FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Rubric Protocol (Echelon Intelligence Group LLC)

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?**

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

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?**

   Our answer: largely yes, with one structural gap that agentic AI sharpens.

   The report addresses the production of oversight records thoroughly — documentation (Sound Practice 3), data provenance and lineage (Sound Practice 7), and traceable logging of agentic decisions (Sound Practice 8). Our comment concerns one property those provisions stop short of: whether such records are independently verifiable — that is, whether their authenticity and integrity can be confirmed by a party other than the institution that produced them. This distinction is latent for traditional AI but becomes load-bearing under the agentic-AI oversight approach the report itself anticipates.

   THE OBSERVATION: PRODUCED AND TRACEABLE IS NOT THE SAME AS INDEPENDENTLY VERIFIABLE

   The sound practices are commendably thorough on record production. Sound Practice 3 sets a documentation mandate; Sound Practice 7 addresses data provenance and lineage explicitly; Sound Practice 8 calls for agentic systems' intermediate decisions and steps to be logged so that humans can evaluate actions after the fact.

   All of these provisions, however, concern records that the operating institution produces, holds, and controls. They establish that a record should exist and be traceable. They do not address a separate property: whether the record's authenticity and integrity can be verified by someone other than its producer.

   Within normal operations, this distinction rarely matters. But oversight records exist for the moments that are not normal — supervisory review, incident reconstruction, dispute, or the

apportioning of responsibility after an adverse outcome. In precisely those settings, the institution holding the record is an interested party. A record that its own keeper could alter, silently and without detection, functions as an assertion rather than as evidence. Traceability within the operator's own systems does not cure this, because the trace itself is part of the same operator-held record.

WHY THE REPORT'S OWN TREATMENT OF AGENTIC AI SHARPENS THE GAP

The report recognises, candidly and in our view correctly, the impracticality of real-time human monitoring of agent decisions as their use scales, and notes that effective monitoring and detection may require augmentation with another AI agent or other forms of AI.

We agree this is the necessary direction. But it relocates the trust question rather than resolving it. Where human oversight is partly supplemented by an automated monitoring layer, the institution's ability to demonstrate responsible oversight comes to rest on the integrity of that layer's records. If those records are verifiable only within the operator's own systems, an external reviewer — a supervisor, an auditor, a counterparty — is asked to accept the monitor's account on the operator's word. Oversight of the agent has been strengthened; verifiability of the oversight has not.

We note the report already gestures toward this theme from the supervisory side: Project Noor, cited in the report, aims to give supervisors independent, practical tools to evaluate bank AI models while preserving privacy. The same principle — independent verification by a party who is not the operator, without requiring full disclosure — applies with equal force to the records of AI oversight itself. Making that principle explicit within the sound practices would align the institution-facing guidance with the direction supervisory tooling is already taking.

SUGGESTED ADDITIONS

1. Within Sound Practice 10, add a consideration that where automated systems supplement or partially replace human oversight, institutions should be able to demonstrate the integrity and authenticity of the monitoring layer's records to relevant third parties, including supervisors, independently of the operator's own systems.

2. Within Sound Practice 8 or 9, distinguish two properties of an oversight record: that it is produced and traceable (captured at the time, with lineage), and that it is independently verifiable (its authenticity and integrity can be confirmed by a party other than the producer). The practices address the former well; the latter deserves explicit mention, proportionate to the materiality and risk of the use case.

3. Note that independent verifiability and disclosure are separable, consistent with the privacy-preserving approach of Project Noor. A record can be made provably authentic and unaltered without exposing its full contents to every party. Strengthening verifiability therefore need not conflict with the report's attention to proportionality or the protection of sensitive information.

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

The distinction raised in our response to Q3 — between records that are produced and traceable versus records that are independently verifiable — is itself a flexibility mechanism. It is technology-neutral and applies uniformly to current and future AI forms, because it concerns the evidentiary property of oversight records rather than the characteristics of any particular model class.

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?**

   The glossary would benefit from distinguishing “traceability” and “auditability” (properties of records within the operator’s systems) from “independent verifiability” (the capacity of a party other than the producer to confirm a record’s authenticity and integrity).

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

   Disclosure: Rubric Protocol builds attestation infrastructure for AI systems — tooling that produces independently verifiable records of what AI agents did. We therefore approach this consultation from the perspective of the verifiability of AI oversight records. We have aimed to make our observations useful on their own terms, and we take no position on any particular technology, including our own.

   The substance of our response is in our answers to Questions 3, 4, and 7. We appreciate the FSB's work and the accelerated timeframe on which it was produced, and would be glad to discuss any of the above with FSB staff.

   Scott Sims

   Founder, Rubric Protocol (Echelon Intelligence Group LLC)