FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

ZWING Intelligence

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 agree with the benefits and risks described. One risk is under-covered. The report treats the evaluation of AI outputs almost entirely as a human activity: review, spot-checking, sampled monitoring. Annex 3 itself notes the difficulty of evaluating LLM output, since "there is not a 'ground truth' against which to compare the output". That is true for language tasks. It is not true for a growing class of financial facts. In digital-asset markets, where software agents already execute transactions without human intermediation, much of the ground truth is machine-readable. Collateral structures, dependency graphs, and administrative control paths are derivable from public ledger state. Other facts, such as custody agreements and the identity of key holders, sit off the ledger and require attestation. The boundary between the two classes is itself knowable and should be stated per fact; our published framework handles this by separating assessment confidence from the risk severity it reports. Where ground truth is machine-readable, an agent's claims can be checked against it deterministically before the agent acts. The under-covered risk is the absence of this layer: financial institutions deploying agents whose factual claims could be verified mechanically but are not, so that oversight relies on sampling and human review even where exhaustive checking is available. Detection finds errors. Only verification against ground truth can establish their absence, and only for the claims checked.

   The same machine-readability serves system-level monitoring. The FSB's report "Monitoring Adoption of Artificial Intelligence and Related Vulnerabilities in the Financial Sector" (October 2025) describes the data gaps and the difficulty authorities face in assessing critical dependencies and concentration. In digital-asset markets that class of evidence is producible today. Structural dependencies can be mapped at scale, our published framework demonstrates this across more than 8,000 protocols, so concentration can be measured rather than surveyed.

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

   One distinction would make Sound Practices 9 and 10 materially clearer. AI outputs divide into two classes with different verification properties. Stochastic outputs, such as free text from an LLM, cannot be re-derived and therefore require corroboration and human review,

as the report says. Deterministic outputs, produced by defined checks over defined inputs, can be re-executed by an independent party against the same recorded inputs; the second run must reproduce the result or expose the discrepancy. The practices currently prescribe the same oversight vocabulary for both classes. Distinguishing them would let institutions allocate scarce human oversight to the work where it cannot be substituted, judgment over stochastic output, while machine re-execution covers what machines can check. This is the proportionality principle the report already endorses, applied to verification itself.

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 agentic-AI additions are the strongest part of the report, and we support them: the boundaries and checkpoints for agent actions, the controls on agents executing transactions, the expectation of "maintaining audit trails of agent transactions", and the expectation of monitoring and documenting agent processes, including "the reasoning at each autonomous decision point". We propose two additions.

   First, pre-action checks should name what the check is made against. The report specifies when a human must approve. It does not specify what the agent's view of the world should be validated against before action. For material actions in machine-readable domains, a sound practice would be a pre-action check of the agent's factual premises against independently observable state, performed by a system independent of the agent itself. Human approval gates scale poorly with agent decision volume; the report's own warning about rubber-stamping acknowledges this. A deterministic pre-action check does not replace the human role; it gives the human-in-command something checked to command over. A pre-action verification layer earns no exemption from this discipline. Sound practice should require that its check definitions be disclosed to those relying on it and that its checks be re-executable by a third party against the same state, so that the layer's reliability can be inspected rather than taken on trust.

   Second, audit trails should record what was known at the time of action, not only what the agent did. A trail of actions establishes sequence. A record that preserves the information the agent relied on, sourced and dated with assumptions stated, establishes defensibility; it allows a reviewer, an auditor, or a supervisor to assess afterwards whether the action was reasonable on the information then available. We suggest the practice language distinguish activity logs from evidence records, and encourage the latter for material autonomous decisions. Records of this kind also make the report's oversight-pattern analysis workable. Rubber-stamping is only detectable if the record shows what the approver could have checked.

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

   The practices will age best where they are stated as properties of outputs and records rather than as properties of named technologies. Whether an output can be re-executed, whether a record carries its sources and dates, and whether a check ran against independently observable state are properties that remain stable across model generations. Technology-

specific guidance dates quickly. We suggest stating the practices in terms of these properties; the report's own materiality-based proportionality principle is the 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.**

   We offer the following case study from our own deployment, for use at the FSB's discretion. A technology provider operates a live verification service for digital-asset exposures, designed to be called before agent-initiated actions. A calling system submits the exposure in question. The service returns a structural report: what the exposure depends on, which dependencies would move first under stress (an ordering derived from the dependency structure, not a statistical forecast), and which facts were checked deterministically against public ledger state as distinct from facts sourced from off-ledger information. Every input carries its source and date. The calling system applies its own policy to the report; the verification layer does not approve or block actions and holds no execution authority, which preserves a clean separation between verification and decision. The pattern generalises: verification as an independent service in front of autonomous execution, with evidence retained per decision rather than per reporting period.

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

   No comment beyond the case study offered under question 5.

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

   Two additions would help. The glossary defines explainability as a continuum but has no term for the stronger property that determines the cost of oversight: whether an output can be checked. We suggest defining "deterministic re-execution" (re-running a defined check that depends only on its declared inputs, so that an independent run must reproduce the result or expose a discrepancy) and "machine-verifiable information" (information whose accuracy can be established by deterministic re-execution against independently observable state). These terms would give Sound Practices 9 and 10 a vocabulary for allocating human oversight to the work machines cannot substitute for. We would also suggest defining "audit trail"; the report uses the term for agent transactions without a definition, and the distinction between an activity log and an evidence record (what was known at the time of action) is operationally significant.

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

   Our assessment methodology is public (arXiv 2605.05145). We would welcome the opportunity to discuss machine-verifiable evidence with the FSB as the final report is prepared.