Dynamics Blindness: When AI Is Locally Correct and Globally Non-Compliant

May 2026

Witold Reichhart and Arnaud Gelas

Five-Paper Programme: Enterprise AI and Organizational Intelligence

This paper is part of a five-paper programme examining why enterprise AI fails in regulated environments, what architecture resolves it, and what emerges when that architecture operates at depth.

Paper A — Dynamics Blindness (Reichhart and Gelas) ← *this paper* Diagnosis. Names the architectural failure mechanism: LLMs process tokens without tracing causal chains through organizational dependencies. Chain-of-thought, RAG, tool use, multi-agent systems, and scaling do not add the missing causal infrastructure.

Paper B — The Predictive Organization (Gelas and Reichhart) Architecture. Specifies the resolution: a tripartite structure — Map (state), Physics (dynamics), Player (agency) — coupling neural perception with symbolic reasoning, operating on claims-based knowledge with prevalence weighting.

Paper C — Build the Medium (Reichhart and Gelas) Theory. Ten independent theoretical traditions converge on what organizational intelligence requires. Introduces the capability/fertility distinction and the autonomy-to-initiative transition as the real measure of AI maturity.

Paper D — Governed Intelligence Architecture (Gelas and Reichhart) Methodology. The practitioner companion: five-stage Governed Intelligence Lifecycle (Ingest, Consolidate, Curate, Expand, Apply) with governance requirements, epistemic immunity, and the path dependency argument for knowledge infrastructure investment.

Paper E — From Autonomy to Initiative (Reichhart and Gelas) Capstone. Three conditions for governed initiative. Graduated immersion systems as institutional analogue. Governance relocation mechanism. Six computational enrichments. Active inference as normative model. The domain graph as missing middle layer.

Causal spine: Enterprise AI fails because of dynamics blindness (A) → the resolution is architectural (B) → the architecture works because ten traditions converge on what living systems require (C) → the practitioner methodology is a governed intelligence lifecycle with epistemic immunity (D) → when the architecture runs at sufficient depth, it produces governed initiative — agents that perceive what matters through immersion, not instruction (E).

Abstract

42% of companies scrapped the majority of their AI initiatives before production in 2025, up from 17% the year before (S&P Global, 2025). Only 6% of organizations qualify as AI high performers (McKinsey, 2025). Gartner projects over 40% of agentic AI projects cancelled by 2027. The failure concentrates in regulated enterprises, and the cause is architectural: large language models are dynamics-blind. They process tokens without tracing causal chains through organizational dependencies — they cannot tell you what breaks three systems away when you change a parameter here.

This paper names and formalizes dynamics blindness as the architectural root cause that persists when data quality, integration complexity, and organizational readiness are addressed. The diagnosis is specific: the problem emerges when the objective shifts from

augmenting human judgment to autonomous action over interconnected systems. Where that shift occurs, chain-of-thought prompting, retrieval-augmented generation, tool use, multi-agent orchestration, knowledge-graph-grounded retrieval, and scale each fail to add the missing causal infrastructure. The resolution requires architectures that provide causal reasoning as a service — coupling neural perception with symbolic constraint evaluation and causal traversal. Neuro-symbolic feasibility is now demonstrated empirically (Gupta et al., 2026) and specified architecturally in the companion paper (Gelas and Reichhart, 2026b).

This is Paper A of a five-paper programme: Paper B (Gelas and Reichhart, 2026b) specifies the architectural resolution — a tripartite structure of state, dynamics, and agency layers; Paper C (Reichhart and Gelas, 2026c) provides theoretical foundations from ten independent traditions; Paper D (Gelas and Reichhart, 2026d) operationalizes the methodology through the Governed Intelligence Lifecycle with epistemic immunity; Paper E (Reichhart and Gelas, 2026e) identifies the conditions under which the architecture produces governed initiative rather than mere autonomy.

Keywords: generative AI, enterprise AI, dynamics blindness, causal reasoning, neuro-symbolic AI, regulated industries, AI governance, model risk

---

1. The Failure That Has a Name

42% of companies scrapped the majority of their AI initiatives before reaching production in 2025 — up from 17% the year before (S&P Global, 2025). Only 7% of organizations have fully scaled AI enterprise-wide; only 39% attribute any EBIT impact to generative AI, with just 6% qualifying as high performers who attribute 5% or more (McKinsey, 2025). 74% of C-suite executives report struggling to achieve value from AI at scale (BCG, 2024). Gartner projects that over 40% of agentic AI projects will be cancelled by 2027 (Gartner, 2025). Cisco classifies just 13% of organizations as structurally prepared to move AI from pilot to production (Cisco, 2025). These are not early-adoption growing pains. The year-over-year trajectory points in one direction: failure rates are rising, not falling.

These statistics establish the scale of the problem, not its cause. Multiple mechanisms contribute to enterprise AI failure — data quality, integration complexity, organizational readiness, misaligned incentives. This paper argues that dynamics blindness is the architectural root cause that persists when others are resolved: the one limitation that better data, cleaner integration, and stronger change management cannot fix, because it resides in the technology's computational structure rather than its operating environment.

The industry's explanation is execution: the data isn't clean enough, the integration is too complex, the

organization isn't "AI-ready." That explanation is wrong — or rather, it describes symptoms while missing the disease. The failures concentrate in regulated enterprises — financial services, healthcare, insurance, pharmaceuticals, critical infrastructure — where decisions carry legal weight, consequences propagate through organizational dependencies, and regulators demand traceable reasoning. In these environments, generative AI fails not because the implementation is poor but because the technology's architecture is categorically mismatched to the domain. The cause is specific, and it has a name: dynamics blindness.

---

2. Dynamics Blindness Defined

A large language model processes tokens. Given a sequence of inputs, it predicts the most probable next output. It does this extraordinarily well — well enough to produce coherent text, answer questions, summarize documents, draft emails, generate code. These are real capabilities with practical value.

What it cannot do is reason about causation.

Consider a concrete scenario in an investment bank operating under MiFID II. An LLM-powered agent receives a routine request from the Head of Compliance: elevate a senior trader's access permissions to include the firm's risk management module, so she can review her desk's exposure data directly. The immediate action is straightforward — a permission flag changes in the identity and access management system. The agent executes it competently.

The cascade begins immediately, and it has three hops.

Hop one: the risk management module shares a data bus with the order management system. Granting read access to the risk module automatically provisions the trader with visibility into pre-trade risk parameters — including position limits and margin thresholds — that feed directly into execution decisions. The trader can now see the constraints under which her own orders will be evaluated. A token-based system processes the access request. It has no model of the data architecture that connects the risk module to the order management system.

Hop two: under the firm's segregation-of-duties controls — the internal equivalent of SOX Section 404 requirements — no individual may simultaneously hold execution authority and access to the risk parameters that govern execution. The trader already has execution authority. She now has visibility into the risk parameters that constrain it. This constitutes a conflict that the firm's control framework explicitly prohibits. A token-based system has no mechanism for evaluating role-permission intersections across systems against control policies it was never asked to check.

Hop three: the conflict triggers a reporting obligation. Under MiFID II Article 16(3), the firm must maintain organizational arrangements to identify, prevent, and manage conflicts of interest — including conflicts arising from the combination of roles and system access within the firm. A segregation-of-duties

failure of this kind is a recordable event under the firm's compliance monitoring framework, and failure to detect and report it constitutes a second, independent regulatory violation. The agent that processed the original request has no awareness that its action created a reporting obligation three systems away.

The agent understood the instruction. It performed the action. It generated a grammatical confirmation message. It was blind to every consequence that followed.

This pattern is not specific to trading floors. In insurance, an LLM-powered claims agent approves a building damage claim and assigns it to a preferred contractor. The approval triggers a reserve adjustment in the actuarial system, which breaches a reinsurance treaty threshold, which obligates the insurer to notify its reinsurance counterparty within 48 hours under the treaty's prompt-notice clause. The claims agent saw a building. The consequence was a contractual obligation to a third party in a different jurisdiction. Same mechanism, different domain.

The empirical evidence for this failure mode is converging from multiple independent lines of research. Chen et al. (2024), in a NeurIPS 2024 paper, demonstrate that LLMs perform only "level-1" causal reasoning — retrieving memorized causal patterns from training data — and fail at "level-2" genuine causal reasoning requiring inference over novel causal structures. Their CausalProbe-2024 benchmark, constructed from sources published after the training data cutoff of all tested models, shows significant performance degradation when LLMs encounter fresh causal scenarios. The finding is structural: the autoregression mechanism of transformer-based LLMs is not inherently causal. Even optimized prompting strategies (Jin et al., 2024) do not resolve the fundamental limitation.

The process mining and business process management communities are reaching the same conclusion from the enterprise direction. Recent work on causally-augmented business process reasoning (Fournier et al., 2024) demonstrates that LLMs fail to trace causal chains through interconnected organizational workflows — precisely the dynamics blindness pattern this paper describes, observed independently in a different research community using different benchmarks.

Gupta et al. (2026) operationalize the failure mode most directly for our purposes in their World of Workflows benchmark, which measures AI systems' ability to trace cascading consequences through interconnected business processes at enterprise scale: 4,000+ business rules, 55 active workflows, and 234 evaluation tasks spanning inventory management, procurement, HR compliance, and financial reporting — each requiring the system to trace causal chains through organizational dependencies and identify all downstream state changes triggered by a single action. Evaluation operates at two fidelity levels: a simplified mode that scores salient consequences, and a full-fidelity mode that requires identifying *all* downstream effects, including those that emerge from rule interactions invisible at the surface level. Under full-fidelity conditions, frontier LLMs consistently fail to predict the invisible, propagating side effects of their actions. No pure language model achieves passing thresholds. The failures are not random: models reliably capture first-hop consequences but miss second- and third-hop effects that arise from constraint interactions across workflow boundaries — precisely the pattern the trading floor example

illustrates.

This is dynamics blindness: the inability to trace how changes propagate through interconnected systems, and to reason counterfactually about what happens when actions interact with rules in dynamic environments.

More precisely: given the organization's current state (its roles, permissions, system connections, and active constraints), its transition rules (regulatory obligations, control policies, contractual clauses), and a proposed action — a dynamics-competent system must correctly identify all reachable downstream states that cross a compliance or control boundary, with completeness guarantees relative to that boundary. The compliance boundary is the set of regulatory, contractual, and internal control obligations the organization must never violate — the hard constraints where an undetected breach constitutes a regulatory event, not merely a quality issue. Different categories of constraint tolerate different levels of assurance: hard regulatory boundaries (embargo lists, capital ratios, segregation-of-duties) demand pre-execution deterministic evaluation; softer operational constraints may tolerate probabilistic monitoring with post-hoc audit. Dynamics blindness is the property of systems that cannot provide either — that produce plausible but incomplete causal chains because they discover consequences by statistical pattern-matching rather than systematic traversal. What matters is whether the system can certify, relative to the applicable compliance boundary, that no relevant chain has been missed.

Dynamics blindness differs from hallucination. Hallucination is a failure of factual accuracy — the model generates content that is false. Dynamics blindness operates on a different axis entirely. An agent can be factually accurate and dynamics-blind at the same time. It can correctly identify the user, correctly parse the request, correctly update the permission — while remaining completely unaware that the action triggered a cascade of consequences through organizational structures it has no model of.

The distinction matters because the remedies are different. Hallucination can be mitigated by grounding the model in verified data — retrieval-augmented generation, fact-checking layers, knowledge bases. Dynamics blindness cannot be mitigated by better data. The problem is not that the model lacks information about the segregation-of-duties policy. The problem is that even with that policy in its context window, it has no mechanism for tracing the causal chain from "permission granted" through "data bus exposure" through "role conflict created" to "regulatory reporting obligation triggered." That chain requires causal reasoning — understanding that A causes B, B enables C, and C under condition D produces E. Token prediction does not perform this operation. It cannot.

Dynamics blindness also differs from the well-documented limitations around reasoning and planning. Those limitations involve the model's ability to chain logical steps within a single problem. Dynamics blindness is about the interaction between an action and a living system — the way consequences propagate through dependencies that exist outside the model's computational frame. An LLM can solve a logic puzzle. It cannot tell you what happens to a regulated organization when you change one parameter in a system of interdependent rules, roles, and constraints.

This makes it a structural blind spot, not a performance gap.

---

3. Why This Is Architectural, Not Solvable by Better Models

The obvious objection is that the next generation of models will fix this. Chain-of-thought reasoning, tool-augmented agents, scaling laws, emergent capabilities — each represents a genuine advance. But none adds the causal infrastructure that regulated enterprises require, and the reasons are specific.

A necessary clarification first. Dynamics blindness is a property of the base model — the token-prediction engine itself. A system that embeds the model in formal causal infrastructure (a dependency graph, a constraint solver, a deterministic traversal engine) is a different artifact with different properties. The claim is specific: the LLM component cannot generate causal reasoning internally, and that current mitigation strategies — the ones actually being deployed — do not add infrastructure sufficient to guarantee completeness. A full neuro-symbolic system with a verified causal model can, in principle, satisfy the requirement. What falls short of that — and why — is the focus of this section.

Chain-of-thought reasoning is the strongest counterargument. Modern LLMs can decompose problems into intermediate steps, producing outputs that look like causal reasoning: "If we change X, then Y is affected, which means Z follows." In many cases these chains are correct. Average accuracy may be fine. The problem is that the model has no mechanism for guaranteeing completeness. In the trading floor example, chain-of-thought might trace the permission change to the risk module (hop one) and even surface the segregation-of-duties conflict (hop two). But the model discovers these steps by statistical salience — by how prominently similar patterns appeared in its training data — not by systematically traversing the dependency graph. Hop three, the MiFID II reporting obligation, arises from the intersection of an internal control failure with a regulatory requirement. If that specific interaction pattern is underrepresented in the training corpus, the model simply will not surface it. The failure mode is silent omission: the chain looks complete, reads well, and is missing the step that matters most. You cannot audit what the model did not consider.

Tool-augmented agents address a different gap. An LLM agent that can query an access management API, look up role assignments, and read policy documents has real operational capability. But the critical question is which tools to call and in what order — and that decision is still governed by token prediction. The agent queries the access management system because the prompt mentions permissions. It does not query the data bus architecture that connects the risk module to the order management system, because nothing in the request signals that this connection is relevant. The dependency is invisible precisely because it operates at the infrastructure level, below the semantic surface where language models work. Tool use gives the agent the ability to act on the organizational environment. It does not give the agent a model of that environment — a representation of how systems, rules, roles, and constraints interconnect.

Without that model, the agent cannot know what it does not know.

The most dangerous variant of this fallacy is the assumption that deploying more agents solves the problem — that multi-agent orchestration, agent swarms, or agentic pipelines overcome dynamics blindness through division of labor. They do not. If each agent in a multi-agent system is individually dynamics-blind, the ensemble is collectively dynamics-blind. The causal gap is not in the coordination between agents but in the absence of a shared causal model that any of them can reason over. One agent processes the permission change. Another monitors compliance. A third handles reporting. Each performs its task competently at the token level. None has a model of the dependency graph that connects their domains. Adding more dynamics-blind agents accelerates action without adding causal awareness, producing faster cascades through unmapped dependencies. This is the Gartner 40%-cancellation forecast made concrete: agentic AI projects will fail precisely because they multiply agency without providing the causal infrastructure that constrains it.

A related objection points to operating model maturity rather than model capability. The most advanced technology organizations have moved toward integrated delivery practices — cross-functional teams that encode security controls and compliance checks directly into their deployment processes, enforcing them automatically before any change reaches production. In the best cases, these organizations maintain layered security harnesses where every change passes through automated policy verification, dependency analysis, and access control validation. This represents real progress and reduces one form of dynamics blindness — the organizational kind, where dependencies and obligations are scattered across departments and artifacts so that no one can see the full cascade. But even in these mature environments, the architectural kind persists. Global constraints in regulated enterprises are temporal (T+1 settlement obligations, reporting deadlines), threshold-based (capital ratios, exposure limits), and cross-domain (segregation-of-duties spanning Risk, Legal, and Finance). These are not pre-deployment checks that automated pipelines can enforce. They depend on runtime state — who holds what position, what the current exposure is, which data feeds are connected — and they propagate across systems that no single team owns. High deployment velocity compounds the problem: each locally validated change accumulates systemic drift against constraints that no automated gate inspects. In the MiFID example, a mature organization might enforce the permission change through automated access-management policy checks. The automation catches simple policy violations. But the reporting obligation triggered by the intersection of execution authority, risk module access, and current exposure is a runtime relationship that requires an explicit causal model of the organizational state — not a pre-deployment gate. Operating model maturity is complementary to architectural resolution, not a substitute for it. And most organizations have not reached even this level of delivery maturity — making the gap between what automated controls can catch and what cross-domain constraints require even wider.

Scaling and emergent capabilities represent the most speculative counterargument — and the most seductive. The reasoning runs: capabilities have emerged unpredictably at scale before (in-context learning,

arithmetic, code generation), so causal reasoning might emerge in the next generation. This misidentifies what emergence has actually produced. Every documented emergent capability is a form of pattern recognition over the training distribution — recognizing structure in inputs the model has seen statistical proxies for. Causal reasoning about a specific organization's dependency graph at a specific point in time is not a pattern in any training corpus. It is a property of a live system whose state changes continuously. No amount of scaling gives a model the causal reasoning infrastructure to evaluate how role assignments interact with control policies across interconnected systems. These are facts about a specific organizational state, not statistical regularities in text.

A more nuanced objection points to hybrid architectures — knowledge-graph-grounded retrieval (GraphRAG), symbolic constraint-checking layers, structured knowledge representations — that embed LLMs in causal or semi-causal infrastructure and recover partial awareness. These approaches represent meaningful progress. Graph-grounded retrieval architectures achieve substantially stronger performance on knowledge-intensive tasks than basic vector retrieval by assembling richer context from structured relationships (Edge et al., 2024). But retrieval quality and causal completeness are different properties. A system that retrieves more relevant context still reasons over that context through token prediction — it has better inputs, not a different reasoning mechanism. A graph-grounded agent that traces causal chains correctly in 95% of cases still cannot identify which 5% it missed. In a regulatory context, this distinction is not a matter of degree. It is the entire question.

The distinction that collapses all these counterarguments is the one regulators already enforce: the difference between producing correct-looking output and guaranteeing auditable traceability.

The Federal Reserve's SR 11-7 guidance requires "conceptual soundness" in models used for decision-making — the model's design must be assessable, and documentation must allow unfamiliar parties to understand how it operates (Federal Reserve, 2011). An LLM can produce a credit risk assessment that states "73% probability of default." SR 11-7 requires something categorically different: here are the factors, here is how they combine under the applicable rules, here is why this classification is correct, and here is the chain of reasoning an independent reviewer can reproduce. The EU AI Act imposes technical documentation and accountability requirements for high-risk AI systems that implicitly demand the same thing — reasoning that is reconstructable, not statistically probable but logically traceable (European Parliament, 2024).

A system that produces the right causal chain 90% of the time remains non-compliant, because the regulator cannot determine which outputs belong to the 90% and which to the 10%. Probabilistic correctness and guaranteed auditability are different things, produced by different architectures.

---

4. What Dynamics Blindness Costs

In unregulated consumer applications, dynamics blindness is a quality issue. A chatbot that misunderstands a request produces a bad experience. The user asks again. In regulated enterprises, dynamics blindness produces three categories of structural cost.

Compliance risk through local correctness. Decisions in regulated environments are not isolated events. They are nodes in a dependency graph where rules, roles, permissions, obligations, and constraints interact. A decision that looks correct in its immediate context — approving a transaction, granting access, classifying a risk — can violate rules that operate three, four, or five hops away in the dependency structure. Dynamics-blind systems produce locally correct, globally non-compliant decisions. The compliance violation is invisible to the agent that caused it because the agent has no model of the system-level dependencies through which its action propagates.

This is not a theoretical concern. The Digital Operational Resilience Act (DORA), which entered into application in January 2025, requires financial entities to maintain ICT risk management frameworks that map systems, identify critical dependencies, and document the relationships between operational components (European Parliament, 2022). BCBS 239 imposes principles for risk data aggregation that demand data lineage — auditable trails from origin through transformations — and governance structures that ensure data quality, ownership, and accountability across the organization (Basel Committee, 2013). Both regulatory frameworks assume an infrastructure capable of representing and reasoning about dependencies. Current AI architectures cannot provide this reasoning. They can retrieve dependency data. They cannot reason about what happens when dependencies interact under changing conditions.

Cascade failures through unmapped propagation. When a change propagates through interdependent systems and no one has mapped the causal chains, failures cascade. A parameter change in one system affects calculations in a downstream system, which triggers a threshold breach in a third system, which generates a regulatory report containing incorrect data. Dynamics-blind AI not only fails to prevent these cascades — it can initiate them, because it takes actions without modeling their downstream effects. The more autonomous the agent, the faster the cascade and the harder it is to trace after the fact.

DORA's emphasis on operational resilience testing, scenario-based assessment, and third-party risk management reflects regulators' recognition that systemic risk arises from interconnection — from the way failures propagate through dependencies, not from isolated component failures (European Parliament, 2022). Operational resilience is, at its core, a causal reasoning problem: given that X fails, what happens to Y, Z, and W? An infrastructure that cannot perform this reasoning cannot deliver operational resilience.

Opportunity cost through the inability to reason counterfactually. Strategic decision-making in regulated enterprises requires counterfactual reasoning: what happens if we change pricing model X while regulatory constraint Y is active and competitor Z enters market W? This is the kind of question that justifies AI investment at the executive level — the ability to simulate scenarios, explore alternatives, and

anticipate consequences before committing resources.

Dynamics-blind systems cannot perform counterfactual reasoning because they cannot model causal relationships. They can generate plausible-sounding scenarios based on patterns in their training data. They cannot simulate what would actually happen under specific conditions in a specific organizational context, because simulation requires a causal model and they do not have one. The most expensive cost of dynamics blindness is not the failures it causes but the strategic capabilities it forecloses. Counterfactual reasoning is a knowledge-generation mechanism: an organization that can simulate how its decisions interact with its constraints produces knowledge it did not previously possess. Without causal infrastructure, this entire category of organizational knowledge — generated through simulation rather than observation — remains inaccessible.

---

5. The Resolution Is Architectural

If dynamics blindness is architectural — if it arises from the nature of token prediction rather than from insufficient data or inadequate training — then the resolution must also be architectural. Better language models will not solve it. Different infrastructure will.

The requirement is specific: the agent needs access to a causal model it can consume as a service — an infrastructure that traces consequences through organizational dependencies deterministically, rather than generating them probabilistically. Multiple implementation patterns can satisfy this requirement: rule engines, constraint solvers, model checkers, simulation engines, policy-as-code frameworks, and neuro-symbolic architectures all provide forms of causal infrastructure. The pattern most directly relevant to LLM-based agents in regulated enterprises is neuro-symbolic — architectures that couple neural perception (reading unstructured inputs, classifying entities, extracting intent) with symbolic causal reasoning (traversing dependency graphs, evaluating constraints, certifying compliance). This is the pattern that handles the characteristic challenge: unstructured input mapped to formal evaluation over organizational state.

The research direction is established and the first empirical results are in. Garcez and Lamb (2023) describe neuro-symbolic AI as "the 3rd wave" — architectures that integrate neural network learning with symbolic knowledge representation and logical reasoning. DARPA's Assured Neuro-Symbolic Learning and Reasoning (ANSR) program funds research into hybrid systems that combine data-driven learning with symbolic reasoning, with the explicit goal of delivering reliable inference, stronger generalization, and verifiable evidence for assurance and trust (DARPA, 2023). The motivation in both cases is the same: neural components alone cannot provide the guarantees that high-stakes applications require.

The feasibility of this approach is no longer theoretical. Gupta et al. (2026) built the World of Workflows benchmark to measure precisely the dynamics blindness this paper describes. In a related line of work,

the same group's CASSANDRA architecture demonstrates the neuro-symbolic alternative: deterministic variables modeled as symbolic rules, probabilistic variables modeled as interconnected neural networks, and causal graph structure binding them together (Gupta et al., 2026). In benchmark simulations measuring enterprise survival over extended horizons, pure neural world models failed catastrophically — becoming insolvent in the majority of runs — while the hybrid neuro-symbolic architecture consistently survived the full simulation period. The performance gap between architectures is not marginal, though the specific thresholds depend on benchmark configuration (number of active workflows, rule complexity, simulation length). What matters for the present argument is that the gap is structural: it reflects an architectural difference, not a tuning difference.

In the context of regulated enterprises, the neuro-symbolic insight translates into a specific architectural division.

Neural components do what they do well: perceive, classify, extract, and generate. They read unstructured data — customer communications, risk signals, behavioral patterns, document content. They identify entities, sentiments, intents, anomalies. This is genuine capability, and no symbolic system matches it.

Symbolic components do what neural components cannot: enforce hard constraints and trace causal chains. Credit limits, embargo lists, regulatory thresholds, segregation-of-duties rules, capital adequacy ratios — these are logical propositions with definite truth values. A transaction either violates an embargo or it does not. A capital ratio either meets the threshold or it does not. These determinations require symbolic evaluation, not probabilistic estimation.

The combination produces something neither component delivers alone: white-box reasoning where every step from input to output can be audited.

Return to the trading floor example. In a neuro-symbolic architecture, the LLM agent still receives the Head of Compliance's request and understands it — that is perception, and the neural component handles it well. But before the permission change executes, the symbolic layer evaluates it against the organizational state: the trader's existing execution authority, the data bus connecting the risk module to the order management system, and the firm's segregation-of-duties controls. The dynamics engine traces the causal chain — risk module access exposes pre-trade parameters, which combined with execution authority constitutes a SOD conflict under Control Policy 4.7.2, which triggers a MiFID II Article 16(3) conflict-of-interest reporting obligation — and returns a constraint violation before the action is taken, not after. The reasoning chain is explicit: here is the action, here are the three affected systems, here is the control conflict, here is the regulatory consequence. An independent reviewer — or a regulator — can examine each step and determine whether the evaluation was correct.

The output is not just an answer — it is a justification graph: a structured, machine-readable chain of reasoning where each node is a state transition or constraint evaluation and each edge is a dependency link. For the trading floor example, the graph reads: *Action(grant risk-module access)* $\rightarrow$ *StateChange(trader*

gains pre-trade parameter visibility) → *DependencyLink(risk module shares data bus with OMS)* → *ConstraintViolation(SOD Policy 4.7.2: execution authority + risk parameter visibility)* → *Regulatory-Obligation(MiFID II Art. 16(3): conflict-of-interest reporting)*. Every node is auditable. Every edge is traceable to a documented dependency. An independent reviewer — or a regulator — can examine each step, challenge any link, and determine whether the evaluation was correct. SR 11-7's conceptual soundness requirement is met by construction.

The infrastructure to connect these components is maturing rapidly. The Model Context Protocol (MCP) provides standardized agent-backend communication with over 1,000 open-source connectors (Anthropic, 2024). The Agent-to-Agent (A2A) protocol, now hosted by the Linux Foundation, enables inter-agent communication across diverse frameworks (Google, 2025). Together, these create the plumbing for architectures where a neural perception layer consults a symbolic reasoning layer via standardized interfaces — precisely the pattern the paper advocates.

This architectural direction points toward a more complete framework that we specify in the companion paper (Gelas and Reichhart, 2026b): the Enterprise World Model — a tripartite architecture comprising state representation, dynamics, and agency.

State representation captures what the organization knows at any point in time — the positions, relationships, constraints, and rules that define its operational reality. This is not a database. It is a formal model of the organization's state space — the set of all configurations the organization can occupy, given its current structure, rules, and commitments.

Dynamics captures how things change under rules. When an action is taken — a clearance is updated, a position is opened, a product is launched — the dynamics layer traces the consequences through the dependency graph, identifying which constraints are affected, which state transitions are triggered, and which compliance boundaries are approached or crossed. This is the layer that resolves dynamics blindness directly. It provides the causal reasoning that token prediction cannot.

Agency defines who acts within those constraints — human operators, automated processes, AI agents — and ensures that every action is taken by an authorized actor operating within the boundaries the dynamics layer enforces. An AI agent in this architecture is sovereign within its authorized domain, because the infrastructure guarantees that its actions have been evaluated against the full causal model before execution. This is not a pipeline but a continuous loop: every action updates the state representation, every state change triggers a dynamics evaluation, every evaluation constrains the next action. The architecture learns — not in the neural sense of weight updates, but in the organizational sense of accumulating state knowledge that makes subsequent evaluations more complete.

The change is structural, not incremental. The infrastructure provides causal reasoning as a service — something any component, including LLM-powered agents, can consume without having to generate it internally.

6. Implications

The argument of this paper has a direct consequence for three audiences.

For CIOs and technology leaders in regulated enterprises, the calculus changes. AI capability investment is already underway and largely justified. What matters now is whether organizational infrastructure can support causal reasoning. Can it trace the consequences of a decision through organizational dependencies? Can it evaluate whether an action complies with rules that operate across systems, roles, and time? If the answer is no, then deploying more capable language models will reproduce the current failure pattern at greater scale and speed. Investment in AI capability without investment in causal infrastructure is investment in faster dynamics blindness.

The good news is that this does not require a greenfield infrastructure build. Regulated enterprises have already invested heavily in the substrate the Enterprise World Model requires — they just built it for compliance rather than for intelligence. BCBS 239 forced organizations to map data lineage and build dependency documentation. DORA mandates ICT risk management frameworks that document system interdependencies. Segregation-of-duties controls require formal role-permission models. Capital adequacy and reporting obligations demand constraint evaluation logic. UK financial services firms alone spent £34.2 billion on financial crime compliance in 2022 (UK Finance, 2023). The dependency maps, control frameworks, system documentation, and constraint logic that this spending produced are not audit artifacts to be filed and forgotten. They are the raw material of the state representation and dynamics layers the Enterprise World Model requires. What is missing is not the data but the architectural frame that makes it computable — that turns static compliance documentation into a live causal model agents can query before they act. The investment case reduces to a single question: why are you treating the infrastructure you have already paid for as dead documentation instead of making it computable?

For regulators, the implications are already embedded in existing frameworks. SR 11-7's conceptual soundness requirement, DORA's operational resilience mandates, BCBS 239's data lineage principles, and the EU AI Act's documentation and accountability requirements all converge on a single demand: decisions must be traceable through an auditable chain of reasoning. As AI agents take on more decision-making authority in regulated environments — a direction the industry is pursuing aggressively — the gap between what regulators require and what current architectures can provide will widen, not narrow. Regulators have already described the standard. The industry has not yet built the architectural frame to meet it — though the compliance investments regulators themselves mandated are, inadvertently, producing the substrate on which that frame can be built.

For the AI industry, the agentic AI wave — the drive to deploy autonomous agents that act in enterprise environments — will hit the same wall this paper describes, unless agents operate within infrastructure

that provides the causal grounding they cannot generate internally. An LLM agent that can browse the web, query databases, and execute multi-step workflows is impressive engineering. An LLM agent that does all of that inside a regulated organization without a causal model of the organization's dependencies, constraints, and rules is a compliance incident in formation. The Gartner projection — 40% of agentic AI projects cancelled by 2027 — is a forecast of exactly this collision.

The capability is not the problem. The architecture is. Paper B in this series (Gelas and Reichhart, 2026b) specifies what to build. Paper C (Reichhart and Gelas, 2026c) provides the theoretical foundations for why this architecture works — grounding the design in the structure of organizational intelligence itself.

---

7. Limitations and Scope

This paper diagnoses dynamics blindness as an architectural property of large language models and argues that the resolution requires causal infrastructure — an Enterprise World Model — that agents can consume as a service, with neuro-symbolic architectures as the most mature implementation pattern. The authors are currently applying elements of the architecture in regulated financial services engagements, with empirical results to be reported in forthcoming papers. Three dimensions of the problem are acknowledged but not fully addressed here.

First, much of the causal knowledge that the symbolic layer requires exists as tacit knowledge held by experienced practitioners. The senior compliance officer who recognizes the SOD conflict in Section 2's scenario possesses knowledge acquired through years of operating within the organizational dependency graph — knowledge that resists codification into formal models. Polanyi's insight that "we can know more than we can tell" (1966) applies directly. Knowledge management initiatives that attempt to make tacit knowledge explicit have failed consistently across three decades of enterprise deployment — a pattern documented extensively in the KM literature (Desouza, 2011; Ragab and Arisha, 2013) and attributable to the fundamental mismatch between tacit knowledge's structure and explicit codification's requirements. The epistemological challenge of externalizing tacit organizational knowledge into formal causal models is at least as significant as the architectural challenge this paper addresses. Paper C in this series (Reichhart and Gelas, 2026c) engages with this dimension through the lens of organizational intelligence and knowledge-as-medium.

Second, the proposed architecture requires organizational capabilities — governance structures, domain expertise, ongoing maintenance — that many enterprises lack. Who encodes the dependency graphs? Who validates the causal models as systems change? These questions of absorptive capacity (Cohen and Levinthal, 1990) and knowledge governance are addressed in the companion papers but remain open challenges for implementation.

Third, leading regulated institutions — notably Palantir, whose Foundry platform explicitly frames the

enterprise AI problem as ontological rather than capability-based — have already begun building causal infrastructure that partially addresses the dynamics blindness described here. The paper's contribution is not discovering a problem no one has noticed. It is naming and formalizing a failure mode that explains why most organizations — the 94% that are not AI high performers — continue to fail, and specifying the architectural properties any solution must have.

References

Anthropic. (2024). *Model Context Protocol: Connecting AI to Everything.* https://modelcontextprotocol.io/

Basel Committee on Banking Supervision. (2013). *Principles for Effective Risk Data Aggregation and Risk Reporting.* BCBS 239. Bank for International Settlements.

Boston Consulting Group. (2024). "AI Adoption in 2024: 74% of Companies Struggle to Achieve and Scale Value." BCG Press Release, October 24, 2024.

Cisco. (2025). *Cisco AI Readiness Index 2025: Realizing the Value of AI.* Cisco Systems.

Chen, L. et al. (2024). "Unveiling Causal Reasoning in Large Language Models: Reality or Mirage?" *Advances in Neural Information Processing Systems* (NeurIPS 2024), 37.

Cohen, W.M. and Levinthal, D.A. (1990). "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly,* 35(1), 128–152.

DARPA. (2023). *Assured Neuro-Symbolic Learning and Reasoning (ANSR).* Defense Advanced Research Projects Agency. https://www.darpa.mil/research/programs/assured-neuro-symbolic-learning-and-reasoning

Edge, D. et al. (2024). "From Local to Global: A Graph RAG Approach to Query-Focused Summarization." arXiv:2404.16130.

European Parliament. (2022). *Regulation (EU) 2022/2554 on Digital Operational Resilience for the Financial Sector (DORA).* Official Journal of the European Union.

European Parliament. (2024). *Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence (AI Act).* Official Journal of the European Union.

Federal Reserve. (2011). *Supervisory Letter SR 11-7: Guidance on Model Risk Management.* Board of Governors of the Federal Reserve System.

Fournier, F. et al. (2024). "Towards a Benchmark for Causal Business Process Reasoning with LLMs." In *BPM 2024 Workshops*, Lecture Notes in Business Information Processing, Springer.

Garcez, A.d. and Lamb, L.C. (2023). "Neurosymbolic AI: The 3rd Wave." *Artificial Intelligence Review,* 56, 12387–12406. https://doi.org/10.1007/s10462-023-10448-w

Gartner. (2025). "Gartner Predicts Over 40% of Agentic AI Projects Will Be Cancelled by End of 2027." Gartner Press Release, June 25, 2025.

Gelas, A. and Reichhart, W. (2026b). "The Predictive Organization: Architecture for Enterprise Intelligence." SSRN Working Paper.

Gelas, A. and Reichhart, W. (2026d). "Governed Intelligence Architecture for Institutional AI." SSRN Working Paper.

Reichhart, W. and Gelas, A. (2026c). "Build the Medium: Why Organizational Intelligence Is Mechanism, Not Metaphor." SSRN Working Paper.

Reichhart, W. and Gelas, A. (2026e). "From Autonomy to Initiative: Enterprise AI's Real Endgame." SSRN Working Paper.

Google. (2025). "A2A: A New Era of Agent Interoperability." Google Developers Blog, April 2025. https://a2a-protocol.org/

Jin, Z. et al. (2024). "CLadder: A Benchmark to Assess Causal Reasoning Capabilities of Language Models." *Advances in Neural Information Processing Systems* (NeurIPS 2023), 36.

Gupta, L. et al. (2026). "World of Workflows: A Benchmark for Bringing World Models to Enterprise Systems." arXiv:2601.22130.

McKinsey & Company. (2025). "The State of AI: How Organizations Are Rewiring to Capture Value." McKinsey Global Survey on AI.

Desouza, K.C. (2011). "An Introduction to Knowledge Management." In *Knowledge Management: An Introduction,* pp. 1–28. Neal-Schuman Publishers.

Ragab, M.A.F. and Arisha, A. (2013). "Knowledge Management and Measurement: A Critical Review." *Journal of Knowledge Management,* 17(6), 873–893. https://doi.org/10.1108/JKM-12-2012-0381

Polanyi, M. (1966). *The Tacit Dimension.* University of Chicago Press.

S&P Global Market Intelligence. (2025). "Voice of the Enterprise: AI & Machine Learning, Use Cases 2025." S&P Global.

UK Finance. (2023). "Financial Crime Compliance Spend in UK Financial Services." UK Finance Research, March 2023.

Correspondence: witold.reichhart@gmail.com, arnaudgelas@gmail.com