FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Meridian Autonomy B.V.

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?**

   The report's identification of risks is comprehensive at the level of the individual financial institution. The substantive risk less fully developed in the report is that several of the named vulnerabilities are properties of the cross-institutional system rather than of any single firm.

   Third-party concentration in cloud, hardware, and foundation-model providers (Section 1.3) is named accurately, but the risk it creates does not exist inside any single institution's risk register. It exists in the joint exposure of many institutions to the same providers, where the joint distribution determines whether a failure or change at one provider propagates as an idiosyncratic event or as a systemic one. The report's reference to "homogeneous behaviours and correlated outcomes" arising from common AI dependencies identifies precisely this property; the inferential step that follows is that the property is observable, in the first instance, only by looking at the cross-firm topology rather than the firm-level disclosure.

   A second risk which is less developed in the report is the speed at which agentic AI operates relative to human intervention. Section 1.3 names the "impracticality of real-time human monitoring of agent decisions as their use scales" as a distinct challenge for human oversight of AI agents. At agentic speeds the human-in-the-loop and human-on-the-loop forms of oversight become post-hoc rather than operative. Specifying the operative control at machine speed is, by extension, a substantive supervisory question worth naming explicitly alongside the broader human-oversight requirement.

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

   The sound practices are clear and comprehensive at the level of the individual financial institution. They translate well into firm-level policy, governance, and risk management work. The issue of comprehensiveness at the level where financial-stability risk actually manifests is separate.

   The sound practices, as written, describe what each financial institution should do. The supervisory consequence is empirical: are financial institutions doing it, in what depth,

across the population, and how are practices evolving over time. The report's description of shadow AI (Section 1.3) is a good example of why this question matters. By definition, shadow AI is the AI the financial institution does not know it is operating, and the firm-internal governance frameworks the sound practices propose are not the mechanism by which it will be detected. Detection at the population level requires external observation built on public evidence, with a documented methodology that does not depend on each institution's self-report.

The same architecture applies to the third-party concentration and correlated-behaviours risks identified in Section 1.3. The sound practices' approach to third-party risk management (Sound Practice 12) is firm-by-firm and contract-by-contract; the system-level concentration risk is not addressed there, because it cannot be, by construction. A complementary practice worth naming would be the use of independent, population-level measurement as a supervisory input alongside firm self-attestation. Such measurement is reproducible cycle to cycle once the methodology is settled and provides a continuing read on the dimensions where firm self-disclosure is structurally limited.

Independent public-evidence measurement should not be viewed as a complete substitute for supervisory returns or firm-internal inventories; its value is as a complementary external observability layer that does not depend on each institution's self-report and that supports cross-firm comparability.

The element worth adding in the next iteration is empirical infrastructure that supports observation of adoption at the cross-institutional level, against measured rather than asserted positions.

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 appropriate at the level of governance and lifecycle management. The element that justifies some further development is the operative control mechanism for agentic AI at machine speed.

The report itself names "Kill switch" in Section 3.7 as one of six labelled common forms of human oversight, alongside Human-in-the-loop, AI-in-the-loop, Human-on-the-loop, Human-in-command, and Contestability. Section 3.7 describes the Kill switch as "mechanisms enabling human intervention to halt or constrain AI operations" ranging from complete shutdown to dynamic or graduated degradation. This categorisation is correct. The development worth pursuing is supervisory specification of this already-named element, given that at agentic speeds the human-in-the-loop and human-on-the-loop forms of oversight become post-hoc rather than operative. The public record does not generally contain sufficient detail concerning the parameters, testing conditions, or coordination of such automated controls to enable supervisory comparison across institutions.

One additive supervisory specification of the Kill switch concerns coordination. Where systemic propagation across institutions is plausible, such as in trading, payments, or other tightly coupled domains, the Kill switch can be specified as capable of coordinated firing across the relevant population, not only as an institution-internal capability. Other aspects

of the Kill switch, including the conditions and parameters under which it fires, the documentation of those conditions for supervisors, and the integration with broader risk management, already sit within Sound Practices 3, 5, and 11. The coordination dimension is the element not currently elaborated in the framework.

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

   The flexibility provisions of the framework, including the acknowledgement that the practices may be refined as AI technologies evolve and Sound Practice 4 on organisational adaptability, are appropriate to the rate at which the technology is changing. The cycle through which refinement occurs is an open question as it concerns the process more than the speed.

   Here there are two considerations worth raising. First, the framework's adaptability depends in practice on the availability of external indicators that capture how AI risks are materialising across the system, not only within individual financial institutions. Periodic supervisory review of adoption patterns, dependencies, and concentration across the population would provide the empirical basis for refining the sound practices in subsequent iterations. Second, the rate of change in agentic AI and in the model architectures underpinning it suggests that the interval between refinements of the practices, and the interval between the supervisory assessments that should inform them, may need to be shorter than has historically been the case for cross-sectoral guidance of this kind. The framework allows for this but the operational practice that supports it remains to be agreed and implemented. The kind of refinement resulting from external empirical indicators differs from refinement informed solely by internal observation and the framework's flexibility becomes more durably anchored when it draws on both.

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.**

   The case studies presented are informative and substantive. As a structural observation, they concentrate on deploying institutions, predominantly large internationally active banks and insurers, and predominantly illustrate the use of AI inside the institution. Following are two perspectives that would broaden the report's coverage, if additional case studies are added in the next iteration:

   (i) Cases from nonbank financial institutions, including specialist asset managers, financial market infrastructures, and payments providers, where the AI adoption patterns and the risk dimensions differ materially from those of universal banks.

   (ii) Cases from external observation activities, including academic research, audit and assurance work, regulatory technical surveys, and independent measurement infrastructure, where AI adoption is being observed across institutions rather than deployed within one. Such cases illustrate complementary supervisory inputs alongside firm self-attestation and could help illustrate Sound Practices 3, 4, and 12 at the cross-institutional level.

One specific example from the category in (ii) above follows.

Meridian Autonomy B.V. is an independent provider of empirical AI-governance measurement based in the Netherlands. The instrument is offered here as an illustration of external observability infrastructure complementary to firm self-attestation and supervisory returns. It takes public evidence as input, including disclosures, regulatory filings, technology contracts, and organisational documents, and produces a documented and reproducible reading of agent-level oversight relationships, control relationships including automated halts, and provider dependencies across financial institutions and jurisdictions. The methodology is published openly under three Zenodo concept DOIs: 10.5281/zenodo.19890626; 10.5281/zenodo.20182998; 10.5281/zenodo.20182177.

The architecture has four documented properties. Entity resolution is rule-based and reproducible against a fixed source corpus. Extraction is model-based and probabilistic, with each data point carrying an evidence tier. Canonicalisation (the reduction of variant representations to a single canonical form) is deterministic in mode, producing reproducible output from sealed inputs. A reference baseline is sealed under a published cryptographic hash, making each result set immutable and auditable. The instrument is longitudinal by design; early-cycle volatility is declared and attributed through a published variance-decomposition method rather than smoothed away. Each sealed baseline marks a point in a continuing measurement programme, with successive baselines recording methodological refinement between cycles and the seal enabling comparison across them.

Construct validity is documented in the methodology papers. At the v13.1.0 sealed snapshot, sector-level mean governance scores reproduce the theoretically predicted gradient of prudential supervisory maturity from public evidence alone, with no sector-level input during scoring. The ordering, from lowest to highest, is pension funds, sovereign wealth funds, reinsurers, investment managers, exchanges, fintechs, insurers, banks; the between-sector mean range of 8.82 points exceeds within-sector dispersion, satisfying known-groups validity in the classical measurement-theory sense (Cronbach and Meehl 1955; Messick 1995).

The April 2026 production baseline (v13.1.0) covers 543 institutions and resolves 95,876 agents and 626,390 relationship edges across multiple jurisdictions. The methodology, including the construct-validity evidence above, is open for inspection; institution-level scores, the dataset itself, and per-institution findings are held privately and shared only under formal supervisory or analytical arrangement.

The architecture sits adjacent to several of the sound practices. The cross-institutional structure provides a population-level vantage alongside the firm-level approach of Sound Practice 3. The longitudinal design supports the external indicators Sound Practice 4 implies. The cross-firm structure makes third-party dependency patterns observable at the level Sound Practice 12 treats firm-by-firm. The instrument does not substitute for firm-internal inventories, supervisory returns, or contractual third-party oversight; it sits alongside them as a complementary observability layer.

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

The case studies are substantive and illustrate AI adoption inside financial institutions usefully. Their actionability is strongest where they specify the supervisory or governance configuration that enabled the outcome, since these are the elements financial institutions can replicate; less strong where outcome metrics are reported without the operational decisions that produced them. The case-study category proposed in the response to Question 5 (external observability and measurement infrastructure) complements rather than substitutes for the deploying-institution cases, and the actionability of the two categories is different in kind. Deploying-institution cases support direct replication; external-observation cases support supervisory practice and cross-firm comparison. External-observation data also offers a useful dimension which helps complete the picture when added to data submitted by deploying institutions based on internal registers and inventories.

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

The definitions in the report, including the Section 3.7 enumeration of human-oversight forms and the Section 1.3 framing of shadow AI, agentic AI, and third-party concentration, align with the methodological vocabulary used in the response to Question 5 and in the published methodology cited there. We have no specific observations on the glossary at this stage.

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

We would emphasise three points in conclusion to this submission:

First, the sound practices describe a firm-internal architecture that is comprehensive at the level of the individual financial institution. The empirical and observability questions sit at the cross-institutional level, where individual firms cannot answer them and where the public record alone does not currently support supervisory comparison at the necessary depth. Adding the cross-institutional measurement layer as a supervisory complement to firm self-attestation would strengthen the framework substantially. This is not a replacement for supervisory returns or for firm-internal inventories; it is the third dimension that makes the framework legible and enforceable at the level where financial-stability risk actually manifests.

Second, the agentic AI section of the report acknowledges that human oversight at machine speed is impractical and lists the Kill switch as one of the labelled forms of human oversight. The supervisory specification of coordinated firing across institutions, where systemic propagation is plausible, is the genuinely additive element the framework currently lacks. Without coordination across institutions, the Kill switch protects the institution but not the system; tightly coupled domains such as trading, payments, and shared infrastructure require the population-level coordination dimension to be named explicitly.

Third, the architecture of independent population-level measurement of AI governance is operable in practice, not only in theory. The instrument described in the response to Question 5 documents one such measurement, with the methodology published openly, validation evidence including known-groups validity across sectors, and a sealed reference baseline that supports cycle-to-cycle comparison. The architecture is reproducible. What

such infrastructure cannot do is substitute for supervisory returns or firm-internal inventories. What it can do is sit alongside them as a complementary external observability layer, providing the cross-firm comparability the firm-level framework cannot achieve by construction.

The supervisory architecture for AI in financial services is being shaped at the highest level of supervision and regulation at this moment. The framework would benefit from naming external observability infrastructure as a recognised supervisory complement alongside firm self-attestation and supervisory returns.