FSB FINANCIAL STABILITY BOARD

Sound Practices for Responsible Adoption of Artificial Intelligence (AI): Consultation report

Response to Consultation

South African Reserve Bank-Prudential Authority

1. **Do you agree with the benefits and risks of AI adoption by financial institutions described in this report? Are there any substantive benefits or risks not covered?**

   Yes, we broadly agree with the benefits and risks described in the report. The report appropriately recognises that AI adoption can support efficiency gains, business model transformation, improved customer service, and enhanced risk management, including cyber defence. It also correctly notes that these benefits may introduce or amplify risks, and that effective risk management is necessary to support sustained value creation and limit financial stability risks. The report also usefully recognises that GenAI and agentic AI are accelerating adoption and may introduce additional concerns linked to complexity, explainability, autonomy and cyber risk.

   Minor additional point on systemic risk (not explicitly in FSB text) flagged for completeness.

2. **Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?**

   Yes, this aligns with international standards and the Twin Peaks approach and appropriately integrates AI governance into existing risk and governance structures. The distinction between organisation-wide governance and AI lifecycle management provides a clear foundation for responsible AI adoption. We suggest slight clarifications (i.e., documentation/validation expectations) to strengthen clear guidance further.

3. **Do the sound practices strike an appropriate balance between managing risks relating to all forms of AI, and addressing some of the risks relating to emerging and new complex forms of AI, such as GenAI and agentic AI?**

   Yes, the report strikes an appropriate balance overall. Its principles-based, technology-neutral structure allows the practices to apply to traditional AI, machine learning, GenAI, and agentic AI, while recognising that certain practices may be more relevant to specific technologies, including risks linked to agent autonomy. Recommends ongoing vigilance by FIs & supervisors as AI evolves.

4. **Are the sound practices sufficiently flexible to accommodate and address newer types of AI and responsible adoption over time?**

Yes, the sound practices are sufficiently flexible. The report explicitly states that the practices are not intended to create an international standard, impose a prescriptive approach, or dictate technology choices, and that they are not exhaustive and may be refined as AI technologies evolve. We encourage periodic review of the sound practices as GenAI, agentic AI, model assurance techniques, AI safety testing, and financial-sector uses mature.

5. **Do the case studies in this report sufficiently highlight how different types of financial institutions can benefit from responsible AI adoption? Are there additional case studies for inclusion in the report? If so, please provide such case studies, particularly for nonbanks.**

Yes, the case studies are useful and demonstrate practical benefits across different use cases. The report includes case studies on credit risk management and GenAI-supported SME credit underwriting, including early-warning credit risk alerts and GenAI-assisted credit proposal generation.

6. **Do the case studies in this report provide actionable insights for financial institutions in their responsible AI adoption?**

Yes, the case studies provide actionable insights. The PA finds them practical and helpful for financial sectors' implementation. It also gives practical examples of AI performance testing, including sensitivity testing, out-of-sample monitoring, bias testing, hallucination checks, stability testing, and sampled expert review.

7. **Are the definitions in the glossary clear and aligned with industry sound practices, including recent developments in AI?**

The PA generally supports the glossary and finds the definitions broadly aligned with emerging international practice. However, consideration could be given to expanding definitions relating to frontier AI, AI agents, AI assurance, model validation, concentration risk within the AI supply chain, and synthetic employees to ensure consistency with evolving supervisory and industry terminology.

8. **Are there any comments you would like to make on this report? Please fill in.**

Page 18 Section 2.1 "Strategic direction and oversight" Relevant text: "The board and senior management align AI adoption and governance with the financial institution's business model, risk appetite, and strategy." The PA recommends adding that boards should formally incorporate AI considerations into enterprise risk management (ERM) and risk appetite frameworks, ensuring that any adoption of AI is explicitly covered by risk appetite statements and business continuity plans. The guidance could also highlight the need for boards to consider emerging systemic risks (i.e., industry-wide dependencies on certain AI technologies) in their strategic oversight, aligning with their financial stability mandate.

Page 20 Section 2.2 "Governance and accountability" Relevant summary: "Financial institutions define clear roles and responsibilities and maintain an appropriate governance framework to enable responsible AI adoption." The PA suggests

emphasising the need for a multidisciplinary governance approach, i.e., involvement of risk management, IT, data science, compliance, and business lines in reviewing AI projects, to avoid siloed decision-making.

Page 21 Section 2.3 “AI risk management framework & documentation” Relevant topic: Incorporating AI-related risks into existing risk management. The PA suggests emphasising the need for robust AI documentation and model inventories, aligned with existing model risk management principles (such as record-keeping of model purpose, design, limitations, validation results, and any independent reviews). The PA suggests that AI inventories be recognised as a key governance and supervisory tool, providing firms with a consolidated view of AI models, systems, use cases, owners, dependencies, risk ratings, validation status, third-party relationships, and materiality classifications.

Page 27 Section 3.2 “Materiality and risk assessment” Relevant wording: “…assess the materiality and risk of AI use cases at the inception stage and thereafter.”

The PA suggests that the FSB consider common criteria or taxonomy for AI use-case materiality, which could help both industries and supervisors in identifying where the most robust controls are needed.

Page 29 Section 3.3 “Selection” Relevant idea: Selecting appropriate AI models/systems considering objectives, needs, and risk of use cases.The PA suggests referencing an organisation’s procurement and outsourcing policies, given that model selection frequently involves the use of third-party products.

The PA recommends additional guidance on change management and version control for AI systems, particularly where models continuously evolve, are retrained, or are updated by third-party providers.

Page 30 Section 3.4 “Data governance” Relevant wording: “…establish appropriate data governance to maintain data that is fit-for-purpose for training, testing, and using AI (i.e. accurate, complete, consistent, reliable, secure).” The PA recommends explicitly noting that data quality, privacy, and bias mitigation are critical not only for model performance but also for regulatory compliance and reputation. The PA suggests that financial institutions consider data residency, cross-border data transfers, and legal jurisdictional risks where AI models, infrastructure, or training environments are located outside the institution’s home jurisdiction.

Page 33 Section 3.5 “Explainability and transparency” Relevant wording: “Financial institutions understand differences in the explainability of various types of AI… If appropriate and feasible, adopt more explainable AI, or consider compensating controls… provide appropriate transparency tailored to different stakeholders.” The PA agrees and suggests providing examples of such controls (e.g., human review of certain outcomes, simplified explanations, or the use of proxy interpretable models) to provide greater clarity. The PA further recommends emphasising that transparency should extend to supervisors, not just customers or internal stakeholders; i.e., firms should be prepared to explain their AI-driven decisions and models to regulators, enabling effective supervisory review and model risk assessment. This is crucial for supervisors to assess a firm’s compliance with prudential and conduct requirements when decisions are increasingly automated.

Page 37 Section 3.6 “Performance management” Relevant topic: Ongoing performance evaluation and monitoring of AI use cases. The PA suggests clarifying that performance management processes should feed into model risk capital assessments or contingency plans; for instance, significant deterioration in the performance of material AI models should trigger predefined remediation actions, enhanced oversight, model redevelopment, contingency measures, or reversion to alternative approaches where appropriate. The PA also suggests a more explicit reference to independent validation and assurance arrangements for material AI models and systems. Independent validation remains key to effective model risk management and should be applied proportionately to AI use cases, particularly where AI is used in critical operations, prudential risk measurement, underwriting, credit decisions, fraud detection, or customer outcomes.

Page 41 Section 3.7 “Human oversight” Relevant wording: “Financial institutions implement appropriate and effective human oversight relevant to the materiality, risk, autonomy, complexity, and explainability of different AI use cases.” This section could explicitly reference the need to prevent “automation bias,” ensuring that human reviewers have the authority, expertise, and willingness to challenge AI outputs rather than rubber-stamping. Furthermore, the human oversight mechanisms should be commensurate with AI autonomy and complexity, i.e., “human-in-the-loop” for credit decisions or high-impact insurance claims, versus “human-on-the-loop” for low-risk, automated processes.

Page 44 Section 3.8 “Cyber and ICT risk management” Relevant wording: “Financial institutions manage AI-related cyber and ICT risks, including… incorporating AI cyber and ICT risk scenarios into tests and exercises, sharing relevant information, and using AI tools in cyber and ICT risk management.” The PA suggests including explicit mention of “AI-specific attack vectors” (i.e., model poisoning, data manipulation, adversarial inputs, prompt injection attacks) and recommending that institutions update their cyber risk assessments and controls to account for these unique threats. The PA also recommends encouraging participation in information-sharing forums. The PA also recommends expanding the guidance to include AI-specific incident management processes, including classification, escalation, root-cause analysis, remediation, and reporting of significant AI-related incidents, model failures, hallucinations, unintended autonomous actions, data leakage events, or customer harm incidents.

Page 48 Section 3.9 “Third-party risk management” Relevant wording: “Financial institutions appropriately manage risks from AI third-party use with a focus on performance, transparency, data quality, supply chain and concentration risks, and business continuity.”

The PA suggests emphasising that contractual arrangements with AI providers should include clear provisions for auditability, data ownership, and contingency plans, to ensure monitoring of vendor performance and has fallback options if the service is disrupted. The PA recommends further emphasis on concentration risk arising from reliance on a small number of hyperscale cloud providers and foundation model providers. Firms should assess concentration exposures across the AI supply chain and consider diversification, portability, substitutability, and exit arrangements where feasible. Additionally, the PA recommends encouraging firms to maintain visibility of the broader AI supply chain, including foundation models, cloud infrastructure providers, data providers, and embedded subcontractors, to better understand systemic and operational dependencies.

General Comments:

• Overall support and alignment: The PA welcomes the FSB's Sound Practices for Responsible AI Adoption and supports its principles-based, technology-neutral approach. The guidance aligns well with South Africa's Twin Peaks framework and appropriately integrates AI governance into existing risk and governance structures. The distinction between organisation-wide governance and AI lifecycle management provides a clear foundation for responsible AI adoption.

• Risk management and proportionality: The PA supports the recognition that AI presents both opportunities and risks, requiring strong governance and risk management. AI-related risks should be embedded within existing frameworks rather than managed separately. The proportional approach is particularly important, ensuring expectations align with the size, complexity, and risk profile of institutions and AI use cases.

• International consistency and coordination: The sound practices are broadly consistent with international standards and promote a harmonised approach to AI governance. The PA encourages continued coordination between prudential and conduct authorities to ensure consistent expectations across sectors and avoid regulatory gaps or overlaps.

• Responsible innovation and SupTech: The PA supports a balanced approach that promotes responsible innovation while managing risks. The guidance could further acknowledge supervisors' own use of AI (SupTech) and encourage authorities to apply similar governance principles to their AI tools, strengthening trust and credibility.

• Key strengths and areas for enhancement: The 12 sound practices provide a comprehensive framework covering the full AI lifecycle. The inclusion of cyber resilience and third-party risk is particularly welcomed. Additional emphasis on independent model validation and assurance mechanisms would strengthen the framework. Consideration of potential systemic risks arising from widespread AI adoption may also be a useful area for future work. Overall, the guidance provides a strong basis for enhancing AI governance across the financial sector.

• Operationalising Proportionality Remains a Key Gap: the largest gap in the FSB paper. The principle is well articulated but not sufficiently translated into practical implementation expectations

• Fairness, Customer Outcomes, and Redress: The paper must add more emphasis on this section, as it was noted to be underdeveloped

• Practical Supervisory Guidance: The paper should relook at the practical supervisory guidance, as it is currently primarily a principles-based governance framework. Recommend that the paper must go further by providing practical supervisory examples, assessment considerations, and monitoring approaches.

• The PA suggests additional guidance on governance controls for highly autonomous and frontier AI systems, including escalation thresholds, capability assessments, kill-switch arrangements, containment mechanisms, and restrictions on autonomous decision-making in critical functions.