FSB Public Consultation on Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

Survey response 1

Respondent's Information

Name of jurisdiction:
United States of America

Name of jurisdiction: [Other]

Please provide your information:
Given name and last name: - Michael Bidun Email address: - mike@bidun.com Name of organisation: - Bidun Group LLC

Do you agree with your responses being made public on the FSB website?
Yes

Questions

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?
This is a supplemental response to the comments I submitted on 15 July 2026. Those comments addressed where controls for agentic AI should attach: at the orchestration boundary, with the delegation tree as the governed unit. This supplement addresses a second question that designing an implementation has since sharpened: how much control should attach, and the property with which it should scale. It also extends the delegation-tree analysis to a case my earlier comments did not reach: delegation between persistent, separately certified orchestrators.

2. Are the sound practices sufficiently comprehensive and clear to enable financial institutions' responsible AI adoption?
No further comment

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?

Supervisory expectations for agentic AI should scale with the delegation tree's maximum entitlement ceiling, not with the binary fact of agentic architecture. The ceiling is the declared and technically enforced boundary of what a tree can reach and effect: the data classes it touches, whether it writes back, whether it acts externally, which tools it holds, and what it may request of other agents. Two trees with identical architecture and radically different ceilings, a news summarizer and a payment executor, present radically different risk, and a control regime that treats them identically will be evaded by one and inadequate for the other. A low ceiling does not mean zero risk: erratic or correlated behaviour at low entitlement remains a matter for the universal monitoring floor, and for the compositional risks addressed under Question 8; what scales with the ceiling is the depth of engineered control.

The scaling has a natural structure, because agentic controls divide into two kinds. The low-cost controls are exactly the ones that should be universal: identity anchored at the orchestration boundary, delegation logging, restrictive entitlement inheritance, and aggregate limits are deterministic, platform-level mechanisms whose marginal cost per agent approaches zero. The costly controls, engineered containment demonstration, red-teaming including agent-to-agent testing, and independent authorisation, should deepen as the ceiling rises toward external actions, customer funds, and customer data. This is the distinction my earlier comments drew between verification and judgment, applied to the cost side: the floor is verification and can be universal; depth is where judgment and expense concentrate, and it should follow the ceiling.

One further distinction protects both supervisability and adoption: registration should be universal and nearly free; review depth is what scales. A complete inventory is what makes unauthorized-use detection possible at all, since a registry is the reference against which shadow deployments are identified. But universality need not mean ceremony: agent classes of uniform risk, defined by purpose, platform, and ceiling, can be certified once, with the class boundary technically enforced by the platform, instances enrolled automatically from platform inventory, and any capability beyond the boundary treated as a material change requiring individual review. In building a reference governance framework to implement my earlier recommendations, the initial design routed all agentic execution to the highest review tier, and it proved unworkable on inspection, before any deployment: a scheduling assistant faced the same committee process as a trading agent, guaranteeing either a flooded oversight body or unregistered agents. Proportionality here is not a concession to convenience; it is what keeps the high-rigour tier credible and the inventory complete.

Proportionality must also extend to where agents begin. Nearly every agent starts as a proof of concept, and a proof of concept should route by its sandbox ceiling (non-production data; no write-back beyond the sandbox; no external actions) and expire unless renewed or promoted, so that it registers cheaply rather than hiding; promotion toward production data or effects is a material change that re-routes at the new ceiling. A regime that prices early-stage agents by their label rather than their ceiling guarantees unregistered experimentation.

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

The ceiling-scaled approach is also what makes the sound practices durable. Architecture will keep changing; what a system can reach and effect is a stable, enumerable property that survives each new pattern. Expectations keyed to the ceiling extend automatically to architectures the report cannot yet anticipate, whereas expectations keyed to any particular structure will require revision as populations and patterns shift.

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 further comment

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

No further comment

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

No further comment

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

My earlier comments addressed delegation downward, from an orchestrator to ephemeral sub-agents. Delegation also occurs sideways, between persistent, separately certified orchestrators, and there it takes the classical form of the confused-deputy problem: an agent of low entitlement achieving, through a peer of higher entitlement, what its own certification forbids. A related path is escalation by grant, where a credential-issuing agent confers entitlements an agent did not hold and was never certified to hold. I recommend the final report state the closing rule: calls between agents carry a record of the originating principal; an effect is permitted only where every party in that chain is within its certified boundary for the effect; effects count against the aggregate limits of each party in the chain; and no agent may acquire, from any source including credential issuance by another agent, entitlements exceeding its certified boundary. These requirements are deterministic and platform-enforceable, and they belong to the same inexpensive floor described under Question 3.

One further channel deserves the same treatment. Even an agent certified for advisory output only can act through its human reader, by rendering a hostile link drawn from the content it ingests: an indirect escalation through the human interface, and a system-output problem, not a user-training one. The remedy is again deterministic and belongs to the same floor: links rendered from inbound external content should display as full literal URLs, pass through the institution's safe-link infrastructure, and be restricted to the source domains the agent is certified to read. This closes the last effect channel of read-only agents without subjecting them to review depth their ceiling does not warrant.

A limit of the chain rule should also be stated plainly: it makes certification compositional for entitlements, not for behaviour. Even where every effect in a chain is within every party's certified boundary, the combination can carry three distinct risks no single certification examined: (i) aggregation, where agents individually certified for separate data classes jointly assemble what none was certified to hold (the mosaic effect familiar from privacy practice, in which records that are individually not personal data become identifying in combination); (ii) feedback, where agents acting on one another's outputs amplify beyond what either review anticipated; and (iii) correlation, where many separately certified agents act on the same signal at once. The whole can exceed the sum of its certified parts. The remedy begins with data these expectations already produce: the delegation records and on-behalf-of chains that traceability mandates are precisely the material from which an institution can derive its inter-agent interaction graph. I recommend the final report expect institutions to monitor that graph for aggregation, feedback, and correlation patterns, and to treat recurring compositions of separately certified agents as units of monitoring and, where material, of review.

One corollary deserves note: the delegation records, on-behalf-of chains, and reasoning traces that traceability expectations produce are themselves sensitive records, and should be classified, access-controlled, and excluded from external exposure; the audit trail must not become the leak.

This is the within-institution form of the cross-institutional question I raised under Question 8 of my earlier response; resolving the vocabulary now would lay the ground for the harder cross-institutional case.

Thank you again for a consultation that has been unusually practical in its treatment of agentic AI. I would be glad to discuss any of the above.