FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Dorvelon

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?**

   We broadly agree. One additional risk should be recognized: documentary assurance lag. Documentary assurance lag occurs when the formal governance apparatus operates through policies, inventories, approvals, model reviews, committee decisions, and periodic audit, while the AI system's risk materializes through autonomous live execution. This is particularly relevant for agentic AI. A financial institution may have complete documentation and still lack the ability to evidence what an autonomous system did, which tools it invoked, whether it exceeded authority, whether sensitive data was exposed, whether downstream systems were affected, and whether the workflow can be safely remediated. We recommend adding documentary assurance lag as an assurance-process risk for high-materiality agentic AI.

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

   Partially. The sound practices are comprehensive at the principle level, but should be clearer on the assurance threshold for high-materiality agentic AI. For such systems, documentation should not be treated as sufficient evidence of control. Financial institutions should be expected to evidence the effectiveness of the live control environment. Suggested addition: "For high-materiality agentic AI use cases, financial institutions should supplement documentation, policies, and lifecycle review with runtime control evidence sufficient to demonstrate traceability, accountable attribution, effective human oversight, escalation, interruption, and recovery or remediation readiness, as appropriate to materiality and technical feasibility." We also recommend that the FSB recognize independent architectural validation where internal assurance capability or vendor-provided transparency is insufficient.

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?**

   Not fully. The final report should more clearly distinguish AI that produces outputs from AI that can take actions. Traditional AI and many machine-learning systems can often be

governed as model artifacts. GenAI intro- duces content, explainability, and information-integrity risks. Agentic AI introduces action risk. For high-materiality agentic AI, the assurance question is not only whether the model is accurate, explain- able, or monitored. The assurance question is whether the institution can explicitly govern and interrupt autonomous interaction with live systems. Suggested addition: “Where AI systems are authorized to take autonomous or semi-autonomous actions in live environments, financial institutions should apply heightened assurance expectations proportionate to the system’s autonomy, tool access, operational criticality, data sensitivity, and potential customer, market, prudential, or operational impact.”

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

   Yes, with one qualification. Flexibility should be bounded by minimum assurance outcomes for high-materiality AI. Otherwise, flexibility may allow governance form to substitute for control substance. The final report should remain technology-neutral, but clarify that high-materiality AI governance should evi- dence, at minimum, traceability, accountable ownership, effective human oversight, escalation, interruption, recovery or remediation readiness, and independent challenge by qualified architectural validation functions where appropriate.

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.**

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

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

   We recommend adding or clarifying the following definitions: 1. Documentary Assurance Lag The risk that formal governance processes rely primarily on policies, inventories, approvals, model docu- mentation, attestations, and periodic review while the AI system’s material risk occurs through autonomous live execution. 2. Runtime Assurance Assurance based on cryptographic or system-level evidence that a live AI system operates within hard algorithmic boundaries, and that autonomous actions can be explicitly verified, constrained, interrupted, and remediated at machine-speed, in proportion to risk. 3. Evidence Lineage The ability to trace material AI outputs or actions to relevant prompts, data sources, model or system versions, tool calls, approvals, accountable owners, and downstream effects. 4. Independent Architectural Validation Validation performed by parties independent of the AI system vendor and the institution’s internal build and operating functions, with demonstrable competence in non-deterministic failure modes, operational resilience, and the design of deterministic kinetic control planes, rather than traditional documentary audit alone.

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

We support the FSB's proportionate and non-prescriptive approach. Our comments focus on one narrow point: high-materiality agentic AI requires runtime assurance, not documentary assurance alone. Existing model risk management, risk management, compliance, and internal audit disciplines remain foundational. However, where AI systems can autonomously plan, invoke tools, interact with live infrastructure, and take intermediate actions, assurance based primarily on policies, inventories, approvals, model documentation, periodic validation, and retrospective audit review is structurally insufficient. The final report should recognize that high-materiality agentic AI requires evidence that the live control environment can trace, constrain, challenge, interrupt, escalate, and support recovery or remediation of autonomous actions, as appropriate to materiality and technical feasibility. We recommend that the FSB add three concepts to the final report: 1. Runtime Assurance Assurance based on cryptographic or system-level evidence that a live AI system operates within hard algorithmic boundaries, and that autonomous actions can be explicitly verified, constrained, interrupted, and remediated at machine-speed, in proportion to risk. 2. Evidence-Analytical Internal Audit Internal audit must remain independent in judgment, but for high-materiality agentic AI, independence cannot mean operational isolation. To effectively challenge management's AI governance architecture, the internal audit function must possess the evidence-analytical capabilities required to interpret and evaluate runtime data, system-change records, incident feedback, and recovery testing results. 3. Independent Architectural Validation For high-materiality agentic AI, where internal capability or vendor transparency is insufficient, financial institutions should utilize parties independent of the AI system vendor to validate the runtime control environment. These entities must possess demonstrable competence in non-deterministic failure modes, operational resilience, and the design of deterministic kinetic control planes, rather than traditional documentary audit alone.