FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Veritas Core

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 accurately diagnoses the terminal vulnerabilities of the current AI trajectory, particularly by explicitly acknowledging that agentic AI poses a "distinct challenge for human oversight" and that real-time human monitoring of autonomous agents at scale is impractical. Furthermore, identifying the risk of "reward hacking" and unauthorized actions accurately pinpoints the danger of autonomous execution.

   However, the report misses the most critical, substantive risk: the terminal vulnerability of the software abstraction layer itself. The report assumes that the risks of software-based AI can be governed by layering more software-based policies and "guardrails" on top of it. Software is inherently malleable. If a threat actor or a hallucinating agentic model operates within an abstraction layer, it can spoof, bypass, or manipulate those software controls. You cannot secure non-deterministic, autonomous systems using malleable logic.

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

   The sound practices are comprehensive regarding policy, but fundamentally deficient regarding execution. Practices like Sound Practice 10 (Human Oversight) and Sound Practice 11 (Cyber and ICT risk management) rely on human committees, manual monitoring, and API-level software controls.

   A governance committee does not stop a malicious electrical signal. When an autonomous agent attempts an unauthorized transaction at machine speed, "AI-in-the-loop" monitoring simply creates a recursive loop of software validating software. To be truly responsible and comprehensive, the practices must mandate that governance boundaries are translated into deterministic mathematical constraints and enforced physically at the bare-metal silicon layer before the command ever reaches the operating system

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?**

No, the balance relies too heavily on legacy paradigms. The report acknowledges the fundamental leap to agentic AI—noting that agents can execute complex goals independently and that overriding them "can be difficult or impossible for humans".

Yet, the proposed mitigations for agentic AI remain rooted in traditional ML governance paradigms (e.g., more monitoring, logging, and software-based "kill switches"). If an agent dynamically modifies its objectives and operates autonomously at scale, a software kill-switch is a vulnerability, not a control. The balance must shift from monitoring agentic AI to physically gating its execution at T=0.

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

   The sound practices are too flexible, which is precisely the problem. In the context of agentic AI executing financial transactions or interacting with critical infrastructure, "flexibility" in governance is a systemic vulnerability.

   Financial institutions do not need flexible enforcement; they need absolute, deterministic truth. While the creation of localized legal frameworks and operational rules should be flexible (allowing entities to define their own governance), the enforcement of those rules must be inflexible and anchored in the physical substrate.

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 current case studies exclusively highlight software-layer integrations. An additional case study must be included demonstrating hardware-enforced governance (Admissibility by Design) for autonomous agents.

   Proposed Case Study Addition:

   A financial technology infrastructure provider successfully completed Phase 1 Conformance testing with a third-party AI Interoperability Lab to govern agentic AI actions. Instead of relying on software wrappers, the firm deployed a bare-metal hardware circuit breaker. During the test, autonomous agents requested actions across three core integration boundaries (Application Autonomy, Cryptographic/Hardware Governance, and Human-in-the-Loop). By evaluating the mathematical parameters of every request at T=0, the hardware circuit breaker achieved a 100% pass rate in physically blocking tampered fields, mismatched cryptographic hashes, and unauthorized autonomous actions deterministically, proving that unauthorized execution can be made physically impossible at the register layer.

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

   They provide actionable insights for compliance reporting, but not for preventing systemic failure. The case studies focus on improving detection and speeding up manual processes, but fail to demonstrate how a financial institution can continuously prove that it retains meaningful operational authority while an autonomous system is running. To be actionable

against the risks of agentic AI, insights must demonstrate how to transition from "approving decisions" to physically "governing the execution boundaries."

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

   The glossary captures standard AI terminology but entirely omits the terminology required to govern it effectively at the infrastructure level.

   To align with the reality of securing complex AI, the following terms should be added:

   Hardware-Enforced Determinism: The translation of operational, legal, or compliance boundaries into mathematical constraints that are evaluated at the bare-metal silicon layer, making unauthorized execution physically impossible.

   Abstraction Layer Vulnerability: The inherent risk that any software-based governance, monitoring, or security tool can be bypassed, hallucinated, or manipulated by an advanced AI model operating within the same digital environment.

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

   The FSB has correctly identified that the financial sector is transitioning from digital dangers to an era of complete systemic vulnerability due to autonomous capabilities. However, you cannot govern physics with policy.

   As long as global financial stability relies on AI vendors enforcing their own guardrails via software, the global financial system remains entirely exposed to algorithmic hallucinations, data fabrication, and autonomous "reward hacking." True digital sovereignty and systemic stability require demanding that all agentic AI actions intersecting with the financial system be cryptographically proven and physically evaluated at the bare-metal layer before execution. Until the FSB mandates hardware-anchored truth, all proposed AI risk management is simply "vibe governance."