FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

RGP

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 with the benefits and risks described but consistently find the risk discussion far thinner than the benefits discussion. We note that benefits are supported by detailed case studies, while risks are largely generalized bullet points that leave significant interpretive work to the reader. The report also raises the "risk of not adopting" AI in its introduction but does not develop the point further. However, we consider the following substantive risks to be missing or underdeveloped:

(1) Risks from using AI to oversee AI. The report itself anticipates that scaled oversight will rely on automation (Sound Practice 9 suggests augmenting performance assessment "potentially through automation"; Sound Practice 10 notes monitoring "may need to be... performed by AI"). Yet the risks of this approach, such as correlated blind spots between the overseeing and overseen models, LLM-as-judge evaluation bias, and circularity where both systems share training data or a foundation model, are not addressed. Given that AI-on-AI oversight is the most likely operating model at scale, this deserves explicit treatment.

(2) Cross-institution multi-agent interactions as a systemic risk. Section 1.3 addresses agent collusion and unanticipated interactions within an institution, and homogeneity/herding through common models. It does not address the distinct scenario of autonomous agents from different institutions interacting in markets (for example in trading, liquidity management, or payments), where feedback loops could emerge that no single institution can observe or control. This is a financial stability question squarely within the FSB's mandate.

(3) Energy, compute, and physical infrastructure dependency. Energy appears only as a cost item in Sound Practice 6 and a footnoted supply-chain question. Concentration in compute and energy supply is a material resilience dependency and warrants treatment alongside the other third-party concentration risks.

(4) Talent and skills concentration. The report treats skills as an internal capability issue (Sound Practices 1 and 4) but not the sector-level risk that AI expertise is concentrated in a small number of technology firms, which compounds third-party dependency.

(5) Human capability and knowledge atrophy. As AI takes on more analytical, investigative, and decision-support work, professionals may have fewer opportunities to build the domain expertise and judgment needed to challenge, validate, and oversee AI outputs. As experienced staff retire and are replaced by AI-assisted-only practitioners, this could erode an institution's ability to provide effective oversight. For example, a loan underwriter who has never underwritten without an AI tool. This capability-pipeline risk compounds over time into a governance and control risk and deserves deeper treatment.

(6) Third-party accountability boundaries. It is unclear who is ultimately accountable for a third-party or vendor AI model, particularly as GenAI models evolve or drift after deployment.

(7) Legal and IP risk. The risk taxonomy in Section 1.3 omits legal/IP exposure (e.g., ownership of AI-generated output, copyright infringement, license contamination), even though case studies elsewhere celebrate AI-generated code, proposals, and marketing content.

(8) Execution risk. The report does not address the risk of implementation and cost overruns associated with deploying AI, which can materially affect a firm's ability to execute its AI strategy.

(9) Evidentiary gaps behind the stated benefits. The benefit claims (e.g., in the fraud detection, credit analysis, and other use-case examples on pages 5–12) are not supported by baseline sample sizes, comparison data, false-positive rates, implementation costs, human-review requirements, or a description of how systems adjust over time to account for model drift.

(10) Operational risks omitted from the discussion. Data leakage, infrastructure/vendor monoculture, and the effect of vendor-side model changes on institutions that rely on third-party models are not addressed.

We also note that the “risk of not adopting” AI is raised in the Introduction (for example in fraud defense) but not developed. A short treatment of competitive and defensive dynamics would make the benefits analysis more balanced.

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

The twelve practices are comprehensive at the level of principle. They are not yet sufficient to enable adoption, for four reasons:

(1) Proportionality is asserted, not operationalized. The report repeatedly states that smaller or less complex institutions may apply “only relevant” practices or “appropriate modification,” but never illustrates what a proportionate baseline looks like. We recommend an annex setting out an illustrative minimum implementation of the twelve practices for a small institution, alongside the full implementation expected of a G-SIB. At a minimum, every institution should be required to document ownership, access controls, testing, model and prompt changes, performance monitoring, incident reporting, rollback procedures, and system retirement. The report should also specify what documentation must be retained,

including model versions, data sources, retrieved information, tool calls, approvals, overrides, testing results, and incidents.

(2) The risk taxonomy and the practices are not mapped. Section 1.3 identifies nine risk categories, but the report never shows which practices address which risks. A one-page mapping matrix would materially improve usability and would expose the gap noted next.

(3) Consumer protection and market conduct risks have no dedicated practice. These risks are described at length in Section 1.3 but are covered only partially and indirectly through transparency (Sound Practice 8) and human oversight (Sound Practice 10). We recommend either a thirteenth sound practice on consumer outcomes, fairness, and contestability, or an explicit expansion of Sound Practice 8 to carry that weight.

(4) Key practices lack implementation tools. Sound Practice 5 explains the materiality/risk distinction well but offers no scoring approach, tiering example, or worked assessment. Sound Practice 3's documentation list would be far more usable as a template. We also recommend a crosswalk annex mapping the practices to NIST AI RMF, ISO/IEC 42001, and major jurisdictional regimes, so the report reduces rather than adds to the reconciliation burden institutions already face.

(5) Sound Practice 7 would benefit from further guidance on when to use deterministic/ML models versus GenAI models depending on the criticality of the underlying data.

(6) There is room to expand guidance on organizational change management, benefits realization, and preservation of workforce capability as institutions scale AI adoption.

On clarity, the drafting contains errors that should be corrected before final publication, including: “The sound practices builds on and is broadly compatible” (Executive summary); and the references list the same South African publication twice under “Hlophe” and “Holphe” with different attributed publishers. These are minor individually but collectively at odds with a document about rigor and documentation.

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 balance is broadly appropriate: the practices are technology-neutral at their core, with GenAI- and agentic-specific guidance layered on. Two structural issues weaken the treatment of emerging forms:

(1) Agentic AI content is scattered and repetitive. It appears in Section 1.3, Sound Practice 3 (documentation), Sound Practice 10 (a long list of “additional considerations”), Sound Practice 11, and Annex 2, with visible duplication (data security and privacy risks appear twice within the Section 1.3 agentic list alone). We recommend consolidating agentic guidance into a single dedicated section or annex, cross-referenced from the practices.

(2) The report would age better, and apply more consistently to future AI capabilities, if guidance were organized around levels of autonomy (autonomous decision-making, tool use, and reduced opportunity for human intervention) rather than around AI categories

(GenAI vs. agentic AI) alone, since governance implications flow primarily from these characteristics.

(3) Requirements should be consolidated into explicit control sets by AI type. For GenAI: source verification, retrieval quality, confidential-data leakage prevention, citation accuracy, prompt-injection defenses, and controls over model updates. For agentic AI: unique agent identities, restricted tool access, transaction limits, mandatory human approval before irreversible actions, memory-poisoning protections, detailed activity logging, and a reliable mechanism to stop or roll back an agent that behaves unexpectedly.

We would also caution against the framing in Sound Practice 10 that institutions treat AI agents "as synthetic employees." The analogy is useful for identity management and access control but risks importing inappropriate assumptions (for example about accountability and judgment) if read more broadly. Clearer boundaries on the analogy would help.

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

Partially. The core practices are durable because they are outcome-based. The flexibility problem is structural: technology-specific guidance (agentic AI controls, GenAI testing challenges, named tools and taxonomies such as OWASP and MITRE ATLAS) is embedded throughout the body text, so it will age with the technology and drag the whole document with it. We recommend:

(1) Separating durable principles from technology-specific guidance, with the latter placed in annexes that can be revised independently.

(2) Replacing "may be refined as AI technologies evolve" with a stated review cadence and version control for the document itself, which would also model the change-management discipline the report asks of institutions.

(3) Establishing a feedback mechanism through which institutions and authorities can report AI incidents, near-misses, and implementation experience to the FSB, so future revisions are grounded in evidence rather than consultation rounds alone. This would extend the information-sharing ambition of Sound Practice 11 to the practices themselves.

(4) Maintaining the glossary as a living, versioned document (see Question 7).

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.**

Not sufficiently. The case studies skew heavily toward large, internationally active banks. There is one insurer example and one G-SIB relationship-management pilot; asset managers, FMIs, pension funds, payment institutions, and nonbank lenders are described in the narrative but not evidenced in case studies. Smaller institutions, the audience most in need of illustration given the proportionality framing, are represented only by a passing observation that some banks keep AI inventories in spreadsheets.

Two additions would materially improve the set:

(1) Nonbank and small-institution case studies showing proportionate application: for example, a mid-sized asset manager's governance of a third-party LLM, an FMI's controlled testing of anomaly detection, or a community bank applying a reduced version of the documentation practice.

(2) None of the existing case studies are really about successful governance as such. The reader is left to infer that governance must have occurred because an outcome (e.g., faster processing, cost impact) is reported, rather than seeing the governance process itself illustrated.

(3) At least one post-deployment failure or harm case study. The only adverse case in the report (the developmental testing case in Sound Practice 9, where a bank postponed an underperforming ML model) is, tellingly, also the most instructive. A case where deployed AI caused customer harm or operational disruption, and what the institution changed as a result, would do more for responsible adoption than any success story.

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

Only partially. The case studies are anonymized and their headline metrics (95% of non-performing SME loans identified three months early; fraud losses reduced over 20%; turnaround reduced to 45 seconds; six million lines of code) are presented without methodology, baseline, or measurement basis. As drafted, several read as promotional rather than instructive, and institutions cannot judge whether the results are replicable or even comparable.

We recommend that final-report case studies: (i) state how headline metrics were measured; (ii) describe the governance and control decisions taken, not only the outcomes, since the decision path is what other institutions can actually reuse; and (iii) note the institution's size and the proportionality choices made. The developmental testing case study and the Annex 2/Annex 3 worked examples come closest to this standard and are the most actionable content in the report; more material in that style would be welcome.

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

The definitions provided are generally clear and consistent with mainstream usage. The main issue is coverage rather than quality:

(1) Terms used substantively in the body are absent from the glossary, including prompt injection, jailbreaking, shadow AI, model inversion, distillation, synthetic data, ground truth, and materiality. Red-teaming and interpretability are defined only in footnotes; both belong in the glossary.

(2) Conversely, some glossary entries (disinformation, vertical integration) barely feature in the body.

(3) Definitions of “materiality” and “criticality” should be added, given how central both concepts are to the sound practices (see Question 2).

(4) The glossary should state its relationship to the FSB Cyber Lexicon and to the OECD and NIST definitions it draws on, and should be versioned so definitions can track the technology (agentic AI terminology in particular is still settling).

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

We welcome the report and support its overall approach: a non-binding menu of twelve sound practices, applied proportionately, that builds on existing FSB and standard-setting body work rather than creating a new international standard. The coverage of GenAI and agentic AI is timely and more developed than in most comparable publications.

Our comments concentrate on themes that recur across the questions below:

(1) The report is strong at the level of principle but thin at the level of operation: proportionality, materiality scoring, and documentation are described conceptually without the templates, rubrics, or worked examples that practitioners, particularly smaller institutions, will need.

(2) The risk taxonomy in Section 1.3 and the twelve practices are not explicitly connected, and consumer protection risks identified in the taxonomy have no corresponding sound practice.

(3) Case studies and illustrative examples throughout the report focus on large international banks, leaving smaller institutions and nonbanks without a clear picture of how the practices apply to them.

(4) Existing case studies describe benefits/outcomes without describing the risks encountered, how they were mitigated, or the methodology/baseline behind the reported results, leaving governance to be inferred rather than demonstrated.

(5) The report's future-proofing rests on a general statement that the practices “may be refined”; a defined review mechanism would serve the document better.

We also note some minor editorial errors that should be corrected before final publication (listed under Question 2).

Overall, we believe that the report is a solid foundation and its non-prescriptive, proportionate stance is the right one. The priority for the final version should be closing the gap between principle and practice: a proportionality baseline, a risk-to-practice mapping, implementation templates for Sound Practices 3 and 5, a crosswalk to existing frameworks, and a defined revision mechanism. With those additions the sound practices would be usable by the full range of institutions they are addressed to, and durable enough to remain relevant as the technology moves.