PUBLISHED · AUGUST 2026

When the Model *Said It Was Fine*

Model Risk Management in financial services: regulatory foundations, AI-era challenges, and cross-sector implications.

Lukas Ziewer · Neil Catton

Journal article · Published 7 August 2026 · Revised 17 August 2026

NEILCATTON.COM

Where validation stops and accountability begins. SS1/23 · Solvency II · SR 11-7 · SR 26-2 · IMAP · ORSA · SMCR

ABSTRACT

Abstract

Model Risk Management frameworks were designed for a deployment environment in which human review sat between model output and consequential decision. Artificial intelligence is changing the scale, speed and opacity of model use in ways those frameworks were not built to handle. Financial services developed the most mature model risk governance discipline of any industry sector — built on decades of regulatory pressure arising from the leveraged nature of the business, the centrality of models to core decisions, and the systemic consequences of failure. This article examines how that discipline is being tested by AI deployment conditions, where the frameworks fall short across UK and European banking and insurance practice, and what the cross-sector diffusion of AI decision-making implies for sectors now encountering equivalent risks without equivalent governance. The comparison with the United States — where SR 26-2, the April 2026 revision of SR 11-7, explicitly excludes generative and agentic AI from scope — functions as the control case: the jurisdiction that has declared the gap openly, where European supervisors have answered it by restating the frameworks they already have. The implications land, across European jurisdictions, on the named individuals who carry personal accountability for model risk in their institutions. The central claim is that AI does not merely strain these frameworks; in practice and in regulatory intent, they do not reach it, and no jurisdiction examined here has yet redesigned for that fact.

> *Exclusion from scope is not exclusion from consequence. The models are deployed. The decisions are being made.*

ONE

The validation gap

A system produces a confident output. The output is wrong. No one in the decision chain has reviewed whether the assumptions beneath the output still hold, or whether the decisions and actions the output suggests are appropriate ones. Both halves of that sentence matter. A model is not correct because its mathematics are correct; it is correct when the decisions taken on the strength of it are sound. A model can be mathematically valid and still be the wrong model for the decision at hand. Using a valid model in circumstances it was never approved for is as much a modelling failure, and as much a matter for validation, as a wrong formula. That is the failure mode this article examines — and it is arriving in financial services, and in every other sector deploying AI to make consequential decisions, faster than the governance around it is adapting.

Financial services has the most developed answer to that failure mode of any industry sector. The answer is called Model Risk Management (MRM), and it was built over three decades of regulatory pressure, crisis response and accumulated governance failure. Its institutions, frameworks and vocabulary are mature. They are also, under current AI deployment conditions, being outpaced.

That is the argument within financial services. This article also makes a cross-sector argument. AI is not deploying only in banks and insurance companies. It is deploying in hospitals, transport systems, benefit assessments, and criminal justice — in every domain where models now inform or take consequential decisions. The governance that follows that deployment, outside financial services, ranges from nascent to absent. The failure modes it produces are, in many cases, familiar to anyone who has spent time in MRM.

The sector that has thought hardest about what happens when models fail is therefore in a position to say something useful to sectors that are now encountering the same problems for the first time. Sections 2 to 3 establish the financial services position: what MRM was built to do and why financial services developed the discipline, and how AI changes the conditions it operates under. Section 4 examines the cross-sector picture: where AI governance in other sectors currently stands, what failure modes it shares with the FS experience, and what each side might learn from the other. Section 5 addresses the accountability question, and the specific individuals who carry it across European jurisdictions, before turning to what the model risk management profession needs to examine next. The argument that runs through all of it is that the frameworks financial services built have not failed on their own terms — they have been left behind by what they now govern, and no jurisdiction has yet redesigned for that fact.

TWO

What Model Risk Management was developed to do

The regulatory genealogy in financial services

MRM did not emerge from abstract governance thinking. It emerged from the discovery, repeatedly and at cost, that models used to make consequential financial decisions could be wrong in ways that were not visible until they were catastrophic.

The earliest formal model requirements in banking arose from the use of internal models to calculate capital for market risk. The 1996 Market Risk Amendment to Basel I permitted banks to use internal Value-at-Risk models to determine their market risk capital requirement, subject to supervisory approval and quantitative standards set by the Basel Committee on Banking Supervision. The requirements went beyond the mathematics. Supervisors expected demonstrable board and senior management oversight of the model programme — not technical competence, but informed accountability for whether the institution had the capabilities required to use internal models responsibly. They required documented backtesting of VaR forecasts against actual profit and loss outcomes, creating an empirical feedback loop that had not previously existed in regulatory capital frameworks. And they introduced stress testing as a formal discipline: systematic examination of how the model performed under conditions it was not calibrated to expect. These three requirements — board accountability, backtesting, and stress testing — became, and remain, the structural foundations of every MRM framework built since.

1996 · MARKET RISK AMENDMENT

Board accountability

Not technical competence — informed accountability for whether the institution could use internal models responsibly.

1996 · MARKET RISK AMENDMENT

Backtesting

Documented testing of VaR forecasts against realised P&L: the first empirical feedback loop in regulatory capital.

1996 · MARKET RISK AMENDMENT

Stress testing

Systematic examination of how the model performs under conditions it was not calibrated to expect.

Fig. 1 — The three foundations of every MRM framework since 1996

Basel II (2004) extended internal model use substantially. The Internal Ratings-Based approach for credit risk and the Advanced Measurement Approach for operational risk embedded models still more deeply into capital adequacy, with validation requirements carrying direct regulatory consequence. For insurance, the development that followed was not incidental. Solvency II — the European regulatory framework for insurance and reinsurance — is the direct extension of Basel II thinking into internal capital modelling. Articles 120 to 126 of the Directive impose ongoing obligations for insurers using internal models to calculate their Solvency Capital Requirement (the buffer an insurer must hold to be confident it can pay future claims): regular validation, performance monitoring, documentation standards, and governance oversight. The requirement that firms demonstrate the model is genuinely used in business decision-making — not merely maintained for regulatory calculation — runs through Solvency II's Internal Model Approval Process in the same spirit as the IRB use test in banking. Both frameworks rest on the same conviction: governance exists only if the model is genuinely embedded in how the institution thinks, not merely in how it reports.

The Global Financial Crisis of 2007–09 was the decisive event. The models that had priced mortgage-backed securities, assessed counterparty credit exposures and estimated tail risks had, in aggregate, produced a structurally unstable financial system while appearing to justify it. The political response was global. At the G20 London Summit in April 2009, the Financial Stability Forum — the existing international body for regulatory cooperation — was upgraded to the Financial Stability Board, with an expanded mandate and membership. The FSB's role was coordination: setting the direction, identifying the gaps, and pressing national regulators to close them — including through the development of general MRM standards — as part of the broader Basel reform agenda.

The substantive early work on MRM specifically was concentrated in the United States, and for reasons that were not accidental. The causes of the Global Financial Crisis were largely located in US financial markets — the origination and securitisation of subprime mortgages, the ratings models that had mispriced the resulting securities, the internal risk models that had permitted institutions to hold inadequate capital against their exposures. The US political environment in the aftermath was supportive of stricter regulation in ways that proved durable, producing Dodd-Frank in 2010 and, in April 2011, Supervisory Letter SR 11-7 — the Federal Reserve and Office of the Comptroller of the Currency's joint guidance on MRM. SR 11-7 was not the only regulatory development of the period, but it was the most comprehensive and explicit statement of what a serious MRM framework requires, and it set the direction that other jurisdictions have followed. It was revised in April 2026 as SR 26-2; that revision, and what it now leaves out, is examined in Section 3.

European supervisors have followed, and in some respects have gone further. The European Central Bank, acting through the Single Supervisory Mechanism, is explicit about the scope of MRM expectations. Chapter 4 of the ECB Guide to Internal Models states that a comprehensive MRM framework is expected to apply to all models across the institution, regardless of whether they are used for regulatory capital calculation. Models supporting business decision-making, financial reporting, asset pricing, or operational automation are all within scope and must be logged and governed within the bank's MRM inventory. This is a significant extension beyond the Basel capital framework's original intention: it applies the Basel disciplines to the full model population, not merely the regulated subset.

The route by which it arrives there is worth noticing. The ECB is the prudential supervisor for banks across the euro area, and the broadest statement of its MRM expectations sits inside a guide written for internal capital models — and, within that guide, the extension to all models in use is made in a footnote. The supervisor has taken the back door. Whether by design or by accident, it is the internal model function that now carries the institution's general model risk expectations, and increasingly its AI expectations with them — a connection this article returns to in Section 3.

A third wave of requirements has emerged not from financial stability concerns but from conduct-of-business ones. The first two waves — Basel internal capital models, then the Global Financial Crisis response — were primarily about institutional soundness and systemic risk. The third is about customer outcomes and the fairness of automated decisions. In Test-Achats (C-236/09, 2011), the Court of Justice of the European Union struck down the use of gender as an actuarial factor in insurance pricing, requiring insurers to abandon a pricing variable that had been standard practice across the industry. In SCHUFA (C-634/21, 2023), the same court held that producing an automated credit score determinative of a lending decision constitutes profiling under Article 22 of the General Data Protection Regulation (GDPR), requiring meaningful human review before the decision is acted upon. Both rulings were made during the UK's membership of the European Union and remain binding precedent in UK law. The EU AI Act, which entered into force in 2024, extends this logic further, imposing risk-based requirements — risk assessment, human oversight, documentation, transparency — on AI systems in high-risk use cases including credit scoring and insurance pricing. The vocabulary is different from MRM — human oversight is effective challenge, documentation is the inventory, risk assessment is validation — but the requirements are the same.

WAVE	TRIGGER	WHAT IT ADDED
One · 1996–2004	Basel I / Basel II internal capital models	Board accountability, backtesting, use testing, stress testing, validation tied to regulatory consequence.
Two · 2007–11	Global Financial Crisis	SR 11-7 (2011): the first comprehensive MRM framework, covering the whole model universe of an organisation, and the FSB's coordinated reform agenda.
Three · 2011–24	Conduct-of-business rulings and the AI Act	Test-Achats, SCHUFA and the EU AI Act extend model governance to fairness and automated decisions — different vocabulary, same concepts.

Fig. 2 — Three waves of MRM requirement, 1996–2024

Why financial services leads

This regulatory accumulation did not happen in financial services by accident. Three features of the industry created both the exposure and the pressure to develop governance responses earlier and more thoroughly than any other sector.

The first is leverage. Financial institutions borrow heavily against relatively thin capital bases to generate returns. The consequence is that model error does not merely produce a bad outcome — it can produce an existential one. A model that incorrectly estimates credit losses by a few percentage points can, in a sufficiently leveraged institution, transform a solvent balance sheet into an insolvent one before management has time to respond. The relationship between model accuracy and institutional survival is direct in a way it is not in most other sectors.

The second is model centrality. Banks and insurance companies do not use models as decision-support tools alongside other sources of judgement. Models are the core of the business. Credit decisions are model outputs. Insurance pricing is a model output. Risk capital is a model output. Trading positions are assessed, hedged and reported through model outputs. The institution's stated financial position at any point in time is, to a significant degree, a function of what its models say. That is a different relationship to models than a hospital that uses an AI diagnostic tool or a government department that uses a predictive algorithm — and it creates a correspondingly different relationship to model failure.

Of the many risk categories where model flaws create problems, liquidity risk has grown most critical. A model that fails to capture the speed at which funding can be withdrawn, or the correlation of liquidity stresses across institutions, can leave management believing they have adequate buffers when they do not. Liquidity crises materialise in hours, not weeks, which means the window between model failure and institutional consequence is shorter than in almost any other risk category. This

was at the heart of the Global Financial Crisis — not just credit model failures, but the liquidity model failures that prevented institutions from seeing how rapidly their position was deteriorating.

The third is systemic and societal risk. The 2008 crisis demonstrated that model failure in financial services does not stay within the institution. It propagates — through interconnections between institutions, through credit markets, into the real economy and onto households that had no exposure to financial modelling whatsoever. That propagation created a regulatory mandate with unusual breadth. For the largest institutions — those classified as systemically important — an additional dimension applies: the implicit expectation, confirmed in crisis after crisis, that the state will not permit them to fail. The implied guarantee that attaches to a systemically important financial institution means that the risk of model failure is ultimately socialised. Regulators govern model risk in these institutions not only on behalf of depositors and policyholders, but on behalf of the taxpayers who stand behind them.

Together, these three features — leverage, model centrality, and systemic risk with its implicit state guarantee — created a regulatory mandate with no equivalent in any other industry. The mandate is what produced the MRM discipline as we have it.

EXPOSURE

Leverage

Model error does not merely produce a bad outcome — in a leveraged institution it can produce an existential one.

EXPOSURE

Model centrality

Credit decisions, insurance pricing and risk capital are model outputs, not model-assisted judgements.

EXPOSURE

Systemic risk

Failure propagates through interconnected institutions and carries an implicit state guarantee for the largest firms.

Fig. 3 — Why financial services developed MRM before any other sector

The architecture that emerged

The foundational text of modern MRM as a governance discipline is SR 11-7, issued by the Federal Reserve and the OCC in April 2011. It matters to a European audience for a specific reason: it is the first and most comprehensive regulatory statement of what a serious MRM framework requires, and it set the direction that the PRA's SS1/23 and the ECB's supervisory expectations have followed. Understanding SR 11-7 is not about following US rules — it is about understanding the conceptual foundation on which the European frameworks rest.

SR 11-7 defines a model as "a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to transform input data into quantitative estimates." Everything else in the discipline — in every jurisdiction that has developed a framework — hangs off that definition. A model, in this conception, is a simplified, relatively static representation of real-world relationships: built by identifiable people, on identifiable assumptions, producing outputs whose reliability can be assessed by examining the construction.

Three distinct but interrelated structures follow from that conception. The first is the model inventory: a comprehensive record of every model in use — methodology, data sources, risk rating, validation status, known limitations. The inventory is foundational to the whole programme, because it is the institution's answer to a question that is simple to pose and hard to answer: what is deployed, to what purpose, on what assumptions? Nothing downstream works without it.

The second is validation. Validation is not the redoing of the work by a different pair of hands, though it can include that. It is the search for tests that reveal where the model's use is not valid — profit and loss attribution, backtesting against realised outcomes, stress testing under conditions the model was not calibrated to expect. For machine learning models specifically, that toolkit extends to explainability techniques such as SHAP, which attribute a model's output to its contributing features, and discrimination metrics such as AUC, which measure how well a model separates the outcomes it is meant to distinguish. That work does not have to be performed by someone independent, and the development team should be doing it first, as a matter of course. Validation asks where this model stops working.

The third is effective challenge: review by parties with the standing, competence and incentive to find what is wrong, separate from the development function. Independence belongs here, to the challenge — not to the definition of validation itself. The two are routinely combined in practice, which is sensible, and routinely conflated in framework discussions, which is not. An institution that treats "validation" and "independence" as the same word has already lost the distinction that makes each of them useful.

STRUCTURE ONE

The inventory

A comprehensive record of every model in use. Foundational: nothing downstream works without it.

STRUCTURE TWO

Validation

The search for tests that reveal where a model's use is not valid — not the redoing of the work by someone else.

STRUCTURE THREE

Effective challenge

Review by parties with the standing, competence and incentive to find what is wrong. Independence belongs here.

Fig. 4 — The three structures of an MRM programme

Within FS practice, model users themselves bear primary responsibility for the models they use. Independent challenge is a second line function — it does not substitute for the responsibility of the business to understand and own its models. The Basel II/Solvency II use test crystallises this: an insurer must demonstrate that the internal model is genuinely embedded in business decisions, not merely maintained for regulatory calculation. Governance without genuine use is not governance.

The United Kingdom arrived at the same architecture through SS1/23, issued by the PRA in 2023. Its formal scope is banks, building societies and designated investment firms with internal model permissions — insurers are not within it. But the PRA was explicit in the accompanying policy statement that all firms, regardless of size or sector, are expected to manage the risks associated with models, and that firms outside the formal scope should review the principles in their entirety. SS1/23 sits in the banking supervisory materials; it functions as the PRA's statement of what managed model risk looks like across its regulated population. And it has one feature that matters for everything that follows: under Principle 2, responsibility for the MRM framework is allocated to a named Senior Management Function (SMF) holder, providing initial and ongoing annual attestations of compliance. Model risk in a UK bank is not owned by a committee. It is owned by a person, by name, on the record. That allocation sits inside the Senior Managers and Certification Regime (SMCR), the accountability architecture the UK built after the financial crisis; European jurisdictions have developed analogous structures. All of them express the same post-crisis principle: named individuals, not institutional processes, carry accountability for the risks that can bring institutions down. The SMCR is the mechanism by which everything in Section 5 of this article lands on a desk.

The EU insurance framework sets this architecture in its most demanding form. Under Solvency II, an insurer using an internal model to calculate its Solvency Capital Requirement must obtain regulatory approval through the Internal Model Approval Process. The Own Risk and Solvency Assessment (ORSA) connects model output directly to the institution's capital decisions — the model is not advisory, it determines the capital position.

UK · SS1/23

Named individual

Principle 2 allocates MRM responsibility to a named Senior Management Function holder, with initial and annual attestations, inside the SMCR.

EU · SOLVENCY II

Named desks

The CRO owns ongoing internal-model validation; the board approves the ORSA. The ORSA connects model output directly to the capital position.

Fig. 5 — Where personal accountability lands, UK vs EU

It is worth being precise about what board engagement in this framework actually means. Getting a board genuinely engaged with complex models is hard, and the expectation is not that board members understand the mathematics. The expectation is three things: that the board can satisfy itself the organisation has the capabilities required to build and govern models responsibly; that the governance structure is robust enough to prevent the biases that can be embedded in model assumptions from passing unexamined through the institution; and that the board understands and owns the business assumptions built into the models — the commercial and strategic judgements that sit beneath the mathematics. The third is the most important and the most frequently underexamined.

Where validation is performed independently, it carries a tension at its centre that is rarely named. Validators need independence from the development team — but they also need to have genuinely engaged with the model, developed their own analytical approaches, and tested changes to modules and assumptions in practice. A validator who has read the documentation but not worked the model is not performing the function the framework requires. A validator who has worked the model closely risks compromising the independence that makes the challenge effective. Both conditions are required simultaneously, and they pull in opposite directions. Institutions that resolve this tension by treating validation as a formal review of completed documentation have not implemented effective challenge. They have implemented the appearance of it.

The question of what validation should actually produce is also underspecified in most framework discussions. Finding errors and faults is part of the function — but they should be the exception, not the purpose. A development process that is producing models with systematic errors is a governance problem that precedes validation. Identifying improvements is legitimate, but a validator who recommends specific changes risks becoming a shadow developer — compromising independence and creating ambiguity about who is responsible for what the model does. The most important output of a validation exercise is a defined and relatively narrow scope of application: the specific conditions, data environments and use cases for which the model is approved. Validation does not certify that a model works. It certifies that a model works in these circumstances, for these purposes, on these assumptions.

Both banking and insurance practice rest on one assumption so basic it is rarely stated: a human being who understands the model's construction stands between the model's output and the consequential decision. Validation assumes a reviewable artefact. Effective challenge assumes a challenger who can interrogate the construction. The inventory assumes the institution knows what it is running. Every mechanism in both is, at bottom, a way of organising informed human review.

THREE

How AI changes the conditions

This section is not an argument that AI is risky in general terms. General unease does not meet the standard of a governance argument. It is a specific examination of five ways AI deployment alters the conditions under which model risk frameworks operate, followed by the gaps those changes open in the frameworks themselves.

CHANGE	WHAT IT BREAKS
Speed & vibe coding	Model-building outpaces the population who can be questioned about construction.
Self-modification	The validated model is not, in any meaningful sense, the model now running.
Direct-to-execution	Closes the human gap every governance mechanism is built around.
Documentation decay	A model can become unreliable with no explicit change to code or parameters.
Classification	“The AI” is not a discrete object — an inventory of parts is not an inventory of the machine.

Fig. 6 — Five ways AI breaks MRM's operating assumptions

Speed and the vibe coding problem

AI development tools have compressed model development timelines from months to weeks, and in some cases to days. Validation cycles have not compressed with them: industry practice remains in the range of six to twenty-four weeks. Meanwhile the number of models in use at financial institutions has been growing at approximately 25 per cent per year. The arithmetic is unforgiving. An inventory and validation regime designed for a stable population of slowly changing models cannot be maintained at the pace of current deployment. The queue grows faster than the review capacity, and institutions face a choice the frameworks never anticipated: slow deployment to the speed of governance, or let governance fall behind the speed of deployment.

6–24 weeks
Industry validation cycle — unchanged

~25%/yr
Growth in models deployed at financial institutions

Fig. 7 — The arithmetic behind the speed problem

The speed problem has a second dimension that compounds the first. The practice now described as "vibe coding" — generating working code through large language model prompts without requiring the developer to understand the code produced — has extended model-building capacity well beyond the population of people who could previously build models. The governance condition the MRM frameworks assume is not merely that models are built carefully; it is that models are built by people who understand them, and can therefore be questioned about their construction by a validator who understands them. Vibe coding breaks that condition before a model reaches the validation queue. There may be no one in the institution who could perform the review the framework assumes.

Self-modification

The traditional validation paradigm assumes a model is a fixed artefact: built at one point in time, validated, and then stable until an explicit change triggers revalidation. Some AI systems violate this assumption in a more fundamental way than mere output variability. They are capable of modifying their own parameters — updating their behaviour in response to new data without any explicit code change initiated by a human developer. The model that was validated is not, in any meaningful sense, the model that is now running.

This property — self-modification in response to environmental inputs or changes — is a design feature, not a malfunction. A model that learns continuously from new data and updates its behaviour accordingly is doing what it was built to do. But it breaks the foundational assumption of point-in-time validation. The validation report attests to a version of the model that may no longer exist by the time the report is filed. For an insurer, the consequence runs directly into the internal model: Solvency II validation

standards were written for systems whose behaviour is stable between validation cycles. A self-modifying system cannot satisfy that assumption by design.

Direct-to-execution

Every MRM framework ever written rests on a single structural assumption: a model produces an output and a human decides what to do with it. The model informs; the human acts. Independent validation, effective challenge, the attestation regime — every governance mechanism in the discipline is built around the existence of a gap between output and consequence, occupied by a human who exercises judgement.

AI systems with agentic capability close that gap entirely. A system that identifies a trading opportunity and executes it, assesses a customer and opens an account, generates a compliance submission and files it — in each case the decision is made and the consequence follows without a human review point. There is nowhere in the process for the effective challenge the frameworks require, because the process no longer contains a human competent to make that challenge, or any human at all.

SR 26-2, in excluding agentic AI from scope, has at least acknowledged this problem explicitly. SS1/23's technology-agnostic framing covers agentic AI in principle — but "covered in principle" and "designed to govern" are different things. The governance mechanisms in SS1/23 all presuppose a human standing somewhere between model output and decision. For agentic systems, that presupposition is false.

Documentation decay and the limits of validation

Under the traditional paradigm, a model validated as sound at deployment stays sound until something identifiable changes — a recalibration, a code change, a data source switch. The validation is point-in-time because the model is point-in-time. AI systems decay differently. A model can become unreliable without any explicit change to its code or parameters, as the world it was trained on drifts away from the world it is scoring. Industry analysis of current practice puts it plainly: AI model documentation may be technically accurate at the moment it is written and meaningfully outdated three months later. A validation regime whose unit of assurance is the point-in-time review cannot, by construction, address decay between reviews. This is the temporal gap, and it is the reason the validation report becomes a statement about the past presented as assurance about the present.

The problem is compounded for large language models specifically. LLMs are non-deterministic: the same input can produce a different output on a different run, so there is no single canonical output for a validator to test against. They are also opaque, and drift between validation cycles in ways that are difficult to detect without continuous monitoring. Explainability techniques such as SHAP, built for models with a stable, attributable feature set, strain against systems that are compound, self-modifying or generative, where there may be no fixed set of features left to attribute an output to. The PRA's roundtable on AI and model risk in October 2025 was unequivocal on this point: some behaviours only emerge in production. As Section 3.3 established for agentic systems, that gap can close entirely even short of full autonomous execution: by the time you find out, the model has already made the decision. At scale. Probably thousands of times. That is a categorically different risk profile to a credit scoring model that produces a wrong output on a single application.

For an insurer, this is not a documentation problem. The ORSA carries model output into the capital decision, which means assumption decay between validation cycles becomes a capital adequacy problem — one that no mechanism in the framework requires anybody to notice, because the validation report is not yet due.

The classification problem

The most basic requirement of model governance is knowing what is a model. Here the discipline is failing at the threshold. The single most common SR 11-7 examination finding for machine learning models in 2024–25 was inadequate model inventory — and the dominant reason was classification. Large language models deployed in customer service, document processing and compliance functions were categorised as "tools" rather than "models" and excluded from the inventory entirely.

The model-versus-tool debate is a symptom of rules-based regulation: where a framework defines its scope through precise definitions, institutions facing ambiguity resolve it in the direction that creates less work. In a principles-based system — which UK and European frameworks are — the obligation is different. The question is not whether a system technically satisfies a definitional threshold, but whether it is doing what a model does: transforming inputs into estimates that inform consequential decisions. A system that does that belongs in the inventory, whatever it is called. There is no published evidence that UK and European institutions are classifying more rigorously than their US counterparts. Breadth of principle on paper does not automatically produce completeness of inventory in practice. That gap is worth sitting with rather than resolving too quickly. A principles-based system puts the interpretive burden on the firm rather than on a precise definition — and that interpretive burden depends on exactly the expertise Section 4.4 goes on to describe as depleting under the same deployment pressure that

is expanding the model estate. Breadth of principle is not self-executing; it requires people able to exercise the judgement it presupposes.

The classification problem has a harder form still, and it is the one that breaks the inventory concept rather than merely straining it. The inventory assumes discrete, identifiable models — things you can point at and enter on a register. AI capability is increasingly deployed as compound systems: a generative layer interfacing with an underlying credit-scoring or pricing model, an orchestration layer routing between several models, retrieval systems feeding context into a language model. Where a generative layer sits on top of a traditional model, the underlying model can be validated — but the compound system as actually deployed is not governed. The component is validated; the system is not. The difficulty is no longer deciding whether a system counts as a model. It is that "the AI" is not a discrete object at all: it is a compound in the model landscape, and an inventory of parts is not an inventory of the machine.

Three postures

The five changes above leave every jurisdiction facing the same question: what does an existing framework say about systems it was not written for? The three major regulatory postures differ in form while converging in outcome.

JURISDICTION	POSTURE	LEFT UNADDRESSED
United States · SR 26-2	Explicit exclusion	Generative and agentic AI excluded from scope; RFI promised, no timeline.
United Kingdom · SS1/23	Technology-agnostic incompleteness	Formal coverage in principle; how it applies to compound, self-modifying systems is being discovered firm by firm.
Europe · ECB & EIOPA	Reabsorption into existing requirements	The qualitative change is named; the remedy is the governance machinery that already exists.

Fig. 8 — Three regulatory postures, one outcome

SS1/23 is technology-agnostic and outcomes-focused, which gives it formal coverage of AI systems. But it was written before widespread generative AI deployment, and formal coverage is not the same as designed coverage. The October 2025 roundtables, spanning banking and insurance Chief Risk Officers, are the PRA working out in conversation what the guidance does not yet say — and their composition is the point: this is not a banking posture but a UK one, applied across the sectors the PRA supervises. The principles apply; how they apply to non-deterministic, self-modifying, compound, agentic systems is currently being discovered firm by firm.

European banking supervision has answered in a different register. "Technology is neutral, governance is not" is the title of a February 2026 keynote by Pedro Machado of the ECB's Supervisory Board — and the speech concedes, in terms, that AI introduces "qualitative changes that cannot be addressed simply by expanding the existing frameworks." The remedy it then prescribes is exactly that expansion: clear accountability for AI-driven decisions, effective senior management oversight, robust challenge through the three lines of defence, consistent with the existing governance guidelines of the EBA (European Banking Authority). "AI does not dilute responsibility. If anything, it raises the bar." There are no AI-specific model risk rules for European banks beyond the internal capital model framework; the expectations are holistic, and they rest on the governance machinery that already exists. The qualitative change is named. The answer is more of what was already required.

European insurance supervision has now spoken in the same register. In August 2025, EIOPA published its Opinion on AI governance and risk management — interpretative guidance that, by its own terms, "does not set out new requirements." The Opinion maps AI back onto the existing sectoral legislation — Solvency II's governance provisions, the Insurance Distribution Directive, and DORA (the EU's Digital Operational Resilience Act) — and places responsibility for the overall use of AI systems squarely with the administrative, management or supervisory body. It is not blind to the new phenomena: the autonomy of an AI system is an impact-assessment criterion, and undertakings are expected to monitor for model drift and data degradation. But its answer to those phenomena is the existing machinery, applied proportionately. And it does not touch the internal model. Nothing in the Opinion addresses generative or self-modifying systems within internal model approval or ongoing validation — the place where Solvency II connects model output to the capital position, and where the temporal gap bites hardest.

The United States has made the gap explicit. SR 26-2 excludes generative AI and agentic AI from scope, promises a Request for Information without a timeline, and narrows its applicability to institutions with over $30 billion in assets — so at the moment the model population is expanding fastest, the population of supervised institutions is contracting. For UK and European readers, the comparison cuts two ways. The US exclusion is more honest about the gap; the principles-based framing of UK and European

frameworks is more demanding about who owns it. A US institution can point to the scope statement. A European senior manager cannot.

Three postures — explicit exclusion in the US, technology-agnostic incompleteness in the UK, reabsorption into existing requirements in Europe — and one identical outcome. The European posture is the one that most needs naming, because it is the least visible: an assumption of analogy. It holds that rules written for one kind of model apply, by extension, to another; that a framework built for simplified, relatively static representations reaches systems that are compound, probabilistic and self-modifying, provided the same governance is applied with more diligence. That assumption is the subject of this article, and our argument is that it does not hold. In each of the three jurisdictions, the governing guidance acknowledges the problem while declining to redesign for it. The frameworks have not failed. They have been left behind, and the leaving-behind has been documented by the regulators themselves.

> *The European posture is the one that most needs naming, because it is the least visible: an assumption of analogy.*

FOUR

Cross-sector implications

The preceding sections have examined how model risk governance is being tested in financial services. That examination is valuable in its own right. But its wider significance depends on a further question: is the financial services experience unique, or is it the leading edge of something that is now spreading to every domain in which AI makes consequential decisions?

The answer matters for the readership of this article specifically. If the failure modes appearing in financial services are recognisable across sectors — if the inventory problem, the temporal gap and the classification problem are arriving in healthcare, transport and public administration in forms a model risk practitioner would immediately identify — then the discipline has something to offer those sectors. And if those sectors are encountering failure modes that FS frameworks have not yet confronted, the exchange runs in both directions.

Where models are making decisions outside financial services

The deployment of AI decision systems beyond financial services has been rapid and, in governance terms, largely ungoverned. Four sectors illustrate the range and the pattern.

SECTOR

Healthcare

AI diagnostic systems are in clinical use for radiology, pathology and ophthalmology. Regulatory posture is nascent, and in the UK the regulator — the MHRA (Medicines and Healthcare products Regulatory Agency) — is building its approach in public. Its AI Airlock, a regulatory sandbox for AI as a medical device launched in 2024, exists precisely because the existing device framework was not designed for probabilistic, self-updating systems; a new AI-specific framework has been promised but not yet delivered. In the EU, clinical AI sits under the Medical Device Regulation and now also the AI Act, two regimes whose interaction is still being worked out. The FDA faces the same strain in the United States. The classification problem is direct — whether a clinical AI is a software-as-a-medical-device, a clinical decision support tool, or a data-processing system determines the governance that applies, and there is no settled standard for making that classification. This is the same classification problem producing inventory failures in banks. The failure mode is identical; the vocabulary to address it does not yet exist in the clinical domain.

SECTOR

Criminal justice

Risk assessment algorithms inform custody, sentencing and parole decisions on both sides of the Atlantic. In the UK, Durham Constabulary's Harm Assessment Risk Tool — a random forest model built on around 104,000 custody events, using 34 predictors including residential postcode data — informed custody decisions from 2016. Its independent validation told a story any model risk practitioner would recognise: overall accuracy fell from 68.5 per cent at construction to 62.8 per cent on out-of-sample data, and for the cases that mattered most — those with actual high-risk outcomes — from 72.6 per cent to 52.7 per cent. The model performed worst precisely where its consequences were largest, and the degradation was invisible from the instrument itself. ProPublica's 2016 analysis of the US COMPAS algorithm, a recidivism-scoring tool used in sentencing and parole decisions, found the same shape of failure: recidivism scores unreliable across demographic groups in ways not visible from the instrument's documentation. This is the validation gap in a different domain — a model producing confident outputs, consequential decisions following, and the assumptions beneath the outputs unexamined at the point of use. The structural failure is exactly what the model risk frameworks were written to prevent. The observation here is about governance failure, not about the broader question of whether risk-scoring tools belong in criminal justice — that is a separate and contested question.

SECTOR

Energy and utilities

The Iberian blackout of 28 April 2025 demonstrated what assumption failure looks like in critical infrastructure. The expert panel convened by ENTSO-E, the European network of transmission system operators, concluded in its final report that the collapse of the Spanish and Portuguese grids was not a one-off failure with a single cause. It was the product of many interacting factors — oscillations, gaps in voltage and reactive power control — and, most tellingly, of recurrent mismatches between the reactive power the system expected users to provide and what those users actually delivered in real time. The failure was not simply one of engineering. It was one of assumptions: the grid was being operated on a model of component behaviour that no longer matched the components. The propagation logic — assumptions inadequately tested against the conditions that materialised, failure arriving faster than response could be organised, consequences borne by households with no involvement in the modelling — will be familiar to anyone who has thought about systemic risk in financial services. The validation discipline is not there. There are engineering safety frameworks and sector regulators, but nothing equivalent to independent model validation with a structured challenge function and a documented inventory.

SECTOR

Public sector

The UK government's use of algorithmic systems in benefit assessment and immigration decisions has produced well-documented failures. The Universal Credit calculation models run by the Department for Work and Pensions (DWP) and the Home Office's visa decision algorithms have each, at various points, produced outcomes that regulators or courts subsequently found to be wrong or discriminatory in ways the decision-making process had not surfaced. The governance picture shares features with the FS inventory problem: models deployed at scale, informing consequential decisions, without systematic independent validation of whether the assumptions underlying those decisions remain valid.

Familiar failures, or new beasts?

The cross-sector survey above suggests a clear answer. The failure modes appearing in other sectors are largely recognisable to financial services practitioners. The inventory problem appears in every sector surveyed. The temporal gap — validation that is point-in-time for systems that change continuously — is structural across healthcare diagnostics, criminal justice risk tools and public sector algorithms. The classification problem is producing the same outcome in clinical AI as in FS: systems that inform consequential decisions sitting outside the governance perimeter.

What differs is not the failure mode but the pace of consequence. In a leveraged FS institution, model failure can become a capital event within hours. In healthcare, a misclassified imaging result harms one patient. In criminal justice, a flawed risk score contributes to one unjust decision. The individual harm may be more direct in the non-FS case; the aggregate and systemic harm is more rapid in FS. Both matter. The governing question — is there a human who understood the model's assumptions, reviewed its outputs, and stood between the output and the consequence? — is the same question in every domain.

PACE OF CONSEQUENCE

Financial services

A leveraged institution can turn model failure into a capital event within hours.

PACE OF CONSEQUENCE

Other sectors

Harm lands on one patient, one unjust decision — more direct, less systemically rapid.

Fig. 9 — Same failure mode, different speed of consequence

There is, however, one failure mode in the non-FS picture that FS frameworks have not yet fully confronted: the erosion of the expertise that was the precondition for governance. In sectors where AI is replacing professional judgement — clinical diagnosis, legal risk assessment, benefit determination — the practitioners whose expertise would have constituted the effective challenge function are being removed from the decision process before governance frameworks are built to replace them. FS MRM practice assumes validators exist who understand model construction. In a sector where models are built without anyone fully understanding them, and deployed into professional domains where practitioners once exercised equivalent judgement, the governance precondition is gone.

What financial services can learn from outside

The exchange is not one-directional. Sectors that have confronted algorithmic governance problems earlier, or in more publicly contested ways, have produced evidence and arguments that MRM practitioners would benefit from examining.

The criminal justice and public sector literature on algorithmic fairness — the regulatory record of bodies such as the UK's Centre for Data Ethics and Innovation, and the analytical work of researchers including Cathy O'Neil — has engaged seriously with questions FS governance practice has tended to treat as separate from MRM. What does it mean for a model to be fair? How should the population affected by a model's decisions be represented in its validation? What are the limits of quantitative model assessment for systems whose social consequences are not capturable in the model's own metrics?

These questions are not foreign to financial services — discriminatory pricing and credit allocation are active regulatory concerns — but governance practice has addressed them through conduct rules rather than through the MRM framework itself. The cross-sector literature suggests they belong inside the governance conversation, not adjacent to it. The EU AI Act, by imposing requirements around bias testing and fundamental rights impact assessment for high-risk AI systems, is already pushing in this direction. An MRM framework that does not engage with these questions will find itself addressed by a parallel regulatory stream that will.

The convergence of model risk and people and culture

There is a risk category dimension to the cross-sector picture that the standard MRM framework does not address, and that the AI deployment environment is making impossible to ignore. Model risk, as the discipline has developed it, is a technical and governance risk. The question it asks is whether the model is doing what it is supposed to do and whether the governance around it is sufficient to know. It does not ask what happens to the people whose work the model is replacing, or to the organisational culture in which the model operates, or to the expertise that was the precondition for governance.

But model risk has a use dimension that is equally important. MRM is not only concerned with whether a model is technically correct — it is concerned with whether the model is being used correctly for any particular purpose. A technically valid model used in the wrong circumstances is a model risk event. The model is right; the application is wrong. Two and two make four, but two oranges and two bananas do not make four sunny days in December. The governance challenge is not just to validate the model, but to define and enforce the conditions under which it may be used — and to maintain the organisational capability to recognise when those conditions are not met.

AI deployment is making both of these challenges harder simultaneously. When a model takes decisions that were previously made by professionals exercising individual judgement, two things happen together. The model carries the technical risk. And the organisation loses, or is losing, the human expertise that would have caught the model's failures — including failures of application, not just failures of construction — before they became consequential.

Technology reinforces this dynamic in ways that are rarely governed. Analytical systems create organisational silos: teams that operate the model without understanding it, separated from teams that built it and no longer maintain it, with the institutional knowledge that connected both groups long since departed. Staff get de-skilled by the systems they run. The focus shifts from producing the right output to keeping the machine operational. A related pattern shows up repeatedly in practice: the function building and deploying the system is often eager to adopt it and largely unaccountable for what it produces, while the function that would own accuracy, validation and suitability — the people closest to the decision — decline the responsibility, judging it too technically demanding for them to carry. Both positions are individually defensible. Together they leave accountability for the model's actual behaviour with no one, which slows deployment as much as any formal governance gap does. The effective challenge function — which MRM frameworks treat as a governance mechanism — was, in practice, also a cultural feature: the willingness and capacity of experienced people to push back on model outputs when those outputs conflicted with their professional judgement. That capacity does not survive the removal of those people from the process.

An institution that has replaced underwriting judgement with a model, credit assessment with a model, and compliance review with a model — across a sufficiently short period that the people who held those capabilities have left — has not merely changed its governance structure. It has changed its culture in a way that governance structure cannot compensate for. The model inventory does not capture the departure of the person who would have caught the inventory error.

> *The model inventory does not capture the departure of the person who would have caught the inventory error.*

Healthcare is encountering this acutely: as AI diagnostic tools become more capable than the average clinician in specific domains, the incentive to train the clinical expertise that would validate the AI's outputs is eroded. The validator who would have provided effective challenge in the clinical domain is not being trained because the model is, on average, better. When the model

is wrong — as models are, including models that are usually better than the average human — there may be no one left who would have caught it. Financial services is not immune to the same progression.

The systems financial services does not build

There is a further respect in which this is not a financial services problem, and it may be the most consequential of all. Most of the AI that will shape financial services over the next decade will not be built inside banks and insurance companies. It will be built by technology companies, data firms and platform providers — organisations outside the FS regulatory perimeter, without the governance culture that perimeter has produced, and selling into financial institutions as vendors.

That is a genuinely novel situation for an industry that has always owned its own models, governed its own assumptions, and validated its own outputs. Every mechanism described in Section 2 assumes the institution can interrogate the construction of what it runs: the inventory assumes a knowable estate, validation assumes tests that can be devised and run against the thing itself, effective challenge assumes a challenger with access. A model bought as a service, built on a foundation model the vendor does not fully explain, trained on data the institution cannot inspect, updated on a release schedule the institution does not set, satisfies none of those assumptions. The governance frameworks were not designed for it. Neither were the accountability structures.

The consequence for the named individual is stark, and it generalises well beyond financial services. A senior manager can be held personally accountable for the behaviour of a system whose development, training data, update cadence and failure modes are determined by a supplier — one who is not within the regulatory perimeter, and who has no attestation to sign. Contractual assurance is not the same as governance. The hospital deploying a diagnostic tool, the police force deploying a risk instrument, the department deploying an eligibility model are all in the identical position: accountable for what the system does, and dependent on an organisation outside their governance regime for the ability to know it. The party carrying the accountability is not the party determining the behaviour. Across every sector examined in this section, that separation is now the normal case rather than the exception.

> *The party carrying the accountability is not the party determining the behaviour.*

FIVE

The accountability question and what the profession needs to examine

Suppose the diagnostic set out in Section 3.6 — that the frameworks have not failed but have been left behind — is correct. Now consider who signs.

Under SS1/23 Principle 2, a named SMF holder attests — initially and then annually — that the institution's MRM framework is in place and effective. The attestation is personal. It sits within the SMCR, the structure the UK built after the financial crisis precisely so that accountability could not dissolve into committees. In insurance, the structure is differently expressed but lands on the same desks: the CRO (Chief Risk Officer) who owns the internal model's ongoing validation, the board that approves the ORSA, the individuals whose names are attached to the capital adequacy conclusion. Across European jurisdictions, analogous frameworks place named individuals on the record for equivalent risks.

Here is the problem the preceding sections create for that signature. An attestation of effective MRM is, implicitly, an attestation about five things: that the institution knows what models it is running; that those models behave as validated; that the validation is current in substance and not merely in date; that no model is taking consequential decisions without a human review point; and that the expertise to perform independent challenge of the model estate remains present within the institution. Sections 3 and 4 give reasons to doubt all five under current deployment conditions.

1 **The institution knows what models it is running**

2 **Those models behave as validated**

3 **Validation is current in substance, not merely in date**

4 **No model takes consequential decisions without a human review point**

5 **Independent-challenge expertise remains present in the institution**

Fig. 10 — What an MRM attestation implicitly claims

The attestation can therefore be sincere and wrong. The SMF holder signs on the basis of a clean validation report, a populated inventory, a completed cycle — every mechanism functioning as designed — while the conditions the mechanisms were designed for no longer hold. This is not a failure of diligence. It is a gap between what the governance measures and what the risk now is. The signature attests to the machinery; the machinery no longer reaches the risk.

The institutional consequences follow the signature upward. Boards rely on the attestation when they approve model risk appetite. The ORSA carries model output into the capital decision; if assumptions have decayed unreviewed, the capital conclusion inherits the decay. Supervisors rely on the attestation regime as the substitute for examining every model themselves — that is the point of an outcomes-focused approach. Each layer of reliance is reasonable on its own terms. Stacked, they mean that one quietly expired assumption can pass through validation report, attestation, board approval and supervisory review without any person along the chain being required to ask whether the model's construction still justifies its confidence.

The UK's position is genuinely distinctive. The US has parked the question — out of scope, Request for Information to follow. UK senior managers do not have that option. The technology-agnostic framing of SS1/23 means the new systems are already inside the attestation perimeter, governed by principles written before those systems existed. The individuals signing are carrying an accountability the guidance has not yet caught up with. The CRO community's posture on attesting over GenAI-era model estates is, at the time of writing, one of the more consequential unanswered questions in European financial governance.

US POSITION

Parked

Out of scope, Request for Information to follow, no timeline set.

UK POSITION

Already inside the perimeter

Governed by principles written before these systems existed — no equivalent opt-out.

Fig. 11 — Parked vs already inside the perimeter

The governance professionals best placed to address this are, by their own account, uncertain how it will work. That uncertainty is honest and important. It is also the starting point for what the profession needs to examine.

Five things follow from the diagnostic, and we put them forward as a work programme rather than as observations. Each is a live question rather than a settled answer, but each requires an answer, and requires one soon. Each proposition also carries a countervailing risk: over-specifying governance for systems whose value lies partly in their adaptiveness could constrain the capability financial services is trying to responsibly deploy. The work programme that follows should be read with that tension in view, not as a case for maximal caution. The people who must ultimately supply it are the people who sign — the SMF holders, chief risk officers and board members carrying personal accountability for model risk. It is not, however, work that belongs to financial services alone. Section 4 established that the same questions are arriving in healthcare, criminal justice, energy and public administration, in sectors with no equivalent discipline to bring to them. We would encourage practitioners, supervisors and professional bodies across those sectors to take the programme up as their own.

A FIVE-POINT WORK PROGRAMME

1	**Renegotiate the definition of a model.**	The definition governs the perimeter, and it was written for artefacts that no longer describe what is being deployed.
2	**Replace point-in-time validation as the unit of assurance.**	The temporal structure of MRM — validation cycle, periodic review, annual attestation — is a design question, not a tooling question.
3	**Extend the inventory to compound systems.**	Validating components does not constitute validating the system they comprise, and compound systems are where the consequential deployments now sit.
4	**Examine the relationship between the attestation regime and the validation regime — before an event tests it.**	Personal accountability and principles-based model risk are each defensible; together, under AI conditions, they place named individuals on the record for risks the machinery beneath them no longer reaches.
5	**Develop a position on the convergence of model risk with people and culture risk.**	The expertise at the centre of effective challenge is depleted by the same deployment decisions that grow the model estate.

Fig. 12 — The profession's work programme

Each of the five is set out below.

The definition of "model" in governance frameworks requires renegotiation. Systems that are dynamic, probabilistic and capable of autonomous adaptation do not fit a definition written for simplified, relatively static representations — and the classification problem documented in Section 3 shows that institutions are already resolving the ambiguity in the direction of less governance. Whether renegotiation proceeds through revised definitions, parallel structures for AI systems, or new principles is the field's live question. That it must proceed is not.

Point-in-time validation is inadequate for systems whose assumptions can drift without code change and whose parameters can self-modify between review cycles. The temporal structure of MRM — the validation cycle, the periodic review, the annual attestation — requires examination as a design question. This is a governance problem, not a technology problem, and treating it as a tooling matter to be solved by monitoring software would repeat the original error of mistaking mechanism for assurance.

The inventory concept must extend to compound systems. Validating components does not constitute validating the system they comprise. Acknowledging the complexity of hybrid architectures — as SR 26-2 has done, partially — is not a governance answer. European supervisors have begun to acknowledge it too: EIOPA expects undertakings to assess the connections between AI systems, and the ECB has named the dependency structure of generative AI as a supervisory concern. But acknowledgment is not an answer either, and the gap is precisely where the most consequential deployments now sit.

The relationship between the attestation regime and the validation regime requires examination before it is tested by an event. The UK chose personal accountability as its post-crisis governance instrument, and chose principles over rules as its model risk instrument. Each choice is defensible. Together, under AI deployment conditions, they place named individuals on the record for risks the validation machinery beneath them no longer fully reaches. Whether the answer is revised attestation language, explicit scope statements, continuous assurance standards, or supervisory clarity about what the signature does and does not warrant — that conversation belongs to the profession, and it is cheaper to have it now than after the first attestation is examined in hindsight.

The MRM discipline needs to develop a position on the convergence of model risk and people and culture risk. The human expertise at the centre of effective challenge is not an infinite resource. It is depleted by the same deployment decisions that grow the model estate. The profession that has thought hardest about what happens when models fail is well placed to say what it means when the capability to govern model failure is itself eroding.

That last point opens onto something larger. The financial services industry has spent thirty years building expertise in how to govern consequential models. That expertise — in validation methodology, in governance architecture, in the cultural requirements of effective challenge — is genuinely transferable. Other sectors now face the same problems, at speed, without the accumulated discipline. Financial services practitioners have something to offer.

But the transfer does not run only in one direction. Section 4.5 set out the harder truth alongside it: financial services is becoming a receiver of systems developed outside its own perimeter, and the expertise it has accumulated will need to work in that direction too — not just sharing practice with other sectors, but learning to govern systems whose development it did not control. Every sector deploying AI at consequence is in the same position, and none of them has yet built the governance for it.

That is the conversation that matters, and it will not resolve itself.

SOURCES

Further reading

- Bank of England, Prudential Regulation Authority, SS1/23 — Model risk management principles for banks, May 2023. https://www.bankofengland.co.uk/prudential-regulation/publication/2023/may/model-risk-management-principles-for-banks-ss
- Bank of England, Prudential Regulation Authority, AI/ML CRO roundtable notes, October 2025. https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/publication/2025/november/ai-roundtable-oct-2025.pdf
- EIOPA, Opinion on Artificial Intelligence governance and risk management, EIOPA-BoS-25-360, 6 August 2025. https://www.eiopa.europa.eu/publications/opinion-artificial-intelligence-governance-and-risk-management_en
- Machado, P., Technology is neutral, governance is not: AI adoption in the banking sector, keynote speech, ECB Banking Supervision, 24 February 2026. https://www.bankingsupervision.europa.eu/press/speeches/date/2026/html/ssm.sp260224~6c5b64a77a.en.html
- Basel Committee on Banking Supervision, Amendment to the Capital Accord to incorporate market risks, January 1996.
- Solvency II Directive (2009/138/EC), Articles 120–126.
- Federal Reserve Board / OCC, Supervisory Letter SR 11-7, Guidance on Model Risk Management, April 2011. https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm
- Federal Reserve Board / OCC / FDIC, Supervisory Letter SR 26-2, April 2026. https://www.federalreserve.gov/supervisionreg/srletters/SR2602.pdf
- GARP, SR 11-7 in the Age of Agentic AI: Where the Framework Holds — and Where It Strains, February 2026. https://www.garp.org/risk-intelligence/operational/sr-11-7-age-agentic-ai-260227
- ValidMind, AI Is Rewriting the Rules of Model Risk Management, 2024–25. https://validmind.com/blog/model-risk-management/
- Domino Data Lab, What Changes with SR 26-2, 2026. https://domino.ai/blog/what-changes-with-sr-26-2
- PwC, Model Risk Management Survey 2024, April 2025. https://www.pwc.com/cz/en/temata/model-risk-management-survey.html
- Oswald, M., Grace, J., Urwin, S. and Barnes, G.C., Algorithmic risk assessment policing models: lessons from the Durham HART model and 'Experimental' proportionality, Information & Communications Technology Law, 27(2), 2018. https://www.tandfonline.com/doi/full/10.1080/13600834.2018.1458455
- ENTSO-E, Final Report on the Grid Incident in Spain and Portugal on 28 April 2025, March 2026. https://www.entsoe.eu/publications/blackout/28-april-2025-iberian-blackout/
- Angwin, J. et al., Machine Bias, ProPublica, May 2016. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- O'Neil, C., Weapons of Math Destruction, Crown Publishers, 2016.
- Centre for Data Ethics and Innovation, Review of Bias in Algorithmic Decision-Making, December 2020. https://www.gov.uk/government/publications/cdei-publishes-review-into-bias-in-algorithmic-decision-making
- Court of Justice of the European Union, Case C-236/09, Association belge des Consommateurs Test-Achats ASBL v Conseil des ministres, 1 March 2011.
- Court of Justice of the European Union, Case C-634/21, OQ v Land Hessen (SCHUFA), 7 December 2023.
- EU AI Act (Regulation (EU) 2024/1689).

About the authors

Lukas Ziewer — Senior Advisor, Risk Strategists, Regulatory Solutions

Lukas Ziewer is an independent risk strategist with over twenty-five years of leadership in financial risk management and actuarial across EMEA and Bermuda. Having begun his career in actuarial roles at UBS Swiss Life and PwC, he progressed through senior strategy consulting roles into C-suite executive leadership.

He has held Chief Risk Officer and Partner roles at Athora Group, MetLife, KPMG, and Oliver Wyman. At Athora Group, he served as Group CRO on the founding executive team, building the group risk function from scratch and navigating major regional acquisitions. Prior to this, he was Regional CRO for MetLife across Europe, the Middle East, and Africa.

His advisory and governance experience spans multi-jurisdictional board leadership — including Non-Executive Director and Risk Committee Chair roles across Ireland, Germany, Belgium, and Luxembourg — and extensive liaison with key supervisors such as the Central Bank of Ireland, BaFin, and the Bermuda Monetary Authority.

He now operates through Fuseki Risk Insights across a portfolio of independent advisory roles, supporting global technology and BPO firms in optimising risk management efficiency, training regulatory authorities on entity risk oversight, and advising on next-generation credit-risk modelling strategies.

He holds an MSc in Theoretical Physics, an LLB (Hons), and qualified as a Chartered Financial Analyst (CFA), PRM holder, and Actuary (DAV).

Neil Catton — Independent Technology Strategist

Thirty-eight years across more than twenty sectors, from computer operator to Group CTO. CTO and director-level roles at BT, Wipro, Fujitsu, Sopra Steria and Capita. Author of the technology ethics trilogy *The Next Evolution*, *The Cognitive Crucible* and *The Shadow System*, and of the standalone *Working as Designed*. He writes regularly at writing.neilcatton.com on the intersection of technology, human consequence and organisational systems.

The career began at British Nuclear Fuels and has returned to regulated, model-driven sectors repeatedly since — most substantially as one of three architecture directors on BT's £1.5bn transformation programme, with £500m of allocated budget spanning digital, data, integration, cloud and AI across a hundred-strong architecture practice. The banking and insurance work behind this article — including two challenger bank builds taken through the financial services licensing process — is where the question underneath this piece, can you evidence that a model-driven decision was made properly, was first asked in earnest.

The practice is a fractional and advisory portfolio spanning public safety, public sector, insurance, health technology and early-stage ventures: board-level technology strategy, independent challenge, and the unglamorous work of making an organisation able to explain itself.

How this article was produced

This article was co-authored for joint publication rather than commissioned as a client briefing, and the production discipline follows from that. The regulatory positions the argument depends on — the EIOPA Opinion, the ECB's February 2026 keynote, the scope of PRA SS1/23, the Durham HART validation study, and the ENTSO-E final report on the Iberian blackout — were each read in full against the primary publication during review, not taken from summaries. Foundational legal and regulatory sources — SR 11-7, the Solvency II Directive, the CJEU rulings, the EU AI Act — are drawn from the established legal record. A small number of secondary industry analyses (GARP, ValidMind, Domino Data Lab, PwC) are cited for context and industry framing; they are not treated as primary evidence for any claim in the article and are identified as such in Further reading.

NOTICE

1. Not advice. This document is general professional commentary. It is not legal, regulatory, financial or actuarial advice and must not be relied on as such.

2. Position and date. The regulatory positions described are current as at publication, 7 August 2026, and this revision, 17 August 2026. Several changed within the year before publication and may have moved again since.

3. Third-party sources. Figures and quotations are drawn from public sources, cited in good faith and listed in Further reading. Verify against the original before relying on them.

4. Named organisations. Organisations are named only where a regulator, court or the organisation itself has placed the relevant information on public record. They are used illustratively; none has been engaged or approached in connection with this article.

5. Publication status. This is the published version of a jointly authored article, revised 17 August 2026. It represents the authors' own analysis and does not represent the position of any organisation with which either author is affiliated.

6. AI assistance. Claude (Anthropic) was used in the drafting, editing and research support of this article. All analysis, argument and factual claims remain the responsibility of the named authors, who reviewed and take ownership of the final text.

7. Independence. Neither author holds vendor affiliations, reseller agreements, referral arrangements, commission interests or equity positions in any product or supplier referenced, and neither has received payment from any organisation named in the article.

lukas.ziewer@fuseki-risk.com · nc@neilcatton.com · neilcatton.com · writing.neilcatton.com