FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Data Sense

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?**

   While I agree with the primary benefits and risks outlined in Section 1, the report underrepresents two critical risk vectors that stem from advanced and emerging AI deployments:

   1) Multi-Agent Interoperability & Networked Risk: The report treats AI adoption largely as isolated, single-institution implementations. It does not sufficiently address the systemic risks arising when AI agents interoperate autonomously, both internally across business functions and externally across financial institutions and third-party ecosystems. Autonomous agent-to-agent interactions introduce non-linear feedback loops, cascading goal misalignments, and rapid propagation of erroneous transactions or market signals that traditional risk models cannot detect or constrain.

   2) Asymmetric Cyber Attack Surfaces vs. Defensive Gains (Page 18): While the report notes that AI enhances defensive capabilities (e.g., fraud detection and identity verification), it should explicitly acknowledge the net security trade-off. Advanced AI introduces novel attack surfaces such as: prompt injection, agentic memory poisoning, model inversion, and automated vulnerability exploitation (which are noted), that compress the timeline from vulnerability discovery to execution. The net impact on security depends strictly on the strength of runtime controls, not the AI technology itself.

   3) Omission Bias & Explainability in Credit Decisioning (Pages 11–12): In credit scoring and underwriting (particularly using Gradient Boosting Machines, Neural Networks, or GenAI reference material), the focus must shift from algorithm selection to feature engineering transparency and omission bias. The report highlights cases where GenAI matched manual review for identifying fraud, but misses the risk of false negatives. Specifically, relevant creditworthiness information that was omitted or overlooked. Without robust proxy testing, explainability, and formal consumer appeal/redress mechanisms, high performance metrics can mask systematic unfairness and automation bias in human reviewers.

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

The sound practices are sufficiently broad, but their high-level nature creates ambiguity that may result in inconsistent implementation across institutions. To ensure clarity and operational utility, the following areas may require refinement:

1) Explicit Inclusion of Auditability & Execution Oversight (Sound Practice 2, Page 16): Sound Practice 2 emphasizes governance, roles, and accountability. However, accountability without immutable auditability and real-time execution oversight (e.g., deterministic API guardrails, transaction threshold caps, and automated kill switches) is ineffective in autonomous environments.

2) Unstructured Data Lineage & Access Controls (Sound Practice 7, Page 36): Sound Practice 7 covers general data governance, but lacks actionable specifics on managing unstructured data flows (e.g., RAG pipelines, vector databases, and LLM context extraction). Guidance should emphasize policy-aware data frameworks by embedding data sensitivity classifications and handling policies directly into metadata, while defining consistent cross-system access controls, monitoring, and audit requirements.

3) Certain text in Section 3.4 (e.g., describing data preparation as "the most resource-intensive stage") adds little supervisory value and should be replaced with explicit expectations around policy-aware data controls and data provenance validation. Footnote 56 provides much more value here.

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?**

No, the report remains disproportionately anchored in traditional, static Machine Learning (ML) models, treating GenAI and Agentic AI as secondary add-ons rather than fundamental shifts in operational risk.

1) Deterministic Gating vs. Post-Hoc Analysis (Pages 31–32, 47): Traditional ML relies on periodic validation and post-hoc log analysis. In an agentic environment where systems make sub-second autonomous decisions, post-hoc monitoring merely records damage after it occurs. The sound practices must explicitly require in-line, deterministic gating (approval checkpoints, dynamic IAM, and hard scope boundaries) for high-risk agentic actions.

2) Reframing Workflow Automation (Page 12): The report characterizes autonomous trade-related task sequencing as a novel application. In reality, workflow orchestration (NLP, signal generation, rule-based execution) has existed for years; what is novel is the speed, lowered engineering friction, and probabilistic nature of agentic orchestration. The risk guidance must focus on runtime execution boundaries rather than treating the underlying task as new.

3) Proposed Governance Amendment for Continuous Evaluation Across the Three Lines of Defense (Page 25): Evaluation should be elevated from a static, downstream second-line duty to a defining governance mechanism that operationalizes continuous feedback loops across all three lines of defense:

a) First Line: Continuous performance and materiality data to refine local controls in real time (or near real time depending on data availability and function).

b) Second Line: Independent evaluation metrics (data quality, explainability, model drift) feeding back directly into the development lifecycle as an ongoing challenge-and-remediation mechanism.

c) Third Line: Consolidated evaluation trails enabling independent verification of overall governance integrity.

This continuous evaluation loop serves as the operational engine for Sound Practice 3, ensuring use-case documentation, prompt versioning, and chain-of-thought logging actively inform real-time risk mitigation.

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

The sound practices are structurally flexible, but they risk being too open-ended to provide actionable guidance for rapidly evolving technologies.

1) Contradiction Between Periodic and Continuous Evaluation (Pages 31–32): The report repeatedly references reassessing materiality and risk at "regular intervals" or "periodic evaluation." For agentic and foundation model deployments, periodic reviews create dangerous risk blind spots. The framework must mandate continuous, event-driven evaluation paired with automated role-based access gates that react more instantaneously to model drift, prompt modifications, or unapproved tool calls.

2) As institutions move from passive assistance tools to autonomous execution agents, guidance that focuses on static lifecycle checkpoints will quickly become obsolete. Future flexibility depends on establishing clear supervisory expectations around dynamic IAM, agent identity management, and tool-access scoping.

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.**

No comment.

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

While the case studies demonstrate clear operational efficiency gains (e.g., turnaround time reductions and fraud detection metrics), they predominantly focus on successful implementation outcomes. To be truly actionable, future case studies should highlight governance failure modes, edge-case remediation, and how human oversight intervened when models/agents degraded or produced false negatives. Additionally, adding case studies on non-bank financial institutions (e.g., asset managers or fintechs deploying multi-agent workflows) would provide broader industry value.

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

   The glossary is generally clear, but definitions for Agentic AI and Data Lineage should be updated to reflect current technical realities:

   Agentic AI: Should explicitly highlight non-deterministic multi-step tool execution, dynamic goal setting, and autonomous inter-agent communication.

   Data Lineage: Should be broadened beyond structured data pipelines to address unstructured data extraction, RAG/vector retrieval context provenance, and policy-aware metadata guardrails.

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

   Financial institutions have governed traditional ML models under established Model Risk Management (MRM) standards (e.g., Federal Reserve SR 11-7 / OSFI E-23) for over a decade. It is somewhat surprising that this consultation report spends significant space reiterating traditional ML sound practices while offering comparatively limited supervisory depth on Generative and Agentic AI.

   The primary value at this point in time should be to get ahead of the shift toward autonomous, agentic, and multi-agent implementations before these systems move en masse from experimental pilot phases to core production infrastructure. I strongly encourage the FSB to rebalance the report by placing Agentic AI governance, continuous in-line evaluation loops, deterministic execution gates, and multi-agent interoperability risks at the center of the sound practices, rather than viewing them through the lens of legacy model governance.