FSB FINANCIAL STABILITY BOARD

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

Response to Consultation

Certavixx Technologies Ltd.

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?**

We broadly agree with the benefits and risks described in the consultation report, and we particularly welcome its treatment of agentic AI, third-party dependency, and the limits of human oversight at scale. We would, however, draw attention to three substantive risks that in our view are under-weighted, and one benefit not fully captured.

First, demonstrability risk. The report catalogues the risks of AI systems themselves but gives less attention to the risk that an institution cannot prove its AI estate is under control. Governance frameworks that exist as policy documents, committee minutes and periodic reports may satisfy internal processes while remaining unexaminable in substance: a supervisor cannot readily test whether stated controls operated at the point a consequential decision was taken. The lesson of earlier governance regimes is that documentation and assurance can diverge - compliance can succeed on paper while failing in operation. We regard the absence of a standard, time-stamped, traceable evidence record for material AI systems as a risk in its own right, because it creates a supervisory blind spot precisely where adoption is fastest.

Second, taxonomy-gap risk. With generative and agentic AI placed outside the scope of revised US model-risk guidance, and analogous perimeter questions arising in other jurisdictions, material agentic systems risk falling between established risk taxonomies - too complex for conventional model inventories yet owned by no single alternative discipline. Fragmented ownership across model risk, operational risk, cyber, data, conduct and third-party functions is itself a source of unmanaged risk: each function sees a component; no function is accountable for the delegated authority of the whole system. Our accompanying paper proposes treating the Delegated Decision-and-Action System, rather than the underlying model, as the unit of assurance for this reason.

Third, oversight-decay risk. The report rightly recognizes that continuous human monitoring becomes impractical at scale. We would add the failure mode that precedes that point: human oversight that formally exists but has degraded into routine approval through automation bias. Where AI is used to monitor AI, a further risk arises if the evaluating mechanism shares models, data or dependencies with the systems it evaluates - effective

challenge is quietly dissolved unless the assurance layer is structurally independent and itself governed.

One benefit not fully captured: responsibly adopted AI can strengthen risk management and supervision itself. Continuous, evidence-led assurance of AI estates - posture that is computed from current system state rather than narrated retrospectively - could give both institutions and supervisors materially better visibility than periodic document-based review, converting a compliance burden into a supervisory capability.

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

The twelve sound practices are comprehensive as a statement of what good adoption involves, and we support their lifecycle structure, their treatment of agentic autonomy, and the deliberate decision not to impose a prescriptive approach. In our view, however, three additions would materially improve their ability to enable responsible adoption in practice, which is the standard the question sets.

First, the practices need a defined object to attach to. As drafted, they apply to "AI use cases" - a term institutions will interpret with wide variance. For material systems that plan, invoke tools, retain state and act, the relevant object is not the model or the use case narrowly conceived, but the whole delegated decision-and-action system: model, prompts, orchestration, memory, tools, permissions, data, human decision points and downstream connectivity. Without a common definition of the unit to which governance attaches, two institutions applying the same twelve practices may govern materially different boundaries, and supervisors will be unable to compare postures. A technology-neutral definition of the assurable system, with standard variables for classifying it - materiality, autonomy, reversibility, connectivity, and velocity of propagation - would make proportionality operational rather than rhetorical.

Second, the practices describe controls but not the assurance of controls. Each practice states what institutions should do; none states how an institution, its board, or its supervisor should be able to verify that the practice operated - at the point of decision, after material change, and under challenge. We would encourage the final report to pair the practices with an expectation of demonstrability: a structured pre-deployment assurance case for high-materiality systems, event-triggered revalidation when defined changes occur, and a current, traceable evidence record linking each practice to the artefacts that prove its operation. Sound practices without an assurance layer risk reproducing a familiar failure mode: frameworks that are adopted in policy and absent in operation.

Third, the menu format needs a floor. Presenting twelve practices as a menu invites aggregate thinking, in which strength in some practices is read as compensating for weakness in others. Certain controls should be identified as non-compensable: unrestricted access to customer funds, unbounded tool use, the absence of an immutable audit trail, or an untested emergency stop cannot be offset by maturity elsewhere. Naming a small set of red lines would sharpen clarity considerably, particularly for boards approving high-autonomy deployments.

With these additions - a defined unit of assurance, an explicit demonstrability expectation, and a non-compensable floor - we believe the sound practices would move from a well-constructed description of responsible adoption to an instrument that enables and evidences it.

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?**

We believe the report strikes a reasonable balance across the forms of AI it addresses - and we would respectfully suggest that the balance question itself is framed on the wrong axis. The material risk gradient in financial institutions does not run between "traditional AI" and "GenAI and agentic AI" as technology categories. It runs along the authority a system is permitted to exercise and the consequences of its failure.

Consider a single foundation model deployed three ways: drafting internal research summaries; recommending retail credit decisions to a human underwriter; and executing payments within treasury mandates. The technology label is identical in all three cases. The risk profiles - and the appropriate control intensity - differ radically. A balance calibrated to technology form will simultaneously over-govern low-authority uses of new technology and under-govern high-authority uses of familiar technology. A balance calibrated to delegated authority, potential harm, reversibility, connectivity, and speed of propagation governs each deployment in proportion to what it can actually do.

Read through that lens, the report's agentic-specific controls - individual agent identifiers, prohibited actions, approval checkpoints, dual authorisation for financial transactions, audit trails, and kill switches - are among its strongest contributions, and we support them. But we would encourage the final report to present them as controls triggered by authority tier, not by the "agentic" label. Two consequences follow. First, the controls extend naturally to any system that acquires action authority, whatever it is called by its vendor or whatever technology succeeds today's agents. Second, institutions cannot avoid the controls by classification arbitrage - describing an action-taking system as a "workflow tool" or "copilot" to escape the agentic category.

This reframing also resolves a tension the current balance leaves open. Practices addressed to "all forms of AI" and practices addressed to "emerging forms" will inevitably drift apart as the technology evolves; a taxonomy of authority does not drift, because it is anchored in what institutions permit systems to do rather than in how the systems are built.

Finally, on the report's recognition that human oversight of agents has limits and that AI may monitor AI: we agree, with one condition that the balance should make explicit. The monitoring layer must be structurally independent of the systems it monitors - separate accountability, no shared incentives, and governed evaluation mechanisms - or the industry will have automated away effective challenge at precisely the authority tier where it matters most.

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

Largely yes at the level of principle - the practices are outcome-oriented, proportionate, and deliberately non-prescriptive, which are the right foundations for durability. But flexibility over time is not only a drafting property; it is an architectural one, and we would highlight three mechanisms that would determine whether these practices age well or require revisiting within a few years.

First, anchor durability in authority rather than technology vocabulary. For the reasons set out in our response to Question 3, practices keyed to today's technology categories - "GenAI," "agentic AI" — inherit the shelf life of those labels. The instructive precedent is model-risk guidance itself: SR 11-7 endured for fifteen years precisely because it was principles-based, yet its 2026 successor still found it necessary to draw a scope boundary using the technology labels of the moment - a boundary that will itself require reinterpretation as capabilities evolve. Practices anchored in what a system is permitted to do - decide, act, transact, self-modify - need no such reinterpretation, because new technologies enter the framework by acquiring authority, not by acquiring a name. Multi-agent networks, self-improving systems, and whatever follows them will all arrive as changes in delegated authority.

Second, make change itself the trigger for reassessment. Calendar-based review cycles are the least flexible mechanism available in a fast-moving field: they are simultaneously too slow for a foundation-model version change that alters system behavior overnight, and wastefully frequent for stable deployments. A flexible framework defines events that invalidate or qualify a prior assurance conclusion - model or vendor changes, new tools or permissions, purpose or population shifts, behavioral anomalies, incidents at common providers - and requires reassessment when they occur. This makes the framework self-updating: it responds to the actual rate of change of each system rather than to a review calendar fixed at drafting time.

Third, hold the practices constant and let the evidence evolve. The most future-proof formulation states the invariant outcomes - delegated authority remains bounded, observable, reversible, and accountable - and treats the artefacts that demonstrate those outcomes as the evolving layer. Testing methods, telemetry, and evaluation techniques will change beyond recognition; the questions a board and a supervisor need answered will not. We would also suggest the glossary be maintained as a living annex on a defined revision cycle, since definitions are where technology vocabulary ages fastest, and that the final report state an explicit intent to review the practices when capability thresholds - rather than dates - are crossed.

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 are useful and appropriately practical, but as a set they share two structural gaps: they skew toward large, sophisticated institutions in advanced markets, and they are exclusively success stories. We would suggest additions on four fronts, three of which respond directly to the request for nonbank coverage.

Insurance. Insurers are material adopters of AI in underwriting, pricing, claims triage and fraud detection, and increasingly of agentic workflows in claims handling - decisions with direct customer and conduct consequences. Public precedent already exists for converting principles into tested assessment methods in this sector: the MAS Veritas consortium's published case studies applied fairness and FEAT assessment methodologies to predictive underwriting and fraud detection with named insurers. A case study showing an insurer applying the sound practices to an agentic claims process - including the human-approval boundary for repudiation decisions - would extend the report's reach to a nonbank sector where supervisory regimes (including the CBUAE's 2026 guidance, which applies to licensed insurers alongside banks) already expect exactly this.

Payments, exchange and remittance providers. These licensed nonbanks are often early adopters of AI for sanctions screening, fraud interdiction and transaction monitoring, operate at machine speed by construction, and typically lack the three-lines apparatus the current case studies presuppose. A case study demonstrating proportionate application - what sound practice looks like without a model-risk department - would serve a large population of institutions the report currently addresses only by implication.

Mid-tier institutions and emerging-market adoption. Responsible adoption looks different where governance capacity, rather than technology access, is the binding constraint. A case study of a mid-sized institution in a jurisdiction with active supervisory AI expectations - the UAE, for instance, where guidance and statutory enforcement frameworks are both newly in force - would show proportionality working under real supervisory engagement rather than in the abstract. We would be glad to contribute observations from the GCC market to such a case study.

Finally, the set needs at least one case study that is not a success story. Every current example illustrates adoption benefits; none illustrates detection and remediation - the discovery of ungoverned or shadow AI during an estate inventory, a near-miss caught by a monitoring threshold, an agent authority boundary that failed and was redesigned. Supervisory literature in other domains has long recognized that anonymized failure cases teach institutions more than exemplars. A "governance uplift" case - from unmanaged estate to evidenced control - would arguably do more to enable responsible adoption than any success story in the report.

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

The case studies are genuinely illustrative - they show that the sound practices can be applied, and by whom. Whether they are actionable is a different test: can a reader take a case study to their own institution and reproduce the result? Against that test, four recurring characteristics limit their value, each straightforward to remedy in the final report.

First, the case studies describe end states, not journeys. They present institutions that have arrived at good practice, with little visibility of the starting condition, the sequence of steps, the internal obstacles - fragmented ownership, legacy inventory gaps, contested accountability - or the time and resource required. A practitioner reading them learns what maturity looks like but not how to get there from a realistic starting point. Actionability lives

in the delta, not the destination. Even two or three sentences per case on "the position before, the sequencing, and the hardest decision" would multiply their practical value.

Second, they stop short of the evidence layer. The cases describe controls and governance arrangements but rarely the artefacts that demonstrated them: what the assurance documentation contained, what the board actually saw, what would have been shown to a supervisor on request. Since the report's own logic runs from principles to practices, the natural completion is practices to proof. A standard closing element for each case - "the evidence this institution could produce" - would convert descriptions into templates, and would quietly seed the common evidence expectations that comparable supervision will eventually require.

Third, the decision points are invisible. The most actionable content in any adoption story is the boundary-drawing: where the institution set the human-approval threshold and why; which actions were designated non-delegable; what the system was prohibited from doing; which proposed use was rejected or descoped. Every case study presents choices already made without the reasoning that made them. Institutions do not struggle to adopt controls in the abstract - they struggle to place the lines. Showing the placement logic, even briefly, is where insight becomes transferable.

Fourth, no case study shows the control environment working under stress. None describes a threshold breach, an override, an escalation, a revalidation triggered by a vendor's model update, or a kill switch invoked - in test or in production. Controls that are only ever described in repose provide weak evidence that they operate. One case tracing a single triggering event end-to-end - detection, containment, decision, remediation, and the evidence trail it left - would be the most actionable page in the report.

7. 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 current industry usage. We offer four suggestions, in descending order of importance.

First, the glossary defines technologies but not the thing institutions must govern. Its entries describe kinds of AI - the model-layer vocabulary - while the report's own sound practices attach to something larger: the deployed system through which a model plans, accesses data, invokes tools and takes actions. That governed object currently has no name in the glossary, which means each institution will bound it differently and no two postures will be comparable. We would propose adding a definition along the following lines:

Delegated decision-and-action system: a socio-technical system - comprising one or more AI models together with their prompts and policies, orchestration, memory, tools and interfaces, permissions, data, and human decision points - to which a financial institution grants authority to interpret objectives, make or recommend decisions, or execute actions within defined boundaries.

The precise term matters less than the presence of a defined, technology-neutral unit of governance; without one, the glossary describes the ingredients of the risk but not its object.

Second, "agentic AI" and "AI agent" should be defined by capability thresholds rather than description. Vendor usage of "agent" now spans everything from a chat interface to an autonomous payments workflow. A definition anchored in observable capabilities - the capacity to plan multi-step tasks, select and invoke tools, maintain state, and act on external systems without per-step human instruction - gives institutions a test they can apply, and prevents classification arbitrage in either direction: escaping controls by re-labelling an action-taking system, or drowning inventories by labelling every assistant an agent.

Third, the human-oversight vocabulary needs disambiguation. "Human in the loop," "human on the loop," and "human oversight" are used across the industry with inconsistent and sometimes contradictory meanings. Since several sound practices depend on which mode applies, the glossary should define each distinctly - approval before action; monitoring with the ability to intervene; and accountable governance of the system overall - because a control stated as "human oversight" is unimplementable and unexaminable until the mode is specified.

Fourth, define the assurance vocabulary the final report will rely on. Terms such as assurance, independent validation, continuous monitoring, and revalidation carry precise, distinct meanings in supervisory practice - respectively an evidenced justified conclusion; an independent evaluation; ongoing observation against thresholds; and a renewed conclusion after material change. Defining them would prevent the most common industry conflation, in which monitoring dashboards are presented as though they were validation.

Finally, we support the glossary being maintained as a living annex, as noted in our response to Question 4 - definitions are where technology vocabulary ages fastest.

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

We commend the FSB on a consultation that moves the international conversation decisively from principle to practice, and we make four closing observations.

First, the historical position of this consultation deserves to be stated plainly. The major governance regimes of the last two decades - data protection, post-crisis prudential reform, model-risk management itself - were each written after the harm they address. This consultation stands in an unusual and valuable position: the governance gap for delegated, action-taking AI has been identified, named in the supervisory record, and opened for comment before the loss event that would otherwise define it. The sequence has been reversed for the first time. That opportunity is perishable, and it places a premium on converting these sound practices into examinable assurance while the gap is still being discussed in consultations rather than in enforcement actions.

Second, the coordination window is now. Jurisdictions are moving on different clocks: US model-risk guidance has drawn agentic AI outside its scope; the EU has deferred its high-risk obligations to end-2027; the UK is extending model-risk discipline through supervisory engagement; Singapore has operationalized assessment methodology; and Gulf supervisors have paired lifecycle AI guidance with strengthened statutory enforcement. Each supplies a fragment - none the whole. The FSB is the only body positioned to align these fragments around a common regulatory object, a shared authority-based taxonomy,

and comparable evidence expectations before national approaches harden into incompatibility. We would encourage the final report to claim that convergence role explicitly.

Third, the test of the final report will be demonstrability. Across our responses we have returned to one theme: the distance between governance that is documented and governance that can be proven - at the point of decision, after material change, and under supervisory challenge. If the final report establishes that institutions should be able to evidence the operation of the sound practices, not merely attest to their adoption, it will have avoided the recurring failure of well-designed frameworks that succeed on paper and fail in operation.

Finally, we thank the FSB and the workstream for the opportunity to contribute. Our supplementary paper elaborates the assurance architecture referenced in these responses, and we would welcome the opportunity to engage further as the final report develops, including through observations from GCC market implementation.

WORKING PAPER · JULY 2026

From Model Validation to Agentic Assurance

What SR 26-2 reveals about the future of AI regulation in banking

A proposed regulatory architecture for delegated decision-and-action systems

Hemant Julka
Founder & CEO, Certavixx Technologies
Dubai International Financial Centre

The proposition

New-age AI regulation should be moulded around the authority a system is permitted to exercise, the consequences of failure, and the strength of its control envelope - not around technology labels alone.

Version 1.0 | Current as at 21 July 2026
This paper is intended for policy and industry discussion and does not constitute legal advice.

Abstract

The April 2026 revision of US model-risk guidance, SR 26-2, draws a deliberate boundary: generative and agentic AI are outside the guidance, while banks are expected to govern them through broader risk-management arrangements. Two months later, the Financial Stability Board proposed twelve sound practices for responsible AI adoption, including specific controls for autonomous agents. The regulatory problem is therefore no longer an absence of principles. It is the absence of a common regulatory object and an assurance architecture that can convert those principles into evidence of control. This paper proposes that regulators and financial institutions treat a Delegated Decision-and-Action System (DDAS) - not the underlying model alone - as the unit of assurance. It introduces an Agentic Assurance Framework built around authority-based tiering, an explicit agentic control envelope, a pre-deployment assurance case, continuous runtime monitoring, event-triggered revalidation, periodic independent assurance, supervisor-ready evidence and macroprudential monitoring of common dependencies. The framework preserves the strongest ideas of model-risk management - proportionality, independent challenge, validation and accountability - while adapting them to systems that can plan, use tools, maintain state and act in the world. The objective is not more AI paperwork. It is a regulatory design in which institutions can demonstrate, and supervisors can test, that delegated machine authority remains bounded, observable, reversible and accountable.

Core thesis

An agent is not simply a model producing a different kind of output. It is a system to which an institution delegates a bounded degree of authority. Regulation should therefore focus on the whole decision-and-action system, the authority granted to it, and the evidence that the authority remains under control.

Contents

Executive summary

SR 26-2 should not be read as a regulatory retreat from AI. It should be read as an acknowledgement that a model-risk framework designed around quantitative models does not, by itself, describe the full governance problem created by systems that can plan, call tools, retain memory, interact with other systems and execute actions. The FSB's June 2026 consultation then moved the debate forward by describing sound practices for the responsible adoption of AI, including agent identity, boundaries on autonomy, human approval for material transactions, kill switches, red-teaming and third-party controls. [1,3]

The remaining gap is narrower but more consequential: neither a list of principles nor a traditional validation report gives a supervisor a coherent view of whether delegated machine authority is under control across a complex institution. The missing layer is an integrated assurance architecture that joins model risk, operational risk, cyber, data, conduct, third-party risk and business accountability around a common unit of analysis.

Seven policy propositions

1. **Regulate delegated authority, not the technology label.** The most important variables are what the system may decide or do, the scale of potential harm, the reversibility of its actions and its connectivity to critical processes.
2. **Make the system, not the model, the unit of assurance.** The relevant object includes the foundation model, prompts, orchestration, memory, tools, permissions, data, human decision points and downstream systems.
3. **Retain independent challenge.** Agentic AI does not make model-risk management obsolete. It makes effective challenge more important and requires it to operate across several risk disciplines.
4. **Separate monitoring from validation.** High-frequency telemetry should support continuous monitoring. Revalidation should be triggered by material change or control events, and independent assurance should remain periodic and accountable.
5. **Require red-line controls.** Some controls must be non-compensable: unrestricted access to customer funds, unbounded tool use, absent audit trails or an ineffective emergency stop cannot be offset by a high aggregate governance score.
6. **Standardise the supervisory evidence record.** Institutions should be able to produce a time-stamped, immutable and traceable record of each material agentic system's authority, controls, exceptions, incidents and approvals.
7. **Add a macroprudential layer.** Supervisors should monitor shared foundation models, cloud and tool providers, correlated behaviours and common-mode failure risks across institutions.

The policy opportunity

The industry does not need a parallel universe of "AI regulation". It needs existing prudential, conduct, resilience and model-risk disciplines to converge around a clearly defined object of delegated authority - and a common language for proving that the authority remains controlled.

1. The regulatory moment: from model risk to delegated authority

The debate has moved beyond whether AI should be governed. The question is how governance becomes examinable assurance.

SR 11-7 established one of the most influential disciplines in modern banking governance. Its enduring contribution was not a checklist. It was the institutionalisation of effective challenge: models should be subject to independent review by people with the competence, authority and organisational standing to question their use. Conceptual soundness, ongoing monitoring and outcomes analysis became the backbone of model validation, supported by inventory, tiering, documentation and board accountability. [2]

SR 26-2 modernises that lineage. It narrows the definition of a model, makes proportionality more explicit and replaces a default annual validation expectation with a risk-based approach. It also places generative and agentic AI outside the guidance while requiring banks to determine appropriate governance and controls through their broader risk-management arrangements. This is not a declaration that such systems are ungoverned. It is a boundary around the domain in which traditional model-risk guidance is considered sufficient. [1]

The FSB's June 2026 consultation is the next important step. Its twelve sound practices cover enterprise governance, materiality assessment, model and system selection, data, explainability, performance, human oversight, cyber and ICT risk, and third-party risk. For agentic AI, it discusses controls such as individual agent identifiers, prohibited actions, approval checkpoints, limits on interactions with external environments, dual authorisation for financial transactions, audit trails, red-teaming and kill switches. The FSB is therefore already moving the international conversation from abstract ethics toward operational control. [3]

Yet a gap remains. Sound practices describe what good may look like; they do not yet define the regulatory object to which all relevant controls attach, how independent assurance should be structured across risk functions, or how a supervisor should compare one institution's posture with another's without reducing complex risk to a superficial score. That is the territory this paper addresses.

EXHIBIT 1 | The real regulatory gap

Layer	What exists	What is still missing
Principles	Model-risk principles, AI governance frameworks, FSB sound practices	A clear unit of regulation for systems that combine models, tools, permissions and actions
Institutional control	Existing first-, second- and third-line functions	An integrated assurance model that resolves fragmented ownership across risk taxonomies
Evidence	Policies, inventories, validation reports and monitoring dashboards	A standard, traceable and time-stamped supervisory evidence record for each material system
Systemic oversight	Third-party and concentration-risk frameworks	Visibility into common models, infrastructure and correlated agent behaviour across institutions

Source: Author's analysis, drawing on SR 26-2 and the FSB 2026 consultation report.

2. The category error: an agent is not simply a model

The break with traditional validation is not stochasticity. It is delegated, environment-coupled authority.

Traditional models are not always static, transparent or deterministic. Banks have long managed recalibrated models, vendor models, probabilistic outputs and systems whose performance changes with data and economic conditions. It would therefore be a mistake to claim that model validation assumed complete stability or repeatability.

The more important distinction is functional. A conventional model transforms inputs into estimates or classifications. An agentic system can interpret an objective, plan a sequence of tasks, select tools, retrieve information, invoke APIs, interact with other agents, maintain state and execute actions. The institution is no longer only relying on an output. It is delegating a bounded degree of decision and action authority to a system operating inside a live environment.

EXHIBIT 2 | From analytical component to delegated decision-and-action system

Dimension	Traditional model-centric view	Agentic assurance view
Primary function	Estimate, rank, classify or forecast	Interpret goals, plan, decide and act
Relevant boundary	Model code, data, assumptions and use	Model plus orchestration, memory, tools, permissions, interfaces and human controls
Change surface	Data, parameters, methodology and use	All model changes plus prompts, toolsets, memory, external systems, policies and permissions
Failure manifestation	Incorrect or misused output	Incorrect action, unauthorised action, cascading interaction or loss of control

Dimension	Traditional model-centric view	Agentic assurance view
Control objective	Reliable model performance and appropriate use	Bounded, observable, reversible and accountable delegated authority
Unit of assurance	Model	Delegated Decision-and-Action System (DDAS)

The distinction is functional rather than technological. A system should be governed as agentic when it is permitted to exercise material discretion or take actions, regardless of the label used by its vendor.

Five properties that change the assurance problem

- **Compositionality.** Behaviour emerges from the interaction of multiple components: the model, prompt and policy layer, orchestration, memory, tools, data and external services.
- **Environment coupling.** The system's behaviour depends on a changing operating environment, including the state of connected applications, permissions, counterparties and other agents.
- **Statefulness.** Memory and intermediate decisions can alter later behaviour, so the relevant state cannot always be reconstructed from the final output alone.
- **Delegated authority.** The system may be permitted to modify records, communicate with customers, initiate workflows or execute financial actions.
- **Machine-speed propagation.** Errors can move through connected processes before periodic human review can intervene, particularly where several agents or external tools are involved.

A more precise formulation

Stochasticity does not make validation impossible. The harder problem is that a material agentic system is stateful, compositional and connected to an external environment in which it may exercise authority. Assurance must therefore evaluate distributions of behaviour, sequences of action and the design of the control environment - not only the accuracy of individual outputs.

3. What should be retained from model-risk management

The answer is evolution, not abandonment.

Agentic AI does not invalidate model-risk management. On the contrary, it exposes why several of its strongest ideas remain essential. The task is to retain the discipline while extending its scope from a component to an operating system of decisions and actions.

MRM principle to retain	How it evolves for agentic systems
Proportionality	Control intensity should reflect materiality, complexity, autonomy, connectivity and potential harm.
Inventory and tiering	Institutions must know which systems exist, who owns them, what they may do and where they are deployed.
Independent challenge	Second-line and independent validation functions must be able to question design, evidence, residual risk and approval decisions.
Conceptual soundness	The purpose, operating assumptions, design choices, constraints and foreseeable failure modes must be coherent.
Ongoing monitoring	Performance, control operation, incidents and usage must be observed after deployment.
Outcomes analysis	Institutions should test realised outcomes, distributions of behaviour, near misses and unintended consequences against defined acceptance criteria.
Change management	Material changes should trigger review and, where necessary, revalidation before the prior assurance conclusion is relied upon.

MRM principle to retain	How it evolves for agentic systems
Accountability	A named business executive should remain accountable for the system's use and consequences; accountability cannot be delegated to the technology.

The proposed framework is intended to extend effective challenge across the whole system, not to create a separate governance silo for AI.

The critical distinction: monitoring is not validation

A common response to fast-changing AI is to call for "continuous validation". The phrase is attractive but imprecise. Automated telemetry can continuously monitor performance, policy breaches, tool use, exceptions and changes. It cannot continuously reproduce the full independence, judgement and accountability of validation.

A more credible assurance cycle has three tempos:

- Continuous monitoring: machine-readable telemetry and controls operate at the speed of the system.
- Event-triggered revalidation: material change, incident or boundary breach reopens the assurance conclusion.
- Periodic independent assurance: a named and independent function assesses whether the full control environment remains adequate and whether management's evidence and residual-risk judgement are credible.

4. The proposed regulatory object: the DDAS

A common unit of assurance is needed before common evidence and supervision are possible.

This paper proposes the Delegated Decision-and-Action System (DDAS) as the unit of assurance for material agentic AI in regulated financial institutions. A DDAS is a socio-technical system to which an institution grants authority to interpret objectives, make or recommend decisions, invoke tools or execute actions within defined boundaries.

The definition is deliberately technology-neutral. It captures an autonomous AI agent, a multi-agent workflow or an AI-enabled orchestration layer when the system exercises material discretion or initiates consequential actions. It does not capture every use of generative AI. A drafting assistant with no access to sensitive data or operational tools may remain a low-risk productivity application. A system that can communicate binding decisions, change customer records or initiate payments should be treated very differently.

EXHIBIT 3 | The agentic control envelope

Component of the envelope	Minimum assurance question
Purpose and prohibited use	What objective is the system authorised to pursue, and what actions or outcomes are explicitly forbidden?
Model and reasoning layer	Which models are used, under what versions and limitations, and how are performance and uncertainty assessed?
Prompts and policy layer	Which instructions, policies and guardrails govern behaviour, and how are they versioned and tested?
Memory and state	What information persists, for how long, who may alter it, and how can material decision paths be reconstructed?
Tools and APIs	Which tools may be called, with what parameters, rate limits and approval requirements?
Identity and permissions	Does each agent have a distinct identity, least-privilege access and enforceable segregation of duties?
Data and context	Which data may be accessed or created, and how are quality, privacy, confidentiality and residency protected?
Human control	Where do humans approve, supervise, contest, override or stop the system, and is the oversight demonstrably effective?

Component of the envelope	Minimum assurance question
Monitoring and response	Which behaviours, actions and control breaches are monitored, and what happens when thresholds are crossed?
Third-party dependencies	Which providers can change the system's behaviour or availability, and what evidence, notification, audit and exit rights exist?

Authority-based tiering

A DDAS should be classified by the authority it can exercise and the consequences of failure, not by whether a vendor describes it as "generative", "agentic" or "autonomous". A practical tiering model can be built from five variables:

- **Materiality:** potential financial, prudential, customer, legal or societal harm.
- **Autonomy:** the degree to which action can occur without prior human approval.
- **Reversibility:** whether an action can be reliably stopped, corrected or compensated.
- **Connectivity:** access to critical systems, data, counterparties, APIs and other agents.
- **Velocity and scale:** how quickly and widely an error or manipulation could propagate.

EXHIBIT 4 | Illustrative authority tiers

Tier	Typical authority	Illustrative use	Minimum posture
A - Assistive	Creates content or analysis; no direct action	Internal drafting or research assistant	Approved data boundaries, output review, basic monitoring
B - Recommending	Influences a human decision; no execution authority	Credit, fraud or advisory recommendation	Independent testing, explainability or compensating controls, decision accountability
C - Executing within limits	Takes reversible actions within pre-set limits	Case routing, record updates, low-value workflow execution	Strong identity, least privilege, transaction limits, event logging, human-on-the-loop
D - Material autonomous	Takes consequential or hard-to-reverse actions	Customer communication with legal effect, payments, treasury or market actions	Pre-deployment assurance case, dual control, non-compensable red lines, kill switch, event-triggered revalidation, senior approval

The tiers are illustrative. Regulators should permit institutions to refine thresholds while standardising the variables and minimum expectations used to justify classification.

5. The Agentic Assurance Framework

Eight principles for turning governance expectations into evidence of control.

Principle 1 - Define the delegated authority.

Every material DDAS should have a precise purpose, authorised actions, prohibited actions, transaction or exposure limits, permitted data, tool boundaries and conditions for human approval. The definition should be understandable to business, risk, audit and supervisors - not only to engineers.

Principle 2 - Build an assurance case before deployment.

Management should demonstrate why the system is suitable for its intended purpose, which failure modes have been considered, what testing has been performed, which controls prevent or contain harm, what residual risk remains and which executive has accepted it. The assurance case should be a structured argument supported by evidence, not a narrative compliance pack.

Principle 3 - Validate at component, system and control-environment levels.

Component testing remains necessary for models, data, prompts and tools. System testing should assess end-to-end behaviour, interaction effects and sequences of action. Control-environment assurance should assess governance, identity, access, monitoring, human control, incident response, resilience and third-party dependencies.

Principle 4 - Engineer runtime control.

High-tier systems should operate under least privilege, tool allow-lists, rate and value limits, segregated duties, immutable logs, anomaly detection, safe fallback, graduated degradation and an effective emergency stop. Control should be designed into the execution path rather than added as retrospective review.

Principle 5 - Monitor continuously and revalidate on events.

Institutions should continuously observe behaviour, actions, policy breaches, permission changes, model or vendor changes, data drift, overrides and incidents. Defined events should invalidate or qualify the prior assurance conclusion until review is completed.

Principle 6 - Preserve functional independence.

The function that challenges a system should have separate accountability from the function that builds or operates it, access to independent test evidence and the authority to constrain or stop deployment. AI-enabled testing tools may be used, but the assurance mechanism itself should be governed, validated and subject to human accountability.

Principle 7 - Make evidence reproducible without manufacturing false precision.

A supervisor should be able to trace an assurance conclusion to current evidence, rules, thresholds, exceptions and named judgement. Computed indicators can improve consistency, but aggregate scores must not conceal red-line failures, uncertainty or stale evidence.

Principle 8 - Connect institution-level assurance to systemic oversight.

Material dependencies on common foundation models, cloud providers, data services, agent frameworks or tool ecosystems should be visible to supervisors. Institutions should support sector exercises and report incidents or material changes that could create common-mode risk.

EXHIBIT 5 | The assurance cycle

Stage	Primary question	Evidence	Decision owner
1. Classify	What authority and harm potential does the DDAS have?	Inventory, dependency map, authority tier, materiality assessment	Business owner with second-line challenge
2. Assure before use	Is the system suitable and controlled for its intended purpose?	Assurance case, test results, control design, residual-risk decision	Independent validation / relevant risk functions
3. Authorise	May the system operate, and under which limits?	Conditions of use, approvals, red lines, expiry or review date	Accountable executive / committee
4. Monitor at runtime	Is actual behaviour within approved boundaries?	Telemetry, actions, exceptions, overrides, incidents and evidence freshness	First line, overseen by second line
5. Revalidate on events	Has change or experience altered the assurance conclusion?	Change assessment, targeted testing, updated risk decision	Independent function
6. Periodically assure	Does the whole posture remain credible and effective?	Independent review, control testing, outcomes and lessons learned	Second/third line with board reporting

Red-line controls and the limits of scoring

A computed posture can support comparability and prioritisation, but it should never imply that all control weaknesses are compensable. Regulators should define, and institutions should expand, a set of red-line conditions for high-tier DDAS. Examples include:

- no distinct agent identity or attributable audit trail;
- unrestricted access to payment, trading or customer-record systems;
- the ability to alter its own permissions or objectives without authorised control;
- no tested method to halt, isolate or degrade the system safely;
- material third-party changes without notification or version traceability;
- use outside the approved purpose or customer population;
- known critical findings without documented acceptance by an authorised executive.

Where a red line is breached, the posture should not be represented as "acceptable" because other controls score well. The appropriate status is constrained operation, remediation, suspension or explicit senior risk acceptance within regulatory limits.

6. The D-SIB and financial-stability perspective

The central risk is not a hallucinating chatbot. It is correlated delegated authority operating through shared infrastructure.

A D-SIB cannot govern agentic AI as a model-risk issue alone. A single DDAS can simultaneously create model, operational, cyber, data, privacy, conduct, financial-crime, legal, third-party, resilience and reputational risk. The first governance challenge is therefore horizontal: who has the authority to form an integrated view when no single risk taxonomy captures the system?

The second challenge is systemic. The FSB has highlighted the possibility that common AI models, data and infrastructure could produce homogeneous behaviour and correlated outcomes across institutions. Shared vendors may also change underlying models without sufficient notice, while concentration in cloud and AI supply chains can create common points of failure. Agentic systems add velocity: a common defect, manipulation or vendor update could influence actions across several firms before traditional supervisory cycles detect the pattern. [3,4]

EXHIBIT 6 | Three levels of regulatory assurance

Level	Objective	Illustrative mechanism
1. Institution-level DDAS assurance	Ensure each material system operates within approved authority and risk appetite	Authority tiering, assurance case, control envelope, telemetry, revalidation and independent review
2. Supervisory evidence interface	Allow supervisors to interrogate posture without waiting for bespoke document production	Standard evidence fields, immutable snapshots, exceptions, incidents, dependencies and named accountability
3. Systemic dependency observatory	Identify concentration, correlated behaviour and common-mode vulnerabilities	Aggregated provider and model dependencies, cross-firm incidents, sector testing and scenario analysis

Third-party minimum rights

A bank cannot reproduce a frontier model's training process, but it can establish minimum evidence and contractual rights before delegating material authority to a third-party system. For high-tier DDAS, regulatory expectations should include:

- advance notification of material model, safety-policy or infrastructure changes;
- version identification and the ability to preserve or roll back approved configurations;
- incident and vulnerability notification within defined timeframes;
- evidence of testing relevant to the bank's use and risk profile;
- audit, access or independent-assurance rights proportionate to materiality;
- clear controls over training, retention and secondary use of institutional and customer data;
- transparency over material subcontractors and concentration dependencies;

- business-continuity, substitutability and exit arrangements tested in practice.

Board-level questions for a D-SIB

- Which systems are authorised to take actions rather than merely make recommendations?
- What is the largest customer, liquidity, market, conduct or operational loss that a single agentic failure could create before intervention?
- Which controls are genuinely preventative and which merely detect harm after the event?
- Which material systems depend on the same foundation model, cloud, data or agent framework?
- Can the institution stop or safely degrade each high-tier system without losing a critical service?
- Where has human oversight become routine approval rather than effective challenge?
- What evidence would the institution provide to a supervisor tomorrow morning after a material incident?

7. Comparative regulatory direction

Different regimes provide complementary building blocks; none yet supplies the complete assurance architecture.

The emerging international landscape is more convergent than it first appears. Most regimes emphasise proportionality, lifecycle governance, accountability, data, testing, human oversight and third-party risk. Their differences lie in legal form, institutional scope and the degree of operational specificity.

EXHIBIT 7 | Selected regulatory contributions

Regime	Current contribution	Important limitation	Implication for the framework
United States	SR 26-2 modernises MRM and explicitly places generative and agentic AI outside its scope while retaining broader governance responsibility	The boundary does not specify an integrated assurance architecture for delegated action systems	Use MRM principles as a foundation, then govern DDAS through an enterprise assurance framework
FSB	Twelve sound practices with concrete treatment of agentic autonomy, human control, cyber and third-party risk	A consultation toolkit, not an international standard; no common supervisory evidence model yet	Translate sound practices into an attestable assurance cycle and comparable evidence
United Kingdom	PRA SS1/23 treats MRM as a risk discipline and is technology-agnostic and outcomes-focused	Its direct scope is prudential and institution-specific; agentic action systems still span wider risk domains	Preserve accountable ownership, tiering and independent challenge while expanding the unit of assurance
European Union	The AI Act creates legal lifecycle obligations for defined high-risk systems, including risk management, logging, human oversight and post-market monitoring	The Digital Omnibus, adopted June 2026, defers Annex III high-risk obligations to 2 December 2027; legal conformity does not by itself equal prudential assurance	Use conformity evidence as an input to, not a substitute for, institution-specific control assurance
Singapore	FEAT and Veritas demonstrate how principles can be converted into assessment methods; MindForge extends focus to GenAI risks	Methods must continue evolving for autonomous, multi-step and action-taking systems	Make principles testable and use-case specific, while adding authority and runtime-control dimensions
United Arab Emirates	CBUAE guidance emphasises governance, fairness, transparency, explainability, human oversight, data and privacy across licensed financial institutions	The guidance is principles-based; the wider statutory reconciliation period and sanctions are not AI-specific	The UAE is well placed to pilot supervisor-ready evidence and outcome-focused assurance without overstating the legal link

The analysis is current as at 21 July 2026. References [1], [3], [5]-[10].

The UAE opportunity - framed with precision

The CBUAE's February 2026 Guidance Note applies responsible-AI expectations across licensed financial institutions and places clear emphasis on consumer protection, accountability, fairness, transparency, human oversight, data management and privacy. The UAE's broader 2025 Central Bank law materially strengthens the supervisory and enforcement environment, but its one-year reconciliation provision is a general statutory requirement and may be extended; it should not be presented as a specific AI compliance deadline. The opportunity is therefore not to claim an automatic enforcement clock for AI. It is to use the UAE's regulatory agility, financial-sector digitisation and SupTech ambition to test a more continuous and evidence-led model of assurance. [9,10,11]

8. Applied banking scenarios

The same model can require radically different assurance depending on the authority and consequence attached to it.

Scenario A - Credit decision agent

A system gathers customer information, requests missing documents, applies policy, analyses affordability and recommends or issues a retail-credit decision.

- **Risk focus:** fair treatment, explainability, data quality, policy compliance, vulnerability, adverse-action communication and contestability.
- **Authority boundary:** whether the agent recommends, conditionally approves or communicates a binding decision.
- **Red lines:** unapproved policy changes, use of prohibited attributes, inability to reconstruct the basis of a decision, or denial without effective appeal.
- **Assurance emphasis:** distributional fairness testing, scenario testing, policy conformance, outcome monitoring, human override analysis and customer-redress evidence.

Scenario B - AML investigation agent

A system gathers alerts and external information, identifies relationships, drafts investigative narratives and recommends escalation or closure.

- **Risk focus:** false-negative risk, evidential quality, confidentiality, legal privilege, explainability, investigator over-reliance and model manipulation.
- **Authority boundary:** the agent may investigate and recommend but should not silently close high-risk cases or file regulatory reports without defined human control.
- **Red lines:** untraceable evidence, unauthorised external data use, suppression of material alerts or inability to identify agent-generated assertions.
- **Assurance emphasis:** traceability to source evidence, adversarial testing, typology coverage, override analysis, sampling of closed cases and independent quality assurance.

Scenario C - Treasury or payments execution agent

A system monitors liquidity or payment obligations and can prepare, route or execute transactions within approved mandates.

- **Risk focus:** financial loss, liquidity, market conduct, fraud, cyber compromise, operational resilience and correlated behaviour.
- **Authority boundary:** value limits, instruments, counterparties, time windows, market conditions and human approval thresholds.
- **Red lines:** direct unrestricted payment access, self-modification of limits, absent dual control for material amounts or an untested kill switch.
- **Assurance emphasis:** pre-trade controls, simulation, stress scenarios, transaction-level logs, segregation of duties, fail-safe behaviour, recovery testing and rapid event notification.

The principle in practice

The same foundation model could sit behind all three scenarios. Its regulatory treatment should nevertheless differ because the delegated authority, reversibility and harm profile differ. This is why technology-category regulation alone is insufficient.

9. A supervisory roadmap

Move from broad principles to controlled experimentation, comparable evidence and systemic visibility.

For regulators and standard setters

1. **Define the unit and taxonomy.** Issue a technology-neutral definition of material delegated decision-and-action systems and standard variables for authority-based tiering.
2. **Specify minimum inventory fields.** Require visibility of purpose, owner, model and vendor versions, tools, permissions, data, autonomy, human controls, dependencies and current assurance status.
3. **Identify non-delegable or restricted actions.** Clarify where human decision or dual authorisation remains mandatory, particularly for customer funds, legal rights, prudential reporting and market-sensitive actions.
4. **Require an assurance case for high-tier systems.** Set outcome-focused expectations for evidence before deployment without prescribing one engineering architecture.
5. **Standardise material-change and incident triggers.** Define events that require notification, revalidation, restricted operation or suspension.
6. **Create a supervisory evidence schema.** Develop a common, secure way for firms to provide time-stamped posture, exceptions and incident evidence without continuous unrestricted supervisory access to operational systems.
7. **Monitor common dependencies.** Collect proportionate information on foundation-model, cloud, data, agent-platform and critical-tool concentration.
8. **Run sector exercises.** Test common vendor failures, malicious agent manipulation, correlated decision errors and the simultaneous loss of a critical AI service.
9. **Use regulatory sandboxes for assurance, not only innovation.** Pilot how evidence, runtime controls, revalidation and supervisory interrogation work before issuing detailed rules.

For D-SIB boards and chief risk officers

- Establish a single enterprise view of material DDAS across business, technology and all risk functions.
- Approve an authority-based risk appetite, including prohibited uses and mandatory human-control thresholds.
- Create a cross-functional assurance decision for high-tier systems rather than sequential sign-offs that leave no one accountable for the whole.
- Require runtime controls and evidence architecture as part of design, not as conditions added immediately before production.
- Test human oversight for effectiveness, including the risk of rubber-stamping and automation bias.
- Map concentration and substitutability across models, cloud, data, tools and agent platforms.
- Provide the board with exceptions, incidents, evidence freshness and red-line breaches - not only adoption counts and aggregate scores.

A pragmatic sequence

Horizon	Regulatory objective	Practical output
0-6 months	Create visibility and common language	DDAS definition, inventory fields, authority tiers, incident taxonomy and pilot cohort

Horizon	Regulatory objective	Practical output
6-18 months	Test assurance and evidence in live use cases	Assurance-case pilots, supervisory evidence schema, sector scenarios and vendor expectations
18-36 months	Embed proportional supervisory expectations	Risk-tiered requirements, thematic reviews, cross-firm concentration monitoring and international alignment

10. Conclusion

The future of AI regulation will be determined by whether delegated authority can be made provably governable.

SR 26-2 is consequential because it draws a clear line around the reach of traditional model-risk guidance. The FSB's 2026 sound practices are consequential because they begin to describe the operating controls required beyond that line. Together, they point toward the next task: defining how a regulated institution proves that an AI system with authority to act remains under control.

The answer is not to discard model-risk management, nor to force every AI system into a model inventory built for a different purpose. It is to preserve proportionality, effective challenge, validation, monitoring and accountability while changing the unit of assurance. For material agentic systems, that unit should be the whole Delegated Decision-and-Action System and its control envelope.

A credible regime would classify systems by authority and consequence; require a pre-deployment assurance case; validate components, system behaviour and the control environment; engineer runtime limits and reversibility; monitor continuously; revalidate when material events occur; preserve independent judgement; and make evidence available to supervisors in a traceable form. For D-SIBs and central banks, it would also look across institutions at shared models, providers, infrastructure and correlated behaviour.

This is the opportunity to avoid two familiar regulatory failures: rules written only after harm, and compliance that succeeds on paper while failing in operation. The standard should not be whether an institution can produce an AI policy. It should be whether the institution can demonstrate - at the point of decision, after material change and under supervisory challenge - that machine authority remains bounded, observable, reversible and accountable.

Final proposition

The future regulatory perimeter should follow delegated authority. The future assurance standard should follow evidence of control.

Appendix A. Minimum supervisory evidence pack

For each material DDAS, an institution should be able to provide a current, time-stamped evidence record containing at least the following fields. The objective is not continuous unrestricted access by the supervisor. It is the ability to produce a trusted and interrogable posture on demand.

Evidence domain	Minimum content
Identity	System name, unique agent identifiers, business owner, technical owner, accountable executive and responsible risk functions
Purpose and authority	Approved purpose, customer or process scope, authority tier, permitted and prohibited actions, value or exposure limits
Architecture and dependencies	Models and versions, orchestration, memory, tools, APIs, data, cloud and material third parties
Risk and approval	Materiality assessment, inherent risk, control assessment, residual risk, approval conditions, expiry or review date
Testing and assurance	Test scope, scenarios, benchmarks, limitations, independent conclusions, open findings and management actions
Runtime posture	Current configuration, permissions, monitoring thresholds, evidence freshness, control status and red-line conditions
Human control	Oversight model, approval points, override and escalation rights, kill-switch ownership and effectiveness testing
Change history	Material changes, vendor updates, prompt or tool changes, revalidation decisions and version lineage
Outcomes and incidents	Performance trends, exceptions, overrides, near misses, complaints, losses, incidents and remediation
Third-party and resilience	Contractual rights, service dependencies, concentration assessment, continuity, substitutability and exit-test status

Appendix B. Illustrative revalidation and escalation triggers

Trigger category	Illustrative events
Model or vendor change	New foundation-model version, safety-policy change, embedding model change, vendor migration or material service update
Purpose or population change	New use, product, geography, customer segment, legal context or decision consequence
Tool or permission change	New API, expanded data access, higher transaction limit, new external system or agent-to-agent interaction
Prompt, policy or orchestration change	Material instruction, workflow, guardrail, routing or decision-policy change
Memory or data change	New persistent memory, data source, retention rule, sensitive-data use or material quality issue
Performance or behaviour event	Threshold breach, unexpected action sequence, instability, bias signal, drift, security anomaly or repeated human override
Incident or near miss	Customer harm, unauthorised action, financial loss, regulatory breach, data leakage, cyber compromise or failed emergency stop

Trigger category	Illustrative events
Control or assurance event	Expired evidence, critical finding, ineffective human oversight, vendor assurance withdrawal or loss of audit rights
Systemic event	Material incident at a common provider, coordinated attack, sector alert or evidence of correlated behaviour across firms

A trigger need not always require full revalidation. Institutions should define which events require targeted review, restriction, suspension, notification or complete reassessment.

Notes and references

The paper is current as at 21 July 2026. Regulatory instruments, consultations and implementation timetables may change. Direct links are included for review.

[1] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency (2026), SR 26-2: Supervisory Guidance on Model Risk Management, 17 April 2026. Source

[2] Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency (2011), SR 11-7: Guidance on Model Risk Management, 4 April 2011. Source

[3] Financial Stability Board (2026), Sound Practices for Financial Institutions' Responsible AI Adoption: Consultation Report, 10 June 2026. Source

[4] Financial Stability Board (2024), The Financial Stability Implications of Artificial Intelligence, 14 November 2024. Source

[5] Prudential Regulation Authority (2026), SS1/23: Model Risk Management Principles for Banks, updated April 2026. Source

[6] European Union (2024), Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence, as amended by the Digital Omnibus on AI (adopted by Parliament 16 June 2026 and Council 29 June 2026, deferring Annex III high-risk obligations to 2 December 2027). Source

[7] Monetary Authority of Singapore (2018), Principles to Promote Fairness, Ethics, Accountability and Transparency in the Use of AI and Data Analytics in Singapore's Financial Sector. Source

[8] Monetary Authority of Singapore, Veritas Initiative and related assessment methodologies and toolkits for responsible AI in financial services. Source

[9] Central Bank of the United Arab Emirates (2026), Guidance Note on the Consumer Protection and Responsible Adoption and Use of Artificial Intelligence and Machine Learning by Licensed Financial Institutions, 23 February 2026. Source

[10] United Arab Emirates (2025), Federal Decree-Law No. 6 of 2025 Regarding the Central Bank, Regulation of Financial Institutions and Activities, and Insurance Business. Article 184 is a general reconciliation provision and may be extended; it is not an AI-specific compliance deadline. Source

[11] Central Bank of the United Arab Emirates (2024), CBUAE enhances SupTech initiative as part of its Financial Infrastructure Transformation Programme, 25 April 2024. Source

Further reading

[12] Gambacorta, L., Lauridsen, N., Kiuhan-Vasquez, S. and Prenio, J. (2025), Making SupTech Work: Evidence on the Key Drivers of Adoption, BIS Working Papers No. 1309. Source

[13] Basel Committee on Banking Supervision (2024), Digitalisation of Finance. Source

[14] Basel Committee on Banking Supervision (2025), Principles for the Sound Management of Third-Party Risk. Source

[15] Board of Governors of the Federal Reserve System (2026), Michelle W. Bowman, Artificial Intelligence in the Financial System, 1 May 2026. Source