FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

MRV Associates

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, but with omissions. The report captures the main efficiency gains — automation of AML/KYC, faster credit underwriting, fraud detection, customer service, and operational resilience — and the key risks including model errors, data bias, explainability gaps, and cyber threats. However, it underemphasizes two systemic dangers.

   The first is herding risk. When many institutions use the same foundation models, the same training datasets, and the same optimization techniques, their decisions may become correlated in dangerous ways. Simultaneous asset sales, similar lending contractions, and synchronized risk-off moves could amplify market stress in ways that no individual institution's risk management would detect. This is the most important financial stability risk that AI creates, and the report treats it as a footnote rather than a centerpiece.

   The second is workforce deskilling. As AI absorbs more decision-support functions, junior analysts may develop fewer independent skills, and the human "challenge function" that catches AI errors may erode quietly. This is especially concerning in credit underwriting, trading, and compliance, where human judgment has historically functioned as a corrective backstop.

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

   Comprehensive in scope, but frequently too vague to be actionable. The framework correctly identifies what institutions should govern — lifecycle stages, data quality, explainability, human oversight — but often stops at "have effective controls" without specifying minimum testing standards, validation frequencies, escalation thresholds, or required documentation. Practitioners implementing these practices against supervisor expectations will face significant interpretive uncertainty.

   GenAI also deserves its own dedicated section. Prompt management, retrieval-augmented generation (RAG), hallucination testing, and human review requirements for generative outputs are substantively different from traditional model validation. Folding them into generic lifecycle guidance undersells the challenge.

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?**

   Mostly yes. The FSB's technology-neutral architecture means it does not write rules only for today's systems, which is the right long-term approach. The explicit treatment of agentic AI risks — unauthorized autonomous actions, erroneous goal pursuit, data breaches, and disruption to connected systems — is commendable. That said, agentic AI still warrants a dedicated chapter covering human intervention requirements, kill-switch protocols, maximum autonomy thresholds, and escalation procedures. The risks of an AI agent that can initiate transactions, move funds, or modify system configurations are categorically different from those of a static machine-learning model, and the framework's general governance principles do not fully bridge that gap.

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

   Yes, structurally. The principles-based approach should accommodate new AI architectures without requiring wholesale revision. The risk is the mirror image of inflexibility: excessive vagueness produces regulatory uncertainty. Institutions may not know what constitutes compliance, and supervisors in different jurisdictions may draw vastly different conclusions from the same principles. The FSB should commit to publishing periodic implementation examples, supervisory FAQs, and emerging risk bulletins that add specificity without calcifying into prescriptive rules.

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.**

   They are useful, but skewed toward large, internationally active banks. The report would benefit from adding cases covering nonbank lenders using AI in credit underwriting, private credit firms applying AI for covenant monitoring and early-warning systems, insurers deploying AI in claims handling and catastrophe modeling, and fintechs operating AI-driven customer service with consumer protection controls. These sectors are growing rapidly and face distinct regulatory environments that the bank-centric cases do not illuminate.

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

   Partially. The existing cases — ML-based SME credit monitoring, agentic fraud detection, AI-powered relationship management — are concrete and credible. But practitioners need more than proof of concept. Each case study should specify the exact controls deployed, the metrics monitored, what failed during development or deployment, how failures were detected, and what was learned. Without those details, the cases risk becoming promotional material rather than implementation guides. A structured template — use case, key risks, controls, metrics, escalation procedures, lessons learned — would significantly improve their utility.

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

Yes, but several terms need refinement. "Agentic AI" should explicitly distinguish between workflow automation, decision-support agents, and fully autonomous agents, as these represent quite different risk profiles. "Human-in-the-loop" requires a definition of what constitutes meaningful review; an employee who clicks "approve" on every AI recommendation in under two seconds is not providing genuine oversight. "Explainability" should differentiate between technical, regulatory, and customer explainability, which have different requirements and audiences. And "AI model" should explicitly cover foundation models, fine-tuned models, RAG systems, and multi-agent architectures, whose risk characteristics differ substantially.

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

The Financial Stability Board (FSB) consultation report, Sound Practices for Responsible Adoption of Artificial Intelligence, outlining twelve principles to guide financial institutions in deploying AI responsibly, arrives at a pivotal moment. Banks, insurers, asset managers, and payment providers are racing to embed AI into everything from credit decisions to fraud detection, while regulators scramble to keep pace. The FSB's framework is a serious and balanced attempt to impose order on that chaos. Unfortunately, there are significant gaps that deserve scrutiny before the July 22 comment deadline.

What the Framework Gets Right

The FSB's twelve sound practices span two broad domains: organization-wide AI governance (practices 1–4) and AI lifecycle management (practices 5–12). The governance pillar rightly places ultimate responsibility on boards and senior management, requiring them to align AI adoption with the institution's risk appetite, establish clear accountability structures, and invest in the skills and culture needed to sustain responsible use. The lifecycle pillar then operationalizes governance through requirements on materiality assessment, model selection, data quality, explainability, performance monitoring, human oversight, cybersecurity, and third-party risk.

Several features deserve credit.

Technology neutrality. The FSB wisely avoids prescribing rules for specific AI architectures. By focusing on governance principles and outcomes rather than today's models, the framework should remain relevant as generative AI (GenAI) gives way to whatever comes next. This is harder to achieve than it sounds; many early AI regulations already look dated.

Recognition of agentic AI. Many regulators are still catching up with large language models. The FSB goes further, dedicating substantive attention to agentic AI — autonomous systems capable of planning, reasoning, and executing multi-step tasks without continuous human direction. The report correctly identifies that agentic AI introduces qualitatively different risks: goal misalignment, reward hacking, emergent behaviors from agent-to-agent interaction, and the near-impossibility of real-time human monitoring at scale. This is forward-looking work.

Third-party and cyber risk. The framework takes seriously the concentration risks created by heavy dependence on a small number of cloud providers, foundation model developers, and AI platform vendors. It also addresses a wide range of AI-specific cyber threats, from prompt injection and model poisoning to deep-fake-enabled social engineering — risks that traditional cybersecurity frameworks ignore.

Proportionality. The report acknowledges that a community bank deploying a chatbot faces different risks than a globally systemically important bank (G-SIB) running agentic trading algorithms. It encourages institutions to calibrate their governance intensity accordingly, an important concession to operational reality.

The Biggest Gap: Systemic Risk from Correlated AI Adoption

The most important enhancement the FSB could make is a dedicated section on systemic financial stability risks from correlated AI adoption. This would address three interrelated phenomena: model monoculture (if a handful of foundation models dominate banking, a single vulnerability or software update could create correlated failures across hundreds of institutions); vendor concentration (heavy reliance on common cloud and AI infrastructure creates single points of failure analogous to the concerns regulators already have with cloud providers); and synchronized decision-making (correlated credit tightening, asset sales, or risk assessments across banks, insurers, and asset managers could amplify procyclicality in ways that dwarf any individual institution's exposure).

This is the dimension where AI moves from an institution-level risk-management challenge to a genuine threat to financial stability — and it is the area where the current framework does the least work.

Concluding Thoughts

The FSB's sound practices earn a solid B+ as a foundation for responsible AI governance in financial services. The framework's technology neutrality, governance emphasis, lifecycle coverage, and explicit attention to GenAI and agentic AI are genuine achievements. Its shortcomings — insufficient specificity for implementation, limited treatment of systemic herding and monoculture risk, thin coverage of nonbank institutions, and case studies that stop short of operational detail — are significant but correctable. The consultation process is an opportunity to address them. Financial institutions, their supervisors, and the broader financial system will be better served if the FSB uses the feedback period to sharpen, rather than merely ratify, what it has produced.