

ARTICLE 01 · DARTMOUTH ADVISORY PARTNERS

DAP

Models are commodities. Context is the only moat left in financial AI.

Asset managers are racing to deploy agents. Examiners are asking a different question: which rule governed the last decision, and can you prove it?

Author

Brendan Sibeth

Published

August 2026

Owner

Dartmouth Advisory Partners

00 / Summary In plain language

Foundation models are becoming commodities. What stays scarce is the institution's own meaning: the mandate as written, the overlay as negotiated, the exception as granted, and the policy version in force when a decision is made. This piece argues that the durable asset in financial AI is a compiled context layer, not the model, and that the test which matters is whether a decision can be reconstructed to the clause that governed it.

Models are rented. Context, if you compile it, is yours.

Author: Brendan Sibeth, Founder and Managing Partner, Dartmouth Advisory Partners, Toronto
Published: 27 August 2026 • Updated: 27 August 2026
Canonical: [dap.solutions](#) • Working paper version on SSRN

01 / The problem A model problem, misdiagnosed

The industry has a model problem it keeps misdiagnosing as a model problem.

Foundation models are getting cheaper, faster, and more interchangeable. That is not a secret. It is the point of a commodity. What is still scarce, and getting scarcer relative to the volume of agents now touching investment process, is the institution's own meaning: the IPS as written, the overlay as negotiated, the exception as granted, the examiner's last letter, the definition of "high risk" that this committee actually uses.

A large language model deployed against a firm's data estate without that layer is a fluent stranger. It speaks finance. It does not speak this mandate.

That gap is now an examination problem, not a research problem.

02 The wrong race

Most AI programs in asset management are still organized around the model.

Which vendor. Which window size. Which retrieval stack. Which copilot seat count. Those choices matter operationally. They do not determine whether the firm can reconstruct a decision.

When an agent proposes an allocation, drafts a report, or clears an exception, the question that survives contact with a regulator is not "was the model state of the art." It is:

- Which policy version was in force?
- Which clause applied to this portfolio, this client, this jurisdiction?
- What was considered and rejected?
- Who was entitled to override, and did they?
- Can that chain be reproduced independently of the vendor that hosted the model that week?

If the honest answer is a narrative assembled after the fact, the firm does not have governance. It has literature.

Picture the ordinary case. On Monday a client overlay is amended to tighten a concentration limit. On Tuesday an agent clears a trade that passes under the old limit and breaches the new one. The examiner asks which version applied. If the answer lives in an email thread and a vendor's log, the firm is guessing. If it lives in a sealed record that names the clause, the version, and the timestamp, the firm is governed.

Mercer's 2026 survey of 131 asset managers, fielded in February and March, found 55 percent already run AI inside at least one investment process, with another 27 percent piloting. Only about 5 percent give AI autonomous or semi-autonomous authority over recommendations or trades. The rest keep a human on the decision. That is the right instinct, and it is exactly the instinct that needs a record.

In the same season, Grant Thornton surveyed roughly 950 senior US executives across ten industries. Seventy-eight percent lack full confidence they could pass an independent AI governance audit within 90 days. Underneath that number: nearly three in four organizations are piloting, scaling, or running agentic AI, and one in five has a tested response plan for when it fails.

The distance between those numbers is not a technology gap. It is a context gap.

03 What context engineering actually is

Context engineering is not prompt craft. It is not "better RAG."

RAG retrieves similar text. A context layer decides what the agent is allowed to treat as true, binding, and current before retrieval and generation happen.

In institutional finance that layer has to carry, at minimum:

- Canonical meaning: what this firm means by mandate, liquidity, restricted list, material exception
- Relationships: how a clause attaches to a vehicle, a sleeve, a client overlay, a jurisdiction
- Constraints: what the agent may decide versus what it must escalate
- Temporality: which version of the rule was live at the moment of the act
- Provenance: the path from source sentence to executable predicate
- Runtime state: what already happened, what was overridden, what is in cooling-off

Without those, an agent is doing similarity search over a filing cabinet and calling it judgement.

The deterministic parts of the institution (accounting, mandate compliance, guideline monitoring) cannot be "probably right." Probabilistic systems can sit next to those parts. They cannot replace the rulebook unless the rulebook has been compiled into something the system can evaluate at decision time.

That compilation is the work.



04 Policy as a file versus policy as a system

Institutions already write the context. They write it as policy.

Thousands of pages. Investment policy statements. Guidelines. Side letters. Risk appetite. Internal manuals. Then the firm hopes the systems follow it.

Hope is not an audit artifact.

Policy management software stores the document and tracks who attested they read it. That is necessary and insufficient. A document describes what should happen. A compiled context layer is the thing that decides what happens, distinguishes the decisions it may take from the decisions it must escalate, regenerates when the source text changes, and writes every outcome back to the clause that produced it.

Three properties separate a living layer from automation theatre:

Continuous. It watches activity against policy as activity occurs, not in a quarterly sampling exercise.

Adaptive. When a mandate or regulation changes, logic is regenerated the same business day as a versioned candidate, not as a six-month IT ticket.

Provable. Every material decision carries a sealed path to the source clause and the policy version in force. Reconstruction does not depend on whoever still works there.

The operating rule underneath all three is short. Models propose. Humans ratify. Only deterministic evaluation over ratified inputs is attested. A model's draft never appears as final, and an empty field stays visibly empty rather than being filled in.

This is why "governance platform" and "context layer" are not synonyms. Most governance platforms document, score, and alert after the fact. A context layer is present at the moment of action. Presence is the difference between a dashboard and a control.

05 Why this is the moat

Models will keep commoditizing. Vendors will keep swapping. Context windows will keep growing. None of that encodes how this institution decides.

The context layer is proprietary for a boring reason: it is made of the firm's actual rulebook, actual exceptions, actual committee behaviour, and actual decision history. A competitor cannot buy that from a foundation-model lab. A consultant cannot leave it behind as a slide. If it is compiled into open, institution-owned artifacts, it compounds. If it lives in a vendor prompt, it walks out with the vendor.

The usual objection is that data, not context, is the moat. Most of what an institution holds is licensed, shared, or purchasable, and a competitor can buy it tomorrow. The data that cannot be bought is the record of how this firm decided: the overlay as negotiated, the exception as granted, the override and who took it, the version in force that day. That record is inert on its own. Compiled with its rules attached, it is the moat. Context is proprietary data made executable.

That is the strategic point most AI roadmaps still miss. They treat context as fuel for the model. In a fiduciary setting, context is the asset. The model is a tenant.

Firms that treat the layer as infrastructure will be able to change models without changing what "compliant" means. Firms that treat the model as the system of record will rediscover their policy every time the vendor ships a new checkpoint.

06 What leadership looks like from here

1 If you run AI inside investment process today, the leadership question is not whether to adopt agents. That decision is already in motion across the industry.

The Grant Thornton data carries a point most boards have backwards. Among organizations that have fully integrated AI, 74 percent are very confident they could pass that governance audit. Among those still piloting, 7 percent. Governance did not slow the leaders down. It is what let them scale. It is the first deliverable, not the retrofit.

The question is whether the firm owns a compiled representation of its own decision rules: queryable at runtime, versioned when the text changes, transferable when the vendor changes, and reconstructable when an examiner asks.

2 That is a build, not a workshop. It starts with source documents, not with a chatbot. It requires a compiler discipline: extract with provenance, emit deterministic artifacts, shadow before cutover, approve with a cooling-off window, reconfigure without erasing the prior chain.

The delivery model that survives contact with a regulated institution is not a perpetual advisory retainer and not a black-box SaaS that owns the logic. It is an embedded architect who installs the layer, runs it against the institution's data inside the institution's boundary, and transfers the compiled representation.

The institution should leave holding capability, not a dependency.

07 The line

Everybody has AI.

Almost nobody can produce the clause that authorized the last material decision.

That is not a model shortage. It is a context shortage. The firms that treat context as the product (compiled, living, owned) will be the ones still standing when the examination calendar catches up to the agent calendar.

4 That calendar has dates on it. In Canada, OSFI's Guideline E-23 brings every model, AI included, into formal model risk management for federally regulated banks and insurers on 1 May 2027. Other regulators are on the same road.

Models are rented.

Context, if you compile it, is yours.

08 / Sources Sources

Mercer, How artificial intelligence is shaping asset management, 2026 AI in Asset Management Survey (131 managers, fielded February to March 2026). <https://www.mercer.com/about/newsroom/how-artificial-intelligence-is-shaping-asset-management/>

Grant Thornton, 2026 AI Impact Survey (approximately 950 senior US executives, ten industries, fielded 23 February to 18 March 2026). <https://www.grantthornton.com/services/advisory-services/artificial-intelligence/2026-ai-impact-survey>

OSFI, Guideline E-23 Model Risk Management (2027), effective 1 May 2027. <https://www.osfi-bsif.gc.ca/en/guidance/guidance-library/guideline-e-23-model-risk-management-2027-letter>

09 / Further reading Further reading

Living Policy Architecture: the case for executable, traceable policy in continuously operating AI systems. <https://ssrn.com/abstract=7108198>

The case for policy-coupled decision attestation. <https://ssrn.com/abstract=7108538>

From frameworks to filings. <https://doi.org/10.2139/ssrn.6935198>

AA-1 Agent Governance Attestation standard (CC BY 4.0). Available from dap.solutions.

AUTHOR

Brendan Sibeth is Founder and Managing Partner of Dartmouth Advisory Partners, a Toronto firm that builds Living Policy Architecture, the context layer for institutional intelligence. Views are his own.