FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Ramasamy S (independent response)

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?**

   I broadly agree with the benefits and risks described.

   On risks, I recommend an addition:

   a) Compositional Risk at Inter-Agent Hand-Offs

   While the report notes that "multiple AI agents may interact in unanticipated ways that create unforeseen behaviors or risks" (Section 1.3), this observation warrants more detailed treatment. In practice, financial institutions are deploying multi-agent systems where AI components handle stages of a workflow that may be sequential, branching, or dynamically determined at runtime. These systems can self-correct, retry, and route tasks autonomously based on intermediate outputs. Each component may individually pass validation, yet the composed system can exhibit emergent failure modes, particularly when one component's confident-but-wrong output cascades through downstream components that lack the context to challenge it.

   The unit of risk in agentic systems is not the model but the hand-off between agents. This is where information loss, context degradation, and accountability fragmentation occur. While model risk frameworks such as SR 11-7 are principle-based and broadly applicable, their practical implementation across financial institutions (model inventories, per-model validation, model-level ownership) tends to treat models as discrete units. Emergent risks at the interaction layer between chained AI components do not yet have a natural home in current model risk management practice.

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

   The practices are comprehensive for traditional ML and GenAI deployments. For Agentic AI, I recommend the following enhancements:

   a) Practice 8 (Explainability and transparency): Distinguish Generated Rationale from Verified Explanation

The report correctly notes that even a system's own developers and deployers may not understand how it maps inputs to outputs (Section 3.5). For multi-agent systems, the explainability challenge is compounded: no single agent's reasoning trace explains the pipeline's end-to-end behavior.

Each component's trace is local to that component. No single agent captures why it received the inputs it did, what upstream agents chose to omit, or how the orchestrator determined the routing. The explainability gap in multi-agent systems exists between agents, not within them. As Sudjianto (2026) demonstrates, prompt-based controls (including instructions to "explain your reasoning") converge to a ceiling set by the system's weakest unbuilt control, not by the sophistication of the prompt.[1]

The report notes that institutions are developing explainability approaches that can "trace and evaluate the full reasoning path" (footnote 61). I recommend the FSB go further by distinguishing between:

- Narrative explainability: an individual agent's account of its own reasoning and actions (accurate for that agent, but incomplete as an explanation of pipeline behavior)

- Forensic explainability: end-to-end reconstructibility of the pipeline's behavior across all agents, including what was passed, omitted, or transformed at each hand-off

Financial institutions should demonstrate the latter for material AI use cases.

Additionally, the report's explainability taxonomy addresses global explainability (overall model behavior) and local explainability (individual predictions) at the single-model level. It does not address pipeline-level explainability: understanding at which inter-agent hand-off an error was introduced or amplified in a multi-agent system. A local explanation for Agent A and a local explanation for Agent B do not compose into a valid explanation for the pipeline A→B. The emergent behavior lives in the interaction: the information loss, context degradation, and confidence-without-warrant that occurs at the seam between components. For multi-agent systems, no existing XAI method produces a faithful explanation of the composed pipeline's end-to-end behavior.

b) Practice 10 (Human Oversight): Define Minimum Conditions for Meaningful Oversight

The report discusses automation bias and rubber-stamping. I recommend minimum conditions for human oversight to be considered "meaningful":

1. The reviewer must have access to 'what the system did not consider' (negative evidence), not merely what it presented

2. The reviewer must be able to independently 'replay the decision path' from primary evidence

3. The reviewer's intervention must be 'operationally feasible' within the time and volume constraints of the process

When these conditions are met, human oversight becomes genuinely effective rather than performative. Absent them, there is a risk that "human-in-the-loop" functions as a compliance artifact rather than a meaningful control.

c) Practice 5 (Materiality and risk assessment): Account for Compositional Materiality

A pipeline of individually low-risk AI components can produce a high-risk composite action. Materiality assessment should consider the end-to-end action authority of a system, not merely the risk classification of each component in isolation.

d) New Concept: Action Authority Envelopes

For Agentic AI systems, I propose the FSB introduce the concept of "Action Authority Envelopes": technically enforced boundaries defining what an AI system is permitted to do autonomously. Actions within the envelope proceed without human intervention; actions that would exceed it require human approval before execution. This operationalizes Practice 10 for agentic systems by replacing the binary question ("is there human oversight?") with a graduated one ("for which actions is human oversight required?"). Critically, Action Authority Envelopes must be inherited by any sub-agents spawned at runtime; the system's permission boundary cannot be expanded by creating new agents.

e) Governance Requirements for Multi-Agent Systems

Complementing Action Authority Envelopes (which constrain what a system CAN do), I recommend five governance requirements for verifying what a multi-agent system DID do. Each addresses a gap the current practices do not fully cover for Agentic AI:

- Auditability: the ability to reconstruct system behavior from evidence, distinct from what the system reports it did

- Traceability: the ability to follow a decision through the full pipeline, identifying which component contributed what

- Task Attribution: clear assignment of accountability to specific agents or hand-offs within the pipeline, enabling Practice 3's accountability requirements at the component level

- Intervention Points: pre-defined locations in the workflow where human review is both operationally feasible and genuinely effective, operationalizing Practice 10 for dynamic workflows

- Normative Compliance: verification that the system's end-to-end behavior satisfies regulatory requirements, not merely that each component individually passes a compliance check

[1] Sudjianto, A., "Unified Prompt and Pray Framework," SSRN No. 6932339, June 2026. Demonstrates formally that prompt-strengthening converges to a ceiling set by the weakest unbuilt control.

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?**

The report maintains appropriate technology neutrality. I recommend it give greater attention to "Agentic AI", that is, systems that autonomously select tools, execute multi-step workflows, and take write actions without requiring human approval for each intermediate step.

This distinction is practical. Agentic AI is already deployed in production across financial services:

Credit & Underwriting:

- Rocket Companies operates "agentic pre-approvals" covering 10% of all pre-approvals with 33% higher conversion.[2]

Trading:

- Robinhood launched Agentic Trading (May 2026), allowing customers to connect third-party AI agents to a dedicated brokerage account and authorize autonomous trade execution via the Model Context Protocol.[3]

- Coinbase launched agent-controlled crypto trading (June 2026) and the first SEC-registered AI investment adviser issuing autonomous buy/sell recommendations.[4] [5]

Regulatory Compliance:

- Agentic compliance platforms (e.g., Sedric) deploy AI agents that autonomously monitor communications, apply policies, identify risk, and generate compliance evidence, operating as first-line-of-defence rather than analyst support tools.[6]

Implication: As financial products increasingly embed AI as the actor rather than the adviser, the practices would benefit from explicitly addressing this category.

I recommend the FSB distinguish between:

- AI that informs human decisions (summarization, analysis, recommendations): the human retains action authority, and existing model risk frameworks largely apply with adjustments for hallucination, completeness, anchoring risks.

- AI that acts autonomously (executing transactions, closing cases, modifying records, routing workflows): the system holds action authority, and the control paradigm shifts from "validate the output" to "constrain the system's operational scope."

[2] Rocket Companies Q1 2026 earnings disclosure, as reported by Zacks, "Rocket Companies' Business Model Explained: AI, Servicing, and Scale," May 14, 2026.

[3] CNBC, "Your AI agent can now trade for you on Robinhood," May 27, 2026.

[4] Coinbase Global, Inc., "Your AI Agent Can Now Trade and Pay with Coinbase," June 11, 2026.

[5] TechTimes, "Coinbase AI Trading Agent Is Now SEC-Registered," June 17, 2026.

[6] Sedric AI, "Sedric Wins Banking Tech Award as Agentic Compliance Reshapes Financial Services," June 2026.

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

The principles-based approach ensures longevity. I recommend the FSB add a forward-looking section on multi-agent systems (Agentic AI), architectures where multiple AI agents collaborate, negotiate, or supervise each other. These introduce governance challenges the practices do not yet address:

- Which agent is accountable when a multi-agent consensus produces a harmful outcome?

- How should institutions validate "AI-supervises-AI" arrangements (e.g., a reviewer agent checking a generator agent)?

- What constitutes "independence" when the supervisor and the supervised share training distributions or are produced by the same vendor?

The report's discussion of "AI-in-the-loop" oversight (Section 3.7), where AI augments human monitoring, opens this question but does not resolve it for arrangements where AI is both the actor and the overseer. These questions will become urgent as orchestration frameworks mature. The FSB has an opportunity to establish principles now, particularly that "accountability must dynamically follow risk concentration" across functional and organizational boundaries, rather than being fixed to a single designated owner regardless of where risk materializes in the pipeline.

The FSB's existing discussion of third-party concentration risks (Section 1.3) would also benefit from extension to shared agent frameworks, where common runtime logic can introduce correlated failure modes independent of the underlying model.

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.**

I recommend additional case studies in three areas:

a) Agentic underwriting at scale: Rocket Companies' deployment of autonomous AI agents for mortgage pre-approval, covering 10% of all pre-approvals, with agentic pre-approvals converting 33% higher than traditional channels.[2]

b) Agentic trading at consumer scale: Robinhood's Agentic Trading platform, illustrating how product design can move beyond the human-oversight assumptions embedded in existing regulatory frameworks.

c) Agentic claims handling in insurance: The report's existing insurer case study (Section 1.2) illustrates GenAI-driven efficiency gains in underwriting quick-quotes. A complementary case study could address agentic claims operations: Suncorp deployed five AI agents for claims sub-processes in June 2026, achieving 99% accuracy on coverage decisions;[7] Zurich scaled Agentic AI to five countries in 90 days, reducing manual risk processing time by 80%.[8] These represent nonbank examples of Agentic AI in production with quantified results.

[7] iTnews, "Suncorp to have AI agents in insurance claims process as soon as this month," June 21, 2026.

[8] FinancialIT, "Zurich Scales Agentic AI to Five Countries in 90 Days, Cutting Manual Risk Processing Time by 80%," 2026.

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

The existing case studies would benefit from the inclusion of structured lessons learned: near-misses, failed deployments, and controls that required strengthening. Financial institutions learn more from structured failure analysis than from success narratives.

To support this, the FSB could consider establishing an anonymized AI incident repository, analogous to aviation's confidential near-miss reporting system (NASA ASRS). Such a repository would generate the practical case study material the report currently lacks while enabling cross-institutional learning without reputational risk to individual firms.

Additionally, the operational resilience use case (Section 1.2) lacks a case study entirely. I suggest including examples of AI-driven resilience engineering (e.g., automated anomaly detection and recovery systems), while noting the recursive dependency risk: when AI is used for operational resilience, the AI system itself becomes a single point of failure requiring its own resilience framework.

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

I recommend the following additions:

- Compositional AI Risk: The emergent risk introduced when multiple AI components are chained in a pipeline, distinct from the sum of individual component risks.

- Action Authority Envelope: The technically enforced boundary on what an AI system is permitted to observe, decide, and modify within an operational context.

- Inter-agent hand-off: The point at which one AI component passes context, decisions, or control to another, identified as the primary locus of risk concentration in multi-agent systems.

- Agent Drift: The gradual, often imperceptible divergence in an Agentic AI system's behavior from its intended operational parameters, occurring without explicit model modification. Unlike model drift (performance degradation due to data distribution changes), agent drift

arises from accumulated changes in routing preferences, inter-agent coordination patterns, and strategy selection across extended operational periods, requiring longitudinal behavioral monitoring distinct from per-episode evaluation.[9]

[9] Chen et al., "Agent Drift in Multi-Agent Systems," arXiv:2601.04170, January 2026.

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

This response focuses on the governance implications of Agentic AI in financial services. I recommend the FSB: (1) give more detailed treatment to compositional risk at inter-agent hand-offs; (2) distinguish between narrative explainability (an individual agent's local trace) and forensic explainability (end-to-end reconstructibility across a pipeline); (3) introduce the concept of Action Authority Envelopes as a mechanism for operationalizing Practice 10 for autonomous systems; and (4) consider adding glossary terms for Compositional AI Risk, Action Authority Envelope, Inter-agent Hand-off, and Agent Drift.

About the respondent: Ramasamy Seranthaiya, Senior Solutions Architect at AWS, specializing in AI architecture for financial services. Presented on agentic AI governance at the FinAI Banking Summit and Wealth Management EDGE. Responding in personal capacity.