FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Foveal Risk Advisory

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 risk taxonomy is comprehensive at the categories it names: governance and accountability gaps, model risk, data quality and governance, conduct risk, explainability, cyber risk, and third-party and concentration risk are all addressed, and the report explicitly discusses unauthorized agentic action as an institutional governance risk, noting that remediation can be difficult or impossible once such action occurs.

   What is not named, at any level of this taxonomy, is the risk that follows once an unauthorized agentic action is a consumer payment: current frameworks do not specify who bears the resulting loss or how the affected consumer's dispute is resolved. This is a distinct risk from the ones the report enumerates, because it sits at the boundary between the institution and its customer rather than within the institution's own governance perimeter. I recommend the FSB add this as a named risk category — external liability arising from unauthorized agentic action — since it falls outside every category currently identified in the report's risk discussion.

   As the industry context above illustrates, even well-resourced card-network authorization infrastructure ends at the network's boundary, and agentic payment activity routinely crosses that boundary today. The sound practices should address this network-perimeter condition directly.

   Recommendation: Add ‘external liability arising from unauthorized agentic action’ as a distinct risk category in the report's taxonomy, and note explicitly that this risk sits at the institution–customer boundary, not only within internal AI governance.

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

   Sound Practice 10, addressing human oversight, is the sound practice closest to this issue, and it appropriately acknowledges the practical limits of continuous human monitoring for agentic systems. However, it is framed entirely as an internal control matter — how an institution monitors its own agents — rather than addressing what happens when an agent-initiated action affects a consumer counterparty's payment rights. I recommend the FSB

clarify, within Sound Practice 10 or as an additional practice, that institutions deploying payment-capable AI agents should maintain records adequate to reconstruct the scope of consumer authorization behind any agent-initiated transaction.

Recommendation: Clarify in Sound Practice 10, or a new practice, that institutions deploying payment-capable AI agents should maintain records sufficient to reconstruct the scope and terms of consumer authorization for any agent-initiated transaction.

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 sound practices address internal governance of agentic AI systems effectively. They do not yet address the external liability gap — the space between what an institution governs internally and how a dispute is actually resolved externally when an agent-initiated transaction produces consumer harm. I recommend the FSB consider a supplemental sound practice specifically addressing agentic AI in consumer-facing payment contexts, covering:

● Explicit authorization scope documentation: institutions deploying AI agents with payment capabilities should maintain documented records of the scope of consumer authorization, sufficient to support dispute resolution.

● Consumer disclosure standards: consumers should receive clear disclosure when an AI agent is the initiating party in a payment transaction, with the ability to contest agent-initiated charges through existing dispute channels.

● Liability framework coordination: the FSB should engage with standard-setting bodies — including CPMI, national regulators administering consumer payment frameworks, and card network operators — to develop guidance on liability allocation for agent-initiated transactions.

The card network trust-layer developments described above are a useful data point for this balance question: they show the industry treats the agentic authorization gap as a genuine risk and has invested substantially in addressing it. They also reveal that capacity's structural limit — industry-led solutions are, by construction, perimeter solutions that govern only transactions on the networks that build them. We recommend the FSB's sound practices distinguish explicitly between network-layer authorization controls and rail-agnostic authorization standards that would apply regardless of settlement substrate. Only the latter can close the governance gap a network perimeter, however well built, leaves open.

A related but distinct balance problem arises on irreversible settlement rails, which I address in Section III below.

Recommendation: Distinguish, in the sound practices, network-layer authorization controls that industry participants are building for their own networks from rail-agnostic authorization standards that apply regardless of settlement substrate — and specify that only the latter closes the governance gap a network perimeter leaves open.

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

Because the sound practices are framed at the level of institutional AI governance, they are largely durable as AI capability evolves. The area least future-proofed is precisely the authorization and liability question raised above: as agentic systems take on longer-running, more autonomous payment-adjacent tasks, the gap between internal governance and external liability frameworks will widen rather than narrow unless it is addressed at the level of definitions and coordination now, while the practices are still in consultation.

This flexibility question also has a data-access dimension that the report does not address. The consultation report is silent on the regulatory status of the data-access frameworks that agentic AI systems depend on to operate — and that silence matters, because any oversight regime for agentic AI implicitly assumes a stable data-access baseline that does not currently exist in either of the world's two most consequential open-finance jurisdictions.

In the European Union, the Payment Services Regulation and revised Payment Services Directive reached final political agreement in April 2026 and are expected to enter into force on a roughly 21-month transition timeline following publication — but these instruments govern payments-focused data sharing only. The Financial Data Access framework, which would extend open-data obligations to investments, pensions, insurance, and mortgages, remains in trilogue negotiations, with realistic implementation timelines pointing to 2029. The compensation mechanism between data holders and data users — arguably the single most consequential design choice for how agentic systems could be authorized to pull data on a continuous basis — remains unresolved.

In the United States, the CFPB's open banking rule under Section 1033 is currently enjoined. A federal court found that the rule's mandate to share data with commercial third parties, rather than only the consumer, exceeded the agency's statutory authority, as did its prohibition on data-access fees. The rule is now being substantively rewritten through a new advance notice of proposed rulemaking. As of this submission, no binding compliance obligation is in effect.

Both frameworks were designed around point-in-time, consent-based data sharing — a human authorizing a single, discrete pull of their own data. Neither anticipated an autonomous agent making continuous, unattended access decisions on a customer's behalf. I recommend the FSB acknowledge this gap explicitly in the final report: sound practices for human oversight of agentic AI in payments and finance cannot be fully specified independent of the data-access regime an agent operates under — a regime that, in both the EU and the US, remains actively unsettled.

Recommendation: Acknowledge explicitly that sound practices for human oversight of agentic AI cannot be fully specified without reference to the data-access regime governing the agent, and invite CPMI and relevant jurisdictions to consider agentic, continuous-access use cases in their open-finance frameworks.

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.

No comment. Our submission does not propose additional case studies.

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

No comment. Our submission focuses on the risk-taxonomy and sound-practice questions addressed above.

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

The report's glossary would benefit from working definitions of three terms that are foundational to how consumer payment liability is currently allocated, and that agentic AI renders ambiguous:

● "Authorization": In Regulation E and card network dispute rules, authorization is the foundational concept determining whether a transaction is valid or disputable. A consumer granting an AI agent general spending authority is legally and operationally distinct from authorizing a specific transaction. That distinction is currently invisible in the report's definitional framework.

● "Transaction initiator": Existing payment frameworks define liability partly by reference to who initiated the transaction. The glossary does not address what it means for an AI agent to be the initiating party, nor how this affects the liability chain between consumer, institution, merchant, and platform operator.

● "Delegated authority scope": The report's discussion of agentic AI implicitly treats these systems as something like synthetic employees operating within defined authorization limits, but the report does not define the bounds of delegated authority or specify how scope limitations should interact with consumer dispute rights when an agent acts outside its intended parameters.

These definitional gaps are not semantic. They are the exact points at which existing consumer payment regulations will produce indeterminate or consumer-adverse outcomes when applied to agent-initiated transactions. I recommend the FSB add working definitions for these terms — or at minimum note that they require definition by relevant standard-setting bodies in the context of consumer payment frameworks — in the final report's glossary.

Recommendation: Include working definitions of 'authorization,' 'transaction initiator,' and 'delegated authority scope' in the glossary, or explicitly task relevant standard-setting bodies with defining these terms for agent-initiated consumer payment contexts.

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

Supplemental Technical Governance Comments: Sound Practices 8 and 10

The comments above address an external liability gap. This section addresses a distinct, internal governance gap in how Sound Practices 8 (Explainability and transparency) and 10 (Human oversight) are currently framed for AI systems operating on irreversible payment and settlement rails.

A. Sound Practice 10 — continuous rather than point-in-time validation

Sound Practice 10 asks financial institutions to implement human oversight calibrated to the materiality, risk, autonomy, complexity, and explainability of an AI use case. This calibration implicitly assumes an oversight architecture that can be assessed and set at a point in time and adjusted periodically thereafter. Adaptive, self-updating agentic systems do not hold still long enough for that assumption to fully apply: an agent's decision pathways can evolve in production in ways that a periodic reassessment cycle will not catch in real time.

The IMF's April 2026 analysis of tokenized finance frames the underlying shift precisely: as legal certainty comes to depend on code execution rather than post-hoc legal remedy, supervisory frameworks need mechanisms that allow contract or transaction execution to be paused under predefined conditions, rather than relying solely on remediation after the fact (Adrian, “Tokenized Finance,” IMF Notes 2026/001, April 2026). We recommend the final report adopt this logic explicitly within Sound Practice 10: institutions deploying agentic AI on irreversible rails should maintain a pre-defined, testable set of intervention conditions — a kill-switch or pause mechanism keyed to specific behavioral triggers — rather than relying on periodic human review alone.

Transaction cloning and replay attacks illustrate the gap concretely. A cloned or replayed transaction reuses a prior, valid authorization signal; the human oversight that covered the original transaction does not extend to the replay, because the replay was never independently reviewed. Detecting this pattern requires continuous, real-time behavioral monitoring calibrated to the specific agent and rail, not a periodic review cadence. We recommend the FSB include cloning and replay detection as an illustrative case study under Sound Practice 10, both to make the continuous-validation requirement concrete and to connect it to a threat pattern institutions can readily test for.

Sound Practice 10 also does not yet distinguish between two materially different oversight architectures for agentic systems. In a human-in-the-loop architecture, a human approves each individual agent decision before execution. In a human-on-the-loop architecture, a human monitors agent decisions in aggregate and can intervene, but does not sit in the approval chain for any single transaction.

This distinction carries direct governance consequences on irreversible rails. In a human-on-the-loop architecture, second-line-of-defense oversight cannot review and reverse an individual transaction after the fact, because the rail itself does not permit it. Oversight must therefore shift from post-execution review to pre-execution behavioral control embedded in the model's own operating parameters.

The 2026 Global AI in Financial Services Report — produced by the Cambridge Centre for Alternative Finance in partnership with the BIS, IMF, World Economic Forum, and World Bank, surveying 628 institutions across 151 jurisdictions — finds that AI vendors serving financial institutions are already distributed across this spectrum: 29 percent operate at

human-in-the-loop, 26 percent at human-on-the-loop, and 24 percent fully autonomous with no human in the loop at all. The same report finds agentic AI adoption among regulators at 28 percent, versus 52 percent across surveyed industry respondents — a gap the report itself characterizes as a potential emergent supervisory risk. The oversight-architecture distinction this comment raises describes deployment patterns that already exist at scale, not a future design question.

We recommend the FSB require, for agentic AI deployed under a human-on-the-loop architecture on irreversible rails, a continuously running behavioral assertion regime: the production model scored in real time against a fixed set of labeled baseline cases, with deviations routed to escalation channels independent of standard transaction monitoring. This is the institutional equivalent of a user-acceptance-test script running in permanent parallel with production, and it operationalizes the report's own observation that continuous human review of individual agent decisions becomes impractical as agentic AI use scales. The IMF's April 2026 analysis reaches a compatible conclusion from a different direction, recommending ex-ante interruption and containment mechanisms — commonly termed “kill switches” — as a necessary complement to human-in-the-loop review once a transaction reaches legal finality and can no longer be reversed.

Recommendation: For agentic AI operating on irreversible rails, Sound Practice 10 should require (i) pre-defined intervention conditions and kill-switch mechanisms, and (ii) continuous behavioral assertions for human-on-the-loop oversight architectures, with cloning and replay detection highlighted as a concrete case study.

B. Sound Practice 8 — pre- versus post-settlement explainability

On irreversible rails, explainability has most of its supervisory value before settlement finality; post-hoc explanations cannot remediate consumer harm once a transaction has settled. Sound Practice 8 treats explainability as a continuum to be assessed against materiality and risk, with compensating controls available where explainability is limited. We recommend the final report distinguish explicitly between explainability assessed before settlement finality and explainability assessed after it. The value of explainability for agentic systems operating on irreversible rails is concentrated almost entirely in the pre-settlement window, which argues for materially different compensating controls than those appropriate for reversible retail payment contexts.

The IMF's July 2026 analysis of tokenization reinforces this concern from a market-structure angle rather than an AI-governance one: as execution, clearing, and settlement converge into simultaneous, software-governed events, risk migrates away from institutional balance sheets and toward the platforms and code managing the transaction (Adrian, “Tokenization Can Change the World's Financial Architecture,” IMF Blog, 2 July 2026). This independently corroborates the case for pre-settlement explainability: the institution being supervised is no longer necessarily where the risk sits.

The cloning and replay pattern described under Sound Practice 10 illustrates this same point from the explainability side: detection has value only in the pre-settlement window, since a settled transaction on an irreversible rail cannot be clawed back regardless of how clearly the agent's behavior can later be explained.

Recommendation: Distinguish pre-settlement and post-settlement explainability in Sound Practice 8, include cloning and replay detection as an illustrative case study, and scale explainability expectations based on the irrecoverability of the underlying rail.

C. Risk-tiered application based on rail irrecoverability

We recommend the final report adopt an explicit risk-tiered application of the sound practices based on the irrecoverability of the underlying settlement rail, layered on top of the proportionality principle already articulated in the report. Two AI use cases with identical materiality and autonomy scores under Sound Practice 5 can carry very different consequences depending on whether the underlying rail allows for reversal, netting, or dispute resolution after the fact.

Recommendation: Add rail irrecoverability as an explicit factor — alongside autonomy, materiality, and complexity — in the risk assessment financial institutions perform under Sound Practice 5, with human oversight (Sound Practice 10) and explainability (Sound Practice 8) requirements scaling accordingly.

IV. Conclusion

The FSB's consultation report represents a significant and well-constructed step toward responsible AI governance in financial institutions. Two distinct gaps remain, in our view. The first is external: closing the authorization gap described in Sections I and II will require coordination between internal AI governance frameworks — which this report addresses well — and external consumer payment liability frameworks that were designed before AI agents existed. The second is internal: Sound Practices 8 and 10, as currently framed, do not yet fully account for the shift from point-in-time to continuous validation that adaptive agentic systems require on irreversible settlement rails, as described in Section III.

I respectfully urge the FSB to address both dimensions explicitly in its final report, and to initiate coordination with relevant standard-setting bodies to develop guidance for agentic payment initiation contexts.

I welcome the opportunity to discuss these comments further and would be pleased to share my practitioner paper, “Who Authorized This?”, upon its completion.

V. Addendum: Concurrent Legislative Developments

Since this consultation opened, U.S. domestic legislation has independently surfaced the same definitional questions raised in Section II above. The Digital Asset Market Clarity Act, pending in the U.S. Senate as of this submission, includes a provision exempting certain non-custodial developers from money-transmitter status based on whether they hold the legal right or practical ability to control, initiate, or execute a user's transaction. Law enforcement organizations and other stakeholders have raised concerns that this control-based test is not yet precise enough to prevent gaps in anti-money-laundering oversight; other stakeholders maintain the test is adequately scoped. I take no position on that provision's merits; I raise it solely to illustrate that the terms this comment identifies as under-defined — control, initiation, and delegated authority over a transaction — are, at this moment, the subject of active legislative negotiation with direct consequences for consumer

protection and financial crime oversight. That concurrent, independent debate reinforces the case for the FSB to treat these definitions as a priority rather than a future refinement.

FOVEAL RISK ADVISORY

Clarity in Every Exposure

Shenard Robinson
Founder, Foveal Risk Advisory
New York, NY

July 2026

Financial Stability Board

Re: Supplementary Submission — Correction to Survey Response 1, FSB Public Consultation on Sound Practices for Responsible Adoption of Artificial Intelligence (AI)

Dear Members of the Financial Stability Board,

I am writing to supplement my survey response submitted via the FSB's public consultation web form (Survey response 1; jurisdiction: United States of America; respondent: Shenard Robinson, Foveal Risk Advisory).

Upon review, I found that my answers to Questions 1 and 3 each reference supporting material — a working definition of the "Authorization Gap" and an industry-evidence discussion of card network authorization frameworks — that was inadvertently omitted from the submitted form. Both answers refer to this material as appearing "above," but the form's structure did not include a field for it. I am providing it here so the submission is complete and self-consistent, and would ask that the FSB read the material below as preceding my answer to Question 1 in the original submission.

I apologize for the omission and appreciate the FSB's consideration of this correction ahead of the 22 July deadline.

I. The Authorization Gap

The consultation report correctly identifies that agentic AI systems can take actions with limited human oversight, and that overriding, redressing, or remediating these actions can be difficult or impossible once they occur. That is an accurate diagnosis at the level of institutional AI governance.

By "Authorization Gap," I mean the absence of clear, rail-agnostic rules for who authorized — and who bears liability for — a payment initiated by an AI agent rather than a human.

The report does not yet address the external liability question that follows: when an AI agent initiates a consumer payment, existing rules for whether that payment is authorized, disputable, or reversible were not built for this scenario. Regulation E and card network dispute rules allocate liability for unauthorized transactions based on authorization, initiation, and scope of authority — concepts that presume a human principal acted or clearly delegated a bounded, specific action. Agentic payment

initiation breaks that presumption, and neither practitioner guidance, supervisory frameworks such as SR 11-7, nor this consultation report has yet resolved how.

This is not only a consumer protection issue. Pre-settlement intervention is often the only viable control point on irreversible payment rails, including stablecoin and other real-time settlement infrastructure. Left unresolved, the authorization gap becomes a financial stability issue as agentic payment volume scales. The report's sound practices are strong on internal AI governance; they do not yet bridge to this external liability dimension.

This gap is no longer only a practitioner observation. The IMF's own April 2026 analysis of agentic AI in payments independently names substantially the same condition as a distinct risk category — what it terms the “instruction gap” between structural and transactional authorization: account holders delegating broad mandates while AI agents execute payments without transaction-level instructions (Davidovic and Tourpe, “How Agentic AI Will Reshape Payments,” IMF Note 2026/004, April 2026). The same analysis separately names “ambiguous liability allocation,” “authorization traceability failures,” and “DLT settlement without legal finality” as related risk categories arising from the same underlying condition. I cite this because it corroborates, from an independent and authoritative source, that the gap described in this comment is a recognized risk category, not a hypothetical one.

Recommendation: I respectfully recommend that the FSB recognize the Authorization Gap as a distinct risk category and initiate work with relevant standard-setting bodies to define authorization, initiation, and delegated authority for agent-initiated consumer payments across both traditional and tokenized rails.

Industry Context: The Card Network Trust Perimeter

Before turning to the consultation questions, it is useful to note that the payments industry has not ignored the authorization problem this comment describes — it has substantially solved it for the rails card networks control. Mastercard's Verifiable Intent is a credential-based framework that cryptographically links an authorizing consumer's identity, an agent's delegated instructions, and the resulting merchant transaction into a single tamper-evident, auditable record. On 10 June 2026, Mastercard extended this framework via Agent Pay for Machines to machine-to-machine settlement across cards, bank accounts, and multiple regulated stablecoins, with more than 30 initial partners. Visa has pursued a parallel path through its Trusted Agent Protocol and Intelligent Commerce initiatives, and industry reporting identifies a consistent set of prerequisites the sector has converged on for agent-initiated transactions: verified agent identity, documented consumer consent, captured transaction intent, and end-to-end traceability.

These are serious, well-resourced technical solutions to the authorization problem. Their jurisdiction, however, ends at the edge of the network that built them. Agent Pay for Machines settles stablecoin transactions within Mastercard's credentialed partner network; it does not extend Verifiable Intent to a transaction that originates in a non-custodial wallet, routes through an independent QR-code merchant flow, and confirms on public blockchain infrastructure outside that network. The authorization gap this comment describes begins exactly where the card networks' jurisdiction ends.

A parallel gap is forming at the regulatory perimeter around stablecoin issuance itself. The Office of the Comptroller of the Currency's June 2026 proposed rulemaking would require Permitted Payment Stablecoin Issuers to comply with the Bank Secrecy Act and related GENIUS Act provisions — but this rule, like the card network frameworks, governs a single point in the chain: the issuer. It does not reach the secondary-market transaction layer where an agent-initiated payment is actually authorized, routed,

and settled. Industry commenters, including the American Bankers Association, have raised this same secondary-market gap in response to the OCC proposal. I raise it here because it is structurally the same authorization gap this comment addresses, appearing at a different layer of the same infrastructure.

This concludes the supplementary material. Questions 1 and 3 of my original submission should be read together with the above.

Respectfully submitted,

Shenard Robinson
Founder, Foveal Risk Advisory

shenardrobinson@hotmail.com

fovealrisk.com

July 2026