Governed Autonomy

How Private Banks Can Industrialise AI Within Regulatory Boundaries

Adrien Pesa

March 2026

Publication date: 25 March 2026 | **Jurisdiction:** Singapore (MAS regulatory framework)

Regulatory status: MAS's proposed *Guidelines on AI Risk Management* (P017-2025) closed for public consultation on 31 January 2026; finalisation is pending as of 25 March 2026. MAS's proposed *Guidelines on Third-Party Risk Management* (P004-2026) are open for consultation until 20 April 2026. Readers should verify the latest status of both instruments against the MAS publications index.

Paper type: Strategic architecture paper grounded in regulatory analysis and supervisory observation. Economic claims are directional.

Provenance: This paper reflects the author's independent analysis of publicly available laws, regulatory publications, academic literature, and industry frameworks.

Table of Contents

Executive Summary

Private banking stands at an inflection point that is architectural, not incremental. Most institutions have deployed machine learning in isolated functions—transaction monitoring, name screening, risk scoring, and in some cases portfolio analytics or client segmentation—and reasonably concluded that AI is useful but peripheral. That conclusion is now outdated. For institutions already deploying AI in compliance and analytics, the strategic question has shifted: not whether to adopt AI, but whether to redesign the operating model around it.

This paper argues that the competitive frontier in private banking is shifting from better models toward governed autonomy: the capacity to embed AI-driven decision support within productised client journeys while maintaining—and indeed strengthening—regulatory defensibility. Achieving this requires a three-layer architecture enveloped by a governance wrapper: a control plane, an LLM workflow layer, and an ontology-bounded autonomy layer—three technical tiers through which regulatory requirements are translated into machine-enforceable boundaries.

The first layer reconceives the machine learning infrastructure banks already operate for surveillance, risk classification, and due diligence as a control plane, a persistent, rules-enforcing layer through which all downstream processes must pass. The second layer introduces LLM workflows—natural language synthesis, document generation, and communication drafting—that can compress multi-day processes into hours within the supervisory expectations for technology risk management.[1] The third layer, and the architecture's most consequential design choice, is ontology-bounded autonomy. Its mechanism is a governed semantic layer that encodes the institution’s clients, portfolios, products, and regulatory obligations as typed objects with machine-enforceable rules, defining the boundaries within which AI agents can act: permissible actions, data entitlements, and escalation triggers built into the architecture itself. This is what separates governed autonomy from both scripted automation and unbounded AI deployment: agents adapt to novel inputs while remaining constrained by invariant, machine-enforceable boundaries that governance defines.[2]

The paper uses the periodic client review as its exemplar productised client journey: a single workflow spanning data gathering, suitability analysis, documentation, client communication, and record keeping that requires all three layers to operate in concert and serves as the lens through which the full architecture is examined.

The governance wrapper that envelops all three layers is not a constraint on innovation; it is the infrastructure that makes innovation defensible, converting regulatory requirements into machine-readable boundaries that accelerate rather than impede deployment. This governance infrastructure reduces the marginal cost of each subsequent AI deployment, but it does not diminish the human

[1] MAS Technology Risk Management Guidelines (January 2021) [SUPERVISORY] establish expectations for technology risk governance, including system security, resilience, and change management applicable to AI-enabled systems. The Project MindForge *AI Risk Management Handbook* [METHODOLOGY], developed by a consortium of financial institutions under MAS's leadership, provides industry consensus guidance for AI risk management in financial services. The Handbook comprises three documents: the Executive Handbook (November 2025), the Operationalisation Handbook (January 2026), and the Implementation Examples (January 2026).

[2] IMDA, *Model AI Governance Framework for Agentic AI*, Version 1.0 (22 January 2026) [METHODOLOGY]. The framework outlines four governance dimensions for agentic AI: assessing and bounding risks upfront, making humans meaningfully accountable, implementing technical controls and processes, and enabling end user responsibility.

oversight obligation for any individual capability; each deployment must still satisfy oversight requirements proportionate to its risk materiality.[3]

MAS's consultation on proposed AI risk management guidelines, closed for public feedback in January 2026, with finalisation pending, signals that supervisory expectations in this domain are likely to formalise. Institutions that embed governance by design now will find themselves positioned to comply; those treating governance as an afterthought will face a compounding deficit of both capability and credibility.

The paper identifies three concrete actions for boards and management committees: assess the institution's current maturity against a diagnostic framework of institutional archetypes (Exhibit 4), commission an architecture review to determine readiness for a phased deployment programme, and establish a governance-first deployment protocol, operationalised through a defined governance gate (Exhibit 6), requiring that governance prerequisites are documented and approved before each stage of capability deployment. These actions are sequential dependencies: skipping the first makes the second premature; skipping the second makes the third impossible.

The paper defines three institutional archetypes—Compliance First, Workflow Augmented, and Governed Autonomy—as a diagnostic framework for honest self-assessment. Together, the three actions produce a gap assessment against these archetypes, a technical feasibility and cost view for the phased programme, and a board-approved governance gate that must be satisfied before any AI capability enters production.

For institutions with sufficient scale and data maturity, the cost of inaction is not stasis. It is a widening gap in service capacity, cycle time efficiency, and regulatory standing that is likely to prove progressively harder to close.

[3] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION]. MAS has proposed supervisory expectations covering AI oversight, risk materiality assessment, lifecycle controls, and a twelve-month transition period following issuance of the final Guidelines.

Glossary of Key Terms

Bounded agent: An AI agent that operates within architecturally enforced constraints defining what data it may access, what actions it may take, and when it must escalate to a human. Introduced in Section 6.

Bounded autonomy: The design principle by which an AI agent adapts to novel inputs while remaining constrained by invariant, machine-enforceable boundaries. Distinguished from both scripted automation (no adaptation) and unbounded autonomy (no constraints). Introduced in Section 6.

Control plane: A persistent infrastructure layer, borrowed from network engineering, that enforces rules, data entitlements, and risk boundaries across all downstream processes. In banking, the control plane maintains continuous state—risk surveillance, client profiling, and policy enforcement—through which all upper-tier capabilities must pass. Introduced in Section 2; developed in Section 4.

Fact-find: The structured record of a client's financial circumstances, investment objectives, and risk tolerance, used as the basis for suitability assessment. Discussed in Section 3.

Governed semantic layer: An ontology-based architectural component that encodes the institution's clients, portfolios, products, and regulatory obligations as typed objects with explicit properties, permitted actions, and access controls. It is the technical mechanism through which governance boundaries are enforced computationally. Introduced in Section 6; mapped in Exhibit 2 as Tier 3.

Governance wrapper: The composite regulatory and policy framework that defines the boundaries encoded in the governed semantic layer. It spans five classification tiers: statutory requirements, supervisory expectations, consultation-stage guidance, industry methodology, and external assurance standards. Mapped in Exhibit 2 as Tier 4; detailed in Section 7.

LLM (large language model): A class of AI model trained on large text corpora, capable of natural language synthesis, analysis, and generation. In this paper, LLM workflows operate as Tier 2 of the four-tier architecture. Introduced in Section 5.

Non-deterministic: A property of systems that do not produce identical outputs from identical inputs across invocations. LLMs are non-deterministic, meaning the same prompt can generate different responses each time it is processed. This creates tension with reproducibility and auditability requirements. Discussed in Section 5.

Ontology: A structured, formal representation of a domain—its entities, properties, relationships, and rules—that makes knowledge machine-readable and machine-enforceable. In this paper, the ontology encodes the institution's operational world: clients, portfolios, products, transactions, and regulatory obligations. Introduced in Section 6.

Persistent state: A system property whereby the control plane continuously maintains an up-to-date view of client profiles, risk classifications, and institutional policies, in contrast to batch processing, which operates on periodic snapshots. This "always-on" characteristic enables downstream AI capabilities to operate on current data and constraints. Discussed in Section 4.

Productised client journey: A standardised, end-to-end service process designed for repeatable, scalable delivery. In this paper, the periodic client review is used as the exemplar productised journey, spanning data gathering, analysis, documentation, communication, and record keeping. Introduced in Section 3.

Prompt governance framework: Version-controlled management of the structured instructions (prompts) that shape LLM behaviour and output format, including systematic output validation and audit trails. Required because prompts directly determine the content and quality of compliance-relevant outputs. Discussed in Section 5.

Risk materiality assessment: A structured evaluation of an AI use case's impact, complexity, and reliance on AI, used to calibrate the depth of governance controls—including validation, monitoring, and human oversight—proportionate to the risk each deployment presents. MAS's proposed AI risk management guidelines and December 2024 Information Paper identify this as a foundational governance discipline. Discussed in Sections 7, 8, and 9.

How to Read This Paper

The paper builds in four movements: the strategic case (Sections 1–2), the architecture (Sections 3–6), the regulatory mapping (Section 7), and the path to implementation (Sections 8–10). Each exhibit is designed as a standalone reference; Appendix A provides the full governance × capability matrix for detailed regulatory analysis.

1 The Inflection Point

The inflection is not technological alone. Three conditions have converged to make this moment structural.

First, language models have matured to the point where complex, multi-document synthesis and analysis is feasible within advisory workflows, not as a research curiosity but as a deployable capability for the periodic client review and comparable productised client journeys. Second, regulators are articulating the governance architecture that makes deployment defensible: MAS's November 2025 consultation on proposed AI risk management guidelines, whose public comment period closed on 31 January 2026, sets out expectations for board-level oversight, risk materiality assessment, and lifecycle controls encompassing machine learning, generative AI, and AI agents. The consultation proposes a twelve-month transition period following issuance of the final guidelines; MAS's response to feedback and finalisation of the guidelines are pending as of 25 March 2026 (see footnote 3). Third, engineering patterns now exist—governed semantic layers, ontology-bounded agent frameworks—that allow institutions to deploy bounded agents within predefined boundaries rather than merely as tools under direct human operation. Each condition alone is noteworthy. Their convergence is structural.

The regulatory signal deserves particular attention. The proposed guidelines are grounded in empirical observation, not conjecture: MAS's December 2024 Information Paper on AI Model Risk Management documented the governance structures, risk materiality frameworks, and lifecycle controls that leading institutions have already operationalised, providing the empirical foundation on which the consultation builds.[4] The good practices MAS observed in examination are converging with the expectations it intends to formalise.

For private banks, this convergence means the supervisory architecture is forming around the very capabilities they need to deploy. Institutions that embed governance by design before these expectations take effect will find themselves ahead of the compliance curve; those that wait will face a compounding deficit, building governance retroactively around systems already in production. Oliver Wyman's wealth management outlook reaches the same conclusion, identifying the imperative to industrialise advice delivery around AI-augmented advisors and unified data architectures as a defining competitive challenge—a call that directly parallels the operating model redesign this paper describes.[5]

---

[4] MAS *Information Paper on AI Model Risk Management* (December 2024), §§1.1–1.4. The Information Paper reports on MAS's mid-2024 thematic review examining how banks in Singapore manage AI model risks across the AI lifecycle. MAS examined practices in governance and oversight (§§3–4), model risk identification and assessment (§5), lifecycle management (§6), and emerging AI trends including generative AI deployment (§7). The Paper does not establish new supervisory expectations; it documents observed practices and identifies good practices and areas for improvement. It is classified in this paper's framework as [OBSERVED PRACTICE], a distinct category from the five regulatory classification tiers.

[5] Oliver Wyman, *10 Wealth Management Trends for 2026* (December 2025). The report identifies the imperative to redesign advice delivery around AI-augmented advisors and unified client data architectures, what it terms the "Unified Client Brain", as a defining competitive challenge and calls for industrialising growth through data-driven engines and disciplined pricing. Oliver Wyman projects that the industry will move from bespoke, manually created financial advice to an industrialised model in which a central automated engine handles the majority of processing.

2 Why Models Aren't Enough

Most private banks have deployed machine learning (ML) in compliance and risk functions: transaction monitoring under MAS Notice 626, name screening, risk classification, customer due diligence. These implementations work. They have reduced false positive rates, improved detection coverage, and satisfied supervisory expectations. They are also architecturally limited. Each model operates within a single function, optimised for a narrow task, validated on its own cycle, and governed through a bespoke oversight regime. The models are effective. The architecture connecting them—or, more precisely, the absence of such architecture—is not. This diagnosis holds regardless of scope: a bank operating machine learning in both compliance and commercial functions such as portfolio analytics, client segmentation, and wealth planning, but governing each deployment independently, faces the same architectural gap as one confined to compliance alone.

This matters because governance cost scales with the deployment pattern, not just the technology. Each new model introduced under the prevailing approach requires its own risk assessment, validation documentation, and monitoring framework. MAS's Technology Risk Management Guidelines expect financial institutions to establish risk management processes encompassing identification, assessment, treatment, and monitoring for their technology systems.[6] The proposed AI risk management guidelines would further expect institutions to maintain AI inventories and implement lifecycle controls for each use case, system, or model.[7] These are sound expectations. But when governance operates model-by-model, its cost scales linearly, or worse, with deployment count. MAS's thematic review of banks' AI governance confirmed the alternative: institutions that established centralised baseline standards applying to all AI across the bank, supplemented by enhanced requirements for higher risk use cases, achieved more consistent risk management than those governing each model independently.[8] For institutions seeking to embed AI across advisory workflows, not merely within compliance functions, this governance trajectory is unsustainable without architectural change.[9]

The alternative is to reconceive the machine learning infrastructure banks already operate, not as a collection of independent models but as the foundation for a control plane: a persistent

---

[6] MAS Technology Risk Management Guidelines (January 2021) [SUPERVISORY], §§3.1.7, 4.1.4, expect financial institutions to establish and maintain a sound and robust technology risk management framework encompassing risk identification, assessment, treatment, and monitoring.

[7] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §§3.4–3.5 (AI inventory), §4.5 (lifecycle controls). MAS has proposed that financial institutions maintain AI inventories capturing key attributes—including purpose, risk materiality, validation status, and lifecycle status—for each AI use case, system, or model, and implement lifecycle controls encompassing data management, evaluation, and ongoing monitoring.

[8] MAS *Information Paper on AI Model Risk Management* (December 2024), §6.1.3. MAS observed that most banks have established baseline development standards that apply to all AI, supplemented by enhanced standards for AI models assessed to have higher risk materiality, including independent validation for higher risk AI and documented peer review for lower risk AI.

[9] McKinsey & Company, "Building the AI Bank of the Future" series (2021), argues that banks need to move from siloed AI deployments to an integrated capability stack delivering "intelligent, personalized solutions and distinctive experiences at scale." The series describes the same architectural gap: legacy systems and disconnected data stacks that prevent AI from operating as enterprise infrastructure rather than a collection of point solutions. See also McKinsey & Company, "Seizing the Agentic AI Advantage" (2025), which calls for organisations to shift "from experimentation to industrialized, scalable delivery" and to scope AI initiatives "around the end-to-end reinvention of a full process or persona journey" rather than isolated use cases.

architectural layer that enforces rules, data entitlements, and risk boundaries across all downstream processes. The concept borrows from network engineering, where a control plane determines how traffic flows, as distinct from the data plane that carries it. Applied to banking, the control plane is the architectural substrate on which models, workflows, and agents operate and from which they inherit governance constraints.[10] The distinction is structural: rather than governing each capability in isolation, the institution builds a common layer that every capability must pass through.

The strategic consequence is direct. A control plane converts governance from a cost that scales with each deployment into infrastructure that accelerates deployment.

Building the control plane itself requires material upfront investment in data integration, policy encoding, and operational tooling. The economic advantage over model-by-model governance materialises only once the number of capabilities deployed on the shared infrastructure exceeds the threshold at which cumulative marginal savings offset that fixed cost. Each new AI capability—a suitability review assistant, a document synthesis engine, a client communication drafter—inherits the control plane's rules, entitlements, and audit architecture rather than requiring its own governance stack. The marginal governance cost of each subsequent deployment declines as a result. This is the mechanism that makes it economically feasible to extend AI beyond isolated compliance functions into the full span of client-facing operations.

> **What this means for leaders:** The strategic question is not how many AI models to deploy, but whether to build the common governance infrastructure that makes every subsequent deployment faster, cheaper, and audit ready by design.

That extension raises the question this paper turns to next: if AI is infrastructure, infrastructure for what? The answer—productised client journeys—is the subject of Section 3.

3 The Productised Client Journey

Of the dozens of workflows that constitute private banking operations, the periodic client review—the reassessment of whether a client's portfolio remains suitable given their objectives, risk tolerance, and circumstances—is among the most consequential and most resource intensive. It recurs across every client relationship on a defined cycle, engaging suitability obligations that MAS's Guidelines on Fair Dealing expect institutions to discharge through periodic review of customer profiles and portfolios, taking into account any change in circumstances and financial objectives.[11] It spans the full operating model: data gathering, analysis, documentation, client communication, and

[10] The control plane concept is borrowed from network engineering, where a control plane determines how traffic flows across distributed infrastructure, as distinct from the data plane that carries it. The pattern has been independently identified as an emerging architectural principle for AI governance in financial services. See Digital Twin Consortium FinTech Working Group, "Financial Services, AI and Complex Systems" (February 2026), describing "a new control plane providing 'Always On' visibility over the source and destination of data" as a prerequisite for AI integration in financial systems; FINOS (Fintech Open Source Foundation), *AI Governance Framework* (2024), applying control plane architecture to AI governance across financial institutions.

[11] MAS Guidelines on Fair Dealing (May 2024) [SUPERVISORY], §3.3.6. MAS expects that representatives should follow through on their commitment to perform periodic reviews of customers' profiles and portfolios, considering any change in circumstances and financial objectives to provide updated analyses and recommendations.

regulatory record keeping. If AI is to be infrastructure rather than a point solution, this is where it must prove its value.

Exhibit 1: The Periodic Client Review (Current State) vs AI-Augmented Journey

Stage	Current State	AI-Augmented State
1. Trigger & Scheduling	Review cycle initiated manually or by calendar prompt. Compliance or operations team maintains spreadsheets of review dates. Scheduling is reactive, with overdue reviews a persistent audit finding.	**Control plane** monitors client risk signals continuously—portfolio drift, life events, market movements—and triggers reviews dynamically based on materiality, not calendar alone. Human role: compliance officer sets trigger thresholds and reviews exception queue.
2. Data Gathering & Profile Assembly	Analyst assembles portfolio positions from custody systems, pulls transaction history, cross-references market data, and retrieves the client's last fact-find—the structured record of their financial circumstances, objectives, and risk tolerance. Typical elapsed time: multiple days. Key pain point: data sits across siloed systems, requiring manual reconciliation.	**Control plane** maintains a continuously updated client profile integrating portfolio, transaction, and market data. **LLM workflow** synthesises these inputs into a structured review package, flagging material changes since the last review. Human role: analyst verifies the package for completeness, flags anomalies the system may have missed. Elapsed time compresses from days to hours.
3. Suitability Analysis	Relationship manager (RM) manually assesses whether the current portfolio remains aligned with the client's investment objectives, risk tolerance, and financial circumstances. Analysis relies on the RM's experience and familiarity with the client, supported by the assembled data. Time intensive for complex, multi-asset portfolios.	**LLM workflow** generates a preliminary suitability assessment, mapping allocations against objectives, flagging concentration risks, and identifying positions that may have drifted beyond tolerance. **Human role remains decisive: the RM reviews the machine-generated analysis, applies contextual judgement, and makes the suitability determination.**
4. Documentation Preparation	RM or analyst drafts the review memo and suitability report manually, often using prior reports as templates. Documentation quality	**LLM workflow** drafts the review memo and suitability report from the structured review package and the RM's determination, applying the institution's documentation

Stage	Current State	AI-Augmented State
	varies by individual. Elapsed time: hours to a full day per client.	standards. **Bounded agent** enforces required disclosures and regulatory language. Human role: RM reviews, edits, and approves the documentation before it is finalised.
5. Client Communication	RM prepares meeting materials, talking points, and a recommendation note. Preparation is largely manual and often compressed into the hours before the client meeting.	**LLM workflow** generates personalised meeting materials, and a draft recommendation note calibrated to the client's communication preferences and the review findings. Human role: RM reviews and approves all client-facing communication.
6. Record Keeping & Audit Trail	Documentation is filed across multiple systems: CRM, document management, compliance records. Completeness depends on manual discipline. Audit readiness is verified retrospectively.	**Bounded agent** routes approved documentation to a staging area and generates a time-stamped audit trail linking every step: data inputs, analysis, RM decisions, client communications. Human role: compliance verifies completeness before the record is finalised.

Exhibit 1 maps the periodic client review across six stages, contrasting the current manual workflow against an AI-augmented design. The transformation is not uniform. At some stages, AI compresses elapsed time dramatically. At others, it changes who does what without eliminating the human role. The architecture's value lies in both.

The most dramatic compression occurs in data gathering and profile assembly. In the current state, an analyst spends days reconciling portfolio positions, transaction histories, and market data from siloed systems. In the augmented state, the control plane has already maintained a continuously updated client profile, and the LLM workflow synthesises the relevant data into a structured review package, flagging material changes since the last assessment. The analyst's role shifts from assembly to verification: confirming completeness and surfacing anomalies the system may have missed. Preparation that consumed days compresses to hours, with the actual compression depending on portfolio complexity, the number of data sources integrated into the control plane, and the maturity of the institution's data architecture. The shift from calendar-driven to risk-driven trigger scheduling introduces its own governance requirement: compliance must periodically perform an affirmative reconciliation, validating that trigger decisions have covered the full client population and that no required reviews were missed due to data gaps, threshold miscalibration, or model blind spots. This completeness check operates as a scheduled assurance control, not a per-trigger approval, ensuring that dynamic scheduling does not introduce systematic false negatives.

The most consequential design choice, however, is at the suitability analysis stage. Here the architecture deliberately preserves the relationship manager's judgement as the decisive act. The LLM workflow generates a preliminary assessment, mapping current allocations against the client's stated objectives, identifying concentration risks, flagging positions that may have drifted beyond tolerance. But the suitability determination itself remains a human decision. This is not a concession to regulatory caution; it is a design requirement. MAS's supervisory expectations, and the institution's own fiduciary obligations, demand that the recommendation to a client rests on a reasonable basis formed by a competent professional.[12] AI compresses the preparation. The relationship manager makes the call.

Downstream, the gains compound. Documentation that took hours of manual drafting is generated from structured data and the relationship manager's recorded determination, then reviewed and approved rather than written from scratch. No AI-generated content—whether a suitability report, recommendation note, or meeting brief—reaches the client without the relationship manager's sign-off. The architecture enforces this as a human-gated action at both the documentation and communication stages. Record keeping ceases to be a manual filing exercise and becomes an automated, immutable audit trail, directly supporting the reproducibility and auditability controls that MAS has proposed in its consultation on AI risk management.[13] Before that trail is finalised as the regulatory record, compliance performs an affirmative completeness and accuracy check on the staged documentation; only after sign-off does the record become the system of record. Ongoing monitoring of finalised records operates on an exception basis; governance is embedded in the workflow at the point of record creation, not applied after the fact.

This end-to-end redesign is not the product of any single AI capability. It requires all three technical layers described in this paper working in concert: the control plane maintaining continuous data infrastructure, the LLM workflow compressing analysis and documentation, and the bounded agent enforcing boundaries, routing decisions, and recording the audit trail. The sections that follow develop each layer in turn, beginning, in Section 4, with the control plane on which the entire journey depends.

> **What this means for leaders:** The periodic client review is not the only journey AI can transform; it is the exemplar. Institutions that redesign one end-to-end productised client journey will develop the architectural pattern for every journey that follows.

[12] MAS Guidelines on Provision of Digital Advisory Services (CMG-G02, October 2018) [SUPERVISORY], §44. Under section 27 of the Financial Advisers Act, advisers must have a reasonable basis for recommending any investment product. While CMG-G02 addresses digital advisory contexts specifically, the underlying suitability obligation applies across advisory models. The Guidelines on Fair Dealing, §3.3.4, further expect institutions to verify that recommendations meet customers' needs.

[13] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §4.17. MAS has proposed that financial institutions document the AI development process to enable reproducibility and auditability, with documentation sufficiently detailed for an independent party to understand and potentially replicate the implementation and its results.

4 The Control Plane

The periodic client review mapped in Section 3 begins before any relationship manager opens a file. It begins with the control plane: the infrastructure layer that maintains a continuously updated client profile, monitors risk signals across portfolios and transactions, and determines when a review should be triggered based on materiality rather than calendar alone. This is not a speculative architecture: leading institutions already operate its constituent services—transaction monitoring engines, risk scoring models, CDD automation—in fragmented form, though the fully integrated design described here remains a target state.[14]

The control plane performs three core functions. First, continuous risk surveillance: transaction monitoring, anomaly detection, and AML/CFT screening that fulfils the statutory obligations of MAS Notice 626, which requires banks to implement systems and processes to monitor business relations and to detect suspicious, complex, or unusually large transactions.[15] The control plane reframe means these operate as persistent services, not batch processes. Second, dynamic client risk profiling: maintaining risk classifications that update as client behaviour, portfolio composition, and market conditions evolve, supporting both the ongoing due diligence that Notice 626 mandates and the commercial imperatives of suitability and personalisation.[16] Third, policy enforcement: encoding institutional rules—investment mandates, concentration limits, cross-border restrictions—as machine-readable constraints that downstream workflows inherit automatically.

The common thread is persistent state. The control plane's value lies not in any individual model's sophistication but in maintaining a continuously current view of risk posture, client profile, and institutional policy. This "always-on" characteristic is what enables the LLM and agent layers described in subsequent sections to operate with confidence in the data and constraints they inherit. It also defines the control plane as critical infrastructure, subject to the system availability, data integrity, and resilience expectations that MAS's Technology Risk Management Guidelines establish for core technology systems.[17] The model risk management disciplines embedded in the control plane—development standards, independent validation, ongoing monitoring, and governance

[14] MAS *Information Paper on AI Model Risk Management* (December 2024), §§2.1–2.4. MAS observed that banks use AI across financial risk management, AML/CFT transaction monitoring, fraud detection, customer engagement and servicing, and internal operational processes. The paper notes that current deployments are concentrated on assisting or augmenting humans and improving internal operational efficiencies rather than direct customer-facing applications (§§2.4, 7.1.5). The constituent services described here—transaction monitoring, risk scoring, and CDD automation—are drawn from these published supervisory observations.

[15] MAS Notice 626 on Prevention of Money Laundering and Countering the Financing of Terrorism—Banks [STATUTORY], §6.22. Banks shall implement adequate systems and processes, commensurate with the size and complexity of the bank, to monitor business relations with customers and to detect and report suspicious, complex, unusually large or unusual patterns of transactions.

[16] MAS Notice 626, §6.24 [STATUTORY]. Banks shall ensure that CDD data, documents and information are relevant and kept up to date by undertaking reviews of existing records, particularly for higher risk categories of customers.

[17] MAS Technology Risk Management Guidelines (January 2021) [SUPERVISORY], §§4.1.2, 8.1.1. MAS expects financial institutions to institute effective risk management practices to achieve data confidentiality and integrity, and to design IT systems to achieve the level of system availability commensurate with business needs.

controls—reflect globally established principles for model risk governance that have shaped supervisory expectations across major jurisdictions.[18,19]

A layered architecture amplifies the consequences of data quality failures. Because the control plane provides the foundational data on which all upper tiers operate, errors or staleness at this layer—a miscategorised risk rating, an outdated client fact-find, a delayed transaction feed—propagate upward into LLM-generated analyses, agent-mediated decisions, and ultimately into client-facing outputs. Data quality controls at the control plane must therefore be treated as critical infrastructure in their own right: anomaly detection to identify values that fall outside expected distributions, freshness monitoring to ensure data feeds reflect current positions rather than stale snapshots, and reconciliation routines that validate control plane state against authoritative source systems on a defined schedule. Without these disciplines, the architecture's productivity gains are built on unreliable foundations.[20]

The control plane excels at deterministic enforcement over structured data. But the productised client journeys it supports do not stop at surveillance and classification. Suitability analysis demands synthesis across investment policy statements, market commentary, and client correspondence. Documentation requires synthesis of unstructured inputs into coherent narratives. These capabilities, operating over language, not rules, require a different architectural layer, one that operates within the control plane's boundaries. That layer is the subject of Section 5.

5 The LLM Workflow Layer

The periodic client review demands exactly this: synthesising an investment policy statement against recent market commentary, interpreting a client email about changed circumstances, drafting a suitability report that integrates quantitative positions with qualitative judgement. These are tasks that require natural language synthesis across multiple unstructured documents, capabilities that large language models introduce and that no prior machine learning (ML) architecture could deliver at comparable quality or scale.

In the context of the Section 3 journey, LLM workflows operate at three distinct stages. At the suitability analysis stage, the LLM performs multi-document synthesis: mapping current portfolio allocations against the client's stated objectives, flagging concentration risks, and surfacing positions that may have drifted beyond tolerance. It does so by drawing on structured data inherited from the

[18] Board of Governors of the Federal Reserve System / Office of the Comptroller of the Currency, *Supervisory Guidance on Model Risk Management*, SR Letter 11-7 / OCC Bulletin 2011-12 (April 4, 2011); Bank of England Prudential Regulation Authority, *Model Risk Management Principles for Banks*, Supervisory Statement SS1/23 (May 2023, effective May 2024). SR 11-7 is the globally foundational framework for model risk management, establishing principles for model development, validation, and governance that have shaped supervisory expectations in every major jurisdiction. SS1/23 extends these principles, requiring banks to treat model risk as a risk discipline in its own right. MAS's own supervisory approach to AI model risk, as articulated in its December 2024 Information Paper, reflects the same foundational principles.

[19] MAS Notice SFA04-N02 on Prevention of Money Laundering and Countering the Financing of Terrorism—Capital Markets Intermediaries [STATUTORY], §§1.1, 6.22. Issued under section 16 of the Financial Services and Markets Act 2022, the Notice applies to all holders of a capital markets services licence under the Securities and Futures Act 2001 and imposes ongoing monitoring and transaction surveillance obligations parallel to those in Notice 626.

[20] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §4.5. MAS has proposed that institutions implement data management controls to ensure that data used across the AI lifecycle is "fit for purpose and representative, of high quality, and subject to robust data governance."

control plane and unstructured inputs—correspondence, meeting notes, external research—that structured models cannot process. At the documentation stage, the LLM drafts review memos and suitability reports from the relationship manager's recorded determination, applying the institution's documentation standards with a consistency that manual processes cannot sustain across hundreds of client relationships. These documentation workflows depend on structured instructions—prompts—that shape LLM behaviour and output format. Because prompts directly determine the content and quality of compliance-relevant outputs, they require version control, systematic output validation, and audit trails—collectively, a prompt governance framework—to ensure that changes to prompt design are traceable and that outputs remain consistent with institutional standards over time. At the communication stage, the LLM generates personalised meeting materials and recommendation notes calibrated to the client's preferences and the review findings. In each case, the human role remains decisive: the relationship manager reviews, edits, and approves before anything reaches the client or the compliance record.

These capabilities are real. So are the risks.[21] LLM workflows in regulated environments face four categories of operational risk that the architecture must address: hallucination, in which the model generates plausible but factually incorrect content that could constitute a compliance failure in a suitability report or client letter;[22] data leakage, where sensitive client data crosses insufficiently isolated model boundaries;[23] non-determinism, in which identical inputs produce varying outputs, creating tension with auditability requirements (see footnote 13); and adversarial inputs, including prompt injection and data poisoning, where malicious or compromised data causes the model to deviate from intended behaviour or degrade output quality systematically.[24] These risks are well documented: industry consensus identifies hallucination, overconfidence, and insufficient model accuracy among the priority concerns for generative AI in financial services, and supervisory observation confirms that institutions deploying generative AI have responded by limiting initial

[21] Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). "Concrete Problems in AI Safety." *arXiv:1606.06565*. This paper provides the foundational articulation of practical AI safety challenges—including unsafe exploration, distributional shift, and reward misspecification—that underpin the risk taxonomy applied to LLM workflows in this section. See also National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1 (July 2024), which provides a comprehensive taxonomy of generative AI risks—including confabulation, data privacy, and information security—that informed the risk constraints described here.

[22] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §2.2. MAS notes that generative AI "introduces risks that are less well understood and harder to mitigate, such as hallucinations that produce convincing but false information." Section 4.15 further proposes that evaluation and testing of generative AI "should cover their key failure modes," including hallucinations and generation of undesirable content.

[23] MAS Consultation Paper P017-2025 [CONSULTATION], §§1.9(b), 4.16. MAS identifies privacy risks "arising from the leakage of confidential or customer data due to the use of third-party Generative AI products or services" and proposes that deployment include data loss prevention controls and network segmentation.

[24] MAS Consultation Paper P017-2025 [CONSULTATION], §4.15. MAS has proposed that evaluation and testing of generative AI "should cover their key failure modes," including vulnerability to adversarial attacks (fn.27). The IMDA *Model AI Governance Framework for Agentic AI* [METHODOLOGY], §2.1.2, identifies prompt injection, tool misuse, and memory poisoning as agent-specific threat vectors requiring design stage mitigation.

scope, building reusable guardrails, and maintaining human-in-the-loop requirements throughout.[25,26] The architecture described in this paper formalises that pattern.

These risks are not reasons to avoid deployment. They are design constraints the architecture must resolve. The control plane provides data integrity. Human oversight provides judgement. But neither solves the boundary problem: an LLM workflow does not inherently know what actions it may take, what data it may access, or when it must stop and escalate to a human. It processes what it is given. It does not govern itself. That governance—encoding permissible actions, data entitlements, and escalation triggers into machine-enforceable boundaries—must come from a dedicated architectural layer. That layer is the subject of Section 6.

6 Ontology-Bounded Autonomy

Exhibit 2: The AI Stack, Four-Tier Architecture for Governed Autonomy in Private Banking

Tier	Representative Capabilities	Governing Instruments	Interaction
4) Governance Wrapper: Regulatory architecture that defines the boundaries encoded in Tiers 1–3	Regulatory mapping and classification; risk materiality assessment; lifecycle controls; external assurance and audit	MAS Notice 626 [STATUTORY]; MAS TRM Guidelines [SUPERVISORY]; MAS Proposed AI Risk Management Guidelines [CONSULTATION]; FEAT Principles / Veritas Methodology [METHODOLOGY]; ISO/IEC 42001 [ASSURANCE]; IMDA *Model AI Governance Framework for Agentic AI* [METHODOLOGY]	Envelops all tiers. Translates regulatory requirements and supervisory expectations into the machine-enforceable boundaries that Tier 3 encodes. Defines *what* the bounds must be.
3) Ontology-Bounded Autonomy:	Agent-mediated documentation routing;	IMDA *Model AI Governance Framework for Agentic AI*	Receives governance boundaries from Tier 4. Constrains all Tier

[25] Project MindForge, *AI Risk Management Executive Handbook* (November 2025), §1.1. The Executive Handbook, developed by a consortium of financial institutions with MAS leadership and formally launched at the 2025 Singapore Fintech Festival, identifies ten priority generative AI risks drawn from the ABS *Handbook on Generative AI Guardrails in Banking*, including hallucination/fabrication/confabulation, overconfidence, and insufficient model accuracy/soundness. The Operationalisation Handbook (January 2026) includes an updated AI risk taxonomy in Appendix B.

[26] MAS *Information Paper on AI Model Risk Management* (December 2024), §§2.4, 7.1.4–7.1.5, 7.1.9, 7.1.14. MAS observed that banks deploying generative AI adopted a more limited scope of initial deployment to understand generative AI's limitations while managing potential risks. Most banks limited generative AI to internal productivity augmentation. Banks invested in reusable enabling modules, including retrieval systems and evaluation frameworks, and maintained human-in-the-loop requirements for generative AI used in decision-making processes. Most banks established input and output guardrails using filters to manage risks relating to toxicity, bias, and sensitive information leakage.

Tier	Representative Capabilities	Governing Instruments	Interaction
Governed semantic layer encoding objects, relationships, permitted actions, access controls, and escalation triggers	permission-bounded data access; boundary-triggered workflow orchestration	[METHODOLOGY] (agent limits, identity, access controls); MAS Proposed AI Risk Management Guidelines [CONSULTATION] (human oversight, contingency plans); MAS TRM Guidelines [SUPERVISORY] (access control, least privilege)	2 outputs and Tier 1 data access through semantically typed permissions and action definitions. Provides the *mechanism* through which bounds are enforced.
2) LLM Workflow Layer: Natural language synthesis and analysis operating within the boundaries Tier 3 defines	Multi-document suitability analysis; review memo and report drafting; personalised client communication generation	MAS Proposed AI Risk Management Guidelines [CONSULTATION] (hallucination controls, reproducibility, data loss prevention); MAS Fair Dealing Guidelines [SUPERVISORY] (suitability basis); Project MindForge *AI Risk Management Executive Handbook* [METHODOLOGY]	Inherits data and risk signals from Tier 1. Operates within baseline data entitlements and purpose constraints (formalised as the governed semantic layer in Horizon 3). All outputs subject to human review before client or compliance delivery.
1) Control Plane: Persistent infrastructure maintaining continuous state, risk surveillance, and policy enforcement	Transaction monitoring and AML/CFT screening; dynamic client risk profiling; investment mandate and concentration limit enforcement	MAS Notice 626 [STATUTORY] (ongoing monitoring, CDD); MAS Notice SFA04-N02 [STATUTORY] (parallel AML/CFT obligations for capital markets intermediaries); MAS TRM Guidelines [SUPERVISORY] (system availability, data integrity)	Foundation layer. Provides structured data, risk classifications, and policy constraints that all upper tiers inherit. Operates as persistent, always-on infrastructure.

Read upward, the architecture builds from Tier 1 (foundation) to Tier 4 (governance wrapper).

The Governance Wrapper (Tier 4) envelops all three technical tiers. Each tier inherits constraints from the tier above and data or capabilities from the tier below. Arrows of dependency run upward for data; arrows of constraint run downward for governance.

The layer that encodes those boundaries is a governed semantic layer, represented as Tier 3 in Exhibit 2. Left unbounded, an LLM processing a suitability review could access any client's data, submit a report without review, or act beyond its scope. The institution would discover what the agent did after the fact. The distinction between a useful capability and an ungovernable one is not intelligence; it is boundary.

An ontology, in this context, is a structured representation of the institution's operational world[27]: its clients, portfolios, products, transactions, and regulatory obligations, together with the relationships between them. Each entity is a typed object with explicit properties, permitted actions, and access controls. A client object links to portfolio objects, which link to products and transactions; each link carries permissions governing who—and what—may traverse it. An agent operating within this ontology can only act on objects it has permission to access, through actions the ontology permits. This is not automation following a script. It is bounded autonomy: the agent adapts to novel inputs while remaining architecturally constrained in what it may do.

Critically, the ontology must define not only permitted actions and data entitlements but also the input conditions under which the agent must escalate rather than adapt, distinguishing inputs that fall within the agent's competence boundary from those that exceed it and require human intervention. All novel input adaptations are subject to post hoc audit, enabling the institution to verify that the agent's behaviour remained within sanctioned boundaries even where the input was unanticipated.

Where an input falls outside the ontology's coverage entirely—a scenario not addressed by any defined object type, relationship, or action permission—the architecture must enforce a fallback procedure that halts autonomous processing and routes the matter to a qualified human, rather than permitting the agent to extrapolate beyond its governed domain.

The distinction from unbounded autonomy is precise. Consider two examples from the productised client journey mapped in Section 3. A documentation agent can draft a suitability report (a permitted action) using the designated client's profile and portfolio data (permitted objects) but cannot access another client's records, because access controls enforce boundaries at the object level. Nor can it submit the report without the relationship manager's approval, because submission is a human-gated action with an explicit escalation trigger. A record keeping agent can route approved documents to the correct systems but cannot modify their content. When either agent

[27] Gruber, T.R. (1993). "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition*, 5(2), 199–220. This is the canonical formulation of computational ontology as a formal, explicit specification of a conceptualisation. Ontology-based governance is already operationalised in financial services through the Financial Industry Business Ontology (FIBO), developed by the EDM Council and standardised by the Object Management Group. FIBO encodes financial entities—instruments, legal entities, counterparties—as typed objects with formal properties and relationships, demonstrating the feasibility of the approach at industry scale. See EDM Council / Object Management Group, *Financial Industry Business Ontology (FIBO)* (first published 2014; continuously maintained), https://spec.edmcouncil.org/fibo/.

encounters conditions outside its defined boundaries, it does not improvise. It escalates. The architecture fails safe, not open.[28]

This approach directly implements governance principles that IMDA's Model AI Governance Framework for Agentic AI identifies as foundational: bounding risks through design by "defining appropriate boundaries at an early stage, such as limiting the agent's access to tools and systems," and ensuring agents are "traceable and controllable through establishing robust identity management and access controls."[29] The MindForge Handbook reinforces these principles from a financial services perspective, recommending least privilege agent access, modular agent certification for reuse across use cases, dynamic data governance over static point-in-time controls, and kill switches for automated interruption of agents taking unintended actions.[30] The governed semantic layer is the technical mechanism through which these principles become enforceable, not as a policy reviewed quarterly, but as a permission boundary enforced computationally at every invocation.[31] Exhibit 2 positions this layer within the full stack: Tier 3 sits between the LLM workflows it constrains (Tier 2) and the governance wrapper that defines its boundaries (Tier 4), drawing structured data from the control plane (Tier 1).

The governed semantic layer must also encode purpose limitation at the object level. The control plane processes personal data continuously under statutory authority; MAS Notice 626 mandates client data collection for AML/CFT compliance, for which the PDPA provides a consent exemption.[32] That exemption covers compliance functions; repurposing due diligence data for personalised investment recommendations operates under a different consent basis. Purpose constraints must therefore attach to data objects, not merely access permissions, so that workflows crossing the

---

[28] Leveson, N.G. (2011). *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press. Leveson's STAMP framework establishes that safety is a control problem, not a failure problem. Systems must be designed so that unsafe states are architecturally unreachable. See also National Institute of Standards and Technology, *Zero Trust Architecture*, NIST Special Publication 800-207 (August 2020). The "fail safe, not open" design principle in this paper draws on both traditions.

[29] IMDA, *Model AI Governance Framework for Agentic AI*, Version 1.0 (22 January 2026), §2.1.2. The Framework recommends that organisations bound risks through design by defining agents' limits and permissions, including limits on the agent's access to tools and systems, the agent's autonomy, and the agent's area of impact.

[30] Project MindForge, *AI Risk Management Operationalisation Handbook* (January 2026), Future Perspectives [METHODOLOGY]. The Handbook recommends restricting the privileges, tool access, data access, and capabilities of agentic systems to the lowest level necessary for each use case (§3.1); tracking agentic-specific identifiers including tools, components, and limitations in the AI inventory (§2.5); implementing dynamic data governance focused on real-time oversight rather than static, point-in-time controls (§3.2); and providing kill switches and automated interruption capabilities proportionate to risk materiality (§3.4). The Handbook treats these as governance considerations mapped to its existing seventeen-Consideration framework, noting that "governance considerations in managing Agentic AI remain an emerging field." The section references in parentheses refer to the Handbook Subsections to which each agentic governance recommendation is mapped, not to separate provisions.

[31] IMDA, *Model AI Governance Framework for Agentic AI*, §2.1.2. The Framework recommends that organisations "define policies that give agents only the minimum tools and data access needed for it to complete its task." The governed semantic layer operationalises this principle through object-level access controls and action-type restrictions.

[32] MAS Notice 626 on Prevention of Money Laundering and Countering the Financing of Terrorism—Banks, §§13.2–13.4 [STATUTORY]. For the purposes of complying with Notice 626, a bank is not required to provide individuals with access to personal data held for AML/CFT purposes or information about how such data has been used or disclosed (§13.2). However, upon request, the bank shall provide access to and correction rights for specified categories of personal data including full name, identification number, residential address, date of birth, and nationality (§13.3). Critically, for the purposes of complying with Notice 626, a bank may collect, use and disclose personal data without the individual's consent (§13.4). Parallel provisions apply to capital markets intermediaries under MAS Notice SFA04-N02, §§12.2–12.4. This consent exemption applies to AML/CFT compliance functions; it does not extend to the use of client data for advisory, commercial, or AI-driven recommendation purposes outside the scope of AML/CFT compliance.

advisory boundary do so on a lawful basis.[33,34] MAS's proposed AI risk management guidelines reinforce this, cross-referencing the PDPC's Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems.[35] For cross-border clients, equivalent data protection regimes constitute an additional constraint.

A reasonable question is what distinguishes bounded autonomy from conventional automation with guardrails. A scripted automation executes a predetermined sequence regardless of input variation; a bounded agent operates—selecting, synthesising, adapting—within constraints that remain invariant at runtime, modified only through governed change management. The principle has a two-decade pedigree in multi-agent systems research, where policy-governed ontologies defined the permissible action space within which agents could operate.[36] The ontology's boundaries are configuration, not physics; updates require the same change management rigour applied to the agents themselves, including versioning, impact assessment, and audit trail. The governed semantic layer provides the mechanism for bounding autonomy. Governed autonomy is a dialectic resolved in architecture: autonomy expands capacity and coverage; governance preserves controllability and accountability. The regulatory architecture that determines what those bounds should be—which obligations flow from statutory notices, which from supervisory guidelines, which from industry methodology—is the subject of Section 7.

---

[33] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §4.5(f). MAS has proposed that institutions implement "appropriate data privacy measures based on relevant regulatory requirements and guidance relating to data privacy; and obtaining permission when using sensitive personal data of customers or employees to train the AI models, or allowing AI to access such data in real-time." Footnote 16 to §4.5(f) cross-references the Personal Data Protection Commission's Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems as the relevant data privacy guidance.

[34] Personal Data Protection Commission, *Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems* (1 March 2024) [METHODOLOGY], §§8–10 (deployment obligations), §§4–6 (development exceptions). Part IV addresses how PDPA consent, notification, and accountability obligations apply when organisations deploy AI systems that process personal data (§§8.1, 9.1–9.2, 10.1–10.3), with the level of governance detail proportionate to the risks in each use case, including the level of autonomy of the AI system. Part III addresses when organisations may rely on the Business Improvement Exception or Research Exception for AI development without consent. MAS's proposed AI risk management guidelines cross-reference this document as the relevant data privacy guidance for AI in financial services (P017-2025, §4.5(f) fn.16). The Advisory Guidelines are issued by the PDPC, not by MAS; they do not carry supervisory force over MAS-regulated financial institutions.

[35] Veritas Consortium, *FEAT Principles Assessment Methodology* (Document 3), Table 5.2 [METHODOLOGY]. The Methodology identifies data protection laws, including Singapore's Personal Data Protection Act (No. 26 of 2012), as a key regulatory consideration at the data input phase of AI use cases: "AIDA applications' reliance on internal and/or external data means that sector agnostic compliance requirements such as data protection regulation remains relevant." The PDPA governs the collection, use and disclosure of personal data by private sector organisations in Singapore. It does not contain AI-specific provisions; its framework applies to AI systems through its general obligations.

[36] Bradshaw, J.M., Jung, H., Kulkarni, S., Johnson, M., Feltovich, P., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Suri, N., Taysom, W., and Uszok, A. (2005). "Toward Trustworthy Adjustable Autonomy in KAoS." In R. Falcone et al. (eds.), *Trusting Agents for Trusting Electronic Societies*, LNAI 3577, pp. 18–42. Springer. The KAoS framework used OWL ontologies to define policy-governed constraints on agent behaviour, limiting what actions agents could take, what resources they could access, and when they must escalate to a human operator. Bradshaw and colleagues subsequently developed the concept of "adjustable autonomy," in which the degree of agent independence is dynamically calibrated based on context, competence boundaries, and policy constraints, the same principle underlying the escalation triggers and fallback procedures described in this section. See Bradshaw, J.M., Feltovich, P.J., Jung, H., Kulkarni, S., Taysom, W., and Uszok, A. (2004). "Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction." In M. Nickles et al. (eds.), *Agents and Computational Autonomy*, LNAI 2969, pp. 17–39. Springer.

7 The Governance Wrapper

The governance wrapper is a composite, polycentric structure: overlapping authorities with defined jurisdiction together govern the system[37], spanning five classification tiers—statutory, supervisory, consultation, industry methodology, and external assurance.[38] As Exhibit 3 summarises, each AI capability intersects with a distinct subset of these instruments, and no single instrument covers the full stack end-to-end.

These tiers carry different binding force, and that hierarchy matters for institutional design. Statutory instruments (MAS Notices) are binding; they define non-negotiable constraints that the architecture must enforce as hard boundaries. Supervisory instruments (MAS Guidelines) define the operating expectations against which institutions are assessed; non-compliance is not a legal breach but invites supervisory scrutiny and remediation. Consultation-stage guidance signals the regulator's direction of travel, not settled obligation; institutions should build with awareness of that trajectory without treating proposed provisions as current requirements. Industry methodologies operationalise good practice and provide implementation frameworks that give practical shape to supervisory expectations. External assurance standards evidence governance maturity to boards and examiners but do not substitute for MAS expectations.

These five tiers define what institutions must or should do. A further source, MAS's December 2024 Information Paper on AI Model Risk Management, documents what leading institutions are actually doing. Based on a mid-2024 thematic review, MAS observed governance structures that this paper interprets as closely aligned with the architecture's governance principles: cross-functional oversight forums, centralised AI inventories with automated lifecycle tracking, and risk materiality assessments calibrated along dimensions of impact, complexity, and reliance on AI, the same dimensions MAS has proposed formalising.[39] The Information Paper does not carry the force of a Notice or Guideline; it reports observed good practice, not binding requirements. But supervisory observations from examination carry distinct practical weight: institutions building governance architecture now would do well to treat them as a leading signal of where formal expectations are likely to settle.

---

[37] Ostrom, E. (2010). "Beyond Markets and States: Polycentric Governance of Complex Economic Systems." *American Economic Review*, 100(3), 641–672. Nobel Prize Lecture. Ostrom's research demonstrated that complex systems are most effectively governed by multiple overlapping decision centres with defined jurisdiction and mutual accountability, a structure the governance wrapper instantiates through its five classification tiers.

[38] The concept of composite, multi-layered governance architecture for AI in financial services has independent precedent. See FINOS (Fintech Open Source Foundation), *AI Governance Framework* (2024), structuring AI governance across identity and access management, data protection, model oversight, and regulatory compliance as overlapping control domains; Financial Services Sector Coordinating Council and Cyber Risk Institute, *Financial Services AI Risk Management Framework* (February 2026), introducing 230 control objectives across governance, data, model development, validation, monitoring, third-party risk, and consumer protection as an integrated operational architecture standard.

[39] MAS *Information Paper on AI Model Risk Management* (December 2024), §§4.1–4.2, 5.2.1–5.2.4, 5.3.1. MAS observed that banks established cross-functional AI oversight forums, maintained centralised AI inventories with automated tracking and approval checkpoints, and conducted risk materiality assessments. See also footnotes 4 and 44.

Exhibit 3: Governance × Capability Summary Matrix, AI Capabilities × Regulatory Instruments

Maps each AI capability against ten instruments across all five classification tiers, plus supervisory observations from MAS's thematic review. Each cell shows only the classification tier of the applicable instrument; a dash (—) indicates the instrument does not directly address that capability. Regulatory density increases as capabilities move closer to client-facing outcomes; suitability, communication, and audit trail carry the broadest governance coverage. The full governance × capability matrix, with cell-level regulatory provisions and section references, is provided in Appendix A. The [OBSERVED PRACTICE] column does not carry the normative force of the five classification tiers; it is included because supervisory observations from examination are a reliable indicator of where future expectations will settle (see footnote 4).

Capability	626	TRM	FD	G02	P017	FEAT	ISO	IMDA	MF	IP
Transaction monitoring	S	SV	—	—	C	M	A	M	M	OP
Client risk profiling	S	SV	SV	SV	C	M	A	—	M	OP
Suitability analysis	—	SV	SV	SV	C	M	A	—	M	OP
Document generation	S	SV	SV	SV	C	M	A	—	M	OP
Client communication	—	—	SV	SV	C	M	—	M	M	OP
Agent routing / audit trail	S	SV	SV	SV	C	M	A	M	M	OP

Key Instruments: **626** = MAS Notice 626 / SFA04-N02; **TRM** = MAS TRM Guidelines; **FD** = MAS Fair Dealing Guidelines; **G02** = MAS Digital Advisory Guidelines CMG-G02; **P017** = MAS Proposed AI Risk Management Guidelines P017-2025; **FEAT** = FEAT Principles / Veritas; **ISO** = ISO/IEC 42001:2023; **IMDA** = IMDA, *Model AI Governance Framework for Agentic AI* (government framework, not a MAS instrument); **MF** = MindForge, *AI Risk Management Handbook* (Executive Handbook Considerations and Practices, Operationalisation Handbook provides detailed guidance behind the same recommendations); **IP** = MAS *Information Paper on AI Model Risk Management* (December 2024).

Tiers: **S** = [STATUTORY]; **SV** = [SUPERVISORY]; **C** = [CONSULTATION]; **M** = [METHODOLOGY]; **A** = [ASSURANCE]; **OP** = [OBSERVED PRACTICE].

The critical insight from this mapping is instrument complementarity. No single regulation governs AI in private banking end-to-end, yet the cumulative effect is dense and directionally convergent, though important elements, notably the proposed AI risk management guidelines, remain in consultation. Regulatory density increases as AI capabilities move closer to customer outcomes. Transaction monitoring draws obligations from statutory and supervisory instruments but falls outside the scope of Fair Dealing and the Digital Advisory Guidelines, which govern advisory and communication conduct. Suitability analysis and client communication drafting, by contrast, attract governance obligations from virtually every tier: statutory record keeping, supervisory conduct expectations, proposed fairness and transparency requirements, industry methodology for bias assessment, and assurance standards for lifecycle management. This convergence is a feature of the

regulatory architecture, not a deficiency: it reflects the proportionality principle that the closer an AI capability operates to a client's financial position, the more governance instruments must constrain it. The governance wrapper converts this landscape from a compliance catalogue into design constraints encoded in the governed semantic layer: permission boundaries, escalation triggers, data entitlements, and audit requirements enforced by construction.

The statutory foundation anchors this architecture with non-negotiable constraints. MAS Notice 626 and Notice SFA04-N02 impose binding AML/CFT obligations on banks and capital markets services licence holders, covering customer due diligence, ongoing monitoring, risk-based profiling, record keeping, and suspicious transaction reporting. These requirements are technology neutral: they apply identically whether the institution deploys machine learning models or manual processes. For the architecture, the control plane must enforce Notice 626 constraints as hard boundaries: risk scoring that cannot be overridden without documented escalation, monitoring thresholds that persist regardless of model updates, and audit trails satisfying the statutory five-year record keeping requirement.[40]

The supervisory layer shapes how AI may augment, but not replace, the relationship manager's judgement. MAS expects institutions to ensure the suitability of recommendations, with representatives fully documenting the basis of each.[41] The Digital Advisory Guidelines extend this to algorithm-driven advice, expecting governance arrangements that mitigate the risk of algorithmic fault or bias affecting clients.[42] For the LLM workflow layer, these expectations mean any AI-generated suitability analysis must preserve a demonstrably human reasonable basis: the system augments, the relationship manager decides, the institution remains accountable.

MAS has proposed a comprehensive AI-specific framework through its consultation on AI risk management guidelines (P017-2025), setting out expectations for lifecycle controls, including data management, fairness, transparency, human oversight, and pre-deployment review proportionate to risk materiality.[43] The proposed guidelines build on what MAS's thematic review observed

[40] MAS Notice 626, §12.3(a) [STATUTORY], requires banks to retain CDD information for at least five years following the termination of business relations or the completion of a transaction; §12.3(b) requires transaction records, including information needed to explain and reconstruct the transaction, to be retained for at least five years following the completion of the transaction.

[41] MAS Guidelines on Fair Dealing—Board and Senior Management Responsibilities for Delivering Fair Dealing Outcomes to Customers, §§2.4, 3.3 [SUPERVISORY]. §2.4.2 expects institutions to adjust marketing approaches to suit target customer segments and to identify customer profiles within a segment for which a product may not be suitable. §3.3.1 expects institutions to demonstrate that scoring and risk profiling methodologies have been "properly designed, tested and validated." §3.3.3(d) expects representatives to "fully document the basis of recommendation." See also footnote 11.

[42] MAS Guidelines on Provision of Digital Advisory Services (CMG-G02), §§27, 44–46 [SUPERVISORY]. §44 addresses the statutory suitability obligation under section 27 of the Financial Advisers Act. §27 of the Guidelines notes that, given the additional risks of algorithm-driven advice, "it is important that digital advisers put in place adequate governance and supervisory arrangements to effectively mitigate these risks." See also footnote 12.

[43] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §§4.1–4.19. Key proposed expectations span data management, fairness, transparency and explainability, human oversight, third-party AI management, evaluation and testing, and reviews, monitoring and change management.

institutions already practising: risk materiality assessments mapped to the depth of validation and monitoring required.[44]

They draw explicitly on the FEAT Principles and the Veritas assessment methodology for evaluating fairness in AI-driven decisions.[45,46] The MindForge Handbook translates these expectations into seventeen actionable Considerations, forming the most comprehensive industry consensus framework for AI risk management in financial services.

Published by a consortium of 24 financial institutions under MAS leadership, the Handbook comprises three documents: an Executive Handbook of Considerations and Practices, an Operationalisation Handbook providing detailed implementation guidance and a mapping to MAS's proposed Guidelines, and Implementation Examples documenting institutional experience. It spans scope and oversight, AI risk management, AI lifecycle management, and enabling capabilities—a four-section structure aligned to MAS's proposed Guidelines. Its scope explicitly covers traditional AI, generative AI, and agentic AI, with a dedicated treatment of agentic governance considerations. MindForge positions governance and adoption as complementary, stating that "widespread, rapid, and useful innovation in AI requires robust risk management and good governance." [47] Institutions adopting the four-tier architecture will find the Handbook's lifecycle framework (Considerations 7–15) a natural operationalisation companion, a governance checklist encodable into deployment gates within the control plane.

Three further elements complete the wrapper. ISO/IEC 42001:2023 provides a certifiable AI management system framework.[48] Certification neither constitutes a MAS requirement nor satisfies supervisory expectations, but it offers a structured mechanism for demonstrating governance maturity to boards and examiners. For bounded agents, IMDA's Model AI Governance Framework for Agentic AI establishes agent-specific principles: limits on access, autonomy, and scope of impact; accountability allocation across the value chain; and meaningful human oversight at significant

[44] MAS *Information Paper on AI Model Risk Management* (December 2024), §§5.3.1–5.3.2. MAS observed that risk materiality assessments were critical for banks to calibrate their approach to AI risk management, including mapping risk materiality to the depth and scope of validation and monitoring required. Most banks considered both quantitative and qualitative risk dimensions grouped into impact (financial, operational, regulatory, reputational), complexity (nature of AI model or system, novelty of use case), and reliance (autonomy granted to AI, involvement of humans in the loop). Most banks also established processes to review that risk materiality ratings remain appropriate over time.

[45] MAS FEAT Principles (Fairness, Ethics, Accountability, Transparency), updated February 2019 [METHODOLOGY]. P017-2025, §2.8 notes that the proposed guidelines "build on the FEAT principles" [CONSULTATION].

[46] Veritas Consortium, *FEAT Fairness Principles Assessment Methodology* (Document 1), §§3.1–3.2 [METHODOLOGY]; *FEAT Fairness Principles Assessment Case Studies* (Document 2), covering customer marketing and credit scoring use cases [METHODOLOGY]. Fairness Principle F1 states: "Individuals or groups of individuals are not systematically disadvantaged through AIDA-driven decisions, unless these decisions can be justified."

[47] Project MindForge, *AI Risk Management Handbook* [METHODOLOGY]. The Handbook comprises three documents: the Executive Handbook (November 2025), the Operationalisation Handbook (January 2026), and the Implementation Examples (January 2026). The Implementation Examples document governance practices across banking, insurance, and investment management, including a private bank. MAS's consultation on AI risk management guidelines cross-references the Handbook's planned release by January 2026 (P017-2025, §2.8 fn.5). The Operationalisation Handbook includes a mapping of MindForge Considerations to the proposed Guidelines (Appendix D). See also footnote 25.

[48] ISO/IEC 42001:2023, *Information technology—Artificial intelligence—Management system* [ASSURANCE]. The standard provides a certifiable management system compatible with ISO/IEC 27001 (information security) and ISO 9001 (quality management).

checkpoints.[49] The IMDA Framework carries no supervisory force over MAS-regulated institutions; it is included as an industry governance framework for the design constraints bounded agents must satisfy.

Data protection obligations intersect every tier of this architecture; the architectural implications, including purpose limitation constraints that the governed semantic layer must encode at the object level, are addressed in Section 6.

A distinct but related challenge is the architecture's reliance on third-party AI infrastructure—foundation models, cloud platforms, and specialist tooling—for which data confidentiality, model integrity, and service continuity depend on controls beyond the institution's perimeter. MAS's Guidelines on Outsourcing expect banks to evaluate the risks and materiality of outsourcing arrangements, with the board approving a risk evaluation framework and setting risk appetite; the TRM Guidelines further expect institutions to manage technology risks arising from third-party services, including those involving confidential customer data.[50] For the four-tier architecture, this means governance constraints must extend across organisational boundaries: the governed semantic layer's permission boundaries apply to data processed by third-party models, not only internally hosted systems. Without prescribing vendor selection, institutions should ensure that AI-related third-party arrangements, particularly those material to client-facing operations, satisfy due diligence, contractual, and board oversight obligations before deployment.[51] Where foundation models are accessed through external APIs, the API boundary should be treated as a governance checkpoint, verifying that data egress controls, output validation, and audit trail requirements are contractually enforceable.

Governance becomes infrastructure. Privacy by Design moved from a voluntary design principle to a binding legal requirement when GDPR Article 25 codified governance by design as law.[52] That trajectory validates this paper's positioning of governance as architecture—strategically sound and

---

[49] IMDA, *Model AI Governance Framework for Agentic AI*, Version 1.0, January 2026, §§2.1.2, 2.2. The Framework is issued by the Infocomm Media Development Authority (IMDA), a separate government agency from MAS; it does not carry supervisory force over MAS-regulated financial institutions and is included in the governance wrapper as an industry governance framework for agentic AI, not as a regulatory instrument. The Framework defines agent limits (access to tools and systems, autonomy, area of impact), agent identity management, and accountability allocation across the agentic AI value chain. See also footnotes 2, 29, and 31.

[50] MAS Guidelines on Outsourcing (Banks) (December 2023) [SUPERVISORY], §§3.1, 3.2. MAS expects the board of directors to approve the outsourcing policy and framework for evaluating the risks and materiality of all outsourcing arrangements; the framework should include risk appetite, risk assessment methodology, and escalation procedures. MAS Technology Risk Management Guidelines (January 2021) [SUPERVISORY], §13.1.1 et seq., expect institutions to assess and manage technology risks arising from the use of external service providers and from internet of things and other connected devices and systems.

[51] MAS Consultation Paper P004-2026, *Proposed Guidelines on Third-Party Risk Management* (March 2026) [CONSULTATION]. The proposed Guidelines extend supervisory expectations beyond outsourcing to all third-party arrangements (§1.2), including third-party AI services. Annex A §2.4 fn. 15 explicitly references the proposed AI Risk Management Guidelines (P017-2025), signalling that third-party AI governance will be addressed through a dedicated supervisory instrument. Annex A §§6.1–6.39 set out lifecycle expectations for onboarding, ongoing monitoring, and termination of third-party arrangements. The consultation closes on 20 April 2026.

[52] Cavoukian, A. (2009/2011). *Privacy by Design: The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario, Canada. Privacy by Design established the principle that privacy protections should be embedded proactively into the architecture of systems and business practices, not added retroactively as a compliance overlay. This design principle was subsequently codified into binding law through Article 25 of the European Union's General Data Protection Regulation (GDPR), which requires data protection by design and by default.

regulatorily anticipatory. The architecture is now defined; what remains is how institutions at different stages of maturity adopt it.

> **What this means for leaders:** The governance wrapper is the design specification for your AI architecture. Commission a mapping of your obligations across all five classification tiers before any capability deployment proceeds.

8 Institutional Archetypes

Exhibit 4: Institutional Maturity Archetypes

Illustrative synthesis based on the public regulatory and industry sources cited in the footnotes to this paper. Not a depiction of any institution's internal operating model or supervisory engagement.

	Compliance First	Workflow Augmented	Governed Autonomy
AI deployment scope	ML confined to compliance functions: transaction monitoring, name screening, sanctions filtering. Models operate as discrete tools within siloed processes.	Control plane operational across client-facing workflows. LLM layer deployed for suitability analysis, document synthesis, and client communication drafting.	Full four-tier architecture (control plane, LLM workflow layer, ontology-bounded agents, and governance wrapper) deployed across the productised client journey. Agents execute multi-step tasks within governed boundaries.
Architecture	Pre-control plane. ML operates as discrete compliance models, not as an integrated persistent state layer. No governed semantic layer. Each model is independently deployed and maintained.	Tiers 1–2 operational (control plane plus LLM workflow layer). Persistent state enables cross-process awareness. Governed semantic layer absent or nascent.	All four tiers operational, including the governed semantic layer. The ontology encodes objects, relationships, permissions, and escalation triggers as machine-readable constraints.
Governance posture	Model-by-model governance. Each deployment requires bespoke validation, documentation, and oversight arrangements.	Governance embedded in the control plane. Policies enforced structurally, through the ML layer, rather than through manual review of each model output. Governance	Governance operates as infrastructure. The governed semantic layer converts regulatory requirements and institutional policies into boundary conditions that

	Compliance First	Workflow Augmented	Governed Autonomy
	Governance scales linearly with model count.	scales with architecture, not headcount.	constrain agent behaviour before execution, not after.
Client experience impact	Clients experience no AI-driven change. Service quality and responsiveness remain a function of individual RM capacity. Periodic reviews follow legacy timelines.	Clients experience faster, more consistent service. Review preparation compresses from days to hours. Documentation quality improves. The RM remains the sole interface, now better equipped.	Clients benefit from continuous, behind-the-scenes preparation. Bounded agents surface portfolio risks, prepare review materials, and draft communications between scheduled touchpoints, but these outputs are queued for RM approval before any client-facing action. No agent-surfaced output reaches or affects the client without the RM's sign-off. The RM's role shifts from preparation to judgement, supported by richer and more timely inputs.
Regulatory standing	Meets current statutory requirements. Well-understood by supervisors. Does not address emerging supervisory expectations around AI risk management or fairness assessment.	Meets statutory requirements and aligns with supervisory expectations for AI governance, explainability, and human oversight. Positioned to accommodate proposed AI risk management guidelines as they formalise.	Designed to satisfy the full composite governance wrapper: statutory, supervisory, and emerging frameworks including bounded agent governance expectations. Regulatory defensibility is an architectural property, not an operational burden.
Defining limitation	Does not deploy AI for client-facing value creation. Cannot compress cycle times, expand RM capacity, or improve service consistency through technology. AI remains a cost of compliance, not a source of	Does not permit autonomous AI action. Every AI output, intermediate and final, passes through a human gate before reaching the client or affecting a decision. Productivity gains are constrained by	Does not operate without constraints. Bounded agents are structurally prevented from acting outside ontology-defined permissions. The limitation is deliberate: autonomy is calibrated to the boundaries that governance defines.

	Compliance First	Workflow Augmented	Governed Autonomy
	competitive advantage.	human throughput at the approval layer.	
Economic profile	AI is a cost centre. Expenditure scales with regulatory scope—more rules, more models, more cost—with no offsetting revenue or capacity effect.	AI compresses cycle times and improves output quality, generating measurable productivity gains. Direct labour cost per client review can decline as LLM workflows reduce manual preparation and documentation effort; total cost of the review depends additionally on infrastructure investment in the control plane and LLM layer. RM capacity expands where time saved on preparation and documentation exceeds the overhead of reviewing AI-generated outputs. Returns are operational, not yet structural.	AI reshapes unit economics. The architecture is designed so that capacity expands as bounded agents absorb preparatory and monitoring tasks that previously required dedicated human effort for each additional client. The cost of serving an incremental client can decline because agent-mediated workflows draw on shared infrastructure (the governed semantic layer, control plane, and governance wrapper) the costs of which do not scale proportionately with client count.

The preceding sections defined what a governed AI architecture looks like and how regulatory instruments map onto it. The question that matters for any leadership team is simpler: where does our institution stand today?

Not every private bank will, or should, pursue governed autonomy immediately. The three archetypes presented in Exhibit 4 represent coherent strategic positions, each with distinct capabilities, limitations, and economic consequences. They are framed here as strategic postures, though in practice many institutions will experience them as transition states along a maturity path. For institutions that lack the fit conditions for governed autonomy—sufficient scale, data maturity, reusable workflow volume, and governance infrastructure readiness—the Workflow Augmented archetype is a legitimate and defensible steady state, not a waypoint.

MAS's thematic review confirmed this spectrum empirically, observing that AI model risk management maturity varies significantly across institutions, from banks still maintaining AI

inventories in spreadsheets to those operating automated lifecycle tracking with cross-functional governance forums.[53]

A candid caveat applies to the Governed Autonomy archetype specifically: its regulatory defensibility depends on frameworks either recently closed for consultation or not yet tested in supervisory practice. MAS's proposed Guidelines on AI Risk Management closed for consultation on 31 January 2026; finalisation is pending. IMDA's Model AI Governance Framework for Agentic AI, published in January 2026, has not yet been tested in MAS supervisory practice. The trajectory is nonetheless grounded in empirical observation: MAS's December 2024 Information Paper documented the governance structures and lifecycle controls leading institutions have already operationalised, the foundation on which the consultation builds (see footnote 4). The proposed guidelines build explicitly on the FEAT Principles (see footnote 45), and the MindForge Handbook, comprising detailed operationalisation guidance and institutional case studies alongside the Executive Handbook's seventeen Considerations, translates these expectations through the most comprehensive industry consensus framework for AI risk management in financial services (see footnote 47). MAS's thematic review observed that generative AI deployment in banks remains at an early stage, concentrated on internal productivity augmentation rather than client-facing autonomy.[54] The MindForge Implementation Examples confirm this pattern: the case studies document governance components—lifecycle controls, risk materiality tiering, human oversight gates—without yet assembling them into the integrated architecture this paper describes. The archetype therefore sits ahead of current industry practice, a target state whose regulatory defensibility will crystallise as frameworks are finalised and tested through supervisory engagement. Institutions should build for the trajectory these instruments signal while maintaining flexibility to adjust as final requirements emerge.

The **Compliance First** institution has deployed ML where regulation demands it—transaction monitoring, name screening, sanctions filtering—and stopped there. This is a stable and supervisory-safe position. It is not, however, a competitive one. AI remains a cost of compliance, and every additional regulatory obligation adds cost without expanding capacity. The architecture described in Section 4 through Section 6 does not exist here; models operate as isolated tools, governed individually. Capacity scales linearly with headcount.

The **Workflow Augmented** institution has taken the decisive architectural step: reconceiving ML as a control plane and layering LLM-driven workflows on top of it. Suitability analysis, documentation, and client communications are AI-assisted. Governance scales through the control plane rather than through manual oversight of each output. This is where the most forward-leaning private banks are converging, and it delivers real gains: compressed cycle times, improved consistency, expanded

[53] MAS *Information Paper on AI Model Risk Management* (December 2024), §§3.2, 5.2.1 fn. 25. MAS noted that AI model risk management maturity varies significantly across different financial institutions. The paper observed that while most banks have established software systems for their AI inventories, a small number still rely on spreadsheets, an approach MAS characterised as more prone to operational issues and unable to support features such as automated tracking and interdependency identification.

[54] MAS *Information Paper on AI Model Risk Management* (December 2024), §§2.4, 7.1.5. MAS observed that the use of generative AI in banks appears to still be at an early stage; see also footnote 26 for observed deployment patterns. Most banks adopted a more limited scope of initial deployment to allow them to understand generative AI's limitations while managing potential risks.

relationship manager capacity. Its defining constraint is equally clear: every AI output still passes through a human gate. Productivity is constrained by human throughput at the approval layer.

The **Governed Autonomy** institution operates the full four-tier architecture: the three technical layers and the governance wrapper that envelops them. Bounded agents execute multi-step tasks—portfolio monitoring, review preparation, risk surfacing—within ontology-defined boundaries. The shift is structural: from "AI assists the human" to "AI operates within governed boundaries; the human oversees and decides." The unit economics follow the same platform logic observed in digital banking and automated wealth platforms. Shared infrastructure absorbs marginal cost per client as the client base grows.[55] The magnitude of the effect will depend on the institution's scale, client complexity, and the maturity of its governed semantic layer. This is the thesis position. It is also the position that demands the most of an institution. The architectural, governance, and cultural prerequisites are substantial, and the emerging regulatory frameworks that would fully validate this posture, particularly proposed AI risk management guidelines, now closed for consultation and awaiting finalisation, have not yet been tested in supervisory practice. Governed autonomy is the defensible frontier, not a near-term default.

> **What this means for leaders:** Honest self-assessment is the prerequisite. Most private banks today sit between Compliance First and Workflow Augmented. The gap to Governed Autonomy is architectural; closing it requires redesigning infrastructure, not adding tools.

The distinction between the two architectures is not merely one of ambition but of value type. The Workflow Augmented archetype captures operational productivity: compressed cycle times, improved consistency, expanded relationship manager capacity. Governed Autonomy is where the architecture claims structural reuse effects and declining marginal cost per client. The distance between these archetypes, and the concrete steps required to traverse it, is the subject of Section 9.

9 Implementation Roadmap

Exhibit 5: Implementation Roadmap, Three-Horizon Phased Timeline

Timeframes are illustrative and directional. Actual sequencing depends on institutional starting point, regulatory developments, and technology maturity. Horizons overlap by design: H2 begins while H1 is maturing; H3 begins while H2 is being deployed. This exhibit is an illustrative synthesis based on the public regulatory and industry sources cited in the footnotes to this paper.

[55] The economic logic is a standard application of platform infrastructure economics, in which shared layers with high fixed costs generate declining marginal costs per additional user or transaction. For the foundational treatment, see Shapiro, C. and Varian, H.R., *Information Rules: A Strategic Guide to the Network Economy* (1998). For its application to financial services platform operating models, see McKinsey & Company, "Platform Operating Model for the AI Bank of the Future" (2021), describing shared infrastructure that minimises duplication by documenting repeatable processes and cataloguing analytical models available for deployment in diverse contexts.

Dimension	Horizon 1 (0–12 months): Industrialise the Control Plane	Horizon 2 (6–18 months): Deploy Governed LLM Workflows	Horizon 3 (12–30 months): Pilot Ontology-Bounded Agents
Focus	Establish ML as persistent infrastructure: unified data layer, model inventory, continuous monitoring	Introduce LLM-assisted workflows within human-approved decision processes	Deploy bounded agents operating within a governed semantic layer for defined, high-volume processes
Milestone 1	Deploy unified client risk profile integrating custody, transaction, and CDD data sources into a single persistent layer	Deploy LLM-assisted document review and summarisation within the periodic client review process, with human approval at all decision points	Define governed semantic layer encoding regulatory boundaries, approval authorities, and escalation rules for agent operations
Milestone 2	Establish comprehensive model inventory and risk materiality classification covering all ML models in production, including transaction monitoring, name screening, and risk scoring	Implement prompt governance framework: version-controlled prompt templates, output validation rules, and audit trails for all LLM-generated content	Pilot a bounded agent within a single high-volume process (e.g., periodic review preparation) with comprehensive logging and human approval gates at every material action
Milestone 3	Implement continuous model monitoring with automated drift detection and performance reporting to board risk committee	Complete AI risk materiality assessment for all LLM use cases, classifying each by impact, complexity, and degree of reliance on AI output	Implement agent identity management and access controls ensuring traceability of all agent actions to accountable human owners
Milestone 4	Formalise technology risk management framework with clear risk ownership, escalation paths, and defined risk appetite for ML-driven decisions	Integrate LLM workflow outputs into the control plane for continuous quality monitoring and anomaly detection	Establish continuous monitoring and testing regime for agent behaviour: automated policy adherence checks, graduated rollback for performance degradation, and kill switch capability for safety-critical failures

Dimension	Horizon 1 (0–12 months): Industrialise the Control Plane	Horizon 2 (6–18 months): Deploy Governed LLM Workflows	Horizon 3 (12–30 months): Pilot Ontology-Bounded Agents
			enabling expeditious deactivation of any agent that exceeds risk tolerances[56]
Governance prerequisites	Board-approved technology risk management framework with defined risk appetite for ML-driven decisions.[57] Designated risk owners for ML models with authority and resources to manage technology risks.[58]	AI inventory and risk materiality assessment framework operational, aligned with emerging supervisory expectations on AI identification, inventory, and risk assessment.[59] Pre-deployment review process established with independence proportionate to assessed risk materiality.[60] Baseline data entitlements and purpose constraints defined for all LLM-accessible data; human approval gates mapped to	Formal agent governance framework defining agent limits, permissions, and human oversight checkpoints, aligned with the IMDA Model AI Governance Framework for Agentic AI.[61] Demonstrated governance maturity across H1 and H2: model monitoring, LLM output validation, and risk materiality assessment all operating effectively.

[56] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §§4.4, 4.23(b). MAS has proposed that institutions should develop contingency plans for AI, especially for high-risk AI, including considering fallback options, such as using humans to replace the function of the AI, to reduce the risk of potential material disruption. Section 4.23(b) proposes that ongoing monitoring processes should "enable the FI to detect any deviation from expected AI behaviour timely and trigger the appropriate response."

[57] MAS Technology Risk Management Guidelines (January 2021) [SUPERVISORY], §3.1.7. MAS expects the board of directors to ensure "a sound and robust risk management framework is established and maintained to manage technology risks" and to approve "the risk appetite and risk tolerance statement that articulates the nature and extent of technology risks that the FI is willing and able to assume."

[58] MAS Technology Risk Management Guidelines [SUPERVISORY], §3.1.2. MAS expects institutions to designate individuals with the authority and appropriate resources to manage technology risks.

[59] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §3.2 (AI identification), §3.4 (AI inventory). MAS has proposed that financial institutions "identify the use of AI in [their] operations and document it in a centralised AI inventory." The AI identification provision encompasses not only AI systems developed internally but also those embedded in third-party products and services.

[60] MAS Consultation Paper P017-2025 [CONSULTATION], §§4.8–4.12. MAS has proposed that the pre-deployment review should include independent review for higher risk AI, with the depth and formality proportionate to the assessed risk materiality of the AI use case. "The FI should conduct pre-deployment review to ensure AI is fit for purpose and can be deployed safely in a production environment."

[61] IMDA, *Model AI Governance Framework for Agentic AI*, Version 1.0, §§2.1, 2.2 [METHODOLOGY]. Horizon 3 governance prerequisites align specifically with the Framework's first and second governance dimensions: assessing and bounding risks through design (§2.1) and making humans meaningfully accountable (§2.2).

Dimension	Horizon 1 (0–12 months): Industrialise the Control Plane	Horizon 2 (6–18 months): Deploy Governed LLM Workflows	Horizon 3 (12–30 months): Pilot Ontology-Bounded Agents
		each workflow decision point.	
Dependencies	Foundational horizon; no prior dependencies	Requires H1 control plane infrastructure and model governance framework	Requires H1 control plane and H2 LLM workflow governance operating at steady state

The institutional archetypes described in Section 8 define where a private bank stands today. This section defines how to move. Exhibit 5 translates the four-tier architecture into a phased programme of work across three overlapping horizons, each mapped to an architectural tier and gated by governance prerequisites that must be satisfied before capability deployment proceeds.

The sequencing is deliberate. Horizon 1 industrialises the control plane: unifying client data, cataloguing models, and establishing continuous monitoring under the technology risk management framework that MAS expects institutions to maintain (see footnote 6). This is an infrastructure commitment that makes all subsequent AI deployment governable. Horizon 2 introduces LLM-assisted workflows within processes that retain human approval at every decision point, governed by the prompt management, audit trail, and risk materiality assessment disciplines that MAS has proposed in its consultation on AI risk management guidelines.[62] Critically, H2 also requires minimum viable semantic controls: baseline data entitlements defining what data each LLM workflow may access, purpose constraints ensuring that data used under one legal basis is not repurposed beyond it, and human approval gates mapped to each workflow decision point. These are precursors to the full governed semantic layer constructed in Horizon 3, not the layer itself; they establish the boundary disciplines that H3 will encode architecturally. The risk materiality assessment framework at the centre of H2 governance prerequisites is not speculative: MAS's thematic review observed banks already assessing AI risk materiality along the dimensions of impact, complexity, and reliance, with the most advanced institutions mapping these ratings directly to the depth of validation and frequency of monitoring required.[63] Horizon 3 pilots ontology-bounded agents, but only where Horizons 1 and 2 have demonstrated governance maturity, and only within boundaries defined by a governed semantic layer that encodes regulatory constraints, approval authorities, and escalation rules.

[62] MAS Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025) [CONSULTATION], §§3.8–3.10. MAS has proposed that financial institutions assess AI risk materiality "considering factors such as the impact, complexity and reliance of AI use." The risk materiality assessment is designed to calibrate the depth and scope of governance measures proportionately.

[63] MAS *Information Paper on AI Model Risk Management* (December 2024), §§5.3.1, 6.4.4. MAS observed that most banks assessed risk materiality across quantitative and qualitative dimensions encompassing impact, complexity, and reliance on AI. Banks adopted a range of approaches for validation requirements: one bank required independent validation for all AI with depth varying by materiality, while most others required independent validation only for higher risk AI, with lower risk AI subject to peer review. See also footnote 44.

Two features of this roadmap merit emphasis. First, the horizons overlap. An institution does not complete H1 before beginning H2; it begins H2 once H1 foundations are operational and begins H3 once H2 workflows are governed at steady state. The illustrative timeframes in Exhibit 5 reflect this staggered parallelism, not a sequential march. Second, the governance wrapper is not a separate phase. In every horizon, governance prerequisites precede capability deployment. This is not because governance slows the programme; it is because governance readiness is what permits deployment at scale without compounding risk.

Actual timelines will vary by institution. A bank already operating a mature model risk framework may compress H1 significantly; one rebuilding its data architecture may require longer. The value of the roadmap lies not in the dates but in the sequencing, the dependencies, and the principle that architecture and governance move together.

The investment profile follows the same horizon structure. Horizon 1 concentrates expenditure on data integration, model inventory tooling, and the governance infrastructure required to operate ML as a control plane. Horizon 2 adds LLM infrastructure—compute, prompt governance capability, output validation tooling—and the risk materiality assessment framework that governs it. Horizon 3 requires ontology construction and agent governance capability: defining the governed semantic layer, encoding permission boundaries, and building the monitoring and kill switch infrastructure that bounded agents demand. Cross-cutting costs span all three horizons: programme management to coordinate sequencing and dependencies, workforce transition to shift operating roles from preparation-intensive to judgement-intensive work patterns, and external advisory for regulatory interpretation and architectural assurance. In aggregate, the programme is comparable in scope to other major infrastructure initiatives private banks have undertaken, core banking platform renewals or enterprise data transformations, though the phased structure allows institutions to stage investment against demonstrated governance readiness rather than committing the full programme cost at inception.

The actions this implies for leadership are the subject of Section 10.

10 Implications for Leaders

The case against acting now is not trivial. The technology remains early stage: large language models hallucinate, produce non-deterministic outputs creating tension with reproducibility and auditability expectations, and remain susceptible to adversarial inputs that architectural boundaries can mitigate but not eliminate (see footnote 21). Agent frameworks lack production maturity. The ontology that bounds agent behaviour is mutable by design; its boundaries will evolve as regulatory requirements, products, and risk appetite change, and each evolution demands the same governance rigour applied to the agents themselves. The regulatory framework supporting governed autonomy is still forming; the caveats detailed in Section 8 apply (see footnotes 2 and 3). The organisational footprint is substantial: redesigning operating models around a control plane requires material investment in data infrastructure, governance capability, and workforce adaptation. A board that pauses to weigh these concerns is exercising proper fiduciary judgement.

For institutions with sufficient scale and data maturity, the risk of prolonged inaction is likely to exceed the risk of disciplined early action, and the gap may widen with time. The convergence

described throughout this paper is structural, not speculative. The Financial Stability Board has identified the progression from traditional machine learning through generative AI to autonomous agents as a systemic development for financial stability, a three-part taxonomy that directly maps to the four-tier architecture described above.[64] Model capabilities are advancing on a trajectory that makes today's limitations temporary. Regulatory expectations are converging towards governance architectures that early movers are already building. Institutions that wait for certainty will not stand still; they will find themselves retrofitting governance onto deployed systems while competitors operate within frameworks designed for governed scale. Delay does not preserve optionality—it compounds the gap in capability, efficiency, and regulatory standing.

Three actions follow for boards and management committees. First, **assess your archetype** honestly. Use Exhibit 4 to locate your institution's current posture and measure the distance to your target state. The assessment should be led by the Chief Technology Officer or Head of Digital working jointly with Compliance, and the deliverable should be a gap analysis mapping current capabilities against each Exhibit 4 dimension benchmarked against what MAS observed in its thematic review of banks' AI governance, not against internal aspiration (see footnote 53). The gap between Compliance First and Governed Autonomy is architectural, not incremental, and recognising that gap is the prerequisite for closing it.

Second, **commission an architecture review**. Determine whether existing machine learning deployments can be reconceived as a control plane or whether the infrastructure requires rebuilding. The review should cover three dimensions: data unification readiness, model governance maturity, and control plane feasibility; and the output should be a technical feasibility assessment with indicative cost estimates for the staggered programme in Exhibit 5. The Chief Technology Officer or equivalent should own the review, with Compliance and Risk as formal reviewers whose sign-off is required before the assessment advances to the board.

Third, **establish a governance-first deployment protocol**. Before any AI capability enters production, require that governance prerequisites—risk materiality assessment, human oversight design, and audit trail architecture—are documented and approved through a defined gate. The minimum contents of the gate pack are defined in Exhibit 6. The gate mechanism should require documented review and approval of all prerequisites before each capability enters production, with the Chief Risk Officer or Head of Compliance holding explicit veto authority to block deployment where prerequisites are not satisfied. Every horizon in the roadmap positions governance prerequisites before capability deployment, because institutions that retrofit governance after deployment are building on foundations that cannot bear the weight of autonomy. The sequencing of these three actions should be governed by completion gates, not calendar timelines: the archetype assessment should be completed and its findings accepted by the board before the architecture review is commissioned; the architecture review's output should be reviewed and endorsed by the board risk committee before the governance-first deployment protocol is established.

[64] Financial Stability Board, *The Financial Stability Implications of Artificial Intelligence* (November 2024). The FSB identifies the progression from traditional machine learning through generative AI to autonomous AI agents as a systemic development for financial stability. The three-stage progression directly maps to this paper's three-layer technical architecture—control plane, LLM workflow layer, and ontology-bounded autonomy.

> **What this means for leaders:** The three actions—assess archetype, commission architecture review, establish governance-first deployment protocol—are sequential dependencies. Skipping the first makes the second premature; skipping the second makes the third impossible.

Exhibit 6: Pre-Deployment Governance Gate for AI Deployment

Illustrative synthesis based on the public regulatory and industry sources cited in the footnotes to this paper. Not a depiction of any institution's internal approval process or supervisory engagement.

Gate Pack Item	Description
Use case description and accountable executive	Named business owner with end-to-end accountability for the capability's performance, risk profile, and regulatory compliance.
Risk materiality rating and rationale	Classification under the institution's risk materiality framework, with documented rationale for the assessed rating (aligned with proposed P017-2025 framework).
Data entitlement and purpose basis	Specification of what data the capability may access and under what consent or legal authority, including purpose limitation constraints for cross-tier data use.
Pre-deployment evaluation and validation evidence	Functional testing, failure mode testing (including adversarial inputs where applicable), prompt or agent evaluation results, and approval thresholds and independence of review proportionate to assessed risk materiality.
Human approval points and escalation triggers	Mapped to the capability's position in the four-tier architecture, identifying where human review is required and what conditions trigger escalation.
Logging, audit trail, and evidence architecture	Design for time-stamped, immutable records of all capability actions, inputs, outputs, and human decisions sufficient to support regulatory examination.
Fallback and kill switch design	Mechanism for expeditious deactivation of the capability, including fallback procedures and defined triggers for activation.
Final approval chain	Named reviewers, approvers, and veto holders, including the Chief Risk Officer or Head of Compliance holding explicit veto authority.

Each item must be documented and approved before the capability enters production. The gate applies to every horizon in the Exhibit 5 roadmap.

The competitive frontier in private banking is shifting from better models toward governed autonomy, embedded within productised client journeys, enforced through architecture, and defended by design.

Note on Scope and Limitations

This paper addresses the regulatory and architectural context for AI deployment in private banking under the supervisory framework of the Monetary Authority of Singapore. The architectures, workflows, maturity archetypes, sequencing steps, and governance checks described here are illustrative design patterns synthesised from the public instruments cited throughout; they do not depict any institution's internal operating model, approval process, or supervisory engagement. Institutions operating across multiple jurisdictions will need to map equivalent obligations in each relevant regime; the governance wrapper described here is illustrative of the approach, not exhaustive of global requirements. All economic claims are directional; actual cost structures, capacity effects, and timeline compression will depend on each institution's data maturity, technology estate, and operating model. MAS's proposed *Guidelines on Artificial Intelligence Risk Management* (P017-2025) closed for public consultation on 31 January 2026; MAS's response to feedback and final guidelines have not been issued as of 25 March 2026. Final requirements may differ from the proposed expectations referenced throughout this paper. Readers should verify MAS's latest response and finalisation status against the MAS publications index. MAS's *Information Paper on AI Model Risk Management* (December 2024) reports observations from a thematic review of selected banks; good practices identified therein represent supervisory observations, not binding requirements, and their applicability to any individual institution will depend on that institution's specific risk profile and operating context. The four-tier architecture describes a target state and a set of design principles, not a product specification or vendor recommendation. The operating model redesign described in this paper implies material changes to the roles of relationship managers, analysts, and compliance professionals, shifting from preparation-intensive to judgement-intensive work patterns. Workforce transition planning, retraining programmes, and change management are material implementation requirements that fall outside this paper's scope but must be addressed in any institution's deployment programme. Similarly, the four-tier architecture must be integrated with each institution's existing model risk management framework; this paper describes target state architecture and governance principles, not the migration path from current model risk management infrastructure to the governed autonomy end state.

Appendix A: Full Governance × Capability Matrix

This appendix provides the detailed regulatory mapping summarised in Exhibit 3. Each cell identifies the specific applicable requirement, expectation, or observed practice grounded in the source instruments cited in the footnotes. The matrix maps six AI capabilities against ten instruments: nine instruments across all five classification tiers, plus supervisory observations from MAS's December 2024 thematic review of banks' AI governance.

Source grounding: All cells are derived from the regulatory and industry documents cited in the footnotes to this paper. MAS's *Information Paper on AI Model Risk Management* (December 2024) is classified as [OBSERVED PRACTICE], a distinct category reflecting supervisory observations from examination rather than binding requirements, supervisory expectations, or industry recommendations. Its cells describe what MAS observed banks doing; they do not establish obligations or expectations. Section references correspond to the published versions of each instrument current as of March 2026. MindForge cells reference the Executive Handbook's Considerations and Practices specifically; the Operationalisation Handbook (January 2026) provides detailed implementation guidance behind the same recommendations. Where an instrument does not directly address a capability, the cell is marked with a dash (—) rather than an inferred requirement.

Transaction Monitoring / AML/CFT Screening

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	A bank *shall* put in place and implement adequate systems and processes to monitor business relations and detect and report suspicious, complex, unusually large or unusual patterns of transactions (Notice 626, §6.22). A bank *shall* pay special attention to all complex, unusually large or unusual patterns of transactions that have no apparent or visible economic or lawful purpose (§6.21). Parallel obligations apply to CMIs under SFA04-N02, §6.21–6.22.
MAS TRM Guidelines	[SUPERVISORY]	The FI *should* establish effective risk management practices and internal controls to achieve data confidentiality and integrity, system security and reliability (§4.1.2). The FI *should* ensure the logging facility is enabled to record activities performed during change processes (§7.5.7). Systems used for transaction monitoring are subject to change management controls (§7.5).
MAS Fair Dealing Guidelines	[SUPERVISORY]	—
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	—
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that an FI *should* put in place and regularly review controls to ensure appropriate human oversight over AI across its life cycle, proportionate to the risk materiality of the AI used (§4.10). For AI applied to AML/CFT monitoring, identified as a high-impact area, MAS has proposed that more robust AI life cycle standards and controls should apply (§1.5, fn. 6). The FI *should* document the AI development process to enable reproducibility and auditability (§4.17).
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that data and models used for AIDA-driven decisions are regularly reviewed and validated for accuracy and relevance, and to minimise unintentional bias (Principle 3). AIDA-driven decisions should be regularly reviewed so that models behave as designed and intended (Principle 4). Firms using

Instrument	Classification	Applicable Requirement
		AIDA are accountable for both internally developed and externally sourced AIDA models (Principle 8).
ISO/IEC 42001:2023	[ASSURANCE]	The standard recommends that the organisation perform AI risk assessments at planned intervals or when significant changes are proposed or occur (§8.2), retain documented information of the results of all AI risk treatments (§8.3), and conduct internal audits to verify conformity and effective implementation (§9.2).
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	Where AI agents are deployed for transaction screening, the framework recommends defining limits on the agent's access to tools and systems, giving agents only the minimum tools and data access needed to complete tasks (§2.1.2). Organisations should implement continuous monitoring after deployment, as agents interact dynamically with their environment and not all risks can be anticipated upfront (§2.3.3).
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends that FIs enhance the organisational risk framework and risk appetite to include enterprise risks, strategies, and KRIs that track, monitor, and mitigate AI-specific risks (Consideration 3). FIs should ensure that AI-specific reviews of AI use cases are conducted periodically post-deployment, with frequency based on the risk materiality of the AI use case (Consideration 5, Practice 6). Ongoing monitoring should report on use case metrics related to AI risks, guardrail effectiveness, and changes in the operating environment (Consideration 14, Practice 1).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that most banks established baseline standards and processes applying to all AI regardless of risk materiality, supplemented by enhanced standards for AI of greater risk materiality or use-case-specific requirements (§6.1.3). Banks maintained continuous monitoring with tiered thresholds, including early warning levels, to pre-empt performance deterioration and trigger early corrective action (§§6.5.3–6.5.10). Models assessed as high risk materiality underwent independent validation; lower risk models were subject to peer review by qualified individuals not involved in development (§6.4.4).

Client Risk Profiling / CDD

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	A bank *shall* perform CDD measures including identifying and verifying the customer (Notice 626, §6.9–6.14B). A bank *shall* ensure CDD data, documents and information are relevant and kept up to date by undertaking reviews, particularly for higher risk categories of customers (§6.24). The Guidelines to Notice 626 specify that for higher risk customers, the bank *should* obtain updated CDD information as part of its periodic CDD review (Guidelines §6-10-4(a)); the frequency of CDD review may vary depending on each customer's risk profile (§6-10-6). Parallel obligations apply to CMIs under SFA04-N02, §6.24.
MAS TRM Guidelines	[SUPERVISORY]	The FI *should* establish a risk management framework to manage technology risks, achieving data confidentiality and integrity (§4.1.1–4.1.2). Systems processing client risk profiles are subject to access controls: the FI *should* implement user access management to restrict access on a need-to-know basis (§9.1).
MAS Fair Dealing Guidelines	[SUPERVISORY]	A financial institution *should* ensure that its fact-find form and risk profiling questionnaire adequately and accurately capture all relevant information about the customer. The FI *should* demonstrate that it has properly designed, tested and validated its scoring and risk profiling methodologies, whether performed in-house or through external expert review and evaluation (§3.3.1).
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	Digital advisers *should* ensure algorithms correctly classify clients according to their risk profiles based on inputs provided; back-testing using hypothetical inputs is expected to ensure risk profiles generated by algorithms are in line with the risk profiling methodology (§31(d)). Controls *should* be in place to detect any error or bias in the algorithms (§32(b)).
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that an FI *should* define what it considers "fair" outcomes and have appropriate controls to identify and mitigate harmful biases across the AI life cycle, calibrated to assessed risk materiality (§4.8). Where fairness considerations are relevant, the FI *should* conduct fairness assessments including defining relevant protected attributes (§4.9). The FI

Instrument	Classification	Applicable Requirement
		should determine the extent of transparency and explainability required, with key considerations including impact on customer outcomes (§4.6–4.7).
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that individuals or groups of individuals are not systematically disadvantaged through AIDA-driven decisions unless these decisions can be justified (Principle 1). Use of personal attributes as input factors for AIDA-driven decisions should be justified (Principle 2). Veritas Phase 2 provides a Fairness Assessment Methodology specifically for assessing alignment of AIDA systems with these principles.
ISO/IEC 42001:2023	[ASSURANCE]	The standard recommends that the organisation perform AI system impact assessments at planned intervals or when significant changes are proposed to occur (§8.4). The AI management system should address processes for the management of concerns related to the trustworthiness of AI systems such as fairness, transparency, and data quality throughout their life cycle (Introduction, §vi).
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	— *(CDD/risk profiling is typically not an agentic workflow; the framework focuses on AI agents that take autonomous actions in dynamic environments.)*
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends that FIs identify and mitigate bias in training and test datasets (Consideration 9, Practice 6) and justify the use of personal attributes in AI use cases (Consideration 9, Practice 2). AI use cases should be designed to operate with a proportionate and practical level of human oversight (Consideration 7, Practice 4). Algorithms should be assessed and selected considering fairness, explainability, performance objectives, implementation complexity, and computational efficiency (Consideration 11, Practice 1).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that banks assessed risk materiality across dimensions of impact, complexity, and reliance on AI, with most mapping materiality ratings to the depth of validation and frequency of monitoring required (§§5.3.1–5.3.2). Risk profiling models affecting customer outcomes were generally subject

Instrument	Classification	Applicable Requirement
		to enhanced standards under banks' tiered governance frameworks (§6.1.3). Most banks established processes to review that risk materiality ratings remained appropriate over time (§5.3.2). Cross-functional AI oversight forums coordinated risk management across business, compliance, technology, and legal functions (§§4.1–4.2).

Suitability Analysis (Preliminary Assessment)

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	*— (Notice 626/SFA04-N02 governs AML/CFT, not investment suitability.)*
MAS TRM Guidelines	[SUPERVISORY]	The FI *should* perform sufficient testing during system development, tracing requirements to test cases, and documenting results signed off by relevant stakeholders (§5.7.2, §5.7.6). Quality assurance *should* be performed by an independent function to ensure deliverables comply with policies, procedures and standards (§5.8.2).
MAS Fair Dealing Guidelines	[SUPERVISORY]	Financial institutions *should* ensure representatives make reasonable enquiries and collect sufficient information to understand and analyse the customer's financial needs and objectives (§3.3.3(a)). They *should* provide advice and present sufficient options that suit the customer's financial objectives, risk tolerance and personal circumstances (§3.3.3(b)). The basis of recommendation *should* be fully documented (§3.3.3(d)). FIs *should* comprehensively and robustly review sales conducted by representatives, verifying that recommendations meet the needs of customers (§3.3.4(a)).
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	As set out under section 27 of the FAA (reflecting the statutory suitability obligation under FAA s.27, articulated through the Guidelines), digital advisers *must* have a reasonable basis for recommending any investment product (§44). Digital advisers *should* ensure the tool collects all necessary information and sufficiently analyses it to make a suitable recommendation, including mechanisms to identify and resolve contradictory or inconsistent responses from clients (§31(b)). Compliance checks on the quality of advice *should* be reviewed by an independent and qualified human adviser (§32(d)).
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that where AI is used in areas with customer impact, such as provision of financial advisory services, AI life cycle standards and controls relating to transparency, explainability, and fairness *should* be applied (§1.5). The FI *should* subject the AI use case to pre-deployment reviews by parties not involved in its development (§4.18).

Instrument	Classification	Applicable Requirement
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that use of AIDA is aligned with the firm's ethical standards, values and codes of conduct (Principle 5). AIDA-driven decisions should be held to at least the same ethical standards as human-driven decisions (Principle 6). Data subjects should be provided, upon request, clear explanations on what data is used and how it affects the decision (Principle 13).
ISO/IEC 42001:2023	[ASSURANCE]	The standard recommends defining and documenting verification and validation measures for the AI system, specifying criteria for their use including testing methodologies, selection of test data, and release criteria (Annex B, §B.6.2.4).
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	*— (Preliminary suitability analysis is typically an AI-assisted, not an autonomous agent, workflow.)*
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends that FIs conduct an AI-specific review based on use case risk materiality prior to deployment to ensure that potential risks are identified and mitigated (Consideration 12, Practice 2). Transparency measures should be evaluated and calibrated based on the use case's risk materiality, degree of autonomy, and intended users, implementing proportionate design features and disclosures (Consideration 11, Practice 4). An inherent risk materiality assessment should determine the risk tiering and guide proportionate governance efforts (Consideration 7, Practice 2).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that banks deploying generative AI limited initial scope to use cases assisting or augmenting human decision-making rather than direct customer-facing applications (§7.1.5). Banks invested in reusable enabling modules, including retrieval systems and evaluation frameworks, to support scalable deployment (§7.1.4). Human-in-the-loop requirements were maintained for generative AI used in decision-making processes (§7.1.9). Most banks established input and output guardrails using filters to manage risks relating to toxicity, bias, and sensitive information leakage (§7.1.14).

Document Generation (Review Memos, Suitability Reports)

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	A bank *shall* maintain records of all CDD data, documents, and information obtained for a period of at least five years (Notice 626, §12; SFA04-N02, §12). Where circumstances warrant consideration of an STR, including where the bank is unable to complete CDD measures, the bank *shall* document the basis for its determination (§14.3).
MAS TRM Guidelines	[SUPERVISORY]	The FI *should* ensure controls are implemented to maintain traceability and integrity for all software codes moved between IT environments (§7.6.2). The FI *should* ensure the logging facility is enabled to record activities (§7.5.7).
MAS Fair Dealing Guidelines	[SUPERVISORY]	Disclosures to customers *should* be readily accessible, written in plain language that avoids technical jargon, presented in a balanced format that highlights key features and risks, and in a format that facilitates ease of reading and understanding (§4.2.1). Representatives *should* fully document the basis of recommendation, stating the customer's objectives, explaining the product recommendation, and highlighting possible risks (§3.3.3(d)).
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	Digital advisers *should* maintain proper documentation on the design and development of algorithms (§29(e)). Digital advisers relying on the FAA-N16 Exception *must* provide a risk disclosure statement to clients alerting them that the recommendation does not take into consideration their financial circumstances (§47(d)).
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that an FI *should* document the AI development process to enable reproducibility and auditability, with documentation sufficiently detailed for an independent party to understand and replicate the AI system's results (§4.17). Documentation standards *should* cover data sources, processing, selection rationale, training procedures, evaluation measures, explainability analysis, and key assumptions (§4.17(a)–(f)).
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that data subjects are provided, upon request, clear explanations on what data is used to make AIDA-driven decisions and how it affects the

Instrument	Classification	Applicable Requirement
		decision (Principle 13), as well as clear explanations on the consequences that AIDA-driven decisions may have on them (Principle 14).
ISO/IEC 42001:2023	[ASSURANCE]	The standard recommends that the organisation document the AI system design and development based on organisational objectives and specification criteria (Annex B, §B.6.2.3). Documented information shall be controlled for availability, protection from loss of confidentiality, improper use, or loss of integrity (§7.5.3).
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	— *(Document generation in this context is typically a generative AI, not an autonomous agent, workflow.)*
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends that FIs document key aspects of the AI build process, including data handling, model training and selection, and evaluation decisions to enable auditability and reproducibility (Consideration 11, Practice 5). Core AI-specific information on AI use cases should be recorded in an AI inventory (Consideration 6, Practice 1), including purpose, scope, types of AI employed, data used, risks and mitigations, and third-party model information. The Handbook identifies hallucination, fabrication, and confabulation as a top AI risk requiring specific guardrails (Section 1.1, Top Risks).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that banks building generative AI capabilities invested in reusable enabling modules, including vector databases, retrieval systems, and evaluation and testing frameworks, to facilitate scalable and governed deployment (§7.1.4). Most banks maintained human-in-the-loop requirements when using generative AI to aid in decision-making (§7.1.9), and established input and output guardrails to manage hallucination, toxicity, bias, and data leakage risks (§7.1.14). Documentation and auditability were supported by centralised AI inventories tracking models through their lifecycle (§§5.2.1–5.2.4).

Client Communication Drafting

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	*— (Notice 626/SFA04-N02 does not directly govern the content of general client communications, though communications must not constitute "tipping off" where AML concerns exist; see §14.4 and Guidelines to Notice 626 §14.)*
MAS TRM Guidelines	[SUPERVISORY]	*— (TRM does not directly govern content of client communications, though systems generating them are subject to general technology risk controls.)*
MAS Fair Dealing Guidelines	[SUPERVISORY]	A financial institution *should* provide customers with clear and relevant information that can be readily accessed and understood, presented in a fair and balanced manner (§4.1.3). Information *should* accurately state what customers can expect of the products and services provided; customers *should* not be led into having unrealistic expectations (§4.2.1). A FI *should* provide timely updates, including after-sales updates on product performance and material developments relating to the financial product (§4.1.3). When serving customers with limited knowledge, additional safeguards *should* be in place (§4.2.3). FIs *should* avoid deploying user interfaces that leverage on human biases to influence customers into choices not in their best interests (§1.3.4, fn. 2).
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	Digital advisers *should* disclose information on how algorithms are used to formulate advice, in a manner that is clear and easy to understand (§36). Any limitations in the algorithm-based advisory process *should* be disclosed (§38). Advertisements and marketing materials *must* comply with applicable guidelines under the FAA and SFA (§52).
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that the FI *should* determine the extent of transparency required, including the need to inform customers of the use of AI and channels for redress (§4.7). Key considerations include the degree of AI autonomy and the level of impact on customer outcomes.
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that use of AIDA is proactively disclosed to data subjects as part of general communication (Principle 12). Data subjects should be provided, upon request, clear explanations on the

Instrument	Classification	Applicable Requirement
		consequences that AIDA-driven decisions may have on them (Principle 14).
ISO/IEC 42001:2023	[ASSURANCE]	— *(ISO 42001 addresses AI management systems broadly but does not contain specific provisions for client-facing communication content.)*
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	Where agents draft client communications, the framework recommends that users should be informed of the agent's range of actions, access to data, and the user's own responsibilities (§2.4). Organisations should consider layering on training to equip employees with the knowledge required to manage human-agent interactions and exercise effective oversight (§2.4).
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends evaluating and calibrating transparency measures based on the use case's risk materiality, degree of autonomy, and intended users, implementing proportionate design features and disclosures to support responsible and informed use (Consideration 11, Practice 4). End users should be provided avenues to enquire, give feedback, or request a review on AI decisions, where applicable, to support continuous improvement and build user trust (Consideration 14, Practice 6). The Handbook identifies toxic and offensive outputs and overconfidence among the top AI risks requiring mitigation (Section 1.1, Top Risks).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that most banks limited generative AI deployment to internal productivity augmentation rather than direct customer-facing applications, adopting a constrained scope of initial deployment to understand generative AI's limitations while managing potential risks (§§2.4, 7.1.5). Input and output guardrails, including filters for toxicity, bias, and sensitive information leakage, were established across generative AI use cases (§7.1.14). Human oversight was maintained throughout generative AI workflows that could affect decision-making (§7.1.9).

Agent Routing / Audit Trail Generation

Instrument	Classification	Applicable Requirement
MAS Notice 626 / SFA04-N02	[STATUTORY]	A bank *shall* maintain an audit function that is adequately resourced and independent, able to regularly assess the effectiveness of internal policies, procedures and controls and compliance with regulatory requirements (Notice 626, §15.12). Record keeping obligations require adequate details to permit reconstruction of transactions (§12). Parallel obligations apply to CMIs under SFA04-N02, §14.12.
MAS TRM Guidelines	[SUPERVISORY]	The FI *should* establish an incident management framework (§7.7.1). Change management controls *should* include clearly defined procedures for assessing, approving, and implementing changes, with authorisers identified (§7.5.6). Segregation of duties in software release processes *should* ensure no single individual can develop, compile and move software codes between environments (§7.6.1). The FI *should* ensure IT audit is performed to assess the adequacy and effectiveness of controls; auditable areas *should* include all IT operations, functions and processes (§15.1.1–15.1.2).
MAS Fair Dealing Guidelines	[SUPERVISORY]	A financial institution *should* have procedures to effectively monitor that its representatives keep proper records of any representations made or advice given to customers (§4.2.2).
MAS Digital Advisory Guidelines CMG-G02	[SUPERVISORY]	Access controls *should* be in place to manage changes to the algorithms whenever necessary (§32(a)). The Board and Senior Management *should* maintain oversight over management of the client-facing tool, including designating appropriate personnel to approve changes to algorithms and having security arrangements to identify and prevent unauthorised access (§29(c)).
MAS Proposed AI Risk Management Guidelines P017-2025	[CONSULTATION]	MAS has proposed that an FI *should* develop and implement comprehensive and robust controls for managing changes to deployed AI, including implementing change control mechanisms, such as human-in-the-loop, to prevent unauthorised alterations (§4.25). Documentation should maintain clear records of monitoring activities, results, identified issues, and subsequent remediation actions for auditability (§4.23(d)). Access to AI models, training

Instrument	Classification	Applicable Requirement
		data, pipelines, and configuration files *should* be strictly controlled and logged (§4.23(d)).
FEAT Principles / Veritas	[METHODOLOGY]	FEAT recommends that use of AIDA in AIDA-driven decision-making is approved by an appropriate internal authority (Principle 7). Firms using AIDA should proactively raise management and Board awareness of their use of AIDA (Principle 9).
ISO/IEC 42001:2023	[ASSURANCE]	The standard recommends that the organisation shall plan, establish, implement and maintain audit programmes including frequency, methods, responsibilities, planning requirements and reporting (§9.2.2). Documented information shall be available as evidence of the implementation of audit programmes and results (§9.2.2). The organisation should control planned changes and review consequences of unintended changes, taking action to mitigate adverse effects (§8.1).
IMDA *Model AI Governance Framework for Agentic AI*	[METHODOLOGY]	The framework recommends establishing robust identity management and access controls for agents to track individual agent behaviour and establish accountability (§2.1.2, "Agent identity"). Organisations should ensure that agents' actions are traceable and controllable (§2.1.2). For agents carrying out high-risk tasks, running agents in self-contained environments with limited network and data access is recommended (§2.1.2). Significant checkpoints in agentic workflows that require human approval, such as high-stakes or irreversible actions, should be defined (§2.2.2).
MindForge *AI Risk Management Executive Handbook*	[METHODOLOGY]	The Handbook recommends establishing an AI change management process to ensure that changes to in-house or third-party use cases are appropriately tracked, reviewed, and approved before implementation (Consideration 15, Practice 1). Proportionate monitoring and analysis should be in place to safeguard against security risks during system usage (Consideration 14, Practice 7). Existing governance processes, forums, assets, and tools should be updated to effectively enable AI governance and risk management (Consideration 1, Practice 3). A monitoring plan and safeguards/contingency measures should be in place prior to deployment, with an

Instrument	Classification	Applicable Requirement
		accountable person designated to address AI risks detected in monitoring (Consideration 13, Practice 1).
MAS *Information Paper on AI Model Risk Management* (December 2024)	[OBSERVED PRACTICE]	MAS observed that banks established cross-functional AI oversight forums coordinating risk management across business, compliance, technology, and legal functions (§§4.1–4.2). Most banks maintained centralised AI inventories tracking models through their lifecycle, with the most advanced institutions using software systems supporting automated tracking, approval checkpoints, and interdependency identification (§§5.2.1–5.2.4). Baseline development, validation, and deployment standards applied to all AI, with enhanced standards for higher risk use cases, including independent validation for high materiality AI and documented peer review for lower risk AI (§§6.1.3, 6.4.4).

References

References are grouped by the paper's five category regulatory classification, followed by supervisory observations, international regulatory references, and academic and industry literature. Full citation details and section-level references appear in the footnotes throughout the paper.

Statutory Instruments

Monetary Authority of Singapore. *Notice 626 on Prevention of Money Laundering and Countering the Financing of Terrorism—Banks*.

Monetary Authority of Singapore. *Notice SFA04-N02 on Prevention of Money Laundering and Countering the Financing of Terrorism—Capital Markets Intermediaries*. Issued under section 16 of the Financial Services and Markets Act 2022.

Supervisory Expectations

Monetary Authority of Singapore. *Guidelines on Provision of Digital Advisory Services* (CMG-G02, October 2018).

Monetary Authority of Singapore. *Technology Risk Management Guidelines* (January 2021).

Monetary Authority of Singapore. *Guidelines on Outsourcing* (Banks) (December 2023).

Monetary Authority of Singapore. *Guidelines on Fair Dealing—Board and Senior Management Responsibilities for Delivering Fair Dealing Outcomes to Customers* (May 2024).

Consultation and Proposed Guidance

Monetary Authority of Singapore. Consultation Paper P017-2025, *Proposed Guidelines on Artificial Intelligence Risk Management* (November 2025).

Monetary Authority of Singapore. Consultation Paper P004-2026, *Proposed Guidelines on Third-Party Risk Management* (March 2026).

Industry Methodology and Governance Frameworks

EDM Council / Object Management Group. *Financial Industry Business Ontology (FIBO)* (first published 2014; continuously maintained). https://spec.edmcouncil.org/fibo/.

Monetary Authority of Singapore. *Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector* (updated February 2019).

Veritas Consortium (MAS and industry partners). *FEAT Fairness Principles Assessment Methodology* (Document 1). *FEAT Fairness Principles Assessment Case Studies* (Document 2). *FEAT Principles Assessment Methodology* (Document 3).

FINOS (Fintech Open Source Foundation). *AI Governance Framework* (2024).

Personal Data Protection Commission (PDPC). *Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems* (1 March 2024).

Project MindForge. *AI Risk Management Executive Handbook* (November 2025). Developed by a consortium of 24 financial institutions under MAS leadership.

Project MindForge. *AI Risk Management Operationalisation Handbook* (January 2026). Part of the MindForge AI Risk Management Handbook. Developed by a consortium of 24 financial institutions under MAS leadership.

Project MindForge. *AI Risk Management Implementation Examples* (January 2026). Part of the MindForge AI Risk Management Handbook. Developed by a consortium of 24 financial institutions under MAS leadership.

Infocomm Media Development Authority (IMDA). *Model AI Governance Framework for Agentic AI*, Version 1.0 (22 January 2026).

Digital Twin Consortium FinTech Working Group. "Financial Services, AI and Complex Systems" (February 2026).

External Assurance Standards

ISO/IEC 42001:2023. *Information Technology—Artificial Intelligence—Management System*.

Supervisory Observations

Monetary Authority of Singapore. *Artificial Intelligence Model Risk Management: Observations from a Thematic Review*, Information Paper (December 2024).

International Regulatory and Institutional References

Board of Governors of the Federal Reserve System / Office of the Comptroller of the Currency. *Supervisory Guidance on Model Risk Management*, SR Letter 11-7 / OCC Bulletin 2011-12 (April 4, 2011).

National Institute of Standards and Technology. *Zero Trust Architecture*, Special Publication 800-207 (August 2020).

Bank of England Prudential Regulation Authority. *Model Risk Management Principles for Banks*, Supervisory Statement SS1/23 (May 2023, effective May 2024).

Financial Stability Board. *The Financial Stability Implications of Artificial Intelligence* (November 2024).

National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST AI 600-1 (July 2024).

Financial Services Sector Coordinating Council and Cyber Risk Institute. *Financial Services AI Risk Management Framework* (February 2026).

Academic and Industry Literature

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., and Mané, D. (2016). “Concrete Problems in AI Safety.” *arXiv:1606.06565*.

Bradshaw, J.M., Jung, H., Kulkarni, S., Johnson, M., Feltovich, P., Allen, J., Bunch, L., Chambers, N., Galescu, L., Jeffers, R., Suri, N., Taysom, W., and Uszok, A. (2005). “Toward Trustworthy Adjustable Autonomy in KAoS.” In R. Falcone et al. (eds.), *Trusting Agents for Trusting Electronic Societies*, LNAI 3577, pp. 18–42. Springer.

Bradshaw, J.M., Feltovich, P.J., Jung, H., Kulkarni, S., Taysom, W., and Uszok, A. (2004). "Dimensions of Adjustable Autonomy and Mixed-Initiative Interaction." In M. Nickles et al. (eds.), *Agents and Computational Autonomy*, LNAI 2969, pp. 17–39. Springer.

Cavoukian, A. (2009/2011). *Privacy by Design: The 7 Foundational Principles*. Information and Privacy Commissioner of Ontario, Canada.

Gruber, T.R. (1993). “A Translation Approach to Portable Ontology Specifications.” *Knowledge Acquisition*, 5(2), 199–220.

Leveson, N.G. (2011). *Engineering a Safer World: Systems Thinking Applied to Safety*. MIT Press.

McKinsey & Company. "Building the AI Bank of the Future" series (2021).

McKinsey & Company. "Platform Operating Model for the AI Bank of the Future" (2021).

McKinsey & Company. "Seizing the Agentic AI Advantage" (2025).

Oliver Wyman. *10 Wealth Management Trends for 2026* (December 2025).

Ostrom, E. (2010). “Beyond Markets and States: Polycentric Governance of Complex Economic Systems.” *American Economic Review*, 100(3), 641–672. Nobel Prize Lecture.

Shapiro, C. and Varian, H.R. *Information Rules: A Strategic Guide to the Network Economy* (1998).