FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Chipo P (independent response)

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 FSB's identification of AI-related risks is comprehensive across most categories — third-party dependencies, model risk, data quality, cyber risk and market correlations. The 12 sound practices address these risks through governance, lifecycle management, explainability, human oversight and third-party risk management.

One substantive risk category is inadequately addressed: the verification architecture gap. The consultation report identifies explainability (Sound Practice 8) and human oversight (Sound Practice 10) as the primary mechanisms for managing AI output quality. Both are administrative responses — they rely on human judgment applied after AI outputs are produced. Neither addresses the risk that AI outputs enter official records, regulatory submissions, compliance frameworks and settlement systems without independent architectural verification at the point of production.

The Microsoft DELEGATE-52 benchmark (April 2026) confirmed the precise mechanism of this risk. Testing 19 AI models across 20 consecutive document editing tasks — the kind of workflow financial institutions use daily for compliance documentation, regulatory submissions, board papers and credit assessments — frontier models corrupted an average of 25% of document content. For average models the corruption rate was 50%. The failures do not accumulate gradually. Models maintain near-perfect accuracy and then fail suddenly and catastrophically. The errors look correct until they do not.

Across South Africa, the UK and the United States in 2026, confirmed AI hallucination incidents have been discovered in legal opinions, government policy documents, audit reports, regulatory submissions and court filings — without exception, discovered by external parties rather than by the institutions that produced them. A researcher at HEC Paris tracking court sanctions for AI-generated legal errors in the United States recorded 10 cases from 10 different courts in a single day. The Oregon federal court imposed a potential record sanction of $109,700 on a lawyer for AI-generated errors. Pinsent Masons — a global top-100 law firm — self-referred to the UK Solicitors Regulation Authority after submitting AI-hallucinated legal authorities in court twice. The correction letter was also hallucinated.

The missing risk is this: financial institutions implementing the FSB's 12 sound practices through AI-assisted governance workflows are creating circular verification systems — AI producing governance documentation, AI reviewing governance documentation. The DELEGATE-52 degradation pattern applies to the verification layer precisely as it applies to the original output. The errors embedded in the governance documentation are verified by the same system that introduced them.

Recommendation: The FSB should add a thirteenth risk category — verification architecture risk — defined as the risk that AI outputs enter official records, regulatory submissions and compliance systems without architectural verification independent of the AI system that produced them. This risk is distinct from model risk (which concerns the accuracy of AI predictions) and data quality risk (which concerns the inputs to AI systems). It concerns the integrity of AI outputs at the point they become institutional facts.

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

The 12 sound practices are comprehensive as a framework for institutional AI governance. Sound Practices 1 through 4 address organisational governance effectively. Sound Practices 5 through 10 address the AI lifecycle with appropriate proportionality. Sound Practices 11 and 12 address cyber and third-party risk. The framework is coherent, proportionate and practically applicable.

Two gaps require attention.

Gap 1 — The Administrative Versus Architectural Distinction

Every sound practice in the consultation report is an administrative response to AI risk — a policy, a process, a governance structure, a human oversight mechanism. All 12 sound practices assume that the institution itself is the verification layer for its own AI outputs. Sound Practice 8 (explainability) asks institutions to understand AI outputs. Sound Practice 10 (human oversight) asks institutions to review AI outputs. Sound Practice 9 (performance management) asks institutions to test AI outputs. In every case the institution that produced the AI output is also the institution responsible for verifying it.

This creates the verification circularity problem. A financial institution that produces AI-assisted compliance documentation and reviews it through AI-assisted audit processes has not created an independent verification layer. It has created a more sophisticated version of the same closed loop. The DELEGATE-52 benchmark confirms that the degradation pattern applies across the full workflow — not just to individual outputs but to the cumulative chain of production, review, amendment and finalisation.

The architectural alternative is verification embedded in the infrastructure beneath the institution rather than administered by the institution itself. Deterministic verification at genesis — where AI outputs are verified against compliance and admissibility conditions before they cross the execution boundary — produces a verification layer that is independent of the AI system that generated the original output. The verification is not an institutional process. It is a structural condition of the output entering the institutional record.

Recommendation: Sound Practice 10 (human oversight) should be expanded to distinguish between administrative human oversight — human review of AI outputs — and architectural verification — deterministic verification of AI outputs at genesis before they enter institutional records. For AI use cases in high-materiality compliance, regulatory reporting and settlement functions, architectural verification should be identified as a sound practice that complements rather than replaces human oversight.

Gap 2 — The Agentic AI Execution Boundary

Sound Practice 2 (governance and accountability) and Sound Practice 10 (human oversight) acknowledge agentic AI as an emerging risk. The consultation report notes that agentic AI systems are capable of planning, reasoning and executing complex high-level goals independently. The FSB identifies this as a risk requiring proportionate human oversight.

The consultation report does not address the most commercially significant vulnerability of agentic AI systems deployed in financial services: the authorisation boundary between the AI agent and the production systems it can execute against. The Meta AI customer support agent breach in June 2026 confirmed the vulnerability precisely. Attackers did not hack the model or the compute infrastructure. They used social engineering — they talked to the AI agent and it complied — transferring over 100 accounts in 48 hours. The vulnerability was not in the model. It was in the absence of a cryptographic deterministic authorisation layer between the AI agent and the execution boundary.

A conversational authorisation layer can be socially engineered. A cryptographic deterministic authorisation layer cannot. The AI agent either meets the admissibility conditions cryptographically or it does not execute — regardless of what it was told. This is not a model training problem. It is an architecture problem. And the FSB's 12 sound practices do not currently address it.

Singapore’s Model AI Governance Framework for Agentic AI — launched by the Infocomm Media Development Authority in January 2026 and updated in May 2026 following feedback from over 60 organisations — is the most sophisticated agentic AI governance framework currently available globally, and it confirms precisely where the gap remains.

The framework structures its guidance across four dimensions: assessing and bounding risks upfront, making humans meaningfully accountable, implementing technical controls and processes, and enabling end-user responsibility. Even this framework — widely regarded as the global benchmark for agentic AI governance — stops at the level of principle and process. Making humans meaningfully accountable is an administrative standard. It asks institutions to design oversight checkpoints and accountability structures around the agent’s actions. It does not require a cryptographic execution boundary that determines, independently of human review, whether an agent’s action is admissible before it executes.

The May 2026 update expanded guidance for bounding agent autonomy in financial services workflows specifically — including source-of-wealth analysis, a function directly comparable to AML and compliance workflows in the FSB’s scope — but the controls described remain checkpoint-based and process-based rather than architectural.

If the world's most advanced agentic AI governance framework, developed with input from over 60 organisations across a full consultation cycle, does not mandate a cryptographic execution boundary, this confirms that the gap identified here is not a deficiency specific to the FSB's consultation report. It is a gap in the global state of the art that the FSB has the standing to close first.

Recommendation: The FSB should add a specific sound practice addressing AI agent action authorisation — defining the requirement for a cryptographic execution boundary between AI agents and production systems in high-materiality financial institution functions. This practice should be distinct from Sound Practice 10 (human oversight) because agentic AI systems by design reduce the human transmission step — the execution boundary cannot depend on human review of every action.

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 consultation report strikes an appropriate balance for most categories of AI risk. Sound Practices 1 through 4 are technology-neutral. Sound Practices 5 through 12 address both traditional AI and GenAI proportionately.

The balance is insufficient for two specific emerging AI risk categories.

First — the document degradation risk specific to GenAI in professional workflows. The DELEGATE-52 benchmark is GenAI-specific. The degradation pattern — sudden catastrophic failure after near-perfect accuracy across 20 consecutive interactions — is a property of the probabilistic output architecture of large language models. Traditional ML models do not exhibit this failure pattern in the same way. The consultation report does not address this specific GenAI risk in the context of professional document production and compliance workflows — which is precisely the category of AI use where financial institutions face the highest regulatory and legal exposure from undetected errors.

Second — the authorisation boundary risk specific to agentic AI. As noted above, the Meta breach confirmed a vulnerability that is qualitatively different from the risks addressed by Sound Practices 1 through 12. The vulnerability is not a governance failure. It is an absence of architectural authorisation control at the execution boundary. The FSB's current framework assumes a human transmission step between AI output and institutional action. Agentic AI removes that step. The framework needs a specific sound practice that addresses the execution boundary in agentic workflows rather than adapting human oversight principles to a context where human oversight is structurally absent.

Recommendation: The FSB should add specific sound practices addressing: (a) GenAI document integrity verification — architectural verification of AI-produced compliance documentation before it enters institutional records; and (b) agentic AI execution boundary controls — cryptographic authorisation requirements for AI agents operating against production financial systems.

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

The 12 sound practices are appropriately flexible for most current AI technologies. The proportionality principle — more robust practices for large, complex, highly connected institutions — is the correct design approach. The practices are sufficiently technology-neutral to accommodate most AI developments that can be anticipated.

One structural flexibility gap exists.

The consultation report's framework is designed around AI systems that produce outputs reviewed by humans before institutional action. The framework assumes that human oversight (Sound Practice 10) is the final verification step before an AI output becomes an institutional fact. This assumption breaks down for two categories of AI that are already commercially deployed: agentic AI systems that execute without a human transmission step, and AI-assisted compliance documentation workflows where the DELEGATE-52 degradation pattern means that human review at the end of a long workflow does not catch errors introduced mid-chain.

The framework needs a structural provision for AI uses where the human oversight model is insufficient — not because human oversight is not applied but because the failure mode is not visible to human review at the final step. This requires a different category of sound practice: architectural verification rather than human oversight.

Recommendation: The FSB should explicitly recognise architectural verification as a distinct category of sound practice — separate from and complementary to human oversight — for AI use cases where the output chain length, the execution autonomy or the production volume makes end-point human oversight an insufficient verification mechanism.

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 in the consultation report are drawn predominantly from internationally active banks, large asset managers and sophisticated trading firms in developed market jurisdictions. This reflects the FSB's membership and the availability of documented AI implementation practices.

A substantive gap exists in the case study coverage: the emerging market dimension of AI governance. Financial institutions in African markets face a specific set of AI governance challenges that differ materially from those of developed market institutions.

First — legacy infrastructure vulnerability. African financial institutions operate predominantly on legacy systems — COBOL-based banking infrastructure, middleware connecting disparate systems, limited real-time data integration. AI applied to document workflows on this infrastructure carries a higher baseline degradation risk than the DELEGATE-52 benchmark suggests for frontier model environments. The errors that look correct until they do not are harder to detect on legacy infrastructure without the monitoring and validation tools that large internationally active banks deploy.

Second — regulatory contestation. South Africa's current legal environment — where the Gauteng Division held in Mangundhla v SARB (1 June 2026) that Bitcoin is both money and capital under the Exchange Control Regulations, directly contradicting Standard Bank v SARB (May 2025) — means that AI-assisted legal and compliance documentation produced in the current period carries specific uncertainty about which legal framework applies. AI systems trained on pre-Mangundhla data will produce outputs that reflect the Standard Bank framework. The compliance exposure from that mismatch is not visible without architectural verification of AI outputs against the current regulatory framework at the point of production.

Third — sovereign compute dependency. Most African financial institutions process AI workloads on infrastructure owned by US-headquartered technology companies — Microsoft Azure, Amazon Web Services, Google Cloud. The EU's Cloud and AI Development Act (June 2026) has confirmed that data processed on infrastructure subject to foreign court orders does not meet sovereign compute requirements. African financial institutions implementing the FSB's 12 sound practices through foreign cloud AI infrastructure may satisfy the administrative requirements while creating structural data sovereignty exposures that the sound practices do not currently address.

Fourth — market structure concentration risk. Several African jurisdictions with the highest digital asset adoption rates globally and substantial domestic technology talent pools have regulatory architectures — including significant risk capital thresholds for virtual asset service provider registration — that create market structures where established fintech incumbents with existing regulatory relationships can enter the digital asset and AI-enabled financial services space while smaller domestic companies building sovereign alternatives cannot. The result is that the AI and digital asset infrastructure beneath some of the world’s fastest-growing crypto markets is being built predominantly by capitalised incumbents — including entities settling on foreign dollar-denominated stablecoin infrastructure — rather than by domestic companies building architecturally sovereign alternatives.

This pattern recurs across the nine African jurisdictions that have enacted formal digital asset regulatory frameworks: South Africa, Nigeria, Kenya, Mauritius, Zimbabwe, Ghana, Morocco, Rwanda and Botswana. The combination of capital thresholds, licensing costs and compliance requirements concentrates AI and digital asset infrastructure deployment in incumbent institutions with foreign technology partnerships. The FSB’s 12 sound practices are designed for institutions that already have the governance infrastructure to implement them. They do not address the market structure dimension — the mechanism by which regulatory capital requirements ensure that AI deployment in emerging market financial services is concentrated in the institutions least likely to build sovereign verification architecture and most likely to deploy agentic AI on foreign infrastructure without an architectural execution boundary.

A live illustration of this tension has emerged in one major African jurisdiction in 2026, where a domestic data localisation directive requiring payment data to remain within national jurisdiction was issued in the same period that a leading incumbent payment platform announced a strategic partnership anchoring its settlement layer to a foreign dollar-denominated stablecoin running on foreign-controlled ledger infrastructure. The two regulatory and commercial positions sit in direct tension — a tension produced by a market

structure that concentrates licensing in incumbents with foreign technology partnerships before data sovereignty mandates are issued.

Recommendation: The FSB should include at least one African or emerging market case study in the final report — ideally addressing the specific challenges of AI governance in legacy infrastructure environments, contested regulatory frameworks and sovereign compute constraints. The Sovereign Rails Curriculum's work on pre-execution compliance architecture for South African digital asset settlement is available as a case study demonstrating how architectural AI verification operates in an emerging market regulatory context.

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

The case studies provide valuable actionable insights for large internationally active financial institutions. The machine learning credit risk case study, the agentic AI fraud detection case study and the relationship management AI platform case study are commercially specific and technically grounded.

The case studies are less actionable for the category of financial institutions that faces the most immediate AI governance risk: mid-sized and smaller financial institutions in emerging markets that are adopting GenAI tools for compliance documentation, regulatory reporting and board paper production without the institutional AI governance infrastructure that the case studies describe.

These institutions — CASPs, regional banks, non-bank financial institutions and fintech payment providers — are producing AI-assisted compliance documentation at scale without dedicated AI governance teams, without model risk management frameworks and without the monitoring infrastructure to detect DELEGATE-52-type degradation. They are the institutions for which the verification architecture gap is most acute and the consequences of undetected errors are most commercially damaging.

Recommendation: The FSB should include a case study specifically addressing the responsible AI adoption challenges of mid-sized and smaller financial institutions — particularly in emerging market contexts — where administrative AI governance infrastructure is limited and architectural verification solutions are the more practical path to responsible AI adoption.

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

The glossary definitions are generally clear and aligned with industry practice. Two additions or clarifications would strengthen the framework.

First — verification architecture should be added to the glossary as a distinct concept from explainability and human oversight. Verification architecture refers to the infrastructure that independently verifies AI outputs at genesis — before they enter institutional records — as distinct from the institutional processes (human oversight, explainability tools, performance management) that review AI outputs after production. The absence of this concept from the

glossary reflects its absence from the 12 sound practices and creates a gap in the FSB's conceptual framework for AI governance.

Second — the definition of agentic AI should explicitly address the execution boundary problem. The current definition — autonomous systems capable of planning, reasoning and executing complex high-level goals independently — does not capture the governance significance of the absence of a human transmission step between AI decision and institutional action. The definition should clarify that agentic AI systems require authorisation architecture at the execution boundary that is qualitatively different from the human oversight requirements for non-agentic AI systems.

RECENT DEVELOPMENTS CONFIRMING THE SUBMISSION'S ARGUMENT

The following developments, which occurred in the weeks leading up to this submission's filing, each independently confirm a dimension of the submission's central argument.

1. The KPMG Authorship Incident — The Feedback Loop Confirmed (June 2026)

KPMG retracted its report titled Redefining Excellence in the Age of Agentic AI in June 2026 after independent investigation revealed that 40 of 45 citations were fabricated. The report's author confirmed that no human at KPMG double-checked the citations, the claims or the sources before publication. UBS, the NHS, Swiss Federal Railways and Transport for London all confirmed that the claims about their AI usage were false. This makes KPMG the fourth Big Four professional services firm caught in the same pattern in months, following EY's retracted cybersecurity report (16 of 27 fabricated citations), Deloitte's refunded government contract and Pinsent Masons' double self-referral to the SRA.

The KPMG incident confirms the submission's most urgent observation. KPMG's hallucinated citations were recycled by industry publications and mainstream media globally before the retraction. Those fabricated facts have now entered the training data of other AI systems. The next AI-generated report that draws on that training data will cite KPMG's hallucinations as sources — confidently, without knowing they were fabricated. This is the verification architecture gap stated at the systemic level. Errors feeding models. Models producing errors. Errors feeding the next generation of models. Without a verification architecture that seals outputs at genesis the loop has no break point. This is precisely the financial stability risk the FSB's framework must address — not as a technology governance question but as a systemic infrastructure question.

2. The EU AI Act Amendment — The Agentic AI Gap Deferred Not Resolved (16 June 2026)

The European Parliament voted — 16 June 2026 — to amend the EU AI Act, passing 423 to 57 in favour of the Omnibus VII simplification package. The amendment defers the compliance deadline for high-risk AI systems listed in Annex III from 2 August 2026 to 2 December 2027 — a 16-month deferral. The stated rationale is the absence of finalised harmonised standards, compliance tools and designated national authorities. The amendment does not address the agentic AI execution boundary gap this submission identifies. It defers the deadline for the administrative compliance framework while leaving the architectural verification gap unresolved. The FSB's consultation closes on 22 July 2026 — six days before the original EU AI Act high-risk systems deadline. The EU has confirmed

that the administrative governance framework is not ready. The architectural verification framework has not been proposed. The FSB's final report is the right moment to name the architectural gap the EU amendment has deferred rather than closed.

3. The Anthropic Export Ban — Frontier AI Weaponised at Government Level (13 June 2026)

On 13 June 2026 the US government issued an export control directive requiring Anthropic to suspend all access to its Fable 5 and Mythos 5 models by every foreign national simultaneously — including Anthropic's own foreign national employees. The stated justification was a narrow jailbreak of Mythos 5's cybersecurity capabilities. Anthropic received the order at 5:21pm ET and disabled the models globally to comply. Mythos 5 had been commercially deployed for approximately three days before the ban. Canadian Prime Minister Mark Carney — former Governor of the Bank of Canada during the 2008 financial crisis and former Governor of the Bank of England — framed the export ban as a systemic warning about concentrated AI infrastructure dependency creating financial stability risks comparable to the systemic linkages that produced the 2008 crisis. He called for redundancy and diversity in AI infrastructure as the post-Lehman lesson applied to frontier model dependency. On 15 June 2026 a coalition of 54 CISOs and cybersecurity practitioners signed an open letter to the US Secretary of Commerce arguing that the ban disadvantages defenders without disadvantaging attackers — the own goal framing.

The Anthropic export ban confirms the submission's third-party dependency risk argument at the most acute level. Financial institutions that have integrated frontier AI models into compliance workflows, credit decisions and payment authorisations on US-headquartered infrastructure are exposed to a single point of failure that can be triggered by a foreign government's decision regardless of the institution's own governance posture. Sound Practice 12 (third-party dependency) addresses vendor concentration risk. It does not address jurisdictional dependency risk — the risk that the jurisdiction controlling the AI infrastructure can remove it from service globally without notice.

3a. The Mythos Red-Team Result — The Speed Problem Made Concrete (11 June 2026)

A development reported recently, attributed to testimony given to the Senate Intelligence Committee and first reported by The Economist, illustrates the underlying capability problem this submission is concerned with in starker terms than the export ban alone conveys.

Senator Mark Warner, the Committee's vice-chair, stated that General Joshua Rudd, who leads the US National Security Agency and US Cyber Command, told him that in a red-team exercise conducted on 11 June 2026, Anthropic's Mythos model breached almost all of the NSA's classified systems, not over weeks, but within hours. The exercise was conducted by the agency against its own infrastructure, not an unauthorised intrusion, which makes the result more significant rather than less: the most hardened, best-resourced signals intelligence infrastructure in the world, when deliberately tested against a frontier model, was found to be comprehensively penetrable on a timescale that no human red team could approach.

This is independently sourced, multiply corroborated reporting, though it remains second-hand testimony relayed by a senator rather than a published government assessment, and should be treated with the corresponding degree of caution appropriate to that source.

What it demonstrates for the purposes of this submission is not a single vulnerability but a capability gap: the speed at which a sufficiently capable model can identify and exploit weaknesses in even the most sophisticated defensive posture has now moved from a theoretical concern to a demonstrated outcome inside the institution best equipped to resist it. A verification and oversight framework calibrated to the pace of human-led security review is not calibrated to this. The FSB's sound practices, framed around governance processes that assume a human-paced threat and response cycle, do not yet account for a world in which the offensive and defensive capability of a single model can outpace institutional review by orders of magnitude within a single business day.

3b. A Regulatory Precedent for Contemporaneous Evidence Over Reconstructed Accounts (24 June 2026)

A separate and directly relevant development confirms that a major prudential regulator has already moved, in a different but closely related domain, toward precisely the standard this submission recommends for AI governance.

David Chaplin, Head of Enforcement and Litigation at the Bank of England, gave a speech on 24 June 2026 at the Fifth Conference on Financial Law and Regulation describing what he termed a sea change in PRA and Bank of England enforcement practice. Under the Early Account Scheme, introduced as part of a 2024 enforcement policy update, firms and individuals that produce a comprehensive, accurate, contemporaneous account of what occurred, supported by senior management attestation as to its accuracy and completeness, and that make admissions well in advance of settlement, are treated materially differently to those that adopt a defensive posture and attempt to reconstruct an account only once an investigation is already under way. Chaplin observed that this is not a policy limited to isolated cases but a sustained shift with the potential to become the new default expectation in enforcement.

The same distinction this submission draws between administrative and architectural verification applies directly to what Chaplin describes. A firm relying on a reconstructed account, assembled after the fact from whatever documentation happened to survive, is in a structurally weaker position than a firm with a genuine, contemporaneous record of what was known, by whom, and when. The Bank of England's own enforcement practice now formally rewards the latter and disadvantages the former.

Applied to AI governance specifically, an institution that has built an architectural verification layer — one that produces a tamper-proof, contemporaneous record of every AI-assisted decision at the moment it is made, independent of the system that produced the output — is the institution best placed to produce exactly the kind of account the Bank of England now formally rewards. An institution relying only on administrative controls, where a record exists only if a human remembered to create one, will be attempting to reconstruct under enforcement pressure something the first institution already has. Early candour, in Chaplin's own words, requires something to be candid about. This submission's recommendation that the FSB name an architectural standard for AI accountability is the precondition for

institutions to be able to meet the standard of regulatory engagement the Bank of England is now actively rewarding in a closely adjacent domain.

4. The American Banker Kill Switch Finding — 72% of Banks Lack Administrative Controls (10 June 2026)

American Banker published on 10 June 2026 that 72% of banks lack AI model kill switches — more than a third of US bankers cannot say with confidence that they have the mechanism to stop a malfunctioning AI model even when they know it is producing errors. A separate report confirmed that an employee at a US bank uploaded customer data to an unauthorised AI application. The bank contacted the vendor before the model trained on the data — the breach was caught after the fact not prevented before it.

Both findings confirm the submission's distinction between administrative intervention and architectural prevention. A kill switch is an administrative intervention after a malfunctioning AI model has already produced outputs. An architectural verification layer prevents non-compliant outputs from crossing the execution boundary before they enter the institutional record. 72% of banks lack the administrative version. The architectural version has not been named as a sound practice in the FSB's framework.

5. The Central Bank of Nigeria Data Localisation Directive — Sovereign Compute as Regulatory Objective (15 June 2026)

The Central Bank of Nigeria issued a directive on 15 June 2026 requiring all banks, fintechs and payment service providers to store all payment transaction data on local servers by January 1 2027. The CBN explicitly named data sovereignty and reducing reliance on offshore infrastructure as the stated objectives.

This is the first African central bank to mandate data sovereignty at the payment infrastructure layer as a regulatory requirement rather than a governance aspiration. The CBN's directive confirms that the jurisdictional dependency risk this submission identifies is being recognised and acted upon at the regulatory level across emerging market jurisdictions. The FSB's final report is addressed to G20 Finance Ministers and Central Bank Governors. The CBN's directive confirms that emerging market central banks are already moving to address the compute sovereignty gap through administrative mandates. The FSB's framework should name the architectural solution — software-defined sovereign compute that satisfies data sovereignty requirements without the capital expenditure and security standards uncertainty of physical data centre migration — as a sound practice that complements administrative data localisation requirements.

South Africa's own central bank confirms the same compute sovereignty and AI governance gaps are live regulatory concerns, on a parallel track. The South African Reserve Bank's National Payment System Regulatory and Oversight Report for the 2025/26 period confirms that the SARB finalised a consultation paper on cloud computing and data offshoring in the national payment system in March 2025, and is currently finalising its consideration of public comments ahead of issuing a directive prescribing requirements for cloud computing adoption and data offshoring by payment system financial market infrastructures and payment institutions. The same report confirms that the South African Financial Sector Conduct Authority and Prudential Authority jointly published a report on AI adoption across

the financial sector in November 2025, and that the SARB's Fintech Division is now collaborating with both authorities on a discussion paper examining policy options for AI in financial services, expected in 2026.

These two confirmed, dated regulatory workstreams confirm that the same underlying concern — that compute sovereignty and AI governance remain administrative matters today, addressed through consultation and discussion papers rather than architectural verification requirements — is live and being actively worked through independently of this submission, by a G20 member's central bank operating under one of the more developed payment system regulatory frameworks among emerging markets, on substantially the same timeline as the FSB's own consultation. This corroborates, without depending on, the submission's recommendation that the FSB name an architectural standard the industry can move toward.

A more direct confirmation followed in the SARB's June 2026 Financial Stability Review, published after the developments described above. The Review names a specific frontier AI model — Anthropic's Claude Mythos Preview — directly, alongside an active Middle East conflict, as a contributing financial stability risk, stating explicitly that "cyber risk has shifted from episodic and largely manageable events to continuous and compounding." This formulation is significant for what it confirms about the nature of the risk, independent of how the two named risk factors are weighted against each other.

A central bank describing cyber risk as having changed in fundamental character, from episodic events a governance framework can respond to individually, to a continuous and compounding condition, is describing precisely the mismatch this submission identifies between human-paced administrative review and machine-paced AI capability.

The International Monetary Fund reached a closely related conclusion in May 2026, reframing the same model as a systemic risk to the financial system generally, rather than an operational matter for individual firms, citing independent evaluations in which the model identified vulnerabilities at a rate and completeness exceeding prior models and, in one assessment, completed a full simulated network compromise that no prior model had completed in full. A South African Reserve Bank deputy governor has since stated publicly that the SARB will not panic over this risk but will assess and adopt mitigating strategies — an appropriately measured institutional response, and one this submission's recommended architectural standard is intended to support, by giving institutions a concrete standard to assess against rather than leaving the mitigation strategy undefined.

These developments — the KPMG feedback loop, the EU AI Act deferral, the Anthropic export ban, the Mythos red-team result, the American Banker kill switch finding, the CBN data sovereignty directive, and the Bank of England's shift toward rewarding contemporaneous evidence over reconstructed accounts — all confirm the submission's central argument from independent directions simultaneously. The verification architecture gap is not a theoretical risk. It is a present condition producing systemic consequences across the global financial system, and a standard regulators in closely adjacent domains are already beginning to reward. The FSB's final report must name the architectural response.

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

The FSB's 12 sound practices represent the most comprehensive international framework for institutional AI governance in the financial sector produced to date. They are proportionate, technology-neutral and practically applicable. This submission does not argue that they are wrong. It argues that they are incomplete in a specific and important way.

The 12 sound practices address the administrative layer of AI governance — what institutions must do to adopt AI responsibly. They do not address the architectural layer — what the infrastructure beneath institutions must be designed to ensure that AI outputs are verifiable at genesis before they become institutional facts.

The distinction between administrative and architectural governance is not theoretical. It is the distinction between institutions that audit AI outputs after the fact and institutions that embed verification as a structural condition of AI outputs entering the institutional record.

The at least five South African AI hallucination incidents confirmed in 2026, the US epidemic of court-sanctioned AI-generated legal errors and the Pinsent Masons double hallucination incident all share one precise feature: the errors entered official records without being caught by the institution that produced them.

Administrative AI governance was in place. Architectural verification was not.

The FSB's final report will be delivered to G20 Finance Ministers and Central Bank Governors in October 2026. It will shape national and regional regulatory frameworks for responsible AI adoption across the global financial system for the next decade. The recommendation this submission makes most urgently is this: the final report should explicitly recognise architectural verification as a distinct and necessary complement to the 12 administrative sound practices — and should provide guidance on what architectural verification looks like for different categories of AI use in financial services.

The FSB was established in response to the 2008 financial crisis — a crisis caused in part by the absence of verification architecture beneath the administrative compliance processes that financial institutions had in place. The AI governance challenge of 2026 has a structural parallel: institutions are adopting AI at scale with administrative governance frameworks that assume the verification layer is adequate. The evidence from 2026 confirms it is not. The architectural response is available. The FSB's final report is the right moment to name it.