. We suggest institutionalising a feedback mechanism for the Draft Guidance to remain current and responsive to advances in AI use cases, capabilities, emerging risks and their management practices.

Key Recommendations

1. **The governance framework may be further strengthened by:**

a. **A formal cross-functional governance and escalation mechanism:** The Draft Guidance requires deployment to be coordinated with relevant stakeholders, including IT and data functions. It also requires validators to assess whether the model is aligned with its intended use. These are welcome initiatives. However, these do not account for the distributed nature of AI supply chains, the nuances of model-use case alignment and the need for different teams to compare notes throughout the model lifecycle.

Experience from other jurisdictions points to the benefits of establishing cross-functional governance and escalation mechanisms to ensure greater cohesion in the model lifecycle and a shared understanding of material riskiness and appropriate mitigants across teams. The Monetary Authority of Singapore (MAS) identifies[1] cross-functional forums involving risk, legal, compliance, technology, data, audit and business functions as a means of ensuring consistent AI governance. The Financial Stability Board (FSB) recommends[2] coordination among business owners, risk managers, technology experts, compliance officers, legal specialists, developers and other relevant stakeholders, supported by common methodologies, regular forums, communication channels and feedback loops. We recommend establishing a standing process through which different teams jointly determine:

- whether AI is necessary, or a simpler model or deterministic process would suffice;
- whether the model is appropriate for the underlying problem;
- whether the proposed data use is lawful and appropriate;
- whether model, operational, cyber, conduct & customer risks have been considered *together*;
- whether the proposed mitigants adequately address those risks; and
- whether the residual risk is acceptable.

Model fitness is not solely a technical question. A model may be statistically accurate yet remain unsuitable because it cannot support legally required explanations, relies on inappropriate data, produces unacceptable customer outcomes, introduces excessive third-party dependence or transfers risk into operational processes. Cross-functional governance is consequently a recurring feature of leading AI and model-risk publications.

A cross-functional and preferably inter-disciplinary Committee, or a functionally equivalent management-level forum may be constituted for material or high-risk AI use cases. It could comprise the business or use-case owner, developers, the independent validation function, legal and compliance, privacy, cybersecurity, operational risk, outsourcing or third-party risk, and customer-protection functions. Internal audit may attend as an observer where this does not compromise third-line independence.

Such a forum would assess use-case and model alignment, identify the dimensions and thresholds against which the model would be evaluated, review mitigants and compensating controls, consider

[1] https://compliance.waystone.com/wp-content/uploads/2026/01/WCS_Consultation-MAS-Guidelines-on-Artificial-Intelligence-Risk-Management.pdf
[2] https://www.fsb.org/uploads/P100626.pdf

customer and legal implications, and recommend deployment conditions or escalation to the Risk Management Committee of the Board (RMCB). To further enhance the effectiveness of this forum, it should lay out clear guidance on resolution, reconnaissance and documentation of divergent views:

(i) *Model developers or business owners should not close or downgrade a material validation finding without review or concurrence from the independent validation function.* This provides further teeth to the requirement that work performed by developers or users remain subject to critical review by an independent party, and that validators have sufficient authority to challenge developers and elevate deficiencies (III.15 of the Draft Guidance). The FSB similarly requires that evaluation, testing and risk-management functions remain functionally separate from development teams. *We therefore recommend that the Draft Guidance should clarify that a material validation finding may not be closed or downgraded by the model owner or developer without documented review and concurrence from the independent validation function. Where concurrence cannot be reached, the matter should be escalated to the designated approval authority.*

(ii) *Unresolved material findings should be escalated to an authority independent of the model owner.* The Draft Guidance requires validation reports containing key findings and recommendations to be placed before the RMCB or a delegated authority, while the RMCB must review high-risk validation reports, breaches and other material concerns (IV(B).33 and II(c).12(1), 12(5)). The Federal Deposit Insurance Corporation (FDIC) additionally expects[3] validators to have explicit authority to elevate findings to a function with sufficient organisational stature to ensure timely and substantive action. *We submit that requiring an escalation of unresolved material findings would offer an additional layer of accountability.*

(iii) *Exceptions should be time-bound and should record the rationale, accountable risk owner, compensating controls and remediation date.* The Draft Guidance already requires exception-approval arrangements to specify approval authorities, thresholds, additional requirements and remediation timelines, and requires the rationale for the exception to be documented. In the same vein, the FDIC says that validation exceptions should be temporary, subject to control mechanisms such as timelines and limits on use and supported by compensating controls where validation remains incomplete. Identifying the accountable risk owner would reinforce the clear allocation of responsibility expected under both frameworks. *We submit that the Draft Guidance may consider clarifying that exceptions be time-bound and document the rationale, accountable risk owner, compensating controls and remediation deadline.*

(iv) *The decision record should preserve the validator's original conclusion alongside management's response and the final approval or exception decision.* The Draft Guidance requires documentation of validation outcomes and approval rationales (Paragraphs 10, 12(3), 34, 45 and 37). This documentation could also track recommendations and attendant responses and reasons that support the final decision. *Recording the decision and the management's position further strengthens transparency and preserves institutional history.*

[3] https://www.fdic.gov/model-risk-management-revised-guidance.pdf

(v) *The validation function should have a direct, documented escalation route to the RMCB for material concerns.* The Draft Guidance requires high-risk validation reports and material concerns to reach the RMCB. Validators may also benefit from a direct escalation route. *We recommend that a direct escalation route may further support validator's independence.*

The Draft Guidance already requires documented exception approvals, approval thresholds, risk mitigants and remediation timelines. The proposed escalation structure would build on these provisions by specifying how an exception is reached when the relevant functions disagree.

b. Indicative principles for ensuring the independence of validators

The Draft Guidance requires every model, including third-party models, to undergo independent validation. Validators may be internal but must remain independent of model development, ownership and use. Validation is required before deployment, periodically thereafter, and following material changes or specified triggers. It must assess data, assumptions, design, performance, limitations and fitness for intended use. Findings must be documented and submitted to the RMCB or its delegated authority, with high-risk models receiving RMCB review (Paragraphs 7(8), 10, 29, 30, 36). We welcome this guidance and note that substantive guidance on independence may help achieve it in practice.

It is imperative that the teams are independent functionally and not just in org-design. The FDIC's model-risk guidance, for example, describes effective challenge as depending on *incentives, competence and influence.* Separation from development is important, but the validator must also have the expertise, authority and organisational standing necessary to ensure that identified issues are addressed. The FSB similarly expects validation, testing, risk-management and audit functions to remain functionally separate from development teams and to be sufficiently empowered and skilled to provide effective challenge.

Validators may be located within the RE, provided they are functionally independent. Relevant safeguards may include:

- reporting lines separate from development and model use;
- remuneration not linked to model approval, deployment or revenue;
- unrestricted access to data, code, documentation and relevant personnel;
- authority to require additional testing and restrict model use;
- adequate staffing and technical expertise;
- conflict-of-interest requirements; and
- periodic internal-audit assessment of the effectiveness of validation.

c. Anchoring the MRMF in a distinct AI policy

The Draft Guidance requires the MRMF to cover model taxonomy, governance, scope of usage, risk tiering, validation, approval, monitoring and business continuity. It also requires institutions to consider fairness, ethical considerations and bias when selecting and developing models (most significantly in Paragraphs 10, 26(2)). These provisions are very relevant and welcome.

Additionally, we recommend grounding the MRMF in a distinct Board-approved AI policy. The MRMF primarily emphasises the proper classification, testing, validation, approval and monitoring of the model (and its behaviour). However, an AI policy grounds this imperative in the prior questions of *why, where and on what terms the institution will use AI*. The AI policy establishes a common institutional position on responsible AI adoption before individual models enter the development and validation pipeline.

The Framework for Responsible and Ethical Enablement of Artificial Intelligence (FREE-AI) Committee, MAS and the FSB each recognise the need for institution-wide AI principles or governance arrangements that align AI adoption with strategy, risk appetite, ethics and customer outcomes. The FSB is considering recommendations for documentation of intended and prohibited applications, affected business lines and customers, accountability, dependencies, mitigants, validation findings and outstanding remediation.

In India, the AI policy could translate the FREE-AI Sutras into operational standards. Alignment should not be demonstrated through general statements of principle alone. REs should maintain evidence appropriate to the materiality and risk of the use case. Such evidence may include:

- **an inventory of AI use cases**, recording their purpose and scope, affected customers and business lines, accountable personnel, third-party providers, dependencies, approved uses, prohibited uses and operational limitations;
- **documented assessments of foreseeable benefits and harms**, including the rationale for using AI, expected benefits, potential adverse outcomes, fairness, ethical considerations and bias;
- **approved thresholds or assessment criteria for performance, fairness, explainability and robustness**, calibrated to the materiality and customer impact of the use case;
- **validation findings, recommendations and outstanding remediation actions**, including the conditions or timelines attached to models approved with exceptions;
- **records of human interventions and overrides, approved exceptions, incidents and near misses**, together with the action taken in response;
- **customer complaints, challenged decisions, redress sought and resolution outcomes**, including customer satisfaction with the resolution where relevant;
- **periodic reporting to the Board or the RMCB on material AI risks**, including high-risk uses, performance, failures, mitigation actions, breaches and models operating under policy exceptions.
- **operational accountability,** in addition to identifying the model owner, it should identify the owner of the AI use case, data, technology, customer process and third-party relationship. It should state who sets risk thresholds, accepts residual risk, approves exceptions, restricts model use and undertakes customer remediation.

The AI policy would supplement the MRMF. While the MRMF would govern model-specific risks, the AI policy would state the organisation's posture with respect to any instance of AI adoption.

2. Customer protection rights may be expanded and made more specific

Models are increasingly being used by REs for the ease and efficiency of administration and improved customer interaction mechanisms, as well as for critical processes such underwriting, especially low-income customers, risk monitoring, and fraud detection.[4] The Draft Guidance defines explainability as, the "property of a model to express important factors influencing its results, in a way that is understandable". The Guidance adopts the principle of proportionality in risk management decisions on explainability – the higher the materiality of risk of the model, the higher the threshold for explainability. Where full explainability may not be possible due to the complexity of the model, higher thresholds must be used for the risk management process ("including enhanced validation and testing, mechanisms to verify and corroborate model outputs prior to their use, frequent validations and continuous monitoring, usage restrictions, and other compensating controls").

While the Draft Guidance provides important customer protections such as provisions for model explainability used by REs, these provisions alone may be insufficient to address customer protection concerns and harms by models. We recommend:

(i) *Need for a wider range of customer protection safeguards in the design, development and validation of models used by RE:* Customer protection safeguards[5], including but not limited to explainability safeguards, must be included in Draft Guidance. This could be through the introduction of customer protection impact assessments which assess potential risks and harms to customer rights. Customer protection impact assessments could help address structural limitations in AI regulatory sandboxes. Structural limitations in this context include edge cases and adversarial human behaviour which could be covered through simulations. Real progress may be seen through the operationalization of AI-specific evaluation through concrete testing protocols, continuous assurance mechanisms and data governance arrangements that regulators can apply. This rights-by-design approach could be embedded into the model design framework, which could enable relationships of trust and assurance between customers and RE.

(ii) *The definition of explainability in the model guidance – specifically the aspect of 'understandability' of explanations – must be tailored to the needs of the person to whom the explanation is being provided:* When providing an explanation about a model's outcome or decision, the model owner must avoid providing a generalized explanation for the sake of compliance. Various international conventions and bodies on model explainability concur that explainability requirements must be determined by what needs to be explained, the target audience, materiality of the use case, and complexity of the model. For instance, auditors and supervisors will require more comprehensive technical information compared with policyholders. Providing context-specific explanations is critical to build an environment of trust between RE, regulators, and customers. As such, human oversight becomes important especially to understand the specific context of the target audience, and to answer any questions that an algorithmic interface (if it exists) may not be able to provide

[4]https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/FREEAIR130820250A24FF2D4578453F824C72ED9F5D5851.PDF

[5] https://www.minister.industry.gov.au/charlton/media/ai-consumer-safety-priorities

an explanation for. We imagine the principle to have baked-in reasonable restrictions, in that, it only obliges the deployers of AI to make the output intelligible by sharing the underlying factors and logic (exogenous explainability) and not lay bare the intricacies of the model itself (decompositional explainability).[6]

(iii) *Customer protection rights, including provisions on model explainability, must be combined with other complementary rights to ensure a holistic approach to potential risks and harms by models:* Model explainability, while important, could be read with a right to information or the right to inquiry, the right to contest decisions made by models, and right to grievance redressal mechanisms.[7] These are provisions that go beyond a one-way provision of an explanation of decisions or outcomes by models and enable a feedback loop between RE and customers. The right to contest explainable models can be defined as the dynamic process which enables users to interact with algorithmic systems, challenge decisions, and obtain revised outcomes, thereby embedding accountability and responsiveness. Contestability upholds human dignity and autonomy by ensuring individuals can challenge impactful decisions, whether made by humans or algorithms.[8] Additionally, India's Digital Personal Data Protection Act[9] and Rules[10] require Data Fiduciaries and Consent Managers to provide adequate measures for grievance redressal mechanisms. The Draft Guidance mention the need to provide for grievance redressal mechanisms but could provide further details about the mechanism. Key considerations on the degree of transparency and explainability required may include reliance on AI for the final decision (i.e., the degree of AI autonomy), level of impact on customer or risk management outcomes. Financial institutions may also consider offering "opt-out" options or channels that allow customers to request a review of AI-generated outputs or human assistance.[11] Keeping in mind the need to balance customer protection rights with adopting complex models for better efficiency, international bodies and conventions advise that RE adopt simple, explainable models. However, should the REs adopt complex models, sufficient guardrails need to be put in place to ensure grievance redressal and contestability, including human oversight mechanisms for models. Finally, customer protection grievances must also include clarity on accountability and liability measures, especially for financial losses that customers may face due to inaccurate or incorrect outputs or decisions made by models.

3. Need for an AI incident management framework and a distinct categorization of harms or risks caused by models.

The Draft Guidance could make explicit a separate category of harms – harms to customers caused by the design or deployment or validation in the model itself. These harms and risks by models could include bias, discrimination, exclusion, misinformation, misuse, or unsuitable financial services. These harms and risks may arise at any point of time in the design, development,

[6] https://dvararesearch.com/wp-content/uploads/2025/09/Responsible-and-Trustworthy-AI-Framework-1.pdf
[7] https://www.fsb.org/uploads/P100626.pdf
[8] https://arxiv.org/html/2506.01662v1
[9] https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf
[10] https://xn--m1bdba5a7gresc7dsa.xn--11b7cb3a6a.xn--h2brj9c/documents/act-and-policies/digital-personal-data-protection-rules-2025-gDOxUjMtQWa?pageTitle=Digital-Personal-Data-Protection-Rules-2025
[11] https://www.mas.gov.sg/publications/consultations/2025/consultation-paper-on-guidance-on-artificial-

validation, and deployment of models, and as such, require a holistic approach in its management. In lieu of this, we suggest:

(i) *The standardization of definitions, taxonomies, and protocols for the management of AI incidents caused by models used by RE and customers:* This would include the development of multi-dimensional taxonomies which can capture incidents of harms or risks by models from a multi-disciplinary standpoint (technical, legal, or social perspectives); impact assessments to identify incident-specific AI harms and how to address them (either through individual grievances or through a system overhaul); guidance for evidence-based harms and risk classifications; and finally, a protocol to identify at which stage of the model lifecycle the harm or risk may have manifested.[12] For this, a centralized taxonomy system could be created that may be used as a reference by RE, who can then build their own taxonomy and protocols. Additionally, RE may decide if a separate system of taxonomy and protocols with a higher threshold should be created for models created by third party entities.

(ii) *REs must be mandated to inform its customers of the potential risk or harm to its customers:* REs must provide customers notice about the potential risk or harm caused by its models to its customers at the time of investigation. The entity must also provide an update about the final decision by the ombudsman (or any other higher authority) about the incident. This reporting through a notice may be mandatory for large scale harms caused by models to customers, similar to incident reporting for data protection harms to the Data Protection Officer. But REs must also be required to make regular publications about incident management, including ways by which future harms or potential risks will be mitigated or avoided.

(iii) *Establish clear resolution mechanisms and evidence-sharing protocols with customers:* The notification mechanisms must be complemented by a robust process for reporting breaches or anomalies as well as for tracking, escalating, and resolving issues or incidents. High-risk customers could also be provided with information about evidence needed to be collected about a potential risk or harm, which is required by the ombudsman to conduct an investigation into the complaint.[13]

4. Acknowledge the practical limitations in independently validating proprietary third-party models and provide guidance on complementary governance measures that RE should adopt

The Draft Guidance places certain additional obligations on REs vis-à-vis the governance of third-party models. For instance, the Draft Guidance subjects third-party models to: (i) independent validation by the RE on the model's soundness, suitability, data and limitations and (ii) places them under oversight by the RMCB. Further, it expects REs to put in place contractual arrangements with third-party service providers that allow them access to technical documentation and exercise audit rights over these models.

These expectations may be difficult to operationalise in practice when REs deploy proprietary

[12] https://cerai.iitm.ac.in/docs/AI_Incident_Reporting_V1.pdf
[13] https://www.mas.gov.sg/publications/consultations/2025/consultation-paper-on-guidance-on-artificial-intelligence-risk-management

third-party models, including frontier Large Language Models (LLMs). For instance, financial institutions deploy generative AI capabilities through enterprise AI platforms offered by global technology providers.

In such arrangements, while the RE develops and governs the downstream banking use case, the underlying foundation model is proprietary and is owned, trained and updated by the third-party model provider, outside the RE's control. REs as the customers of AI providers may have limited bargaining power[14], and the information related to model architecture, training data and methodology may not be available to the RE due to intellectual property protections. Moreover, third-party model providers may expressly limit[15] contractual warranties regarding accuracy or fitness of the model outputs, which constrain the ability of the RE to obtain more assurances from the third-party model providers.

Thus, while RE remains responsible for the third-party model outcomes, it may be important to acknowledge the inherent information asymmetry between the third-party model providers and the RE. This asymmetry can present unique challenges for risk assessment of these models and implementing appropriate safeguards. The REs may not possess the information necessary to independently validate the model to its satisfaction or obtain the breadth of technical documentation and audit access contemplated under the Draft Guidance. Robust model validation depends[16] on access to model details and underlying data, which is challenging for models procured from third-party service providers.

Accordingly, we recommend the Draft Guidance clarify regulatory expectations in scenarios where REs are unable to obtain sufficient information to independently validate and audit third-party models. The Draft Guidance may consider supplementing the existing guidance with compensating governance measures that should be employed by REs to demonstrate oversight. We believe this would preserve the principle that accountability for model outcomes rests with the RE while acknowledging the operational realities of deploying third-party models. The Draft Guidance may consider the following:

i) **Independent assurance from third-party model providers:** When direct access to technical information is not possible, REs may obtain alternate forms of assurance regarding the model's risk and governance. This should complement RE's own assessment. This may include certifications like ISO/IEC42001[17] regarding the vendor's model governance and risk management process, ensuring the design choices[18] of the third-party model provider allow the RE to interact with models and query it.

ii) **Enhanced post-deployment controls:** REs may implement additional safeguards before deployment, including restricting the use of such models to lower-materiality or lower-risk use cases, more intense human oversight arrangements such as lower thresholds for human

[14] https://openai.com/en-GB/policies/services-agreement/
[15] https://openai.com/en-GB/policies/services-agreement/

[16] https://www.oecd.org/content/dam/oecd/en/publications/reports/2026/01/supervision-of-artificial-intelligence-in-finance_1295e5e2/92743dc1-en.pdf
[17] https://www.iso.org/standard/42001
[18] https://assets.kpmg.com/content/dam/kpmgsites/sg/pdf/2025/05/6-key-challenges-of-auditing-ai-and-how-to-approach-them.pdf

intervention and escalation, periodic performance review of the model, to mitigate risks arising from information asymmetry in the case of third-party models.

5. Provide guidance on monitoring and mitigating concentration risks arising from reliance on a limited number of AI model providers

The Draft Guidance recognises additional risks arising from dependence on a limited number of AI model providers including supply chain risk, limitations in independent validation, and changes in model behaviour or capabilities resulting from provider-driven updates. The Draft Guidance would further benefit from clarifying corresponding governance expectations from REs to monitor and mitigate concentration risk.

In a deeply interconnected financial system like ours that includes RE, fintechs, cloud providers, digital platforms and other third-party vendors, reliance on a small number of AI and cloud providers expands the surface area of risk beyond the RE's direct control. As REs increasingly rely on the same vendors for models: operational disruptions, cyber incidents or any action of a single provider could simultaneously affect multiple REs with potential implications on financial stability[19]. Given how it is important for the RE and the regulator to understand where the most important concentrations exist and the potential operational and systemic impact of disruptions affecting AI providers in their operations, it would be beneficial to have guidance on appropriate governance measures for REs to monitor and mitigate concentration risks.

We recommend the Draft Guidance to consider providing more granular details on:

i) **Identification of concentration risk:** REs may be expected to assess concentration risks before entering into contract with the AI model provider and periodically thereafter. The same may be assessed through evaluating factors like: i) whether multiple models[20] are provided by the same vendor, ii) the ability/ inability to substitute[21] the model provider, iii) potential impact in event of a disruption or failure from the provider end[22], iv) the criticality[23] of business functions that are dependent on the model, v) whether other REs have similar dependence on the same provider.

ii) **Documentation of concentration risk using AI inventory:** The model inventory and documentation that is required by the Draft Guidance to be maintained by each RE should include their exposure to AI model providers, dependencies on common providers. At the supervisory level, the AI inventories may be leveraged[24] to develop a sector wide view of concentration risks, consistent with recommendation **23** of the FREE AI Report to establish an AI repository.

[19] https://www.bankofengland.co.uk/financial-stability-in-focus/2025/april-2025
[20] https://eur-lex.europa.eu/eli/reg/2022/2554/oj/eng
[21] https://eur-lex.europa.eu/eli/reg/2022/2554/oj/eng
[22] https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/approach/critical-third-parties-approach-2024.pdf
[23] https://www.fsb.org/uploads/P101025.pdf
[24]https://rbidocs.rbi.org.in/rdocs/PublicationReport/Pdfs/FREEAIR130820250A24FF2D4578453F824C72ED9F5D5851.PDF

iii) **Exit Preparedness:** The REs may be expected to incorporate[25] concentration risk considerations in their business continuity plans and exit strategies for AI models. The strategy may account for alternative third-party AI providers, specify exit triggers, possible migration options, data portability and interoperability, stress testing exit strategies to minimise disruption if continued reliance on a provider is no longer viable.

6. Institutionalise a feedback mechanism so that the Draft Guidance is responsive to advances in AI use cases, capabilities and emerging risks

The Draft Guidance attempts to strengthen governance and risk management of models used by REs. It will benefit from building institutional capacity for continuous and systematic feedback from the REs to ensure that the Guidance remains relevant over time and evolves alongside advancements in AI capabilities, adoption trends in the industry and manages to cater to the risks emanating from model adoption.

In this dynamic and rapidly evolving AI environment, where new use cases and versions of the technology emerge every few months, staying at the frontier of managing associated AI risks would require continuous and systematic efforts. For instance, the advancements in Agentic AI, especially since the debut of Claude's Mythos Preview, may have lowered the barriers for attackers by making cyber-attacks cheap, scalable, swift and accessible. They challenge traditional risk management practices, which often rely on periodic risk assessment, patch management and post-incident response and are suited to an environment in which banks had sufficient time to identify and remediate vulnerabilities before they were exploited. Any model risk management measure and control would also need to account for risks generated by the newer developments in technology.

In this context, the Draft Guidance must stay responsive and adaptive to the threat landscape emerging around AI systems. The Draft Guidance would benefit from establishing a feedback loop through which evidence can be systematically gathered from REs, periodically on:

a) emerging use cases and capabilities,
b) emerging threat landscape and areas of concern and
c) emerging risk management practices

This evidence can be used to refine supervisory expectations by identifying areas warranting additional safeguards, institutional clarity or relaxations from existing expectations, thereby making the guidance responsive. Moreover, this evidence would form the empirical basis for assessing when and which model risks warrant guidance, beyond what is already available.

For REs themselves, this evidence, when made available, would offer near real-time industry benchmarks and allow industry bodies to set baseline standards. For instance, the MAS conducted a thematic review[26] of AI (including generative AI) model risk management practices in mid-2024 of selected banks. The objective of this exercise was to gather good practices for sharing across the industry. This exercise resulted in an *Information Paper on AI Model Risk Management* that sets out risk practices across: (i) Governance and Oversight, (ii) Risk Identification, inventorisation and materiality assessment, (iii) AI Development and Deployment, (iv) Generative AI and (v) Third-party AI.

[25] https://www.fsb.org/uploads/P100626.pdf
[26] https://www.mas.gov.sg/publications/monographs-or-information-paper/2024/artificial-intelligence-model-risk-management

About Dvara Research

Dvara Research is an independent, non-partisan, not-for-profit policy research institution based in India. Its mission is to ensure that every low-income household and every small enterprise has complete access to suitable financial services and social security through a range of channels that enable them to use these services securely and confidently.

Since 2008, Dvara Research has deeply analysed, and carefully written about, financial inclusion and social protection in India from policy, regulatory, and practitioner perspectives that are anchored to its mission. Its work has gained the admiration and respect of policymakers and regulators, and since its inception, Dvara Research has been a research-partner of choice for such key policy-making bodies as the Reserve Bank of India, Securities and Exchange Board of India, Pension Fund Regulatory and Development Authority etc.

Website: https://dvararesearch.com