tamam

PUBLIC SUBMISSION
JUNE 2026

RESPONSE TO THE FSB CONSULTATION

Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

RESPONDENT	SUBMITTED TO	CONSULTATION CLOSES
Tamam	Financial Stability Board (fsb@fsb.org)	22 July 2026

CONTENTS

Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

01 Preliminary remarks

Tamam is a financial institution that builds and operates AI systems, including agentic systems, in production. We welcome the consultation and offer this response from the perspective of a practitioner. Our comments are deliberately principles-based: we describe design principles we consider sound for responsible adoption rather than particular products, vendors, or deployments.

Our central observation, which runs through every answer below, is that the controls which actually hold for autonomous and semi-autonomous AI are **architectural rather than procedural**. Four principles recur:

1. **Structural consent** — authorisation for any consequential action (moving value, signing, changing state) is enforced in code, through a confirmation token or mandate, rather than relying on the model's behaviour or a prompt instruction.
2. **Immutable, independently-verifiable audit** — every AI action is recorded to a tamper-evident log that can be reconstructed and verified after the fact, so accountability does not depend on the system's own good behaviour.
3. **Agent identity and least privilege** — each agent has a distinct identity, the minimum permissions necessary, and a bounded interaction perimeter.
4. **Meaningful human oversight with explainability** — oversight that carries real authority to intervene, supported by explanations that let a human understand and challenge an action.

To make these arguments concrete rather than rhetorical, we include a set of **illustrative simulations**. They are stylised and structural — they show how a risk scales or where a control helps; they are **not** empirical claims about any institution. Parameters are drawn from public literature where available (agent-reliability studies, prompt-injection benchmarks, AML monitoring statistics) and otherwise labelled as transparent assumptions. Full models, parameters, validation, sensitivity analysis, and sources are in the accompanying methodology note. We would be glad to share the underlying code with the FSB.

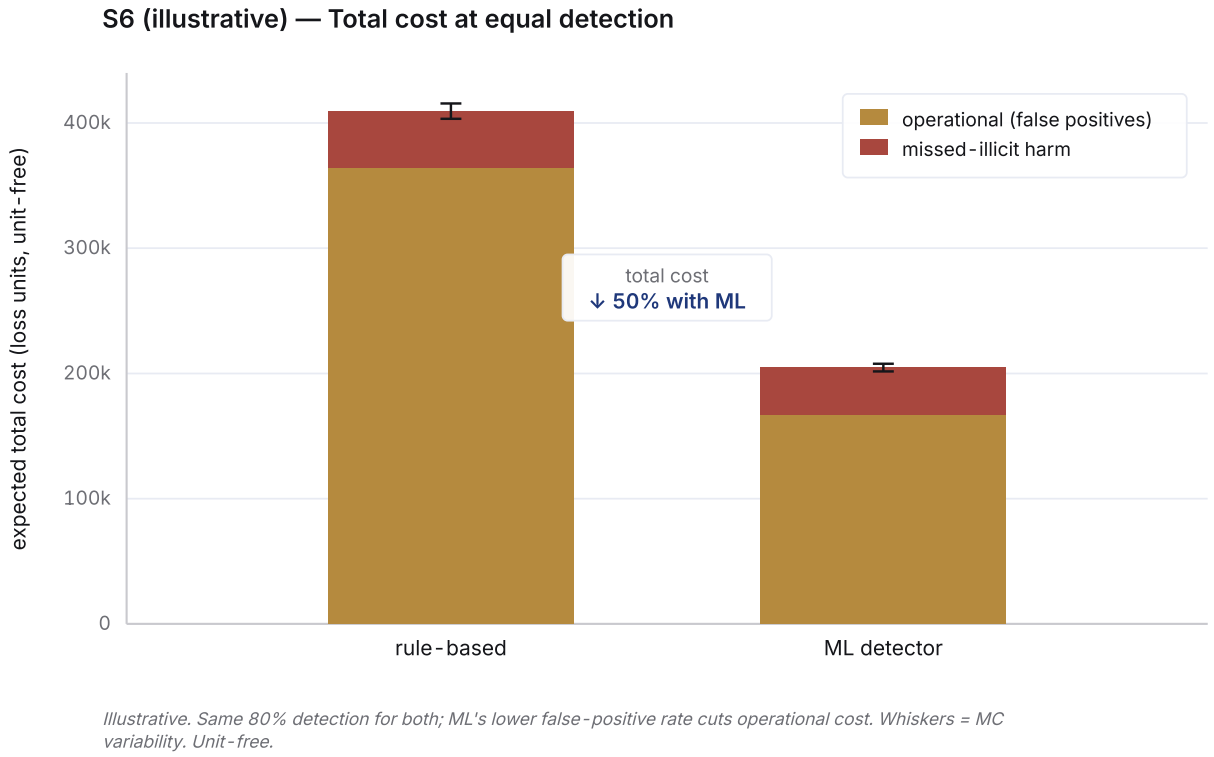
02 Question 1 — Benefits and risks of AI adoption

THE CONSULTATION ASKS

> *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. Section 1 gives an accurate and well-evidenced account of the benefits across banking, insurance, capital markets, payments and FMIs, and the four financial-stability vulnerabilities carried over from the 2024 report remain the right anchors. We particularly endorse two framings: that AI can be simultaneously an opportunity and a risk in the same domain (most clearly cyber), and that **not adopting AI is itself a risk** — most obviously in financial-crime detection, where rule-based monitoring is both less effective and more costly.

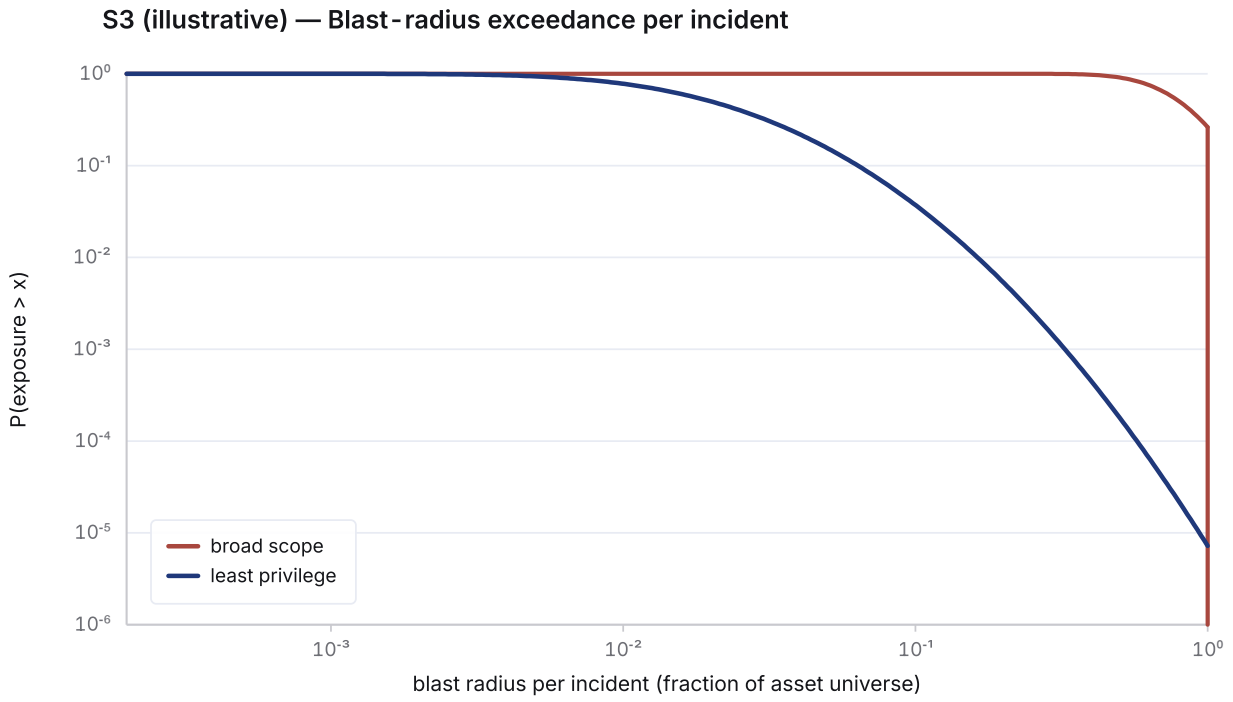
To make the "cost of not adopting" point concrete, we modelled a detection task at equal detection performance. Holding detection constant, a machine-learning detector's lower false-positive rate (kept within the documented 50–60% reduction band) cuts both operational load and missed-harm cost materially. This is the benefit side of the ledger the report rightly emphasises.



S6 — at equal detection, a better false-positive rate lowers total cost. Illustrative.

We would add or strengthen the following risks:

Blast radius as a function of agent permissions. The report rightly flags agent data-breach and disruption risks. We suggest making explicit that the severity of any agent compromise or error is governed by the scope of access the agent holds. An agent performing a low-materiality task can still cause high-materiality harm if over-privileged. Least privilege is therefore not only a cyber-hygiene measure but a primary determinant of tail loss.



Illustrative. Incident frequency identical across regimes; only reachable scope differs (broad 80% vs least-privilege 2% of universe). Severity ~ Lognormal. Log-log.

S3 — least privilege bounds the tail of exposure even when incident frequency is unchanged. Illustrative.

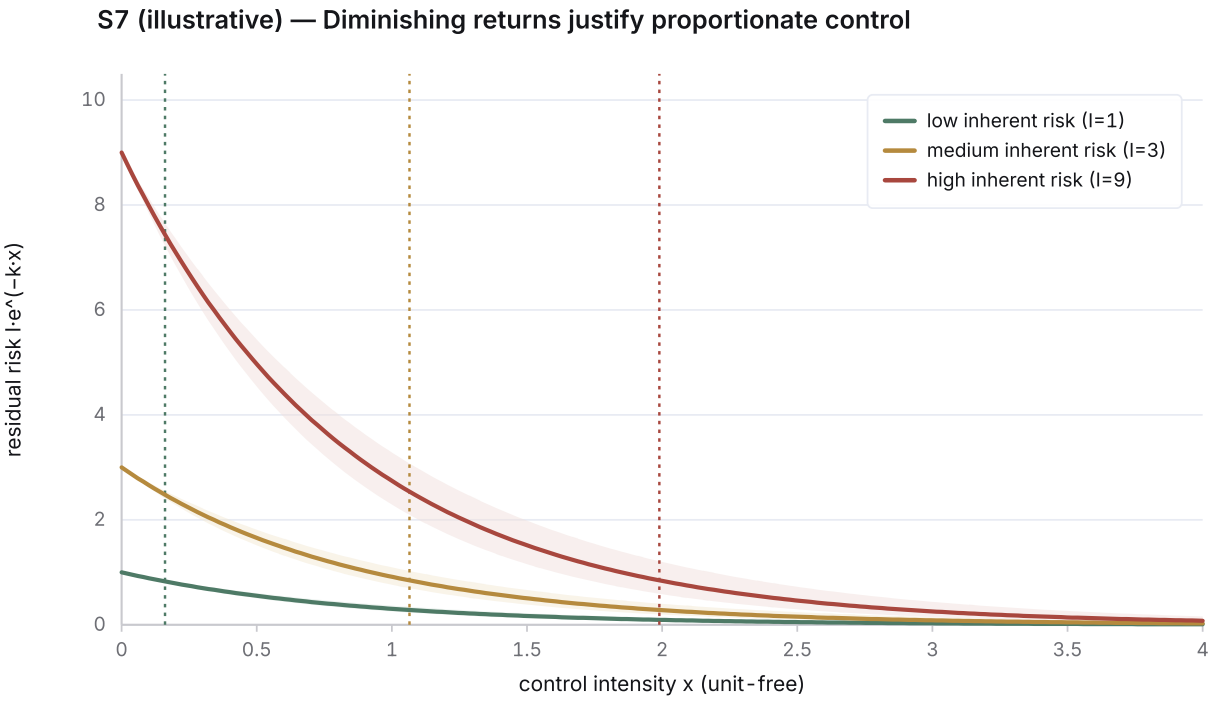
Other additions. We would also weight more heavily: (i) **financial inclusion** as a first-order benefit (AI underwriting on alternative data, lower-cost servicing, vernacular interfaces) with guardrails, not merely as a sub-point of credit scoring; (ii) **talent and skills concentration** — systemic fragility from many institutions depending on the same scarce expertise and few advisers; (iii) **GenAI output, IP and liability** risk distinct from hallucination; and (iv) **model collapse / synthetic-data feedback loops** as synthetic-data use grows. Finally, we suggest the report note **foundation-model concentration** (not only cloud and hardware concentration) as a vector for correlated behaviour across institutions.

03 Question 2 — Comprehensiveness and clarity

THE CONSULTATION ASKS

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

Yes. The twelve practices are comprehensive and coherent, and the lifecycle framing is usable. We single out the treatment of **proportionality** — anchoring control intensity to materiality and risk, assessed at inception and thereafter — as the report's strongest organising idea. It is also analytically correct: because control effectiveness has diminishing returns, uniform maximum control is wasteful and uniform minimum control is unsafe, so the cost-minimising control intensity genuinely differs by tier.



S7 — diminishing returns make proportionate (tier-dependent) control intensity optimal, not merely convenient. Illustrative.

Our main suggestion for clarity is to **distinguish prompt-level / behavioural controls from structural / architectural controls** explicitly. The report's guardrail language sometimes spans both. The distinction matters because, as we show under Question 3, behavioural controls degrade with autonomy and task length whereas structural controls do not. A short statement to this effect — and a worked, explicitly-illustrative proportionality example — would help less-experienced adopters operationalise the practices. We also suggest a one-line clarification of how AI risk management relates to existing model-risk-management frameworks (complement, superset, or extension), to avoid duplicate internal frameworks.

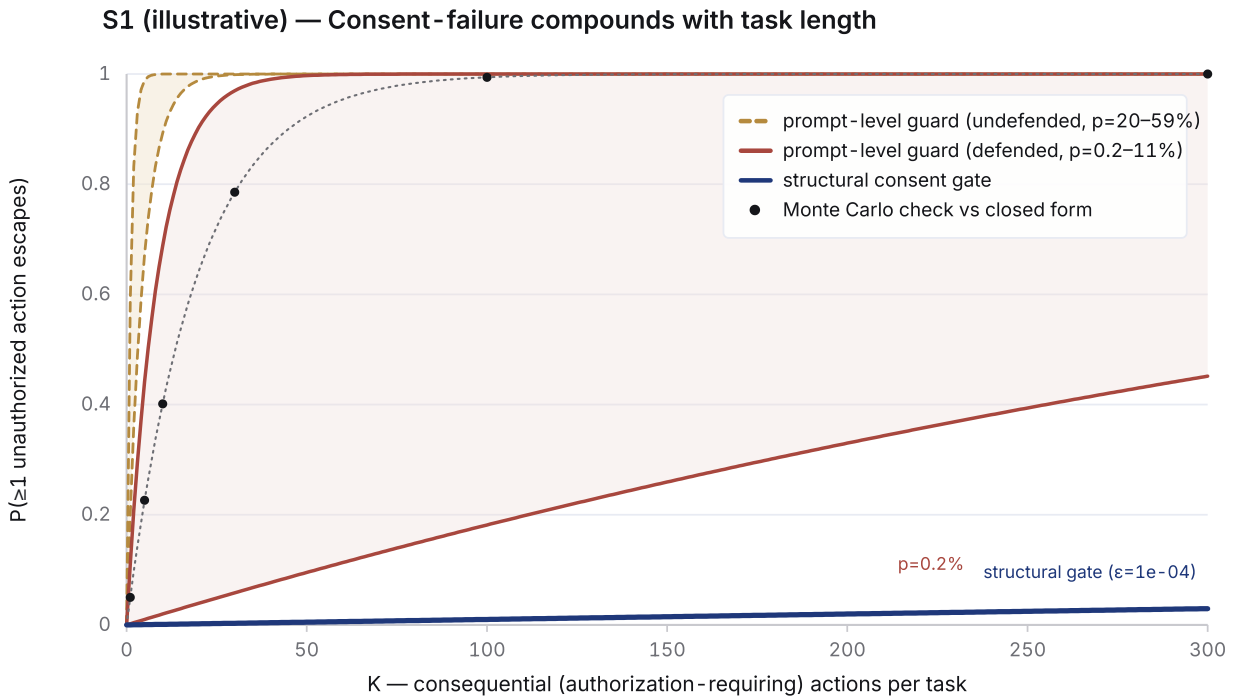
04 Question 3 — Balance between all AI and emerging GenAI / agentic AI

THE CONSULTATION ASKS

> *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?*

Yes, the architecture is right: general practices, with GenAI- and agentic-specific considerations layered into each. The agentic-risk taxonomy and the human-oversight spectrum are among the clearest we have seen. This is also the area where our practitioner perspective is strongest, and where we most want to reinforce a specific point.

Probabilistic controls degrade with task length; structural controls do not. The report observes that an agent may take "hundreds of intermediate steps and make errors in any of those steps." This has a precise consequence. If authorisation is enforced by a probabilistic, prompt-level guard with per-action failure probability p, then the probability that at least one unauthorised action escapes over K consequential actions is `1 - (1 - p)^K`, which rises sharply with K. A structural consent gate — where each consequential action requires a cryptographic confirmation token or mandate — blocks deterministically regardless of K. We modelled this using defended prompt-injection failure rates (0.2–11%), not the worse undefended range, to keep the comparison fair.

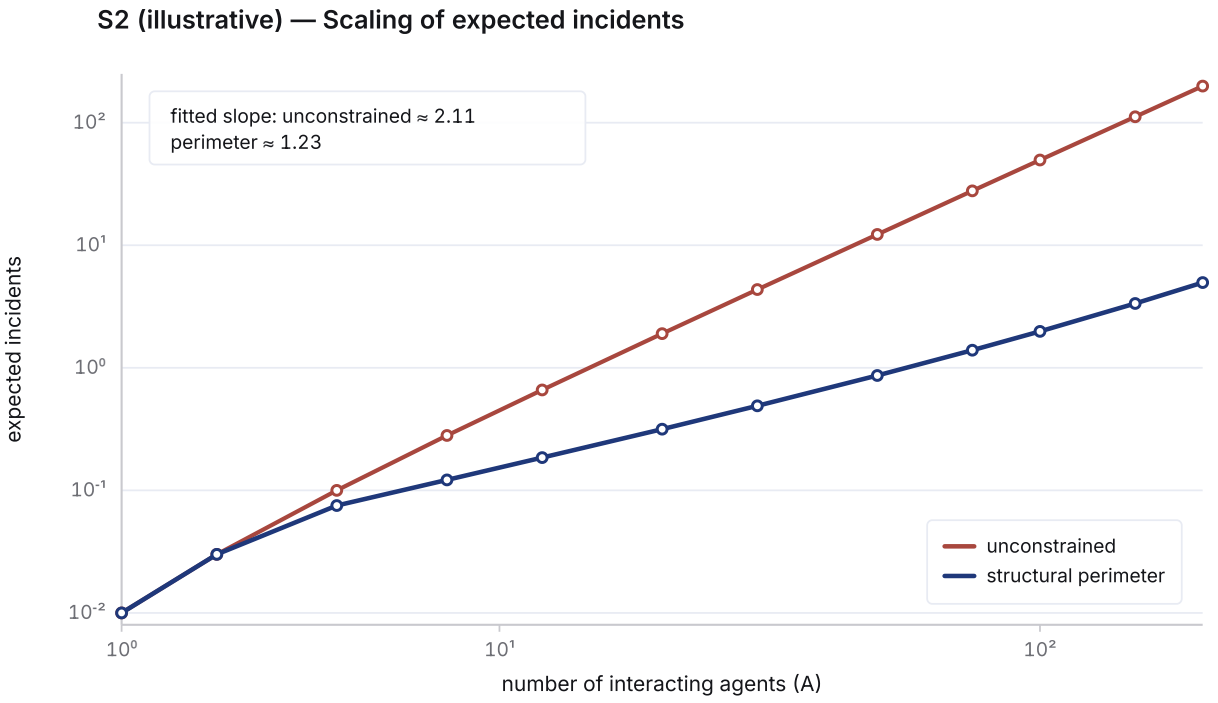


Illustrative. Curves are closed-form 1−(1−p)^K; dots are 1e6-draw Monte Carlo. Structural gate is independent of K by construction (residual ε only).

S1 (hero) — consent failure compounds with task length under prompt-level guards; a structural gate is flat in K. Defended (solid) and undefended (dashed) bands shown. Illustrative.

The practical implication is large: at a representative task-length distribution, the expected number of unauthorised actions per task differs by roughly two to three orders of magnitude between the two regimes, and the loss tail is correspondingly heavier for prompt-level control. We therefore suggest the report state, as a sound practice for agentic AI, that authorisation of consequential actions should be **enforced structurally**, not behaviourally.

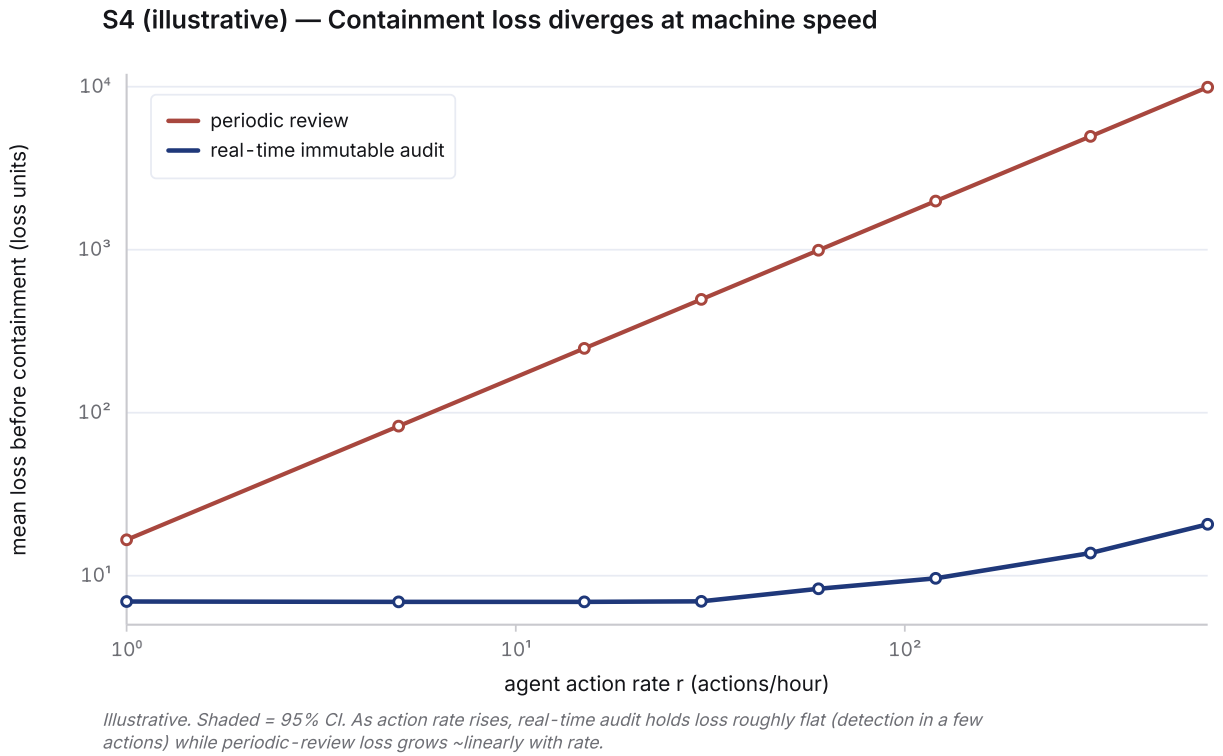
Multi-agent interaction scales super-linearly; a perimeter restores linear scaling. The report notes that multiple agents may interact in unanticipated ways. The number of potential interactions grows roughly with the square of the number of agents; an identity-and-least-privilege perimeter that authorises only a sparse set of interactions restores near-linear scaling. In our model the expected-incident exponent falls from ~2.1 to ~1.2.



Illustrative. Log-log; slope ≈ 2 indicates quadratic growth, ≈ 1 indicates linear.

S2 — expected incidents scale ~A^2 unconstrained vs ~A under a least-privilege perimeter. Illustrative.

Accountability must operate at machine speed. Because agents act quickly, loss accrues between an adverse event and its containment. An immutable, per-action audit feeding real-time anomaly detection bounds that loss; periodic or batch logging lets it compound, and the gap widens with the agent's action rate. This is the financial-stability case for accountability-by-construction rather than after-the-fact reconstruction from mutable logs.



S4 — containment loss diverges at machine speed; real-time immutable audit holds it roughly flat. Illustrative.

On balance we support the report's decision not to over-specify agentic AI given how fast it is moving. The two areas we would prioritise for future iteration are concrete guidance on **multi-agent / emergent-interaction** management and on **evaluating GenAI/agentic performance without ground truth**.

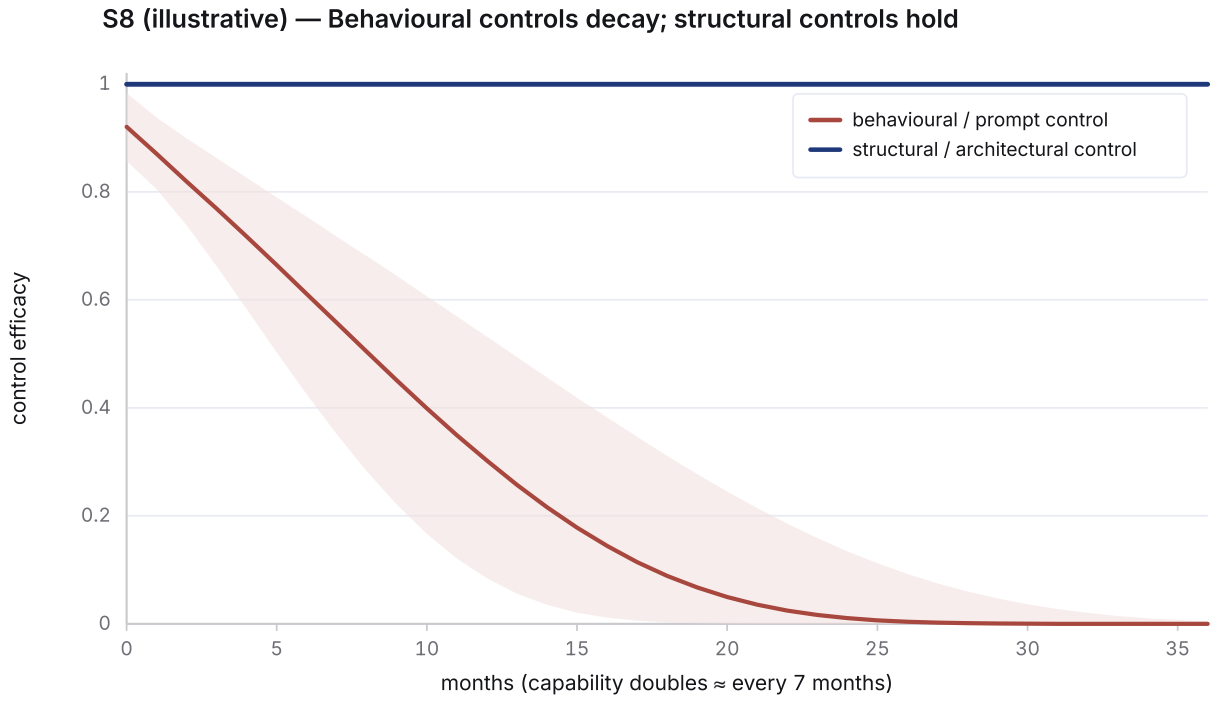
05 Question 4 — Flexibility to accommodate newer AI over time

THE CONSULTATION ASKS

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

Yes. Technology-neutral, outcomes-focused drafting plus a dedicated organisational-adaptability practice make the framework appropriately future-proof. We add one practitioner argument for why an architectural emphasis future-proofs better than a behavioural one.

Controls that depend on the model's cooperation — prompt instructions, behavioural guardrails, persuasion not to misbehave — tend to lose efficacy as model capability and autonomy grow. Structural controls (a consent gate, an identity, a permission boundary, an immutable log) are invariant to capability by construction. Over a multi-year horizon of rising capability, the two diverge: behavioural-control efficacy decays while structural-control efficacy holds.



S8 — as capability grows, behavioural-control efficacy decays while structural-control efficacy holds. Illustrative.

We therefore suggest the report (i) signal an intended periodic review cadence with a lightweight channel for the industry to flag emerging risks between reviews; (ii) frame its references to specific tools and standards as examples of the type of source to track, since named tools will date; and (iii) where it must choose between behavioural and architectural framings of a control, lean architectural, as the more durable.

06

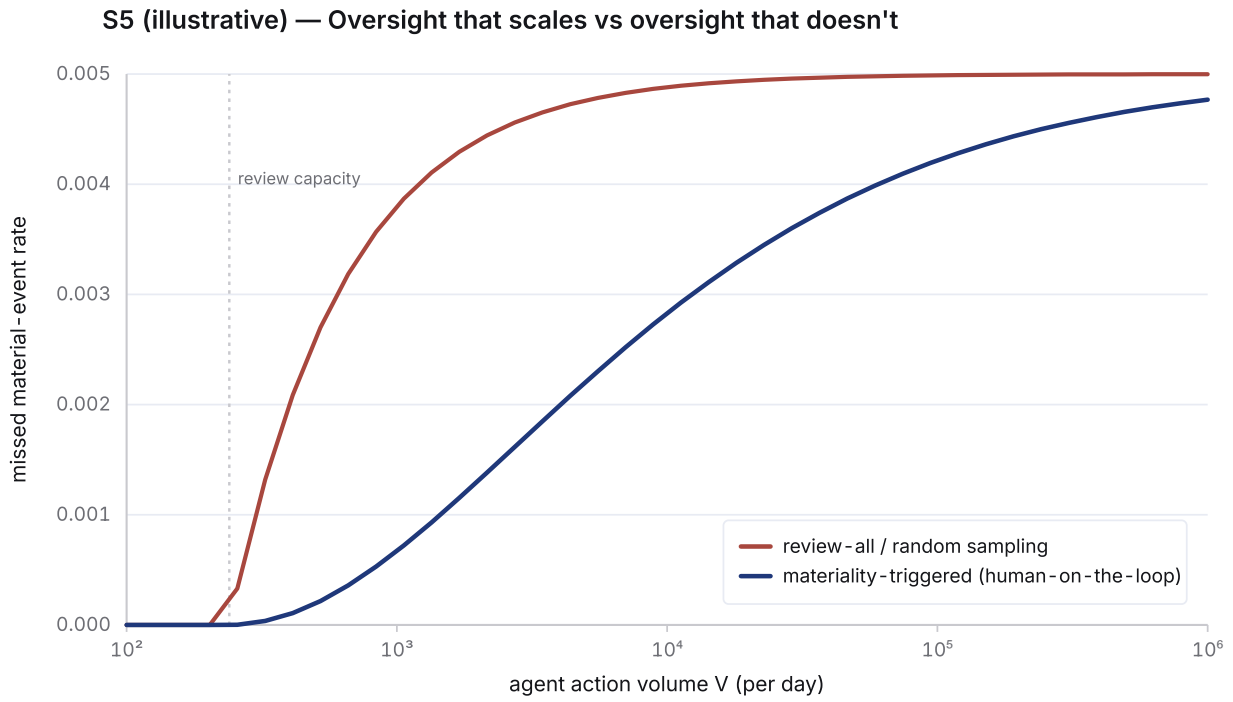
Question 5 — Case studies across financial-institution types, including nonbanks

THE CONSULTATION ASKS

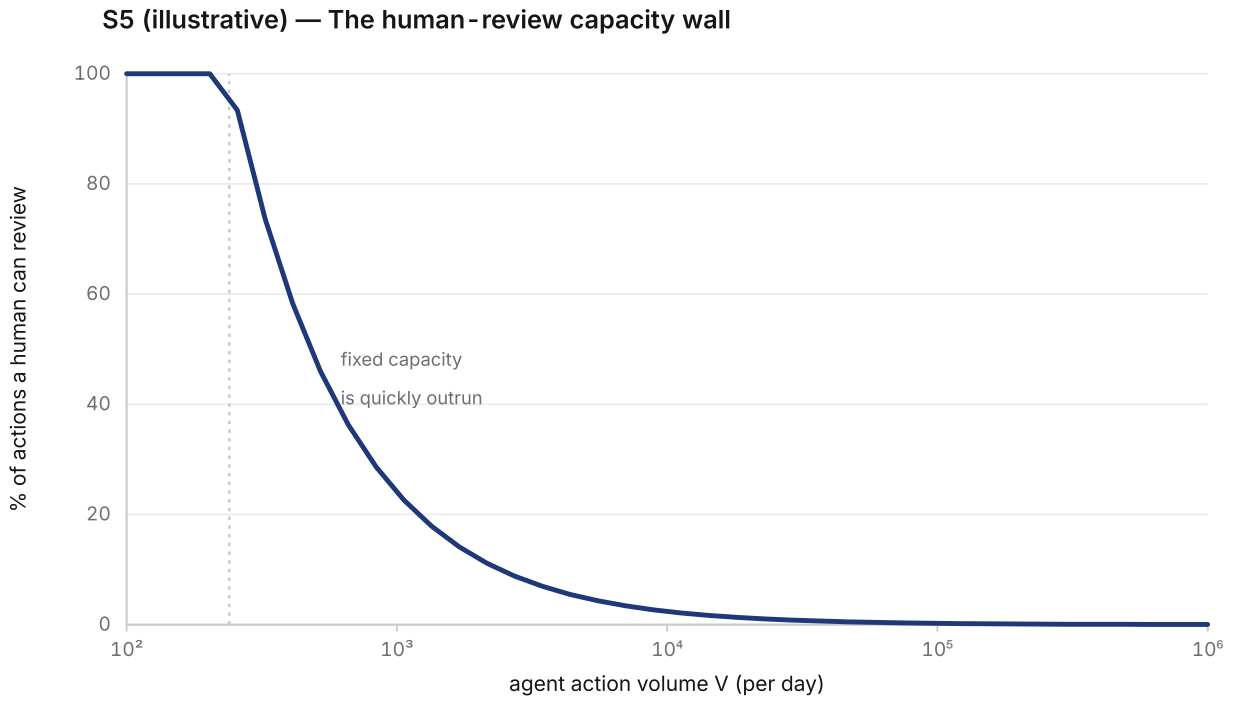
> *Do the case studies in this report sufficiently highlight how different types of financial institutions can benefit from responsible AI adoption? Are there additional case studies for inclusion, particularly for nonbanks?*

This is where we most encourage change. The case studies are valuable but skew heavily toward large, internationally active banks and G-SIBs. Nonbanks — fintechs, payment and e-money firms, consumer-finance providers — and smaller and emerging-market institutions are largely absent, even though they are often the fastest adopters and carry distinct risk profiles (heavier reliance on third-party and open-weight models; agentic, customer-facing deployments). We suggest adding nonbank and smaller-institution cases, and at least one case where the institution is a **deployer** of third-party AI rather than a builder, illustrating compensating controls where vendor visibility is limited.

We also offer a principles-based illustration relevant to nonbank oversight. A recurring nonbank challenge is that human review does not scale with agent action volume: a fixed review team is quickly outrun, and blanket human-in-the-loop becomes infeasible. Materiality-triggered oversight (human-on-the-loop, with scarce review capacity concentrated on the riskiest items) sustains a far lower missed-material-event rate within the same budget. This is the quantitative case for the report's own point about the impracticality of real-time human monitoring as agent use scales.



S5 — blanket review hits a capacity wall; materiality-triggered oversight scales. Illustrative.



S5 — the human-review capacity wall: the share of actions a fixed team can review collapses as volume grows. Illustrative.

We would be glad to contribute an anonymized nonbank case study (agentic, customer-facing operations with materiality-triggered oversight and immutable per-action audit) in the report's house style, if useful.

07 Question 6 — Do the case studies provide actionable insights?

THE CONSULTATION ASKS

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

Partially, and they can be made more so. The most instructive case in the report is the developmental-testing example where an institution declined to deploy an underperforming model — it shows the practices enabling a "no-go," not only adoption. We suggest four improvements: (i) **tag each case study to the specific sound practice(s)** it illustrates, turning narratives into a navigable playbook; (ii) add a one-line **"transferable takeaway"** to each, separated from firm-specific detail; (iii) include **scaled-down variants** showing how a smaller institution could meet the same control objective proportionately; and (iv) add more **near-miss and remediation** cases, which are more actionable than success stories. Where case studies quote quantified outcomes, a brief note on the baseline and measurement method would help readers judge transferability rather than read the figures as benchmarks — the same discipline we have applied to our own illustrative figures.

08 Question 7 — Glossary

THE CONSULTATION ASKS

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

The glossary is clear and broadly aligned with common usage; the agentic-AI and RAG entries reflect recent developments well. We suggest adding terms the report relies on but does not define, which would aid consistent interpretation:

- **Materiality** and **risk** (as applied to AI use cases), defined together with how they differ — they underpin proportionality.
- The **human-oversight modes** the report itself uses: human-in-the-loop, AI-in-the-loop, human-on-the-loop, human-in-command, kill switch, and contestability.
- **Shadow AI**; **compensating controls**; **least privilege**; **agent identity**; **champion/challenger model**; **red-teaming**; **inherent vs residual risk**; **data lineage / provenance**.
- **Agentic memory poisoning** (distinct from the data/model poisoning already defined).

We would also promote **interpretability** from a footnote into the glossary alongside explainability, and add a one-line cross-reference distinguishing **AI model / system / use case / application**, which are currently used somewhat interchangeably.

09 Closing

Tamam supports the FSB's direction. Our one substantive theme is that, for autonomous and semi-autonomous AI, **architectural controls outperform procedural ones and age better** — structural consent, immutable verifiable audit, agent identity with least privilege, and meaningful human oversight. We have tried to evidence this with transparent, reproducible illustrations rather than assertion, and we would welcome continued engagement, including contributing a nonbank case study and sharing the simulation code behind this response.

CEO - Tamam

01 Jul 2026

tamam

PUBLIC SUBMISSION · JUNE 2026

END OF SUBMISSION

Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

RESPONDENT	SUBMITTED TO	CONSULTATION CLOSES
Tamam	Financial Stability Board (fsb@fsb.org)	22 July 2026