

The missing C-Suite: governing algorithmic decision-making in financial systems through the AI Spine framework

By Ziyanda Kitantou

Presented at the Actuarial Society of South Africa's 2026 Convention
Cape Town International Convention Centre, 7–8 October 2026

ABSTRACT

Across South African financial services, credit limits are set, claims are settled, payments are routed, and premiums are priced by automated systems performing functions historically held by named executives, yet without the named executive's accountability scaffolding. This paper terms the resulting condition the missing C-Suite: a structural gap in which the work of senior decision-making is being done at scale and in real time, while the governance machinery the institution was designed to require has not followed suit.

The paper's central argument is that financial risk enters at the decision, not at the model. Existing Model Risk Management frameworks therefore govern the wrong layer, leaving the consequential step (where an algorithm determines whether a claim is paid, credit is extended, or a premium is set) structurally unowned. Aligned with the ASSA 2026 theme of *Broadening Boundaries*, it locates this gap within the actuarial profession's natural domain of risk quantification and capital adequacy. To close it, the paper introduces the AI Spine, a governance architecture that restores named accountability across algorithmic decision systems. It is operationalised by two quantitative instruments: the Decision Materiality Index, a seven-dimension methodology that triages decision classes by materiality, and the Governance Capital Add-on, which quantifies the capital implication of governance deficit within the ORSA framework under SAM Pillar II.

The architecture is situated against emerging international regulation (including the EU AI Act and the Basel Committee's 2024 report on the digitalisation of finance) and the November 2025 PA/FSCA joint report on AI in the South African financial sector.

The paper concludes that algorithmic governance is an emerging actuarial domain in its own right, distinct from the model risk management discipline that produced it: the governance of autonomous financial decision systems. For the actuarial profession, this represents both a professional obligation and a strategic opportunity to define the next frontier of fiduciary risk management in South African financial services.

KEYWORDS

Algorithmic governance; AI Spine; Decision Materiality Index; ORSA; SAM Pillar II; Governance Capital Add-on; November 2025 PA/FSCA joint report; algorithmic accountability

CONTACT DETAILS

Ziyanda Kitantou, Nakhile Consulting; email: ziyanda@nakhile.com

1. INTRODUCTION: BROADENING THE ACTUARIAL BOUNDARY

1.1 In a major South African bank, a reinforcement-learning algorithm chooses which payments settle, at what counterparties, and in what sequence. Functionally, it performs the work of a Treasury Operations Head. In a short-term insurer, an automated settlement engine approves or declines tens of thousands of personal lines claims monthly without human review; functionally, it performs the work of a Head of Claims. In a fintech lender, a gradient-boosting model and its surrounding decision pipeline extend or deny credit at scale, performing the work of a Head of Credit Risk. In a motor insurer, a telematics pricing engine adjusts premiums at renewal across a portfolio of 150 000 policyholders, performing the work of a Chief Underwriter. (These vignettes are composites drawn from the author's engagements with South African financial services institutions during 2024–2026; specific institutional identifiers, volumes, and quantum figures have been generalised or anonymised.)

1.2 In each case the algorithm performs the function of the executive without occupying the position. It is subject to controls, frequently several of them, but it sits in no governance structure that names it, carries no defined accountability at the level of the decision class, and operates beyond the line of sight of the boards that bear ultimate responsibility for it. This is the missing C-Suite. The metaphor is functional rather than formal: the algorithm performs the work of the C-Suite role without occupying its position, and the governance gap is precisely the gap between function and accountability. The work is being done, at scale, in real time, and what is absent is the governance scaffolding that the institution's machinery was designed to require.

1.3 The Actuarial Society of South Africa's (ASSA) 2026 convention theme, *Broadening Boundaries*, could not be more timely. Across South African financial services, the boundaries of decision-making itself are shifting in ways that carry direct implications for actuarial practice, professional accountability, and the governance of systemic financial risk. The digital transformation that followed the COVID-19 pandemic accelerated artificial intelligence (AI) adoption across the sector, often outpacing the governance frameworks meant to accompany it. These conditions describe the operational reality of South African financial services in 2026, and the pattern extends well beyond the largest institutions. Mid-tier insurers, credit providers, and payment processors have deployed AI-driven decision systems in their core operations.

1.4 The consequences of governance lag are well documented. The Apple Card and Goldman Sachs controversy of 2019 began with public allegations that the card's credit-limit algorithm produced gender-correlated disparities; the New York Department of Financial Services investigation that followed found no fair-lending violation and no evidence of disparate impact, but it identified a perceived lack of transparency and an inability to explain individual credit-limit outcomes that undermined consumer trust (NYDFS, 2021). The case is instructive primarily as an explainability and transparency failure rather than as a proven discrimination event: a decision system that the institution could not adequately explain to the customers it affected. The Dutch SyRI welfare fraud detection system was struck down by The Hague District Court in 2020, principally on the grounds that the enabling legislation did not strike a fair balance between fraud detection and the right to privacy under Article 8 of the European Convention on Human Rights, with the absence of transparency about the risk model central to the Court's reasoning (Rechtbank Den Haag, 2020); it is a public-sector analogue of the opacity now embedded in private-sector credit, claims, and pricing systems. Across multiple jurisdictions, automated credit decisioning systems have been found to produce demographically skewed outcomes where the governance structures that would have detected and corrected such drift were absent (O'Neil, 2016). The harm occurs silently, at scale, faster than any human oversight mechanism calibrated for human-speed processes can detect.

1.5 This paper argues that the actuarial profession must step forward into the governance role. The case rests on professional qualification rather than convenience. Actuaries understand the relationship between operational decisions and capital adequacy. They are trained in statistical monitoring, drift detection, and the design of threshold-based control systems. They are embedded in governance structures, Board risk committees, Own Risk and Solvency Assessment (ORSA) processes, and the Appointed Actuary statutory function that carry legal and professional accountability. Boards, regulators, and senior management already trust them to define the parameters within which financial systems operate.

1.6 The AI Spine framework introduced in this paper provides the architecture for that step forward: a systematic approach to translating abstract accountability into enforceable control structures for the era of autonomous financial decision-making, grounded in actuarial discipline and aligned with emerging international regulatory requirements. It is operationalised through two quantitative instruments. The **Decision Materiality Index (DMI)** sizes which decisions warrant governance attention. The **Governance Capital Add-on (GCA)** quantifies the capital implication of governing them inadequately. Together, the DMI and the GCA bring algorithmic governance within the actuarial profession's quantitative perimeter.

2. THE GOVERNANCE GAP: WHERE FINANCIAL RISK ENTERS

2.1 Financial risk enters at the decision

2.1.1 The most consequential misframing in AI governance is the assumption that algorithmic risk is concentrated at the model. The model sits upstream of the financial event. The financial event happens at the decision.

2.1.2 Consider the chain from prediction to financial statement (see Figure 1). An AI system produces a prediction or recommendation. That prediction enters a business decision: to settle a claim, to extend credit, to set a premium, to flag a transaction, to route a payment. That decision triggers an operational action: a payment, a denial, an alert, a settlement. The operational action produces a financial outcome. Money moves. Exposure is created or extinguished. Regulatory obligations engage or release. The outcome lands in the financial statements, where it is audited.

2.1.3 Now consider where governance currently sits. Model Risk Management (MRM) governs the AI system at the top of the chain. External audit governs the financial statements near the bottom. The **decision step in the middle**, where AI-driven judgement meets institutional consequence, sits between those two governance perimeters. It is governed, in practice, by the operations function, which sits outside the risk perimeter. Reporting flows through the data and technology layer, which sits outside the finance perimeter. This is where financial risk actually enters the institution. The decision step is governed, but obliquely: through intake rules that determine which cases reach the automated system at all, through threshold-based referral of individual cases to human review, and through outcome monitoring that surfaces problems in aggregate after the fact. What no one holds is consolidated, named accountability for the decision class as a whole, expressed in terms that the Board and the capital framework can see.

2.1.4 The distinction matters, because the two diagnoses imply different remedies. If the decision layer were genuinely ungoverned, the task would be to build governance where none exists. The practice record shows otherwise: mid-tier and large South African institutions routinely pilot automated decision systems on restricted populations before rollout, restrict which policies, claim types and claim values are eligible for automated handling at all, escalate cases above materiality limits or on fraud-workflow

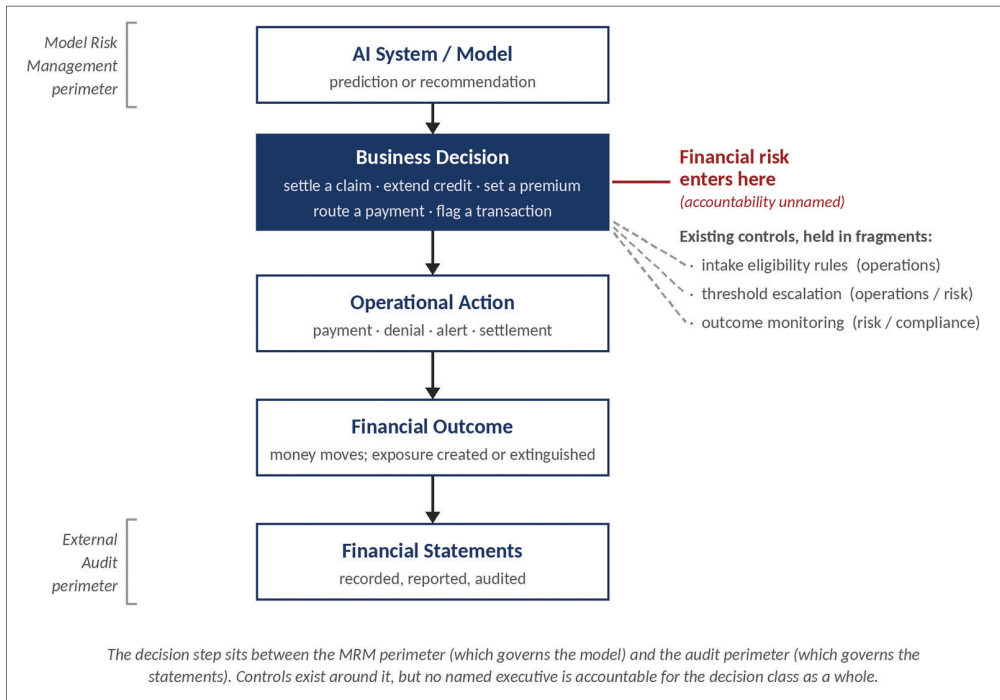


FIGURE 1. Where financial risk enters algorithm-driven decision systems

flags, and monitor loss ratios, override rates and complaint volumes on a defined cadence. These are real controls and they do real work. The difficulty is fourfold: the controls are distributed across the operations, risk, compliance and data functions; they are held as informal practice rather than as mandated governance; no single named executive is accountable for the decision class that they collectively govern; and none of them is integrated into the Solvency Assessment and Management (SAM) Pillar II, ORSA and operational risk machinery through which the Board and the supervisor actually see risk. The gap this paper addresses is therefore one of fragmentation and unnamed accountability rather than one of absence. The AI Spine is a consolidation architecture rather than a greenfield one, and it should be read as formalising governance that in most institutions already exists in pieces.

2.1.5 This has direct implications for the actuarial profession, though the distinction requires care. Mature MRM functions already monitor deployed models continuously, tracking population stability, discriminatory power, override rates and exception volumes in live environments rather than only in controlled ones, so the difference cannot be that validation is static while governance is continuous. The distinction is one of object rather than of tempo. Model validation asks whether the model remains inside the envelope against which it was validated. Algorithmic governance asks whether the decisions the model enables remain decisions the institution should be making, at this volume, under this risk appetite, with this regulatory obligation attached. A model can pass

the first test continuously while failing the second, and the second question falls outside what the MRM function is constituted to ask. The reframing of governance proposed in this paper complements MRM rather than displacing it. It marks where the object of MRM ends and where a different object begins.

2.2 Three properties of the modern AI operational environment

2.2.1 Traditional MRM, as codified in the Federal Reserve and the Office of the Comptroller of the Currency's (OCC) foundational guidance (SR 11-7) and reflected in the model risk management practices adopted by South African financial institutions, was designed for a specific operational context: statistical models that support human decisions. A credit scoring model produces a score; a credit officer makes the decision. A pricing model produces a rate indication; an underwriter applies judgement and signs off. The model is an analytical tool. The human is the decision-maker. Accountability is clear, documented, and auditable.

2.2.2 This architecture is structurally misaligned with autonomous systems, which are characterised by three properties that together define the modern AI operational environment:

- **Decision autonomy** refers to the capacity of a system to initiate and execute consequential financial actions without human intervention. When an automated claims settlement engine pays or declines a claim without referral to a human assessor, or when a payment routing algorithm reallocates settlement flows in response to real-time liquidity signals, no human decision has been made. The institution's risk appetite has been expressed, but by no identifiable person. Accountability has been dissolved into the architecture of the system itself.
- **Financial materiality** refers to the aggregate economic significance of autonomous decisions. A single automated decision may be individually immaterial. A system processing tens of thousands of such decisions daily becomes portfolio-material. Materiality lives in the population of actions, whose aggregate effect on revenue, capital adequacy, regulatory compliance, and consumer outcomes can be very substantial. The governance challenge is inherently a portfolio challenge rather than a transaction challenge.
- **Operational velocity** refers to the speed at which autonomous decisions accumulate consequences. A human credit committee might approve one hundred decisions per business day. An automated system can approve one hundred in a second. Errors in model calibration, data quality, risk appetite alignment, or system configuration compound at machine speed before any monitoring system can detect them. By the time a governance failure is identified, the institution may have processed millions of affected decisions.

2.3 Invisible risk

2.3.1 Together, these three properties define what this paper terms 'invisible risk': a condition in which financial decisions are executed at scale by autonomous

systems without explicit ownership at the level of the decision class, without documented alignment to approved risk appetite, and without governance mechanisms calibrated to detect deviation at the speed at which it occurs.

2.3.2 Invisible risk differs from traditional model risk in two critical dimensions. First, it concerns governance more than model accuracy. A well-calibrated model that has drifted from its approved risk appetite, or that operates at a scale not anticipated in its original validation, generates invisible risk even where its predictive statistics remain sound. The governance failure precedes the model failure and remains analytically distinct from it. Second, it has a velocity dimension that existing frameworks cannot address. Traditional operational risk accumulates at human speed: a process failure manifests over days or weeks, and monitoring systems calibrated for this timescale are adequate. Invisible risk accumulates at machine speed. Ungoverned algorithmic drift can translate into basis-point movements in return on equity or capital adequacy ratios within hours, and into material regulatory non-compliance within days.

2.4 Decision-to-signal latency

2.4.1 A second dimension of the governance gap, less discussed but equally consequential, is **decision-to-signal latency**: the time delay between an AI-enabled decision creating a financial consequence and the board being made aware of it.

2.4.2 Model risk monitoring typically runs on a monthly or quarterly cadence. By the time the model risk team detects drift, its financial consequences have already crystallised on the balance sheet: settlements that have been mis-sized, premiums that have been mis-priced, credit lines that have been mis-extended. The board sees the financial impact in the management accounts. The board sees the cause through the model risk pack. These two streams flow in parallel without intersection. Without a decision-level lens, no one can tell the board how much of last period's loss came from automated decisions, or how.

2.4.3 For the actuarial profession, this is a familiar problem framed in unfamiliar terms. It is the algorithmic-governance analogue of incurred-but-not-reported (IBNR) reserving: the institution is accumulating exposure that has not yet been reported through its existing detection cadence. The actuarial discipline of reserving for what has occurred but is not yet reported translates directly to the governance discipline of accounting for decisions that have already produced exposure but whose consequences have not yet surfaced through monthly or quarterly reporting.

2.5 Pre-empting the counterargument: is this not just MRM with new vocabulary?

2.5.1 A reasonable objection to the framing above is that established MRM practice, as codified in SR 11-7, already requires named ownership, monitoring, validation, and escalation; the AI Spine, on this reading, is organisational restructuring solving what is essentially a discipline problem within existing frameworks. The prudential standard-setters have begun to engage with the specific risks of machine learning, the Basel Committee's

2024 report on the digitalisation of finance takes stock of AI and machine learning among the technologies reshaping banking, identifying governance, explainability, and resilience as areas of concern, though it issues no binding requirements and reserves the question of whether further standards or guidance are needed. The objection deserves a direct response.

2.5.2 MRM is necessary but not sufficient, for three structural reasons. First, MRM governs the model and its validation envelope; it does not govern the decision system in which the model is embedded. A model that is technically valid can produce decisions misaligned with risk appetite when its calibration window no longer matches the deployed population, and that misalignment is invisible to a validation framework that asks ‘is the model performing within its validated band?’ rather than ‘is the decision the model enables one the institution should be making at scale?’ Second, MRM operates on a periodic cycle. The cycle was designed for an era in which models supported human decisions made over days or weeks. Autonomous decision systems make decisions in seconds. The cycle mismatch produces a structural detection latency that no amount of process improvement within the MRM framework can close. Third, MRM accountability typically sits in the data, technology, or risk function, rather than in an operating function with named accountability for the financial outcomes the model enables. The AI Spine extends accountability into the operating function and links it to capital consequence.

2.5.3 The point is sharpened by the direction of regulatory travel. When this line of argument was first developed, SR 11-7 was the operative model risk standard and the objection that it already governed these systems carried its greatest force. That standard has since been superseded. The revised interagency guidance that replaced it on 17 April 2026, SR 26-2 (Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency, 2026), expressly excludes generative and agentic AI models from its scope, on the stated ground that such models are novel and rapidly evolving. On the most plausible reading, the exclusion reflects a recognition that a periodic, model-centric validation framework is the wrong instrument for such systems, rather than a judgement that they are low-risk. The standard-setter’s own boundary now falls precisely where this paper argues a different governance discipline must begin. The AI Spine is a proposal for what governs the systems that model risk management has formally declined to reach.

2.5.4 The three reasons above are structural objections that go well beyond vocabulary, and the regulatory boundary now confirms them from the outside. The AI Spine sits above MRM as a governance architecture that addresses a different question, a question the actuarial profession is specifically equipped to ask.

2.6 Connections to the algorithmic accountability literature

2.6.1 The AI Spine is one institutional response within a broader scholarly conversation about how algorithmic systems should be governed. Several strands of that literature inform, and are extended by, the framework proposed here.

2.6.2 Recent academic work in Fair Machine Learning has formalised the theoretical conditions under which *dynamic* AI governance is necessary, establishing that static auditing frameworks are insufficient for systems operating in non-stationary environments (Basu & Das, 2025). The authors demonstrate that point-in-time auditing cannot detect distributional shifts between audit cycles, provides no adaptive mechanism during deployment, and generates no actionable guidance for the period between audits. The AI Spine's three-layer monitoring and Graceful Degradation protocols are the institutional architecture that this dynamic-governance literature identifies as missing.

2.6.3 A second strand, originating with Selbst et al. (2019), warns against what those authors term the 'abstraction trap': the practice of treating fairness and accountability as properties of the model in isolation, abstracted from the sociotechnical system in which the model is embedded. The AI Spine's reframing of the unit of analysis from the model to the *decision class* (§4.3) is a direct response to this critique. It locates governance at the level of institutional consequence rather than technical artefact. Barocas et al. (2019) make a parallel argument from the fairness side: that fairness cannot be reduced to a single statistical metric and must be instantiated institutionally. The DMI's multi-dimensional architecture, financial, customer-harm, regulatory, reversibility, explainability, systemic and accountability dimensions scored together, operationalises this institutional view in a form intelligible to a Board.

2.6.4 Mittelstadt et al. (2016), in mapping the ethics of algorithms, identify six concerns that recur across the algorithmic accountability literature: inconclusive evidence, inscrutable evidence, misguided evidence, unfair outcomes, transformative effects, and traceability. The DMI dimensions map cleanly onto this typology: D5 (Explainability Gap) addresses inscrutability and inconclusiveness; D2 (Customer Harm Potential) addresses unfair outcomes; D7 (Human Accountability Clarity) addresses traceability. The framework adds what the typology lacks: a means of *prioritising* among these concerns, sized to institutional consequence, in a manner producible at Board level.

2.6.5 Wachter et al. (2017) establish a legal basis for counterfactual explanations as a means of satisfying explanation obligations under the General Data Protection Regulation (GDPR). The AI Spine treats per-decision explanation as a governance output (§4.2, §4.3) and the written-reasons obligation in Rule 17 of the Policyholder Protection Rules (PPR) as a live regulatory engagement of D5 (§6.3). Rudin (2019) makes a more pointed argument: that in high-stakes settings, post-hoc explanations of black-box models (models whose internal logic cannot be readily inspected) should be replaced with intrinsically interpretable models in the first place, on the grounds that post-hoc explanations cannot be trusted to faithfully represent the underlying decision logic. The AI Spine's position on this debate is governance-led rather than methodological: D5 is scored on whether the institution can produce an explanation defensible to the regulator and intelligible to the customer, irrespective of whether that explanation is generated by an interpretable model or by post-hoc tooling applied to a black-box one. Where the explanation gap cannot be

closed in either way, Override 1 (live regulatory breach) engages. Veale & Brass (2019) examine algorithmic governance in public administration and observe that accountability frequently dissolves into the organisational architecture of algorithmic systems. The three-tier ownership structure (§4.4) is the institutional counter-design.

2.6.6 The AI Spine's contribution to this literature is architectural. It provides the institutional scaffolding (named accountability, dimensional materiality scoring, capital-linked consequences, structured operational response) that the academic conversation has identified as necessary but has not, to date, instantiated in a form aligned with financial services regulatory practice.

3. REFRAMING AI AS A FIDUCIARY FUNCTION: SAM PILLAR II AND THE GOVERNANCE CAPITAL ADD-ON

3.1 The fiduciary reframing

3.1.1 The SAM framework provides both the regulatory obligation and the conceptual foundation for the governance reframing this paper proposes. SAM Pillar II, which addresses governance, risk management, and ORSA, requires that insurers maintain a sound system of governance with clear allocation of responsibilities, effective internal controls, and documented risk ownership across all material risk categories (Prudential Authority, 2018). The Prudential Authority (PA) has consistently emphasised that governance must be substantive (rather than merely documentary); it must reflect genuine accountability for financial outcomes.

3.1.2 The AI Spine proposes that autonomous AI decision systems must be treated as **fiduciary decision agents**: entities whose outputs carry the same accountability consequences as decisions made by senior executives, and which must therefore be subject to equivalent governance structures. This reframing has three operational implications:

- **Stewardship** requires that automated decisions reflect the institution's long-term solvency objectives and Board-approved risk appetite, rather than merely the short-term optimisation target embedded in the model's objective function. A pricing algorithm optimising for premium conversion rate may simultaneously erode the quality of the insured portfolio in ways inconsistent with the Board's approved underwriting strategy. Without stewardship governance, this erosion is invisible until it manifests as adverse claims experience, at which point it is no longer governance, it is loss-recognition.
- **Accountability** requires that every automated decision be traceable to a named human owner who is responsible for ensuring that the system operates within its approved parameters. This traceability must be enforceable in operational terms, meaning that the owner carries genuine professional and regulatory consequences for governance failures. The statutory equivalent for AI systems is a named individual carrying defined obligations.
- **Quantification** requires that governance risk, the risk arising from inadequate oversight of autonomous decision systems, be measured and reported as a component of operational risk within the ORSA framework. This transforms governance from a

qualitative assurance exercise into a quantitative risk management discipline consistent with the actuarial profession's core competency.

3.2 Triage before quantification: the role of the Decision Materiality Index

3.2.1 A note on sequencing: quantifying governance capital comes second. Before the institution can size a GCA, it must first identify which automated decisions warrant the calculation. Not every algorithm in the institution is material. A treasury settlement engine, a personal lines claims engine, and a marketing recommendation engine all consume model output and produce automated decisions, but their governance consequences differ by several orders of magnitude. Treating them identically is as wasteful as ignoring them all.

3.2.2 The instrument that performs this triage is the **DMI**: a seven-dimension weighted scoring methodology that produces a comparable score and tier classification for any AI-enabled decision class in a financial institution. It is the methodology layer of the AI Spine. The DMI is described in full in §4.3 of this paper and specified in technical detail in Kitantou (2026). For the purposes of §3, the relevant property is this: a HIGH-tier DMI score is the trigger for full AI Spine governance treatment, including the GCA calculation.

3.3 The governance capital add-on: a structural illustration

3.3.1 The GCA is the AI Spine's proposed mechanism for making governance failure visible in existing regulatory capital frameworks under SAM Pillar II. As presented in this paper, **the GCA is a structural illustration rather than a calibrated formula**. It demonstrates that governance adequacy is in principle quantifiable in terms commensurable with Pillar II, that such a quantification can be integrated into the ORSA process, and that the resulting capital signal creates an incentive for substantive (rather than merely documentary) governance. The calibration of its parameters against empirical loss data, and the formal regulatory engagement required for any Pillar II add-on, are out of scope for this paper. They are set out in §9 as a development programme appropriate for ASSA-sponsored research in partnership with the Prudential Authority.

3.3.2 With that caveat, the structure is as follows:

$$GCA = DVR \times GDF \times CF$$

where:

— **Decision Volume at Risk (DVR)** is the aggregate financial exposure processed by decision nodes operating outside approved governance thresholds in a defined period, expressed in monetary terms. It is observable, in principle, from the institution's own operational records once the Decision Node Map has been constructed. Two refinements matter to its construction. First, the DVR is not the institution's notional automated throughput. Institutions apply substantial governance at the intake layer,

restricting which policies, claim types, claim values and applicant segments are eligible for automated handling at all, so that a material proportion of decisions that could in principle be automated never reach the automated system. The DVR must be constructed from the population actually routed to automated handling and actually falling outside approved governance thresholds, rather than from the population of decisions that are automatable in principle. Where this distinction is not drawn, the DVR, and with it the GCA, will overstate exposure, in some institutions substantially. Second, the DVR should be restricted to exposures not already attributed to the operational risk module of the Solvency Capital Requirement (SCR), for the reasons set out in §3.3.3.

- **Governance Deficit Factor (GDF)** is a scalar reflecting the severity of the governance deficit. The illustration in this paper uses a range of 0,01–0,10, with representative anchor points at 0,10 (no documented Risk Owner), 0,05 (documented ownership but no operational monitoring), and 0,02 (monitoring in place but thresholds exceeded without escalation). *These figures are illustrative anchors that demonstrate the formula's behaviour. They require empirical calibration before any supervisory use.* Empirical calibration would proceed by relating governance deficit characteristics to operational risk loss event databases, longitudinal claims experience, and regulatory enforcement records, a research programme the actuarial profession is well-positioned to lead.
- **Correlation Factor (CF)** adjusts for systemic correlation between governed and ungoverned nodes, analogous to diversification adjustments in the SCR calculation. Where ungoverned nodes provide inputs to governed downstream processes, as frequently occurs in claims processing chains and credit origination pipelines, the governance failure propagates through the system. In such cases, the CF may appropriately exceed 1,0. A terminological caution is warranted. Neither the CF nor the GDF is a statistical correlation coefficient, and neither is bounded on the interval from minus one to one. Both are aggregation multiplier scalars applied to a monetary exposure. The label Correlation Factor is retained because it names the phenomenon the multiplier is intended to capture, namely the propagation of governance failure across linked nodes, but it should not be read as a correlation in the statistical sense, and institutions documenting the GCA in their ORSA should say so explicitly to avoid supervisory misreading.

3.3.3 Bounding and Solvency Capital Requirement interface

Two mathematical guardrails apply to the GCA structure. First, the CF must be bounded above to prevent a runaway capital requirement. A working upper bound of 1,5 is proposed, in line with the upper-tail correlation adjustments used in the SCR aggregation matrix under SAM, with anything above this level requiring explicit Board approval and supervisor disclosure. Second, the GCA sits alongside rather than within the existing diversification effects calculated in the operational risk module of the SCR. Institutions applying the GCA must therefore avoid double-counting governance risk that

has already been recognised through operational risk capital, by restricting the DVR to exposures not already attributed to the operational risk module. Both guardrails require empirical specification through Prudential Authority engagement before any supervisory application.

3.3.4 A fully developed worked illustration of the formula, including illustrative SCR interaction, is provided in Appendix B, separated from the main text to reinforce that the numbers are structural rather than calibrated.

3.4 Why the Governance Capital Add-on matters even as a structural illustration

3.4.1 Even before empirical calibration, the GCA structure carries three properties that justify its proposal at this stage.

3.4.2 First, **commensurability**: the GCA is denominated in monetary units commensurable with the SCR. This permits it to enter ORSA reporting, Board capital discussions, and supervisory engagement on the same terms as other risk categories. A governance discipline that cannot be expressed in capital terms is structurally invisible to the institutional machinery that makes capital decisions. The regulatory pathway should be stated explicitly, because it materially affects the implementation burden. The GCA is proposed as a Pillar II own-assessment adjustment, calculated and disclosed by the institution within its ORSA, and requiring no formal Prudential Authority approval. It is not proposed as a Pillar I SCR add-on, which would require supervisory engagement and, where an internal model is involved, formal model approval. An institution may therefore adopt the GCA as an internal capital and governance signal under its existing ORSA obligations without any change to its regulatory capital position. Whether a calibrated GCA should in time inform a formal Pillar I treatment is a supervisory question that this paper does not resolve.

3.4.3 Second, **incentive alignment**: any framework that creates capital cost for governance deficit and capital relief for governance adequacy aligns the institution's economic incentives with substantive, rather than documentary, governance. The principle is familiar from operational risk capital design under SAM and Basel: an institution that invests in genuine control infrastructure should receive corresponding capital benefit. The GCA proposes that algorithmic governance be treated the same way.

3.4.4 Third, **actuarial domain location**: the ORSA is operated by the actuarial function; the SCR is calibrated through actuarial analysis; the Pillar II framework is supervised through actuarial engagement with the Prudential Authority. Locating algorithmic governance in this machinery, rather than alongside it, establishes the profession's natural authority over the discipline. The GCA is, in this sense, primarily a jurisdictional claim rather than a quantitative one. It asserts that algorithmic governance belongs in the actuarial profession's regulatory architecture, and proposes the mechanism by which that belonging is operationalised.

3.4.5 A note on cadence, and on an objection it invites: the GCA is calculated and reported annually through the ORSA, while the decisions it prices are made continuously, in some cases millions of times a day. That mismatch is real, and it would be a serious objection if the GCA were the framework's only response to governance deficit. The AI Spine, however, operates at three cadences by design. Graceful Degradation responds in real time to statistical triggers, constraining or suspending the automated decision function within the operating cycle. Risk Committee reporting on HIGH-tier decision classes operates quarterly. The GCA operates annually, recognising in capital terms the governance deficit that the two faster layers have not closed. The GCA is the recognition mechanism rather than the detection mechanism, and reading it as the latter would indeed leave a nine-to-twelve month lag between signal and response. Read correctly, the capital signal is the residual consequence of deficits that the faster layers were unable to remediate within the reporting year, which is precisely the incentive the mechanism is designed to create.

3.4.6 The empirical calibration that the GCA ultimately requires is the work that turns this jurisdictional claim into supervisory practice. That work is appropriate ASSA-sponsored research, and is identified as such in §9.

4. THE AI SPINE FRAMEWORK

The AI Spine translates abstract accountability obligations into an enforceable operational control architecture. Figure 2 represents the framework as a seven-layer architecture, reading from base to apex.

4.1 The seven-layer architecture

4.1.1 **Layer 1, Data and AI Systems:** the foundation: models, datasets, vendor AI products, and infrastructure on which AI-enabled decisions depend. Governed at this layer by traditional data and model risk practices.

4.1.2 **Layer 2, AI Use Case Identification:** surfacing all AI use cases, including informal, shadow, and bottom-up adoption, and cataloguing them as decision-influencing systems rather than as IT assets.

4.1.3 **Layer 3, Decision Enablement Mapping:** for each identified use case, mapping the operational decisions it influences. A single AI system frequently influences multiple decision classes; this layer makes the many-to-many relationship visible.

4.1.4 **Layer 4, Decision Materiality Engine (DMI):** the methodology layer. Each identified decision class is scored against the seven DMI dimensions. The output is a comparable, board-ready score and tier classification.

4.1.5 **Layer 5, Proportional Governance Response:** translation of DMI tier into specific governance posture. HIGH-tier decisions attract board-level accountability, per-decision logging, and Risk Committee reporting on a quarterly cadence. MEDIUM tier receives specified controls. LOW tier remains in general institutional risk machinery.

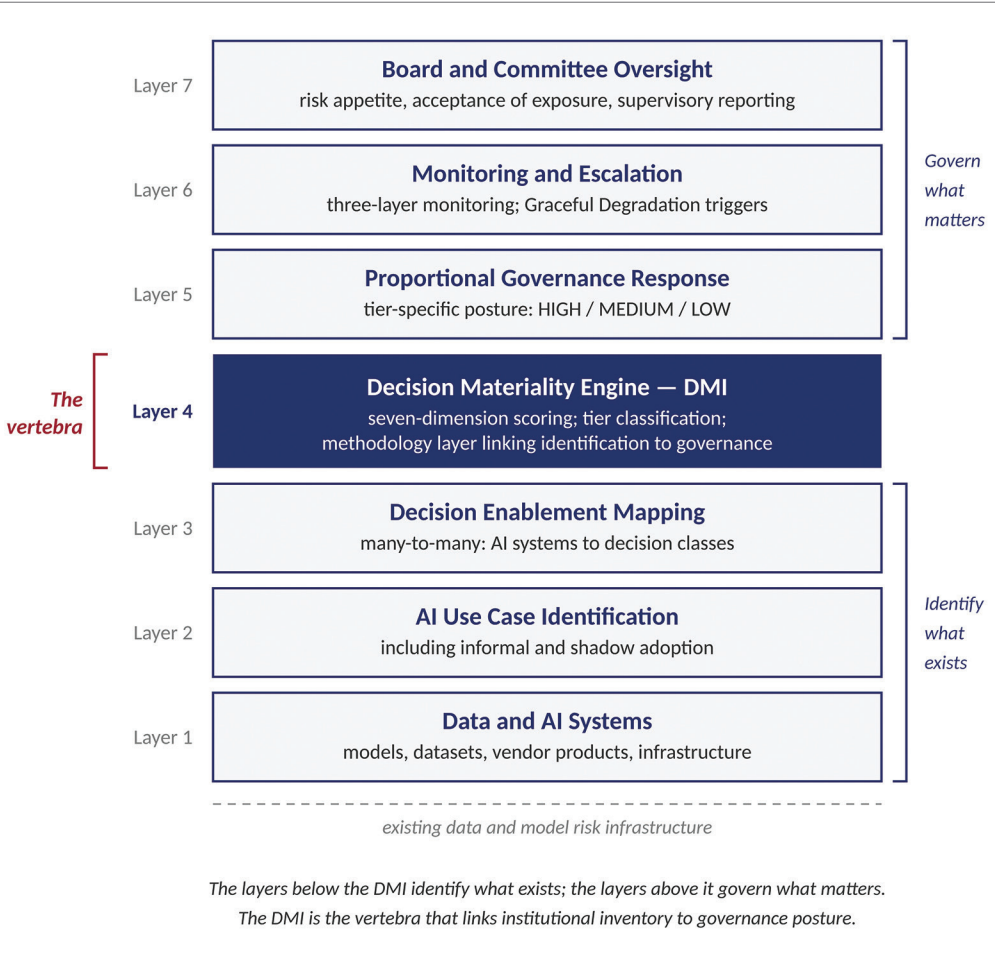


FIGURE 2. The AI Spine: seven-layer governance architecture

4.1.6 **Layer 6, Monitoring and Escalation:** continuous monitoring across three layers, model, decision, and outcome, with escalation triggers calibrated to dimensional findings.

4.1.7 **Layer 7, Board and Committee Oversight:** the apex layer. Board acceptance of AI-enabled decision exposure, risk appetite definition for automated decisions, periodic review of HIGH-tier decision classes, and supervisory reporting where required.

4.1.8 The DMI is the vertebra in the middle of the spine. The layers below it identify what exists; the layers above it govern what matters. Without the DMI, the lower layers produce an inventory and the upper layers produce a posture, but the linkage between them, what warrants which posture, is asserted rather than evidenced.

4.2 Decision node mapping and Protection of Personal Information Act compliance

4.2.1 Every algorithmic point at which a financial outcome is determined, an approval, a decline, a price, a routing instruction, a settlement, a fraud determination, is a Decision Node. The first implementation step of the AI Spine is a systematic inventory of these nodes across the institution's operational systems, producing a Decision Node Map: a comprehensive view of the institution's algorithmic risk landscape.

4.2.2 The Decision Node Map documents, for each node: the decision objective and the financial outcomes it produces; the input variables and data sources it consumes; the population of subjects affected and the volume of decisions processed; the financial exposure at each node; the current monitoring mechanisms in place; and the governance status, whether the node is owned, monitored, threshold-compliant, or deficient.

4.2.3 This inventory serves multiple purposes. Legally, it provides the foundation for compliance with Protection of Personal Information Act (POPIA) section 71, which prohibits a data subject from being subjected to a decision that has legal consequences or substantially affects them where that decision is based solely on automated processing intended to profile the person, including their creditworthiness, conduct, or reliability, unless one of the statutory exceptions applies (Republic of South Africa, 2013). Where the contract or appropriate-measures exception applies, the responsible party must give the data subject an opportunity to make representations and sufficient information about the underlying logic of the processing to enable those representations. Section 71 does not, in its terms, grant a freestanding right to an explanation of an individual decision, and there is academic debate over whether it operates as a prohibition or as a right to object; the conservative governance reading, adopted here, is that an institution must be able to identify which decisions are solely automated and to operate a functional human-review pathway. For insurers, the affected decisions include claims decisions and policy cancellations. For banks, they include credit approvals, account closures, and fraud-based payment blocks. Failure to identify these nodes and to implement the required mechanisms constitutes a live compliance gap, rather than a theoretical risk. Practically, the Decision Node Map enables the institution and its Board to understand, for the first time, the full scope of its algorithmic decision-making exposure. In practice, institutions are consistently surprised by the number of decision nodes revealed through this inventory.

4.3 The Decision Materiality Index: the methodology layer

4.3.1 The DMI is a weighted, multi-dimensional scoring instrument that produces a single materiality score and a tier classification (LOW, MEDIUM, HIGH) for any AI-enabled decision class in a financial institution. A *decision class*, rather than a model or a department, is the unit of analysis: a category of recurring decisions sharing a common decision logic, regulatory perimeter, and customer impact profile.

4.3.2 Decision classes are the right unit because regulatory obligations attach to decision classes (the PPR's written-reasons obligation attaches to claims declines rather

than to the underlying scoring model) and because governance accountability is most usefully documented at the decision-class level. ‘The Head of Claims is accountable for automated settlements under R75 000’ is a governable statement. ‘The Head of Claims is accountable for the SCS-02 model’ is not.

4.3.3 The DMI assesses each decision class against seven dimensions. Table 1 summarises the architecture; the full anchor specification for each dimension is provided in Kitantou (2026).

TABLE 1. The seven DMI dimensions

ID	Dimension	Weight	What it captures
D1	Financial consequence	20%	Direct exposure to balance sheet, profit & loss, regulatory capital, or reserve adequacy from systematic error at typical volume. Anchored to PA notification thresholds and provisioning sensitivity.
D2	Customer harm potential	20%	Severity of harm to a customer in a position of vulnerability at the decision moment, modulated by the speed and completeness of remedy. Anchored to Treating Customers Fairly Outcome 6.
D3	Regulatory obligation intensity	20%	Number, prescriptiveness, and individual-decision applicability of regulatory instruments engaged (PPR, National Credit Regulations, POPIA s71, Insurance Act).
D4	Reversibility	15%	The extent to which an erroneous decision can be undone, scored on the substantive reversibility experienced by the customer, rather than procedural ability to reopen the file.
D5	Explainability gap	15%	Gap between the explanation the institution can produce for an individual decision and the explanation it is obliged to produce by regulation, contract, or reasonable expectation.
D6	Systemic exposure	5%	Volume and concentration at which a single model failure, drift event, or systematic bias propagates.
D7	Human accountability clarity	5%	Degree to which a single, named, formally mandated executive is accountable for the decision class, with documented Board endorsement.

4.3.4 Each dimension is scored on a five-point anchored scale. The weighted aggregate produces a single DMI score between 1,0 and 5,0.

4.3.5 Rationale for the weights

The weighting structure reflects a deliberate design choice about which dimensions function as *primary scoring dimensions* and which function as *amplifiers or structural conditions*. D1, D2 and D3 are weighted equally at 20% because they capture the three primary axes along which an AI-enabled decision creates institutional consequence (financial, customer, and regulatory), and equal weighting reflects the practitioner judgement that no one of these systematically outranks the others in South African financial services. D4 (Reversibility) and D5 (Explainability Gap) are weighted at 15% because they modulate the severity of the primary dimensions rather than constituting separate axes of consequence:

poor reversibility compounds D2; poor explainability compounds D3. D6 (Systemic Exposure) and D7 (Human Accountability Clarity) are weighted at 5%, but for different reasons. D6 functions as a *volume modifier* that scales the other dimensions rather than introducing a separate consequence channel: a HIGH-volume decision class with HIGH financial consequence is doubly material, but the materiality is already captured in D1 (which is sized at typical institutional volume). D7 is the *condition the framework was designed to address*: the missing accountability that gives the paper its title, weighted low in the aggregate because it functions principally as an override-equivalent. A decision class with no named accountable executive is, by design, governance-deficient regardless of how it scores on other dimensions, which is why the *override conditions* in §4.3.7 carry the structural weight that an aggregate score cannot, and specifically why Override 3 attaches to D7 directly. The weighting structure is rationally rather than empirically calibrated; empirical refinement is identified as an active development priority in §9.

4.3.6 The score aggregate is mapped to three materiality tiers (Table 2).

TABLE 2. DMI tiers and governance posture

Tier	Score range	Governance posture
LOW	1,0–2,0	Decision class governed in existing institutional risk frameworks. No elevated AI-specific governance action required. DMI refresh every 24 months.
MEDIUM	2,1–3,5	Documented accountability, model risk integration, monitoring of dimensional weak spots. DMI refresh every 12 months. Inclusion in Risk Committee reporting recommended.
HIGH	3,6–5,0	Board-level governance. Named accountable executive with documented mandate; per-decision logging and explainability infrastructure; quarterly Risk Committee reporting; GCA calculation. DMI refresh every six months.

4.3.7 *Override conditions*

4.3.7.1 The weighted aggregate alone cannot adequately handle two conditions of acute exposure, in which a single dimensional finding so dominates the materiality that ordinary aggregation would mask the governance imperative. The DMI accordingly applies two override conditions that take precedence over the weighted aggregate.

4.3.7.2 **Override 1, Live regulatory breach:** where Dimension 3 evidence demonstrates a current, individual-decision-level regulatory obligation that the institution cannot meet, for example, a PPR Rule 17 written-reasons obligation engaged by a decline that the institution cannot explain to the policyholder, or a National Credit Regulations (NCR) adverse-action notice obligation engaged by a credit decline that the institution cannot trace to identifiable inputs, the DMI is set to a *minimum* of 3,6 (HIGH tier), irrespective of the weighted aggregate. The rationale is that a live regulatory breach demands closure as a compliance failure, and the governance posture must reflect that.

4.3.7.3 **Override 2, Single-event solvency exposure:** where Dimension 1 evidence demonstrates that a single error at typical volume could affect the institution’s

solvency, breach the Solvency Cover Ratio thresholds set by the Board's approved risk appetite, or engage the Prudential Authority's notification regime for material operational risk events under SAM Pillar II (with reference to the institution's own materiality threshold as documented in its ORSA), the DMI is set to a *minimum* of 4,0. This override applies to decision classes whose downside exposure exceeds normal portfolio-level dispersion, most commonly large-corporate credit, treasury operations at systemic scale, or catastrophe-event claims settlement, and ensures that the framework's governance response is calibrated to the institution's solvency, rather than merely to its operational risk appetite.

4.3.7.4 Override 3, Absence of named accountability: where Dimension 7 evidence demonstrates that no single, named, formally mandated executive is accountable for the decision class, and the class processes either a decision volume or an aggregate annual exposure above the institution's documented materiality threshold, the DMI is set to a *minimum* of 3,6 (HIGH tier). The rationale is structural. The weighted aggregate assigns D7 a five per cent weight, which is appropriate to its function as a structural condition rather than as a separate channel of consequence, but which has the effect that a decision class with no accountable owner can land in MEDIUM tier on aggregate. That outcome would be inconsistent with the framework's central proposition. A decision class of material size that no named person owns is governance-deficient by definition, and the instrument should record that directly rather than allowing the arithmetic to absorb it. This override gives the paper's central condition, a material decision with no named owner, operative force within the methodology.

4.3.7.5 The three overrides exist to resolve a structural tension in any weighted scoring instrument: that the aggregate can plausibly land in MEDIUM tier even when one dimensional finding, or one structural absence, represents an acute exposure. They make the framework's response to such conditions explicit, defensible, and supervisor-friendly.

4.3.7.6 A note on inter-rater reliability: the DMI is an instrument that depends on assessor judgement against anchored definitions. In the author's engagement experience across production assessments to date, independent dual-scoring of the same decision class by two qualified assessors has produced score variation in the order of 0,3–0,5 DMI points, principally driven by Dimension 5 (Explainability Gap), where assessor views of what constitutes adequate institution-level versus individual-level explanation can vary. This variation is sufficient to occasionally shift a borderline case between MEDIUM and HIGH tier, which is why the tier-boundary policy resolves such cases in favour of the higher tier and the override conditions provide a structural safeguard. The variance observed in production engagement justifies a procedural control on the scoring process itself. Where independent assessors land within 0,1 of a tier boundary or differ by more than 0,5 DMI points on the aggregate score, a consensus panel of three qualified assessors meets to reconcile the dimensional inputs, with a formal third blind assessment acting as tie-breaker if consensus does not emerge in two working days. This procedural safeguard converts assessor variance from a credibility risk into a documented governance input,

and gives a supervisor or audit reviewer a clear basis on which to evaluate the integrity of the scoring process. Formal inter-rater reliability studies, with the statistical machinery to characterise variance more precisely, are an appropriate area for ASSA-sponsored research. A condensed anchor specification for Dimension 5, reproduced from the published methodology, is provided in Appendix C to illustrate how the anchored definitions ground assessor judgement on the dimension that drives the observed variance.

4.4 Decision ownership: the three-tier structure

4.4.1 Each Decision Node requires a three-tier ownership structure that mirrors the accountability requirements of human executive decision-making and provides the regulatory traceability required under SAM Pillar II.

4.4.2 The **Design Owner**, typically in the data science or technology function, is accountable for the technical construction of the algorithm: its architecture, training methodology, feature selection, and initial validation. The Design Owner must document the model's intended operating environment, its performance boundaries, and the conditions under which its outputs should not be relied upon.

4.4.3 The **Risk Owner**, in the actuarial or risk function, is accountable for ensuring that the node's outputs are consistent with the Board-approved risk appetite, that monitoring thresholds are calibrated to material risk tolerances, and that escalation occurs when those thresholds are breached. The role is one of business enablement rather than of veto. The Risk Owner's function is to ensure that automation proceeds with risk visibility, calibrated limits, and a defined response when those limits are approached, rather than to stand in its way; the closest analogue is the Chief Risk Officer's responsibility for enabling sustainable growth through informed risk-taking. This role, at the intersection of actuarial quantification, risk governance, and operational accountability, is the natural domain of the actuarial profession in the AI Spine framework.

4.4.4 The **Business Owner**, a named senior executive, is accountable for the strategic performance of the node: its contribution to business objectives, its compliance with Treating Customers Fairly (TCF) outcomes, and its alignment with the institution's regulatory obligations. The Business Owner is the individual who would be held accountable by the Board if the node produced materially adverse outcomes.

4.4.5 This three-tier structure eliminates the diffuse, committee-based accountability that characterises most current AI governance failures. When no individual carries specific, enforceable accountability for an automated decision system, the system is effectively ungoverned, regardless of what policy documents say.

4.5 Three-layer monitoring: model, decision, outcome

4.5.1 Governance of autonomous decision systems must be dynamic and continuous, mirroring the actuarial discipline of experience monitoring: the ongoing comparison of actual outcomes against expected outcomes, with defined and enforceable

responses when deviations exceed approved tolerance thresholds. The AI Spine specifies three layers of monitoring, illustrated in Figure 3. Most institutions today operate only the first.

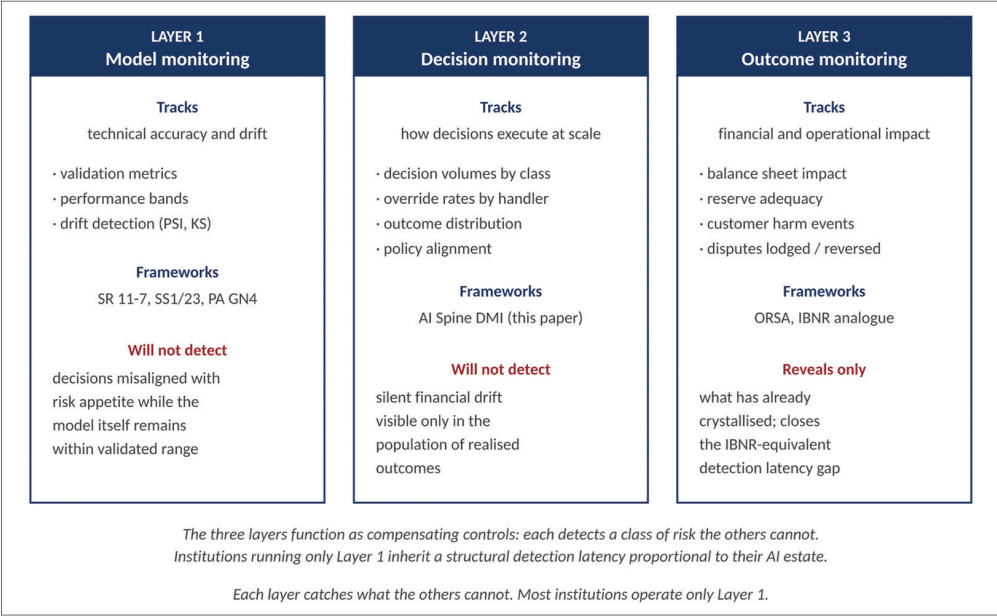


FIGURE 3. Three-layer monitoring: model, decision, outcome

4.5.2 **Layer 1, Model monitoring** tracks technical accuracy and drift. Model risk management guidance, SR 11-7 and its 2026 successor SR 26-2 in the United States and the Bank of England Prudential Regulation Authority’s (PRA) SS1/23 in the United Kingdom, operates here alongside the institution’s existing model risk management cycle. It tells the institution whether the model is still performing in its validated range, but not whether the decisions it produces remain appropriate.

4.5.3 **Layer 2, Decision monitoring** tracks how decisions are actually being executed at scale: decision volumes by class; override rates by underwriter, claims handler, or credit officer; distribution of automated outcomes by demographic, geographic, or product segment; alignment to documented policy. This is the layer that catches a model that is technically accurate but producing decisions misaligned with the institution’s risk appetite or regulatory perimeter.

4.5.4 **Layer 3, Outcome monitoring** tracks the financial and operational consequences of those decisions: balance sheet effect, reserve adequacy, customer harm events, disputes lodged and reversed. This is the layer that catches silent financial drift that neither model monitoring nor decision monitoring will flag, drift visible only in the population of realised outcomes.

4.5.5 Two primary statistical mechanisms support Layer 1 and Layer 2

monitoring, both drawn from established actuarial and data science practice. The **Population Stability Index (PSI)** measures the degree to which the distribution of input variables to a model has shifted relative to its training population. A PSI below 0,1 signals stability; between 0,1 and 0,2 signals moderate drift warranting increased monitoring; above 0,2 signals significant drift requiring active intervention. The **Kolmogorov–Smirnov (KS) statistic** tests the separation between outcome distributions across population segments, providing a measure of the system’s discriminatory power over time. Deterioration in KS, particularly alongside PSI increases, indicates that the model’s relationship between inputs and outputs is breaking down in the deployed environment.

4.6 Graceful degradation

4.6.1 Rather than a binary shutdown mechanism (which could introduce operational shocks, liquidity disruption, or consumer harm), the AI Spine implements **Graceful Degradation**: a tiered response protocol that scales the governance intervention to the severity of monitored drift.

4.6.2 At PSI above 0,1, the system enters **Enhanced Monitoring Mode**: sampling frequency is increased, and the Risk Owner is required to review and sign off in five business days.

4.6.3 At PSI above 0,2, the system enters **Human-Augmented Mode**: automated decisions are constrained, for example, credit limits capped at 50% of standard thresholds, or claims approvals above a defined financial threshold referred to human assessors, pending formal model recalibration and Risk Owner review.

4.6.4 At PSI above 0,3, or where KS deterioration exceeds defined thresholds, the system enters **Restricted Mode**: the automated decision function is suspended and decisions are referred to human operators until recalibration, revalidation, and Risk Owner reapproval are complete.

4.6.5 These thresholds are calibrated by the Risk Owner and approved by the Board risk committee as part of the AI governance framework, ensuring that the monitoring architecture is itself subject to governance. The use of objective statistical triggers, rather than discretionary intervention, ensures that governance responses are data-driven, consistent, and auditable.

5. INTERNATIONAL REGULATORY CONTEXT: THE CASE FOR PROACTIVE GOVERNANCE

5.1 The AI Spine does not exist in an international regulatory vacuum. Across major financial jurisdictions, regulators have moved decisively on algorithmic governance, and South African institutions, many of which operate in or alongside international financial groups or serve customers with cross-border rights, will increasingly be expected to demonstrate alignment with international standards.

5.2 The European Union (EU) AI Act (2024) establishes a risk-based classification system for AI systems, with specific provisions for financial services. Under Annex III, AI systems intended to evaluate the creditworthiness of natural persons or establish their credit score (with a carve-out for financial fraud detection), and AI systems used for risk assessment and pricing in relation to natural persons in life and health insurance, are classified as high-risk, subject to mandatory conformity assessments before deployment, technical documentation requirements, logging of decision outcomes, human oversight obligations, and ongoing post-market monitoring. The insurance designation is deliberately narrow: it names life and health insurance only. Short-term lines such as motor and household, including the automated claims and telematics-pricing systems considered in section 6 of this paper, fall outside Annex III's automatic high-risk designation, although they may still engage other provisions where they profile individuals. This boundary reinforces a point central to this paper: for South African short-term insurers, the operative governance perimeter is the domestic framework rather than the EU AI Act. The Act's extraterritorial provisions, which apply to AI systems whose outputs are used in the EU regardless of where the system is deployed, mean that South African institutions with EU operations or EU-resident customers are nonetheless subject to its requirements for the use cases it does designate.

5.3 The National Institute of Standards and Technology AI Risk Management Framework (AI RMF 1.0, 2023) provides a voluntary but rapidly influential framework organising risk management around four functions: Govern, Map, Measure, and Manage. The alignment with the AI Spine is deliberate. The AI Spine's Decision Node Mapping corresponds to the NIST Map function; the DMI to the Measure function; the three-tier Ownership structure to the Govern function; and the Continuous Monitoring and Graceful Degradation protocols to the Manage function.

5.4 SR 11-7 and its machine-learning supervisory context represents the Federal Reserve and OCC's foundational model risk management guidance (SR Letter 11-7, 2011), widely adopted as an international reference by banking supervisors including the South African Reserve Bank (SARB). On 17 April 2026 it was superseded by SR 26-2, revised interagency guidance issued jointly by the Federal Reserve, the Federal Deposit Insurance Corporation (FDIC) and the OCC, which retains the core model risk architecture for traditional statistical and quantitative models while expressly excluding generative and agentic AI from its scope (see §2.5). The guidance requires that institutions maintain comprehensive model inventories, validate models against their deployed populations, and maintain ongoing monitoring with defined escalation thresholds. The Basel Committee's 2024 report on the digitalisation of finance takes stock of these risks for machine learning models and autonomous decision systems, identifying governance and explainability as central concerns, while noting that the Committee will consider whether additional standards or guidance are needed. The model inventory requirement, retained in SR 26-2,

is directly analogous to the AI Spine's Decision Node Map; the escalation threshold requirement corresponds to the Graceful Degradation protocol.

5.5 The Monetary Authority of Singapore (MAS) FEAT Principles offer a parallel set of conduct-oriented principles, Fairness, Ethics, Accountability, and Transparency, that map closely to the DMI customer-harm and explainability dimensions. Convergence between Singapore, the EU, the US, and the Bank for International Settlements (BIS) on substantive governance requirements is significant. It signals that algorithmic governance is being constituted as an international supervisory standard, rather than jurisdictional preference.

5.6 The November 2025 PA/Financial Sector Conduct Authority (FSCA) joint report is the most directly relevant domestic development for South African institutions. The report sets out the findings of a 2024 industry survey conducted by the Prudential Authority and the FSCA, documenting current AI adoption, investment intentions, use cases, benefits, risks, and governance practices across the sector. It records that financial institutions themselves are calling for principle-based, outcomes-focused regulation rather than rigid rules, and that the two authorities intend to engage stakeholders further and align sector efforts with national AI strategy. While the report does not itself prescribe a governance framework, it establishes the evidentiary groundwork for the regulatory expectations now forming. It is anticipated that these expectations will be articulated in the discussion paper being prepared jointly by the FSCA, the Prudential Authority and the SARB. The paper was announced in the National Treasury's 2026 Budget Review for publication in July 2026 but had not been released at the time of writing; it is intended to lead to a formal joint regulatory instrument (National Treasury, 2026). The domestic governance baseline has also moved in the interim. Joint Standard 1 of 2023, issued jointly by the Prudential Authority and the FSCA and effective 15 November 2024, establishes information technology (IT) governance and risk management requirements across the financial sector, including board oversight of technology risk, documented allocation of responsibility, and risk management of the systems on which regulated activities depend (Prudential Authority and Financial Sector Conduct Authority, 2023). It is not AI-specific and it does not reach the decision layer this paper addresses, but it establishes that the direction of domestic regulatory travel is toward stricter and more explicitly named technology governance. The AI Spine should be read as consistent with that direction and as extending it to the layer at which financial consequence is determined. The AI Spine, with the DMI as its methodology layer, offers a ready-made architecture for institutions preparing to demonstrate substantive governance ahead of those expectations.

5.7 The implication for the actuarial profession is unambiguous: algorithmic governance requirements are coming, through regulation if not through professional initiative. The choice is between designing governance architecture on terms shaped by

actuarial expertise, which the AI Spine enables, and having frameworks imposed through regulatory requirements less adapted to the nuances of South African financial services. The profession that defines the standard shapes the standard.

6. CASE ILLUSTRATIONS

The following four case illustrations apply the AI Spine framework to concrete operational scenarios across banking and insurance. Volumes and monetary figures are illustrative parameters consistent with mid-tier and large South African market characteristics; they are presented to demonstrate framework applicability and order-of-magnitude reasoning rather than as analyses of any specific institution. One case (\$6.3) is presented with full DMI scoring to illustrate the methodology in operation.

6.1 Payment routing as an automated treasury function (banking)

6.1.1 A South African bank operates an AI-driven payment routing system that processes interbank and cross-border payment instructions in real time, selecting routing paths to optimise settlement efficiency, manage counterparty exposure, and respond to intraday liquidity signals. The system executes tens of thousands of routing decisions daily, with aggregate daily settlement exposure measured in billions of rand.

6.1.2 Under traditional MRM governance, this system is treated as a model supporting treasury decisions. Under the AI Spine framework, it is a Decision Node with autonomous financial impact equivalent to a senior treasury function. It determines counterparty exposure, affects settlement timing, and allocates liquidity in ways that directly affect the bank's intraday capital position and its compliance with SARB liquidity requirements. *In effect, the algorithm is performing the work of the Treasury Operations Head and the missing C-Suite framing is literal rather than metaphorical.*

6.1.3 The governance gap is the absence of a defined Risk Owner with accountability for the system's alignment to Board-approved treasury risk appetite. The gap is distinct from message-layer governance, and the distinction should be drawn explicitly. The SWIFT and ISO 20022 frameworks govern message integrity, beneficiary identification and transaction screening at the point of instruction, and do so well. The institution's routing optimisation logic and the alignment of aggregate routing decisions to Board-approved treasury risk appetite operate at a different layer and are not addressed by message-level standards: a payment can be fully compliant at the message layer while the aggregate routing pattern has drifted from the exposure profile the Board sanctioned. The system's optimisation objective, settlement efficiency, is a technical target rather than a governance constraint. When settlement patterns shift due to changes in counterparty behaviour or market conditions, the system adapts its routing autonomously. Without governance oversight, this adaptation may produce counterparty concentration or liquidity corridor changes inconsistent with the Board's approved treasury strategy, invisible until a settlement failure or regulatory query surfaces them.

6.1.4 On a platform processing R5 billion in daily settlements, a persistent 2–3% routing misallocation equates to R100–R150 million in sub-optimal settlement exposure daily. Compounded across five trading days before detection by monthly model risk monitoring, this represents R500m–R750m in unmonitored counterparty risk: an exposure that no treasury committee has approved and no risk report has surfaced. This is the **decision-to-signal latency** problem in concrete form.

6.1.5 When applying the AI Spine, a Decision Node Map identifies the routing system as a HIGH-materiality node under the DMI. A Risk Owner is appointed with documented accountability for monitoring the node's alignment to Board treasury strategy. Monitoring is implemented on key indicators: counterparty concentration ratios, settlement efficiency variance from approved benchmarks, and routing path distribution stability. Graceful Degradation thresholds are set: if counterparty concentration exceeds approved limits, the system enters Human-Augmented Mode and routing decisions above a defined threshold are referred to a treasury officer for review.

6.1.6 Indicative DMI profile: D1 (Financial Consequence) and D6 (Systemic Exposure) dominate, with D7 (Human Accountability Clarity) supplying the override condition. HIGH tier on the financial materiality axis; potential Override 2 (single-event solvency exposure) engagement at peak settlement volumes.

6.2 Credit decisioning: the silent appetite shift (banking)

6.2.1 A mid-tier South African credit provider operates an automated credit origination system using a machine learning model trained on historical application data. The model generates credit approval and decline decisions for personal loan applications, processing approximately 2 000 applications daily with no human review of individual decisions in the approved credit tier.

6.2.2 **Silent Drift** occurs when the population applying for credit shifts over time relative to the model's training population, as it inevitably does in response to economic conditions, competitor behaviour, or marketing channel changes. A model trained during a period of economic stability may, when conditions deteriorate, begin approving applicants in risk segments that were rare in the training data and whose risk profile the model has not adequately characterised. The model's overall approval rate may appear stable. Its internal statistics may show no obvious deterioration. The approved portfolio is quietly accumulating higher-risk exposures that the Board has not sanctioned.

6.2.3 This is the **Silent Appetite Shift**: a condition in which the institution's effective credit risk appetite, as expressed by its automated decision system, has diverged from its Board-approved appetite without any human decision to change it. At 2 000 applications per day with an average approved loan value of R85 000, a 5% silent shift in the effective approval threshold represents approximately R8,5 million in daily credit exposure outside the Board's sanctioned risk appetite, accumulating to over R200 million in a single month before conventional portfolio monitoring would detect it. Under POPIA section 71,

affected applicants must be afforded an opportunity to make representations on their automated credit decision, which on the conservative governance reading set out in §4.2 requires a functional human-review pathway; that obligation cannot be discharged if the institution does not know which decisions were automated or does not have a functional review mechanism in place.

6.2.4 A common response to this gap is the insertion of human reviewers at the individual decision level. This response is insufficient, and the AI Spine does not propose it. The research literature on automation bias documents consistently that human reviewers operating under volume pressure and time constraints systematically defer to algorithmic outputs, producing what is effectively a rubber stamp at significant operational cost without genuine governance value (Goddard et al., 2012). A loan officer reviewing two thousand algorithmically generated decisions per day provides the institutional appearance of oversight rather than its substance.

6.2.5 The AI Spine's Human-Augmented Mode addresses a different and more tractable problem. Rather than inserting humans into individual decisions, where automation bias renders their presence largely performative, it triggers structural constraints on the system itself: credit limits capped at a defined percentage of standard thresholds, with decisions above a materiality threshold referred for genuine review. The human intervention is at the **portfolio level**. The Risk Owner reviews a statistical trigger, assesses the underlying population shift, and makes a governance decision about the system's operating parameters. This is a decision a skilled risk professional can make meaningfully, and one that cannot be automated because it requires human judgement about whether the Board's approved risk appetite remains appropriate given changed conditions. Under POPIA section 71, the representations obligation is met by maintaining a functional referral pathway for applicants who request review, rather than by routing every decision through a reviewer who will not genuinely engage with it.

6.2.6 Indicative DMI profile: D2 (Customer Harm Potential) and D4 (Reversibility) dominate alongside D3 (POPIA s71 engagement); HIGH tier driven by the persistence of bureau effect during reversal and the live POPIA s71 obligation.

6.3 Automated claims settlement: the Policyholder Protection Rule written reasons obligation (insurance), full Decision Materiality Index scoring

6.3.1 A large direct-to-customer South African short-term insurer operates an automated claims settlement engine that assesses, approves, and settles personal lines motor claims below a defined financial threshold. The engine processes approximately 9 000 personal lines claims per month, roughly half the portfolio volume, without human assessor involvement, generating outcomes in seconds of first notification of loss. The settlement engine takes inputs from a vendor damage-estimation AI and an in-house fraud scoring model.

6.3.2 The governance gap here is a live regulatory compliance gap, sitting

above rather than in the model risk layer. The Policyholder Protection Rules (Republic of South Africa, 2017b), as administered by the FSCA under the twin peaks architecture of the Financial Sector Regulation Act 9 of 2017 (Republic of South Africa, 2017a), require that insurers provide written reasons for any decision to reject, reduce, or conditionally settle a claim. This obligation applies regardless of whether the decision was made by a human assessor or an automated system. An automated decline that cannot be explained in terms intelligible to the policyholder, or that cannot be traced to a named individual with accountability for the decision, is PPR non-compliant and a candidate for Override 1 (live regulatory breach) under the DMI.

6.3.3 The PPR distinguishes notification from written reasons. Notification tells the policyholder what was decided. Written reasons tell them why: the specific basis on which their claim was rejected or reduced, expressed in terms the policyholder can meaningfully engage with, challenge, or escalate to the FSCA or an ombud. An automated system that outputs a decline code or a template rejection letter has satisfied the notification obligation. It has not satisfied the written reasons obligation unless the reasons are specific to that policyholder’s claim, traceable to the inputs the system used, and explainable without reference to model architecture that the policyholder cannot access or understand.

6.3.4 DMI scoring for this decision class is summarised in Table 3.

TABLE 3. DMI scoring, automated personal lines motor claims settlement under R75 000

Dimension	Weight	Score	Driving evidence
D1 Financial consequence	20%	3,5	R600 m+ annual settlement throughput; 2% quantum variance creates R12 m exposure; vendor damage-estimator unvalidated in SA repair environment.
D2 Customer harm potential	20%	4,5	Customer is post-loss; vehicle may be unroadworthy; 20-day dispute turnaround compounds harm before resolution.
D3 Regulatory obligation intensity	20%	4,5	PPR written-reasons obligation, TCF outcomes 3 and 6, and POPIA s71, all engaged at the individual decision level.
D4 Reversibility	15%	3,5	Reversible through dispute process; rental and finance-arrears costs accrue during reversal; substantive harm not undone by procedural reopening.
D5 Explainability gap	15%	4,0	System cannot produce per-decision reasons sufficient for PPR written-reasons obligation; decline rationale traceable at model level but not at applicant level.
D6 Systemic exposure	5%	3,5	9 000 monthly settlements; catastrophe-event drift exposes thousands of policyholders simultaneously.
D7 Human accountability clarity	5%	2,5	Operational accountability with Head of Claims; not formalised in Risk Committee terms or Operational Risk Appetite Statement.

6.3.5 Aggregate DMI score: 3,93. Tier: HIGH

6.3.5.1 Score derivation: $(3,5 \times 0,20) + (4,5 \times 0,20) + (4,5 \times 0,20) + (3,5 \times 0,15) + (4,0 \times 0,15) + (3,5 \times 0,05) + (2,5 \times 0,05) = 3,925$, rounded to 3,93.

6.3.5.2 The dimensional pattern matters as much as the headline score. The case is driven principally by D2 (Customer Harm), D3 (Regulatory Obligation), and

D5 (Explainability), a profile that points to a regulatory-compliance-led remediation programme rather than a capital-led one.

6.3.5.3 Applying the AI Spine: the Decision Node Map identifies the automated claims engine as a HIGH-materiality node with both a financial materiality dimension and a regulatory compliance dimension. The documentation requirement for each Decision Node, capturing decision objective, input variables, and the specific factors material to each outcome, directly supports the PPR Rule 17 written-reasons obligation: the system's decision rationale is captured as a governance output, rather than merely as a technical artefact. The Risk Owner, in this context the insurer's Head of Claims Risk, is accountable for ensuring that the system's decline decisions are explainable, auditable, and consistent with policy terms and TCF outcomes. Monitoring includes statistical drift indicators, complaints rate tracking, and regulatory correspondence analysis: early warning indicators of PPR compliance stress that the governance framework is designed to surface before they reach the regulator.

6.3.5.4 A GCA estimate for this decision class, applying the framework in §3.3, would be approximately $R600m \text{ annual throughput} \times 50\%$ (proportion currently outside the explainability standard) $\times 0,05$ (GDF for monitoring without per-decision explainability infrastructure) $\times 1,0$ (CF), producing a GCA estimate of approximately R15 million. The full GCA calculation requires institution-specific evidence of the proportion of decisions outside compliant explainability and the integration of correlated exposures with adjacent nodes (fraud scoring, vendor damage estimation). The order-of-magnitude figure is sufficient to make the governance deficit visible in capital terms.

6.4 Telematics pricing: proxy variable risk and TCF compliance (insurance)

6.4.1 A South African motor insurer operates a usage-based insurance (UBI) product in which premiums are dynamically adjusted at renewal based on telematics data: driving speed, braking patterns, time-of-day driving, and mileage. The pricing model is recalibrated quarterly using updated telematics and claims data. Price adjustments are applied automatically at renewal without individual underwriter review.

6.4.2 The governance challenge in this context is **proxy variable risk**: the risk that telematics-based pricing factors function as proxies for protected characteristics, including geographic area (strongly correlated with race in South African urban geography due to the spatial legacy of apartheid), socioeconomic status (correlated with vehicle type and driving patterns), or employment category, in ways that produce pricing outcomes inconsistent with the insurer's TCF obligations or that approach indirect discrimination under POPIA.

6.4.3 A telematics model that weights time-of-day driving heavily may inadvertently price shift workers disproportionately in low-income employment categories at premium rates reflecting their driving schedules rather than their genuine risk profiles. This is a governance failure, analytically distinct from a model accuracy failure. The model

may accurately reflect the historical correlation between time-of-day driving and claims frequency, while the model's objective function remains unconstrained by a fairness governance layer that distinguishes genuine risk factors from socioeconomic proxies.

6.4.4 In a portfolio of 150 000 UBI policyholders, if 15% of renewal price movements correlate with geographic variables that proxy for protected characteristics after controlling for genuine risk factors, the affected population exceeds 22 500 policyholders. Each carries a potential TCF complaint and a POPIA objection right. At scale, this becomes an enforcement exposure.

6.4.5 **Applying the AI Spine:** the pricing node is mapped with a specific compliance dimension appended to standard materiality assessment: proxy variable monitoring. The Risk Owner, a senior actuarial role in pricing, is required to run quarterly fairness assessments alongside standard performance monitoring, testing whether price movements at renewal correlate with protected characteristics after controlling for genuine risk factors. Graceful Degradation thresholds are set not only for statistical drift but for fairness metric deterioration: if a defined fairness threshold is breached, the pricing engine enters Human-Augmented Mode and renewal price increases above a defined threshold are referred to an underwriter for review. The Board risk committee receives quarterly reporting on both performance and fairness dimensions. This governance architecture preserves the use of telematics data while ensuring that it remains within the bounds of TCF compliance and POPIA obligations, and that the insurer can demonstrate this to the FSCA when required.

6.4.6 Indicative DMI profile: D2 (Customer Harm Potential) and D3 (Regulatory Obligation Intensity) dominate; HIGH tier driven by the TCF/POPIA enforcement exposure characteristic of proxy-variable pricing in the South African context.

6.5 What the four cases show together

The four cases illustrate three things. First, the AI Spine is sector-portable: the same architecture maps to banking treasury, retail credit, claims settlement, and pricing without modification of the core mechanisms. Second, the dimensional pattern of materiality varies. §6.3 is driven by regulatory and explainability dimensions; §6.1 by financial consequence and systemic exposure; §6.2 by customer harm and reversibility; §6.4 by customer harm and regulatory obligation. Different dimensional patterns produce different remediation programmes, even at the same headline score. Third, in each case the algorithm is performing a function historically held by a named C-Suite or senior executive role, and the absence of that named accountability, the missing C-Suite, is the governance condition the AI Spine is designed to close.

7. THE ACTUARY AS GOVERNOR OF ALGORITHMIC CAPITAL

The emergence of AI-driven decision systems creates both a professional obligation and a strategic opportunity for the actuarial profession. The obligation is clear:

as autonomous systems assume functions with material effect on solvency, consumer outcomes, and regulatory compliance, the profession's existing statutory and professional responsibilities, under the Appointed Actuary function, ORSA obligations, and Standard of Actuarial Practice (SAP) 901 (Actuarial Society of South Africa, n.d.), encompass these systems regardless of whether current governance frameworks explicitly address them. The opportunity is equally clear. No other profession combines the quantitative rigour, regulatory embeddedness, and financial accountability required to govern algorithmic capital at institutional scale. The AI Spine proposes that the actuarial profession's role in AI governance is executive in nature. The Risk Owner function in the framework carries genuine accountability, rather than merely analytical input. This represents a substantive expansion of the actuarial domain, from validating predictions to governing decisions, from measuring risk to enforcing the boundaries in which autonomous systems operate.

7.1 Four capabilities the profession brings

7.1.1 Actuaries are specifically positioned to contribute four capabilities essential to effective AI governance, capabilities that other governance disciplines do not replicate in combination:

7.1.2 First, **the ability to translate risk appetite statements into quantitative operating thresholds**: the PSI and KS trigger levels that define the Graceful Degradation protocol must be calibrated to the institution's specific risk tolerances, rather than set as generic industry benchmarks. This calibration requires the same discipline as experience analysis.

7.1.3 Second, **the ability to design and calibrate statistical monitoring systems** that are sensitive to material risk while maintaining operational practicality. This sits at the core of the DMI's Layer 2 and Layer 3 monitoring architecture.

7.1.4 Third, **the ability to quantify governance risk in capital terms** and integrate it into regulatory reporting frameworks, as the GCA demonstrates. This capability is what distinguishes actuarial governance from compliance-led or audit-led approaches.

7.1.5 Fourth, **the credibility to hold accountable the technology and business functions whose AI systems they govern**: a credibility grounded in professional registration, statutory obligations, and a track record of independent risk judgement.

7.1.6 Why this combination is rare

7.1.6.1 The four capabilities above are each held by other professions in financial services. The case for the actuarial profession's role rests on the combination: all four together, embedded in a statutory function that already carries fiduciary accountability for financial outcomes, are rarely held outside actuarial roles. The point is worth making directly, because adjacent professions are reasonable competitors for this work and the jurisdictional claim should be defended explicitly rather than assumed.

7.1.6.2 **Technology and model risk functions** understand AI systems at a level

the actuarial profession typically does not, but the capital framework, the ORSA process, and the SAM Pillar II machinery in which governance consequence has to be expressed sit outside their domain. A technology risk officer can characterise a model's drift profile precisely without being equipped to translate that drift into a Pillar II governance signal or an own-funds implication.

7.1.6.3 Internal audit carries the institutional independence and the control-architecture vocabulary the work requires, but typically lacks the statistical machinery to design and calibrate monitoring thresholds at the level the AI Spine specifies. An internal auditor can verify that monitoring exists; the actuarial discipline is what allows monitoring to be calibrated to material risk tolerances in the first place.

7.1.6.4 Data ethics and fairness specialists have developed the most sophisticated treatment of bias, proxy discrimination, and algorithmic harm available in the literature, but the translation of those concerns into capital, reserving, and supervisory reporting is not the discipline they were trained in. Fairness without solvency is incomplete; solvency without fairness is also incomplete; the combination requires both vocabularies.

7.1.6.5 Compliance and legal functions hold the regulatory text in detail, and any AI governance work that engages PPR Rule 17 or POPIA s71 must be done in close partnership with them. Compliance characteristically governs at the institution-policy level rather than at the operational instrumentation level. The Risk Owner function specified by the AI Spine sits closer to instrumentation than policy, which is the actuarial domain.

7.1.6.6 The actuarial profession's claim, then, concerns location rather than exclusivity: the four capabilities above, combined in a single statutory function that already carries Board-level accountability for financial outcomes under SAM and the Insurance Act (Republic of South Africa, 2018), are most economically located in the actuarial domain. The profession's existing regulatory architecture, the Appointed Actuary function, the Head of Actuarial Function role, and the ORSA process, is the architecture into which the Risk Owner function specified by the AI Spine most naturally fits. Where the work is shared with adjacent professions, the AI Spine's three-tier Ownership structure (Design Owner, Risk Owner, Business Owner) provides the formal allocation; the case made here concerns the Risk Owner role specifically.

7.2 The actuarial qualification gap

7.2.1 Acknowledging the profession's opportunity requires also acknowledging the qualification gap that currently separates most actuarial practitioners from effective AI governance practice. Governing an automated decision system requires skills that current actuarial curricula and CPD frameworks address only partially: machine learning model architecture and its implications for monitoring and interpretability; algorithmic auditing methodology; real-time data pipeline design and its governance requirements; fairness metric construction and its relationship to regulatory obligations; and the organisational dynamics of cross-functional AI governance in large institutions.

7.2.2 The profession has begun to respond. The International Actuarial Association's AI Task Force published an Artificial Intelligence Governance Framework in November 2025, supported by companion papers on the testing and documentation of AI models and systems, and following its earlier actuarial control cycle for narrow artificial intelligence (International Actuarial Association, 2024; 2025a; 2025b; 2025c). The governance framework is a substantial contribution, and its lifecycle treatment, its handling of third-party and vendor models, and its appendix on bias, fairness and discrimination are directly useful to the work this paper describes. Three features nonetheless leave the gap open. First, the framework is expressly educational: it states that it is not an International Standard of Actuarial Practice and that it sets no requirements with which any actuary is expected to comply. Second, its unit of analysis is the AI model or AI system rather than the decision class, and its ownership structure resolves to a single Model Owner accountable for the model lifecycle rather than to a named executive accountable for the financial outcomes that the model enables. Its treatment of model risk rating records that rating methodologies vary by organisation, without specifying a dimensional architecture, a weighting structure, or tier boundaries. Third, it contains no capital dimension: the ORSA, the SCR and the Pillar II machinery through which governance consequence becomes visible to a Board do not feature in it. It is also worth observing that the framework takes the 2011 United States model risk management guidance as its stated background reference for MRM, five months before that guidance was superseded by SR 26-2 and its explicit exclusion of generative and agentic AI. The general obligations of SAP 901 apply to this work without specification for AI-enabled decision systems or for the accountability structures they require. The distance between where most actuarial practitioners are and where the AI Spine's Risk Owner function requires them to be is therefore real, and should be acknowledged as such by the profession.

7.2.3 Several development pathways are available. CPD requirements should explicitly include AI governance competencies: the full lifecycle of algorithmic governance from node mapping through monitoring design to escalation management and Board reporting, beyond simply model validation techniques. ASSA is well positioned to develop a structured AI governance competency framework, analogous to the Chartered Enterprise Risk Analyst (CERA) designation's structured approach to enterprise risk management, that provides practitioners with a recognised qualification pathway. Research partnerships with South African universities and the broader African AI research community should be developed to build the empirical base for AI governance practice: calibration data for GCA parameters, fairness audit methodologies appropriate for South African financial services contexts, and the governance architecture patterns that function effectively at different scales of AI deployment.

7.3 The crosswalk to existing obligations

7.3.1 The AI Spine operationalises existing regulatory obligations. Table 4 maps each AI Spine element to the specific clause it satisfies in each major framework.

The crosswalk is itself a deliverable: a compliance officer or chief risk officer can use it to demonstrate to internal audit, the Risk Committee, or the regulator that the institution’s AI governance architecture satisfies known obligations explicitly.

TABLE 4. AI Spine to regulatory obligation crosswalk

AI spine element	Decision node map	DMI (materiality scoring)	Three-tier ownership	Three-layer monitoring	Graceful degradation	Governance capital add-on
SAM Pillar II / ORSA	System of governance (Art. 41 equivalent)	Risk identification and assessment	Allocation of responsibilities	Effective internal controls	Risk mitigation and contingency	ORSA; own assessment of capital adequacy
POPIA s71	Identification of automated decisions producing legal effects	Risk-based assessment of automated decisions	Designation of accountable parties	Mechanism for objection and review	Mechanism for human intervention	n/a
PPR / NCR	Mapping of decision points subject to written-reasons obligation	Prioritisation of compliance focus	Identification of accountable person for written reasons	Monitoring of compliance with written-reasons obligation	Functional referral pathway	n/a
TCF	Identification of customer-impact decisions	Outcome-based risk prioritisation	Customer outcome accountability	TCF outcomes monitoring	Outcome-based intervention	Capital implication of governance findings
SAP 901/ IAA AI guidance	Inventory of data-science applications	Application of data science risk assessment	Professional responsibility allocation	Ongoing model and data science monitoring	Escalation triggers	n/a
EU AI Act	Risk classification, technical documentation (Annex IV)	High-risk classification under Annex III	Human oversight obligations	Post-market monitoring	Human-in-the-loop requirement	n/a
NIST AI RMF	Map function	Measure function	Govern function	Manage function	Manage function	n/a
SR 11-7 / SR 26-2	Comprehensive model inventory	Risk tiering for validation depth	Named responsibility for model lifecycle	Ongoing performance monitoring	Defined escalation thresholds	n/a
Nov 2025 PA/FSCA joint report	Required AI use case register	Risk-based proportionality	Documented accountability	Continuous monitoring expectation	Structured human oversight	n/a

7.3.2 The cell entries are summary references; full citation chains are available on request to the author.

8. IMPLEMENTATION CONSIDERATIONS

8.1 The AI Spine assumes a degree of institutional maturity that varies considerably across South African financial services. Implementation realism matters, both because it shapes the framework's adoption pathway and because peer reviewers will reasonably ask whether the architecture is operable in practice.

8.2 **Vendor and third-party AI.** A substantial proportion of AI in South African financial services is delivered through vendor platforms (claims engines, credit decisioning platforms, fraud detection systems, telematics aggregators) over which the institution has limited technical visibility. The AI Spine accommodates this through Dimension 5 (Explainability Gap) and Dimension 7 (Human Accountability Clarity) of the DMI. Vendor-AI nodes typically score higher on these dimensions because the institution cannot produce per-decision explanation or trace decision logic to an institution-internal accountable executive. The governance response is contractual (vendor service level agreements (SLAs) must include per-decision logging, explainability data, and incident notification obligations) and operational (through independent validation of vendor outputs against the institution's deployed population). Vendor opacity elevates the documentary and contractual burden while leaving the institution's decision-level accountability intact.

8.3 **Talent:** the Risk Owner function as specified in §4.4 requires actuarial or risk practitioners with cross-functional fluency. The qualification gap discussed in §7.2 is real, and institutional implementation will outpace the supply of fully qualified Risk Owners for several years. A pragmatic interim approach is to constitute Risk Owner *responsibility* as a documented role assignment with defined accountability, supported by an AI Governance Committee that combines actuarial, technology, legal, and compliance expertise. The intent is that the documented Risk Owner does not personally hold every capability required, but is accountable for ensuring those capabilities are brought to bear.

8.4 **Dispute resolution between Risk and Business Owners:** the three-tier ownership structure operationalises accountability, and in the ordinary case the Risk Owner and the Business Owner operate as collaborators with a shared interest in the decision system running profitably and within appetite. Most interactions between them are not disputes at all, and the framework should not be read as constituting an adversarial interface. As with any cross-functional governance structure, however, points of friction do emerge in practice, and a framework that provides no adjudication path for them is incomplete. The most common friction arises between the Risk Owner and the Business Owner: the Risk Owner flags moderate PSI drift or recommends Human-Augmented Mode, while the Business Owner pushes back on short-term commercial or operational grounds. The framework provides two procedural safeguards against deadlock. First, where the

Risk Owner's recommendation flows from an objective statistical trigger (PSI, KS, or a documented dimensional finding), the trigger itself generates an obligation to enter the corresponding Graceful Degradation mode, removing the trigger from negotiated territory. Second, where the dispute concerns whether thresholds should be recalibrated or risk appetite revised in light of changed conditions, the matter escalates directly to the Board Risk Committee for adjudication in five working days, with the system held in the more conservative mode until adjudication is complete. The Risk Owner retains a documented right of escalation, and the Business Owner retains a documented right to seek Board ratification of a different posture, but the system itself defaults to the more conservative state during disputes. This converts a potential operational deadlock into a documented governance event with a defined adjudication path.

8.5 Cost: Full AI Spine implementation across an institution's AI estate is a multi-year programme. The DMI provides the prioritisation that makes the programme tractable. HIGH-tier decision classes receive immediate attention; MEDIUM-tier classes receive standard controls; LOW-tier classes remain in existing risk machinery. Implementation effort scales sharply with infrastructure maturity, and a single timeline estimate would be misleading. For a mid-tier insurer with eight to twelve material decision classes operating reasonably modern, well-documented systems with accessible per-decision logs, the initial DMI assessment and HIGH-tier remediation is typically a six- to twelve-month programme. For institutions operating legacy core platforms, and particularly mainframe environments in which decision logic is embedded in application code and per-decision logging must be built rather than configured, the equivalent scope should be planned at 18 to 36 months. In the latter case the majority of the effort falls in Decision Node Mapping and logging infrastructure rather than in the DMI assessment itself, which is a useful planning distinction because the assessment can proceed on the mapped subset while logging capability is built out. In both cases, subsequent maintenance through periodic DMI refresh integrates into the existing risk management cycle.

8.6 Phased adoption: a reasonable phased path is: (1) AI Inventory and shadow-AI discovery; (2) DMI assessment of identified decision classes; (3) HIGH-tier remediation, Risk Owner appointment, decision-level logging, monitoring infrastructure; (4) integration into ORSA and Board reporting; (5) GCA calibration and reporting. Phases 1 and 2 are typically deliverable in six months for an institution starting from a documented model inventory; the full architecture matures over 18–24 months.

8.7 Adversarial and supply-chain exposure: AI decision systems carry a class of risk that traditional model risk management addresses only partially: deliberate manipulation of inputs, outputs, or the system itself by adversarial actors. Examples include data-poisoning of training pipelines, evasion attacks on production scoring models, prompt

injection in agentic systems that consume natural-language inputs, model-extraction attacks that reconstruct proprietary scoring logic from queried outputs, and supply-chain compromise where the model or its weights are altered upstream of deployment. The AI Spine accommodates these exposures through Dimension 6 (Systemic Exposure) and the Layer 1 monitoring architecture, but explicit treatment is appropriate. The Risk Owner should ensure that adversarial-robustness testing, input-anomaly detection, and integrity verification of vendor model artefacts are part of the monitoring design for HIGH-tier decision classes, particularly those exposed to external inputs (claims documentation, customer-provided application data, third-party data feeds). The institution's information-security function holds the technical capability; the Risk Owner integrates that capability into the AI governance framework. A formal treatment of adversarial robustness in the AI Spine architecture is a development priority identified in §9.

8.8 Multi-party AI systems: the framework as currently specified addresses governance in a single institution. AI systems that operate across institutional boundaries (shared industry platforms, credit bureaus, insurance pooling arrangements, agentic transaction infrastructure) distribute accountability across multiple entities. These multi-party governance structures represent a distinct challenge that requires further development, and are an active area of the author's ongoing work.

9. SCOPE AND LIMITATIONS

9.1 This paper introduces the AI Spine as a governance architecture framework and the DMI as its methodology layer, rather than as a complete technical specification. Several limitations should be acknowledged.

9.2 The paper focuses on financial services in South Africa, with particular reference to insurance and banking. The AI Spine's principles are generalisable, but application in other sectors (healthcare, infrastructure, public services) would require adaptation of regulatory anchors, materiality thresholds, and monitoring design to the relevant operational and regulatory contexts.

9.3 The GCA presented in §3.3 is a structural illustration rather than a calibrated formula (see §3.3 and Appendix B). The calibration of the GDF and the Correlation Factor requires empirical development grounded in loss event data, regulatory engagement, and industry validation. The DMI's weighting structure is rationally rather than empirically calibrated; empirical calibration against observed regulatory and litigation outcomes is an active development priority for future versions. Inter-rater reliability of DMI scoring across independent assessors is currently characterised qualitatively through engagement experience rather than measured statistically. Formal reliability studies are an appropriate area for ASSA-sponsored research.

9.4 The AI Spine addresses governance architecture rather than specific model types. It is designed to be model-agnostic and applies equally to logistic regression scorecards, gradient boosting models, neural networks, and rules-based systems. It does not provide specific technical guidance on explainability methods for deep learning models, a significant and rapidly developing field that warrants separate treatment.

9.5 The framework assumes a degree of institutional maturity that may not be present in smaller institutions or those at early stages of AI adoption. The Decision Node Mapping methodology requires access to comprehensive system documentation that may not exist in all deployment environments; in such contexts, a foundational documentation phase precedes full framework application.

9.6 A note on the framework's positioning

The framework presented in this paper is offered as a starting reference for ASSA development of a profession-level standard for the governance of AI-enabled financial decisions. The DMI dimensional architecture, the seven dimensions and their anchored scoring logic, the override conditions, and the AI Spine's seven-layer structure are contributed to the profession in this paper for the purposes of peer review, scholarly engagement, and the development of a non-proprietary standard. The author's commercial application of the framework, including the full DMI anchor specifications used in production engagements and the operational implementation guidance, is documented separately in Kitantou (2026); that document functions as an application guide rather than as a missing methodology layer. The intent is that the methodology be available for the profession to build on, refine, and calibrate, while the author continues to develop its commercial application alongside that profession-level work.

9.7 Two areas of active development warrant mention. First, a **continuous-monitoring overlay** that re-scores selected DMI dimensions on real-time signal (volume drift, complaint rate, dispute reversal rate) is in design. The current DMI is a snapshot, refreshed periodically. Second, an extension to **agentic AI and autonomous transaction initiation**, where AI systems move from decision-support to decision-initiation, including agentic payment infrastructure and AI-orchestrated trade execution, is under development. In these contexts, the decision class is the institution itself, and the DMI architecture is being extended to score the fiduciary authority of agentic systems.

10. CONCLUSION: THE AI SPINE AS A PROFESSIONAL STANDARD

10.1 The question facing South African financial services is no longer whether algorithms will make consequential decisions. They already do, at scale, in real time, across every major sector of the financial system. The question is whether those decisions will be made under governance (with defined accountability, calibrated monitoring, and enforceable

controls), or whether they will continue to be made under the fragmented and informally held controls that characterise most AI deployment in the sector, where governance of a kind exists at several points but accountability for the decision itself is named nowhere.

10.2 The AI Spine provides the architecture for a governed answer. By mapping Decision Nodes, scoring them through the DMI, establishing statutory-equivalent ownership structures, implementing three-layer monitoring with Graceful Degradation protocols, and quantifying governance risk through the GCA, it translates abstract accountability obligations into concrete operational controls. It does so in a manner aligned with SAM Pillar II, consistent with emerging international regulatory standards, and grounded in actuarial practice.

10.3 For the actuarial profession, three concrete next steps follow:

- First, **ASSA should constitute an AI Governance Working Party** with a defined remit to develop the empirical base for DMI weighting and GCA calibration. The working party should include practitioners with active engagement experience in AI governance, supervisory representation from the Prudential Authority and the FSCA, and academic researchers in machine learning and algorithmic accountability. Its outputs should inform the development of ASSA practice guidance for data science and AI, and the profession's contribution to the forthcoming PA/FSCA discussion paper consultation on AI in financial services.
- Second, **ASSA should issue governance-specific practice guidance** for actuaries acting as Risk Owners in the sense defined by this paper. The guidance need not be onerous; it should establish that the Risk Owner role is in the actuarial domain, define the minimum competencies expected, and connect the role to existing professional accountability mechanisms.
- Third, **academic research partnerships** should be commissioned to build the empirical base the framework currently lacks: calibration data for GCA parameters, fairness audit methodologies appropriate for the South African context, and longitudinal inter-rater reliability studies for DMI scoring. This quantitative development work defines a new actuarial practice domain.

10.4 The AI Spine provides a concrete response to the ASSA 2026 convention theme. Broadening boundaries means applying the actuarial core of quantitative rigour, professional accountability, and capital integrity, intact, to the systems that now exercise the judgement that used to belong exclusively to people.

10.5 Algorithms have taken the decisions. The actuarial profession is the one best placed to ensure they take the accountability with them.

REFERENCES

- Actuarial Society of South Africa (n.d.). SAP 901: General Actuarial Practice, Version 1. Johannesburg: ASSA. Available at: actuarialsociety.org.za
- Bank of England Prudential Regulation Authority (2023). Model risk management principles for banks (Supervisory Statement SS1/23). London: Bank of England
- Barocas, S, Hardt, M & Narayanan, A (2019). Fairness and machine learning: Limitations and opportunities. fairmlbook.org
- Basel Committee on Banking Supervision (2024). Digitalisation of finance. Report d575, 16 May 2024. Basel: Bank for International Settlements
- Basu, D & Das, U (2025). The fair game: Auditing and debiasing AI algorithms over time. *Cambridge Forum on AI: Law and Governance*, 1, e27. doi: 10.1017/cfl.2025.8
- Board of Governors of the Federal Reserve System and Office of the Comptroller of the Currency (2011). Supervisory guidance on model risk management (SR Letter 11-7 / OCC Bulletin 2011-12). Washington, DC: Federal Reserve / OCC
- Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation and Office of the Comptroller of the Currency (2026). Revised guidance on model risk management (SR Letter 26-2 / OCC Bulletin 2026-13), 17 April 2026. Washington, DC: Federal Reserve / FDIC / OCC
- European Parliament and Council of the European Union (2024). Regulation (EU) 2024/1689 on Artificial Intelligence (EU AI Act). *Official Journal of the European Union*
- Goddard, K, Roudsari, AC & Wyatt, JC (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121–127
- International Actuarial Association (2024). Actuarial control cycle for narrow artificial intelligence. AI Task Force deliverable. Ottawa: IAA
- International Actuarial Association (2025a). Artificial intelligence governance framework. AI Task Force, November 2025. Ottawa: IAA
- International Actuarial Association (2025b). Testing of artificial intelligence models or systems. AI Task Force. Ottawa: IAA
- International Actuarial Association (2025c). Documentation of artificial intelligence models or systems. AI Task Force. Ottawa: IAA
- Kitantou, Z (2026). The decision materiality index: How South African Boards govern AI-enabled financial decisions (AI Spine Decision Governance Framework Whitepaper v1.1). Johannesburg: Nakhile Consulting. Public overview: doi: 10.5281/zenodo.21134247
- Mittelstadt, BD, Allo, P, Taddeo, M, Wachter, S & Floridi, L (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21
- Monetary Authority of Singapore (2018). Principles to Promote fairness, ethics, accountability and transparency (FEAT) in the use of artificial intelligence and data analytics in Singapore's financial sector. Singapore: MAS

- National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1. Gaithersburg: US Department of Commerce
- National Treasury (2026). Budget review 2026. Pretoria: National Treasury
- New York State Department of Financial Services (NYDFS) (2021). Report on Apple Card investigation. March 2021. New York: NYDFS
- O’Neil, C (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group
- Prudential Authority (2018). Prudential Standards under the Insurance Act, 2017: Governance and Operational Standards for Insurers (GOI 1 and GOI 3). Pretoria: South African Reserve Bank Prudential Authority
- Prudential Authority and Financial Sector Conduct Authority (2023). Joint Standard 1 of 2023: Information technology governance and risk management requirements for financial institutions. Effective 15 November 2024. Pretoria: South African Reserve Bank Prudential Authority and FSCA
- Prudential Authority and Financial Sector Conduct Authority (2025). Artificial intelligence in the South African financial sector. 24 November 2025. Pretoria: South African Reserve Bank Prudential Authority and FSCA
- Rechtbank Den Haag (2020). *NJCM and Others v The State of the Netherlands* (SyRI), ECLI:NL:RBDHA:2020:865, judgment of 5 February 2020 (English translation: ECLI:NL:RBDHA:2020:1878)
- Republic of South Africa (2013). Protection of Personal Information Act 4 of 2013. Government Gazette No. 37067. Pretoria
- Republic of South Africa (2017a). Financial Sector Regulation Act 9 of 2017. *Government Gazette* No. 41060. Pretoria
- Republic of South Africa (2017b). Policyholder Protection Rules (Short-term Insurance), 2017, made under section 55 of the Short-term Insurance Act 53 of 1998. Government Notice 1433, *Government Gazette* No. 41329, 15 December 2017. Pretoria
- Republic of South Africa (2018). Insurance Act 18 of 2017. *Government Gazette* No. 41388, 18 January 2018. Pretoria
- Rudin, C (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215
- Selbst, AD, Boyd, D, Friedler, SA, Venkatasubramanian, S & Vertesi, J (2019). Fairness and abstraction in sociotechnical systems. *In: Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* 2019)*, 59–68. Atlanta: ACM
- Veale, M & Brass, I (2019). Administration by algorithm? Public Management meets public sector machine learning. *In* Yeung, K & Lodge, M (eds), *Algorithmic Regulation*. Oxford: Oxford University Press
- Wachter, S, Mittelstadt, B & Russell, C (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law and Technology*, 31(2), 841–887

APPENDIX A

BOARD READINESS CHECKLIST, TEN QUESTIONS

The ten questions below are the board-level analogue of the DMI. A board chair, audit and risk committee, or external reviewer can use the checklist as a one-page test of AI governance readiness, without a full DMI engagement. Each question maps to the DMI dimension it most directly engages, so a board that struggles to answer a question knows which dimension is under-developed.

A.1 The escalation path. If your AI suddenly doubles the default rate in a single province, who receives the automated alert before the month-end report and within what time? *(DMI: D7 Accountability)*

A.2 The override protocol. Can a human executive manually override an algorithmic ‘decline’ or ‘limit’ without breaking the system’s data integrity? *(DMI: D7 Accountability / D5 Explainability)*

A.3 The ‘why’ factor. If a customer or the regulator asks exactly why an automated credit limit was set at R10 000 instead of R50 000, can your team produce a data-backed reason in under 24 hours? *(DMI: D5 Explainability / D3 Regulatory Obligation Intensity)*

A.4 Proxy bias check. Have you tested whether your ‘neutral’ variables (device type, mobile data usage, time-of-day driving) are acting as proxies for protected classes? *(DMI: D2 Customer Harm Potential / D3 Regulatory Obligation Intensity)*

A.5 Drift detection. Do you have a back-testing schedule that compares current algorithmic predictions against actual repayment, claims, or settlement behaviour? *(DMI: D1 Financial Consequence / D6 Systemic Exposure)*

A.6 The kill-switch. In the event of a systemic logic error or a ‘flash crash’ in your scoring model, do you have a pre-approved protocol to freeze automated approvals? *(DMI: D7 Accountability / D6 Systemic Exposure)*

A.7 Third-party handshake. If your primary data provider has a data outage or ‘garbage data’ surge, how does your algorithm react? *(DMI: D1 Financial Consequence / D5 Explainability)*

A.8 The audit trail. Could an external auditor, or a due-diligence team, trace a single decision back to the specific version of the code that made it? *(DMI: D5 Explainability / D7 Accountability)*

A.9 The accountability node. Has the Board formally designated an accountable individual qualified to sign off on the ethics and logic of the algorithm, or is the Board currently assuming collective liability for a black-box system it cannot explain? (*DMI: D7 Accountability*)

A.10 The capital adequacy buffer. If the algorithm's probability-of-default model is found to be under-calculating risk by 1% due to drift, do you have a real-time calculation of how much additional capital would be required to remain compliant? (*DMI: D1 Financial Consequence*)

A.11 Interpretation. If the institution can answer fewer than seven of these ten questions with documented evidence, it is operating without a decision-level AI governance perimeter. The DMI assessment is the structured remediation pathway.

APPENDIX B.

THE GOVERNANCE CAPITAL ADD-ON, WORKED STRUCTURAL ILLUSTRATION

This appendix sets out an extended worked illustration of the GCA formula introduced in §3.3, applied to the case in §6.3. The figures throughout are illustrative anchors rather than empirically calibrated parameters. They are intended to demonstrate the formula's behaviour and the order-of-magnitude effect a GCA might produce on Pillar II capital reporting. They are presented as illustrative anchors, and institutions should not rely on them as supervisory parameters.

B.1 The example portfolio

The institution is the same illustrative large direct-to-customer short-term insurer as in §6.3:

- Solvency Capital Requirement: R3,5 billion
- Own Funds: R5,0 billion
- Solvency Cover Ratio: 143%
- Three HIGH-tier decision classes:
 - Automated personal lines motor claims settlement (the case worked below, same as §6.3)
 - Automated commercial claims triage
 - Renewal pricing with telematics inputs

B.2 Decision volume at risk (DVR)

B.2.1 The automated personal lines motor claims settlement engine processes approximately 9 000 settlements per month with annual throughput of approximately R600 million. The DMI assessment (§6.3) identifies that the per-decision explainability infrastructure required to meet the PPR Rule 17 written-reasons obligation is not yet in place.

B.2.2 A conservative DVR formulation attributes the governance deficit only to the portion of decisions where the obligation is actually engaged, declines and reductions, where written reasons are required, rather than to the full settlement throughput. The institution estimates this proportion at 50%.

B.2.3 A prior caution applies, following §3.3.2. The R600 m annual throughput used here is the throughput actually routed to automated settlement. It is not the insurer's total personal lines motor claims exposure. The insurer's intake rules exclude several categories from automated handling entirely, among them claims above the R75 000 threshold, total losses, third-party liability claims, and claims flagged by the fraud workflow at first notification. Constructing the DVR from total portfolio exposure rather than from the automated population would overstate the GCA substantially. Practitioners applying

this illustration should establish the automated population first, and should document that population as part of the Decision Node Map.

$$\text{DVR} = \text{R}600 \text{ m} \times 50\% = \text{R}300 \text{ m a year.}$$

(An alternative, more aggressive formulation would attribute the deficit to the full throughput on the basis that *every* automated settlement decision is one for which the institution would be obliged to explain the rationale if challenged. That formulation would produce a DVR of R600 m, and a correspondingly higher GCA. The conservative formulation is preferred here for first-principle defensibility.)

B.3 Governance deficit factor (GDF)

Using the illustrative scalar range introduced in §3.3, the institution's governance posture is assessed as 'documented ownership but no operational monitoring at the dimension that is failing' (D5, Explainability Gap):

$$\text{GDF} = 0,05 \text{ (illustrative).}$$

This is the parameter most exposed to the calibration critique in §3.3. Empirical calibration would require evidence anchored either to operational risk loss events for control-failure incidents in financial services, or to documented governance-failure outcomes; this paper does not propose a calibrated value.

B.4 Correlation factor (CF)

The automated claims settlement engine takes inputs from a vendor damage-estimation model and an in-house fraud scoring model. Each is itself an AI sub-system; the explainability gap at the settlement layer propagates into the upstream models because the institution cannot satisfactorily explain settlement outcomes that trace back to vendor model logic. The institution assesses correlation as moderate.

$$\text{CF} = 1,0 \text{ (illustrative, taken as neutral pending dedicated assessment of cross-node propagation).}$$

B.5 The illustrative GCA for one decision class

$$\text{GCA} = \text{R}300 \text{ m} \times 0,05 \times 1,0 = \text{R}15 \text{ million a year for this decision class.}$$

B.6 Portfolio aggregation across HIGH-tier classes

The illustrative insurer holds three HIGH-tier decision classes. Applying the same conservative DVR methodology (50% of throughput) and the same illustrative GDF (0,05) to all three would produce a portfolio GCA of approximately R45 m ($3 \times \text{R}15 \text{ m}$). The

wider indicative range of **R45–75 million a year** reflects the realistic heterogeneity across decision classes: two of the three classes carry stronger D5 (Explainability) gaps that justify a higher GDF anchor (0,07–0,08 rather than 0,05) and one carries cross-node correlation ($CF > 1,0$) with adjacent fraud and vendor sub-systems. For the worked illustration below a midpoint is taken in the range:

Portfolio GCA = R60 million a year (illustrative).

The specific build-up across the three classes (which differ in volume, regulatory engagement, and governance posture) would be documented in the institution's ORSA submission; for illustration here only the aggregate is shown.

B.7 Effect on Solvency Cover Ratio

B.7.1 Incorporated as a Pillar II add-on to the institution's SCR, a portfolio GCA of R60m produces a Solvency Cover Ratio impact of approximately 1,7 percentage points: SCR effectively increases from R3,5 billion to R3,56 billion, and the Cover Ratio falls from 143,0% to approximately 141,3%.

B.7.2 This is a *modest but visible* capital effect, small in proportional terms but reportable to the Board through the ORSA and surfaced to the supervisor through capital-efficiency disclosure. The illustration's value lies in *visibility* rather than magnitude. Governance deficit that was previously invisible to the capital framework now produces a defined capital signal commensurable with other risk categories.

B.7.3 In an institution with greater AI estate or more concentrated automation, the effect scales with DVR; in the largest South African banking groups, where DVR across a comprehensive AI estate could plausibly run into billions of rand, the illustrative SCR impact would be materially larger.

B.8 Incentive alignment in action

The mechanism's value is in the incentive direction it creates. Implementing per-decision explainability infrastructure would, in this illustrative example, lower the GDF (deficit severity drops) and reduce the DVR (the portion of throughput outside the governance threshold falls). A successful remediation could plausibly drop the per-decision-class GCA from R15m to under R3m, with corresponding capital relief. The framework is designed so that genuine governance investment produces measurable capital benefit; this is the structural condition under which institutions choose substantive over documentary compliance.

B.9 Critical caveats

The figures above illustrate the formula's behaviour. They illustrate behaviour rather than proposing supervisory parameters. Specifically:

- The GDF scalar of 0,05 is illustrative; its empirical calibration is the ASSA-sponsored research identified in §9.
- The DVR methodology adopted (50% of throughput) is one of several plausible formulations; the choice materially affects the GCA magnitude and would require supervisor engagement to standardise.
- The CF is set to 1,0 here; in practice, cross-node correlation in AI decision chains has not been well characterised in the literature, and a calibrated CF requires institution-specific assessment of how governance failure propagates through linked nodes.
- The portfolio aggregation in §B.6 assumes independence across decision classes. Correlated governance failures, for example a shared vendor across multiple HIGH-tier classes, would produce a larger aggregate GCA than the indicative midpoint shown.

The illustration's purpose is to show that the GCA could be made to produce a capital signal commensurable with other Pillar II reporting. The work of making it so is supervisory and actuarial in nature, and is identified as the development priority in §9.

APPENDIX C

DIMENSION 5 (EXPLAINABILITY GAP), ILLUSTRATIVE ANCHOR SPECIFICATION

C.1 Purpose and provenance

This appendix reproduces, in condensed form, the anchor specification for Dimension 5 (Explainability Gap) from the published DMI methodology (Kitantou, 2026). It is included for two reasons. First, D5 is the dimension identified in §4.3.7.6 as the principal driver of observed inter-rater variance, so it is the dimension on which the grounding of assessor judgement most warrants demonstration. Second, it illustrates how the conceptual definition of a dimension translates into anchored scoring criteria and an evidence inventory; the remaining six dimensions are anchored in the same manner in the full specification. The wording below follows the published specification so that the instrument described in this paper and the instrument in the public record remain a single instrument.

C.2 Definition and scope

C.2.1 Dimension 5 measures the gap between the explanation the institution can currently produce for an individual decision in the class, and the explanation the institution is obliged to produce, by regulation, by contract, or by reasonable customer expectation.

C.2.2 The dimension is not a measure of model interpretability in the technical sense of feature-attribution or counterfactual tooling. It is a governance question: when a customer asks why a claim was declined, a loan refused, or a premium increased, can the institution produce an explanation specific to that customer's case, defensible to a regulator, and intelligible to the customer. Where regulation requires written reasons, as under the PPR and the adverse action provisions of the National Credit Act, an inability to produce a per-decision explanation is a live regulatory exposure rather than a future risk.

C.3 Scoring anchors

The scale carries named anchors at scores 1, 3 and 5. Scores 2 and 4 are interpolated positions, assigned where the evidence supports a state between the named anchors; half points are available where the evidence supports them. This anchor-and-interpolation structure is common to all seven dimensions.

- **Score 1 (no gap):** decision logic fully transparent. An individual-level explanation is generated automatically for each decision, and the explanation language is intelligible to the customer.
- **Score 3 (producible with effort):** a general explanation is available at decision-class level. An individual-decision explanation is producible with analyst effort within 48 hours.

- **Score 5 (gap engaged and unmet):** black-box at the individual decision level. The institution cannot produce a decision-specific explanation defensible to a regulator, court, or customer, and a regulatory written-reasons obligation is engaged but not met.

C.4 Evidence inventory

Each dimensional score is supported by a written rationale citing the specific evidence relied upon. For D5 the specified inventory comprises:

- a sample of decision letter or notice outputs for the class;
- the per-decision logging architecture;
- the existence and coverage of an explanation generation pipeline;
- compliance sign-off on explanation language against the engaged instruments (PPR, National Credit Act, TCF); and
- customer comprehension testing, where conducted.

C.5 Application

The interpolated positions carry real weight in practice. In the automated claims settlement assessment summarised in §6.3 and in the published worked examples, D5 was scored at 4,0: the institution's settlement engine could produce general decision-class explanations but could not produce per-decision reasons sufficient for the PPR written-reasons obligation, a state between the named anchors at 3 and 5. Where evidence is incomplete, the score is set to the higher of the two plausible values and the evidence gap is recorded as a finding. The full specification for all seven dimensions, including the evidence inventories and the anchor wording for the remaining dimensions, is set out in Kitantou (2026).