From Autonomy to Initiative: Enterprise AI's Real Endgame

May 2026

Witold Reichhart and Arnaud Gelas

Five-Paper Programme: Enterprise AI and Organizational Intelligence

This paper is the capstone 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) Diagnosis. Names the architectural failure mechanism: LLMs process tokens without tracing causal chains through organizational dependencies.

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. Five-stage Governed Intelligence Lifecycle with epistemic immunity. The practitioner methodology for building and maintaining the substrate.

Paper E — From Autonomy to Initiative (Reichhart and Gelas) ← *this paper* 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

Current agentic AI discourse often treats autonomous execution as the apex of maturity — agents that operate independently within assigned parameters. We argue this is the wrong endpoint for enterprise AI in high-consequence environments. Autonomy is independence in *how* to work while remaining dependent on humans for *what* to work on. Initiative — the capacity to identify and prioritize action opportunities aligned with organisational purpose through immersion rather than instruction — is a categorically different capacity, and the one that determines whether enterprise AI generates value or merely generates output.

We define organizational intelligence as a system property: the capacity to perceive institutionally relevant patterns, infer their consequences, and select action opportunities under changing constraints. Governed knowledge is not intelligence itself — it is the substrate on which organizational intelligence operates. On this basis, we propose three conditions for governed initiative — initiative that is auditable, correctable, and institutionally legible: fertile pattern density above a perceptual threshold (Condition 1), a constraint-legible action space where governance rationale is causally transparent (Condition 2), and progressive governance relocation as the institutional world model deepens (Condition 3). We examine the professional consulting pyramid as a mature institutional analogue — one of several graduated immersion systems, including medical residency and military officer development, that satisfy all three conditions through accumulated exposure, legible constraints, and graduated relaxation of control as judgment develops.

For AI agents, each condition transforms in implementation but the functional requirements are identical. The central engineering claim is a three-step governance relocation mechanism: as the institutional world model deepens, governance enforcement migrates from explicit pre-action constraint checking (early lifecycle) through substrate-resident causal structure (mid lifecycle) to risk-based monitoring and audit sampling (mature lifecycle). We identify a substrate gap in the current literature — world models without governance and governance without world models each satisfy only a subset of the conditions — and propose six architectural enrichments to the Governed Intelligence Lifecycle (Paper D) that close it. We introduce the domain graph as the missing middle layer between the foundation model and application context: a governed, reusable, compounding substrate for domain-specific organizational intelligence.

Active inference (Friston, 2010) provides a normative model for describing why agents operating on deep, governed institutional world models may move from autonomous task execution toward institutionally legible initiative. We address the strongest counterargument — that scale alone may produce initiative without governed architecture — by scoping the claim: in regulated industries and enterprise deployments where agent actions carry material consequences, initiative must be auditable, correctable, and governable. Scale may produce initiative-like behaviour in low-consequence settings; governed architecture is necessary where initiative must be institutionally legible.

This is the capstone of a five-paper programme: Paper A (Reichhart and Gelas, 2026a) diagnoses dynamics blindness as the failure mechanism; Paper B (Gelas and Reichhart, 2026b) specifies the architectural resolution; Paper C (Reichhart and Gelas, 2026c) provides theoretical foundations from ten independent traditions; Paper D (Gelas and Reichhart, 2026d) operationalizes the governed intelligence lifecycle with epistemic immunity; this paper identifies the conditions under which that architecture produces governed initiative — agents that perceive what matters through immersion, not instruction.

Keywords: initiative, autonomy, enterprise AI, active inference, governed world models, knowledge graphs, organisational intelligence, governance relocation, domain graph, regulated industries

---

1. The Wrong Metric

The agentic AI industry is optimising along the wrong dimension. Paper C (Reichhart and Gelas, 2026c) introduced the capability/fertility matrix — capability measuring what AI can do, fertility measuring the organisational rate at which existing knowledge generates new knowledge, new capability, and new possibilities. Current maturity frameworks measure capability. They do not measure fertility. The consequence is visible in the market data: substantial deployment, minimal value.

	Low Fertility	High Fertility
High Capability	Powerful Tools	Living Medium
Low Capability	Early Stage	Fertile Ground

Many influential AI maturity narratives — Gartner's AI Maturity Model, McKinsey's QuantumBlack maturity scales, MITRE's AI maturity dimensions — treat their highest level as pervasive autonomous capability. Level 5 is the organisation where AI operates independently across processes, requiring minimal human intervention. This is the capability axis at full extension.

The market data tells us something about what this achieves. McKinsey's 2025 global survey found that only 7% of organisations have scaled AI enterprise-wide, and only 17% attribute more than 5% of EBIT to generative AI. While alternative interpretations exist — early-stage maturity, measurement difficulty, non-EBIT value capture — the structural pattern is consistent: this is not primarily a deployment problem. Deployment is widespread. It is a value problem. Organisations have built sophisticated agent architectures that execute well on assigned tasks — and generate marginal returns. They occupy the Powerful Tools quadrant: high capability, low fertility.

The distinction between these quadrants is not capability but alignment. The organisations that generate value have not merely deployed more capable agents. They have built conditions where AI perceives through accumulated organisational context — where the substrate is fertile, knowledge compounds through operation, and patterns propagate across boundaries. They are in or approaching the Living Medium quadrant.

Autonomy is a capability-axis achievement. It is independence in *how* to work — selecting methods, decomposing tasks, choosing tools — while remaining dependent on humans for *what* to work on. An autonomous agent executes assigned objectives efficiently. It does not perceive which action opportunities serve organisational purpose. That perception requires a different capacity entirely.

Initiative is a fertility-axis achievement. It is the capacity to identify and prioritize action opportunities aligned with organisational purpose through immersion — perceiving what matters through accumulated exposure to the organisation's dynamics, constraints, and values. A categorical shift from executing assigned objectives to perceiving which action opportunities serve organisational purpose in the first place.

This distinction has a precise analogue in professional practice. A first-year consultant executes assigned tasks autonomously — she can run analyses, build decks, manage workstreams without supervision. She has autonomy. She does not have initiative. She works on what she is told to work on. A senior partner walks into a client meeting and perceives the real problem — the one the client has not articulated, the one that will surface in six months if the current trajectory continues. She perceives through fifteen years of accumulated exposure within a governed professional structure. Her initiative is a different kind of capacity entirely, developed through a different kind of mechanism.

The capability/fertility matrix from Paper C exposed the strategic error: organisations invest in capability while neglecting fertility. Paper C established that path dependency makes the sequence consequential — capability-first investment progressively forecloses fertility, making initiative harder to achieve over time. This paper asks: if initiative requires fertility — a living, generative substrate — what mechanism converts fertility into initiative? What must the substrate satisfy for initiative to emerge?

Knowledge, intelligence, and initiative

The programme this paper belongs to is titled "Enterprise AI and Organizational Intelligence." The term requires definition — not as a philosophical claim about the nature of mind, but as an operational concept that distinguishes what the architecture produces from what it contains.

Knowledge is structured content: validated claims, causal relations, provenance, confidence, temporal validity, scope, and governance status. Knowledge is what the system holds. A governed knowledge graph is a knowledge infrastructure.

An institutional world model is knowledge organized with causal, temporal, and action-relevant structure — not merely what is true, but what causes what, what changes when, and what follows from action or inaction.

Organizational intelligence is the system-level capacity to perceive institutionally relevant patterns, infer their consequences, and select or recommend action opportunities aligned with organizational purpose under changing constraints. Intelligence is not a property of the model alone. It is a property of the agent-substrate system — the interaction between inferential capability, governed knowledge, and the institutional constraints that define what "aligned" means. A rich knowledge graph without an agent exercising inference over it is not intelligent. A capable model without a governed substrate to reason over is not organizationally intelligent.

Autonomy is organizational intelligence applied to assigned objectives — independence in how to work, while remaining dependent on humans for what to work on.

Initiative is organizational intelligence applied to the identification and prioritization of action opportunities before they are explicitly assigned — perceiving what matters through immersion rather than instruction.

Governed initiative is initiative whose reasoning remains auditable, correctable, and institutionally legible. In regulated and high-consequence environments, ungoverned initiative is not a lesser form of the same capacity — it is a different and potentially dangerous one.

This hierarchy matters because it prevents two errors. The first is treating knowledge infrastructure as intelligence — a governed knowledge graph is not intelligent; it is the substrate from which intelligence can be exercised. The second is treating model capability as organizational intelligence — a capable model without governed domain knowledge may be perceptive in the abstract but has no mechanism for determining which patterns matter to the institution.

The question must also be stated carefully. "Initiative" in this paper is a computational-organisational construct — the system-level capacity for agents to identify and act on action opportunities aligned with organisational purpose through immersion. It is not the organisational behaviour construct of "personal initiative" (Frese et al., 1996), which describes individual proactive behaviour in workplace settings. Personal initiative is a property of individuals. The initiative this paper specifies is a property of the agent-substrate system — closer to Hutchins's (1995) distributed cognition than to individual psychology.

---

2. Three Conditions for Initiative

Governed initiative — initiative that is auditable, correctable, and institutionally legible — requires three formal conditions. This paper does not claim that governed architecture is the only path to initiative-like behaviour. It claims that governed architecture is necessary for institutionally legible initiative in high-consequence environments. The conditions are stated here independent of any particular institution or technology, then tested against an established institutional analogue. The situated learning literature (Lave and Wenger, 1991; Wenger, 1998) has long established that immersion matters. What it has not provided is a formal specification of what the exposure architecture must satisfy. That specification is the contribution.

Condition 1: Fertile pattern density above perceptual threshold

Exposure must compound — and the compounding must be generative, not merely cumulative. Each layer of accumulated pattern must enable recognition of patterns invisible without that layer as substrate. This is Cohen and Levinthal's (1990) absorptive capacity formalised as a structural requirement: prior related knowledge determines what new knowledge the system can perceive.

But density alone is not sufficient. The substrate must also be fertile in Paper C's precise sense: generating new patterns through operation, not just accumulating existing ones. The distinction matters. A deep-but-dead substrate — extensive, well-governed, but static — might cross the perceptual threshold for historical pattern recognition. The agent would recognise what mattered last quarter. A fertile substrate crosses the threshold for what is emerging now, because the substrate itself produces new connections, new cross-domain edges, new pattern combinations through ongoing operation. Paper C's six fertility dimensions (knowledge flow rate, contextual richness, pattern diversity, temporal depth, integration maturity, governance completeness) are the measurable properties that distinguish dead density from living density.

The formal requirement: the system's perceptual capacity P at time t is a function of accumulated exposure E(t) and the substrate's generative rate G(t), where P(E, G) exhibits threshold behaviour — below a critical density E*, novel patterns in the domain are unrecognisable regardless of processing capability. Above E*, each additional exposure increment enables recognition of previously invisible patterns at a rate that increases with both density and generativity. E* is not proposed as a universal constant. It is an empirically estimable threshold for a given domain and action class, observable when increases in substrate depth begin to produce discontinuous improvements in pattern recognition, escalation quality, or action-opportunity detection. A substrate above E* with $G \approx 0$ produces historical initiative (recognising patterns that existed). A substrate above E* with $G > 0$ produces living initiative (recognising patterns that are forming).

This is not a learning curve. It is a perceptual threshold with a fertility multiplier. The partner sees what the junior cannot not because she processes faster but because her accumulated, living substrate — continuously enriched through ongoing practice — supports pattern recognition the junior's substrate cannot yet sustain.

Condition 2: Constraint-legible action space

The system must operate within boundaries that encode organisational values, and those boundaries must be transparent — the system can trace why a constraint exists, what it protects, and what violating it would cause. Opaque rules produce compliance. Legible constraints produce judgment.

The formal requirement: for every governance constraint C, the system has access to a causal model M(C) that traces C to its institutional rationale — the value, risk, or obligation it protects. Action selection within the constraint space is therefore informed by the constraint's purpose, not merely by its boundary.

A system that perceives the causal consequences of violation within its world model — rather than merely checking a rule — produces judgment, not compliance. And only judgment scales to novel situations where the rule set is incomplete, which is to say, every situation of consequence in a regulated environment.

Condition 3: Progressive governance relocation

The operational locus of governance must shift as the substrate deepens. At early exposure, governance is external — explicit rule checks, human oversight, hard boundaries enforced before every action. As the institutional world model deepens, the causal rationale those constraints were protecting becomes structurally represented in the substrate itself. The agent's reasoning naturally traverses the consequence structure that governance was encoding. Governance does not disappear. Its enforcement migrates from synchronous pre-action gating to the substrate's own causal architecture, with explicit checks moving from universal requirement to risk-based monitoring and audit sampling.

This is not internalisation in the psychological sense — the model's weights do not change. It is governance relocation: the constraint's causal rationale moves from an external lookup to a substrate-resident structure that the agent's inference process encounters during normal reasoning. The agent does not become dispositionally different. It operates over a richer external substrate where governance rationale is structurally accessible at the right point in the reasoning chain.

The formal requirement: for a given action class A, there exists a substrate depth D*(A) beyond which the agent's reasoning over the institutional world model reliably traverses the causal structure that governance constraints for A were protecting, with probability above a threshold τ. Below D*(A), explicit pre-action governance checking is necessary. Above D*(A), governance can shift to asynchronous verification and periodic audit sampling for A because the substrate makes the constraint's causal rationale available during reasoning. D*(A) is proposed as an empirically estimable threshold, not a universal constant — it is reached when governance intervention rates for action class A decline while decision quality, auditability, and outcome alignment remain stable or improve.

The relocation is measurable: the rate of governance interventions (overrides, vetoes, corrections) for action class A should decrease as substrate depth for A increases past D*(A). If it does not — or if rates decline only because the agent suppresses action rather than exercises better judgment — the substrate is insufficient for governance relocation in that action class.

The institutional analogue: graduated immersion systems

Several institutional traditions reliably produce initiative under governance at organizational scale: medical residency, military officer development, judicial clerkships, doctoral apprenticeship, and professional consulting pyramids. These are graduated immersion systems — institutional structures that produce judgment through progressive exposure under structured control. The consulting pyramid is the focal example here because it is especially relevant to enterprise advisory work, but it belongs to a broader family. The convergence of multiple independent institutions on the same structural pattern strengthens the case that the three conditions capture something real. The consulting pyramid satisfies all three conditions.

Condition 1 — fertile pattern density: Exposure accumulation follows a logarithmic curve. Year ten's patterns are invisible without years one through nine as substrate. The partner's perceptual capacity is a function of cumulative exposure, with threshold behaviour visible at approximately the senior manager / junior director level — the point at which professionals begin to perceive patterns that span departments, client contexts, and regulatory domains. The substrate is fertile because each engagement produces new connections, not just more data — the partner's accumulated experience generates new pattern combinations through each new context it encounters.

Condition 2 — constraint legibility: Professionals do not just follow rules — they understand why the rules exist. The control network (review structures, quality gates, escalation paths) encodes organisational values transparently. A senior consultant does not merely avoid a regulatory breach; she perceives the causal chain that makes it a breach — client obligation, regulatory requirement, enforcement consequence, reputational impact. The constraints are legible because the institutional structure makes their rationale accessible.

Condition 3 — progressive governance relocation: The junior is on a short leash. The partner operates with almost no external constraints — but this is not because she has fewer rules. It is because the causal rationale those rules encode has become structurally accessible through fifteen years of immersed practice within a graduated control network. The control network does not disappear. Its operational locus migrates from external checkpoints to judgment exercised through accumulated domain structure. That migration is the autonomy-to-initiative transition.

The consulting pyramid does real explanatory work here — it demonstrates that the three conditions are jointly satisfied in a well-documented institutional context that reliably produces initiative at scale. The same structural pattern appears in medical residency (dense clinical exposure, legible institutional constraints, graduated reduction of supervision as judgment develops) and military officer development (accumulated operational depth, transparent chain-of-command rationale, progressive command authority). Without institutional analogues, the conditions are a theoretical specification. With convergent evidence from independent traditions, they are a specification supported by decades of organisational evidence. The question becomes: can the same conditions be satisfied computationally?

Differentiation from adjacent traditions

The three conditions draw on but are not reducible to established frameworks. Situated learning (Lave and Wenger, 1991) shares the emphasis on immersion over instruction — learning through participation rather than transfer. But situated learning has no formal specification of what the exposure architecture must satisfy, no governance dimension, and no computational operationalisation. The three conditions are what situated learning describes but does not specify.

Absorptive capacity (Cohen and Levinthal, 1990) is formalised in Condition 1 as a structural requirement with threshold behaviour. But Cohen and Levinthal's framework addresses knowledge recognition and integration — not the constraint-legibility (Condition 2) or governance relocation (Condition 3) required for aligned initiative.

Communities of practice (Wenger, 1998) describe the social infrastructure of legitimate peripheral participation. The three conditions specify the formal properties that infrastructure must possess — abstracting from the specifically human social mechanisms to conditions any system, human or computational, must satisfy.

---

3. The Compressed Pyramid — and the Governance Relocation Mechanism

For AI agents, each condition transforms in its implementation — but the functional requirements are identical. The critical question is Condition 3: how does governance relocation work computationally?

Conditions 1 and 2 have relatively straightforward computational analogues. Condition 3 — the operational locus of governance shifting from external rule enforcement to substrate-mediated reasoning — is the hard problem. In humans, this happens through neuroplasticity and years of embodied practice. In AI agents, a different mechanism is required. That mechanism is the paper's central engineering claim.

Condition 1 compressed: substrate depth replaces experiential accumulation

An agent operating on a mature governed knowledge graph has the equivalent structural depth — not the same experience, but the same pattern density — that a fifteen-year practitioner carries. The compression is possible because the substrate is computationally traversable. A human partner accumulated pattern density through sequential exposure across hundreds of engagements. The agent traverses a graph that encodes the distilled output of those engagements — validated claims, causal edges, cross-domain connections, confidence distributions — in seconds.

The question is not "how fast" but "how deep." Depth of the substrate determines whether the perceptual threshold E* is crossed for a given action class. Depth is measurable: number of validated claims, edge density between claims, confidence distribution, domain coverage, and — critically — generative rate. A domain graph with 50 validated claims about settlement penalty mechanics does not cross E* for penalty-related initiative. A graph with 5,000 claims, richly connected, with cross-domain edges to collateral management and regulatory reporting, plausibly does. A graph that generates new connections through each engagement that operates on it satisfies the fertility dimension that distinguishes living density from dead weight.

Condition 2 compressed: governance-as-causal-model replaces institutional socialisation

Paper D's four governance authorities (domain, regulatory, operational, technical) make constraints legible by construction. Each governance constraint in the system carries its causal rationale — not just "this is forbidden" but "this is forbidden because violating it causes X, which triggers Y, which breaches Z." The neuro-symbolic architecture specified in Paper B provides the structural basis: a deterministic constraint layer as shell, probabilistic reasoning within it as pearl. The constraints encode what the organisation values; the agent can perceive through them because it can trace their causal structure within the governed knowledge graph.

A human consultant learns constraint legibility through years of institutional socialisation — watching senior colleagues, absorbing the reasons behind firm policies, absorbing the logic of quality gates. The governed architecture provides the same legibility through structure: the rationale is encoded in the graph, not hidden in institutional culture.

Condition 3 compressed: the governance relocation mechanism

This is the section that must answer the question: how does governance migrate from external rule enforcement to substrate-mediated reasoning?

The mechanism is not weight updating. The foundation model's weights do not change. Governance relocation is not fine-tuning. It is a structural property of the institutional world model's depth and the agent's inference process operating on that depth. The agent does not become dispositionally different — it operates over a progressively richer substrate where governance rationale is structurally accessible during reasoning. Three steps describe the migration:

Step 1: Governance as explicit constraint. At low substrate depth (early lifecycle, sparse graph), the agent's action selection is governed by explicit constraint checking. Before executing action A, the system queries the governance layer: "Does A violate any active constraint?" The governance layer is external — a lookup, not a perception. The agent complies without the substrate providing the causal rationale for compliance.

Step 2: Governance as substrate structure. As the institutional world model deepens (more claims, richer edges, causal chains between actions and consequences), the substrate increasingly encodes the same information the governance constraints protect. The agent's reasoning about "what happens if I do A" now traverses causal consequences that governance was guarding against — the governed knowledge graph contains the consequence structure that makes those outcomes visible through normal inference. The agent encounters the constraint's rationale during reasoning, because the substrate makes the consequences structurally accessible.

Step 3: Governance as substrate-mediated alignment. At sufficient depth $D^*(A)$, the agent's policy selection for action class A reliably avoids governance violations — because the institutional world model makes the consequences of violation as salient during reasoning as the consequences of the action itself. The agent does not avoid the violation because a rule says no. It avoids it because its reasoning — traversing the governed substrate — encounters the violation's downstream consequences as predicted undesirable states. Governance enforcement can shift for A because the substrate provides the causal rationale that explicit checks were encoding.

What "shift" means operationally is not constraint removal. It is relocation of the enforcement locus:

At depth below D*(A): explicit constraint check required before action. Governance is synchronous overhead. At depth near D*(A): constraint check runs in parallel but rarely triggers. Governance is asynchronous verification. At depth above D*(A): constraint check frequency can decrease to audit sampling rather than per-action checking. Governance is periodic assurance.

The transition is gradual, per-action-class, and reversible. If the substrate degrades — claims decay, edges become stale, the epistemic immunity mechanisms described in Paper D fail — governance must re-tighten for affected action classes. The decay monitoring built into the Governed Intelligence Lifecycle (Paper D, Curate stage) is the mechanism that detects when governance relocation is eroding.

The relocation rate for action class A is measurable: the ratio of governance interventions to actions over time, plotted against substrate depth for A. A declining curve with consistent substrate depth growth and stable or improving decision quality indicates governance relocation. A flat or rising curve indicates the substrate is insufficient for A — either too shallow, too sparse, or not causally connected enough to support the reasoning shift that Condition 3 requires. A declining curve with declining decision quality indicates weakening governance, not relocation — an important distinction.

---

4. The Substrate Gap

Two independent research programmes have each specified half the substrate — and neither alone satisfies the three conditions for initiative in regulated industries.

The world model community (Pavlyshyn, 2026; LeCun, 2022; Ha and Schmidhuber, 2018) has specified what agents need computationally: memory taxonomies, causal models, temporal reasoning, social coordination protocols, graph-based world models. Chu et al. (2026) provide the definitive survey of this landscape, synthesising over 400 works into a capability taxonomy — L1 Predictor (one-step transitions), L2 Simulator (multi-step rollout respecting domain laws), L3 Evolver (evidence-driven model revision) — across four governing-law regimes: physical, digital, social, and scientific. Their framework is comprehensive within its scope, but it does not address the organisational regime where governing laws are institutional, regulatory, and emergent rather than physical, programmatic, or empirically discoverable. The gap is not an oversight — it reflects the field's orientation toward environments where transition dynamics are either analytically characterisable or experimentally measurable. Enterprise regulated environments are neither. This is the substrate for Condition 1 — pattern density. The governance community (compliance vendors, regulatory frameworks like the EU AI Act, DORA, MiFID II) has specified what regulated industries need: documentation, traceability, human oversight, audit trails. This is the substrate for Condition 2 — constraint legibility.

Neither has addressed what happens at the intersection. Both gaps are the same gap viewed from opposite directions.

World models without governance satisfy Condition 1 (pattern density) but cannot satisfy Conditions 2 or 3. The agent develops perception but has no mechanism for aligning that perception with organisational purpose. Active inference on a shallow, ungoverned world model produces agents that reduce their own prediction error — not agents that perceive what the organisation needs. Without governance constraints encoding organisational values (Condition 2), the agent's initiative is self-directed, not organisation-directed. Without a governance structure whose enforcement locus can relocate (Condition 3), there is nothing to migrate from external rule checking to substrate-mediated reasoning. The result is autonomous agents that are perceptive but unaligned — they see patterns but have no mechanism for determining which patterns matter to the organisation.

Governance without world models partially satisfies Condition 2 (constraints exist) but at the wrong level — constraints remain opaque rules rather than causal models the agent can trace. And Conditions 1 and 3 are impossible without a world model: no substrate for pattern accumulation (Condition 1), no depth-dependent governance migration (Condition 3). Static compliance produces constraint without the perception that makes constraint productive. The governance functions as overhead friction rather than the constitutive kind — compliance without judgment.

The gap is structural. Both programmes are building half of the same architecture. The substrate that produces initiative requires governed world models — world models deep enough for perception (Condition 1), governed transparently enough for judgment (Condition 2), with depth-dependent governance relocation (Condition 3). Neither programme alone can produce initiative because initiative requires all three conditions simultaneously.

This gap has a practical consequence. An organisation that invests in world models without governance builds perceptive agents that drift from organisational purpose. An organisation that invests in governance without world models builds compliant agents that cannot perceive. Both are stuck in the Powerful Tools quadrant of Paper C's capability/fertility matrix. The Living Medium quadrant requires both halves simultaneously.

---

5. Governed World Models — The Architecture

Six specific architectural enrichments from the world model literature, applied to the Governed Intelligence Lifecycle (Paper D), satisfy the three conditions for compressed exposure within governance constraints. Each enrichment fills a gap that neither programme addresses alone. Together they specify the governed world model — the substrate on which the three conditions are met and initiative can develop.

Pattern density and its temporal integrity

5.1 Temporal interval reasoning (Curate stage). Allen's (1983) interval algebra provides formal tools for reasoning about overlapping validity windows. Pattern density is only as good as the patterns' temporal integrity. A compliance claim's validity is bounded by the intersection of the regulation's effective period, the assessment's validity window, and the operational state's currency. Without temporal reasoning, accumulated patterns silently expire — the substrate rots beneath the agent's perceptual apparatus. This gives the Curate stage formal tools for evaluating when knowledge dies, preserving Condition 1's integrity as the graph grows.

5.2 Graph delta detection (ΔG) as curation trigger. Change is the curation signal. Instead of periodic full-scan review, the system detects graph deltas and triggers re-validation of dependent claims. This makes the substrate self-maintaining — preserving Condition 1 by ensuring pattern density is genuine (validated, current) rather than nominal (large but stale). Without ΔG, the substrate inflates without deepening. Condition 1 requires real density, not dead weight. The epistemic immunity framework in Paper D — with its six failure modes including staleness, pollution, and fragmentation — provides the threat model that ΔG operationally addresses.

Making constraints causally transparent

5.3 Causal reasoning (Expand stage). Pearl's (2000) Level 2 (intervention) and Level 3 (counterfactual) reasoning give the agent the inference machinery to understand why constraints exist, converting compliance into judgment. Constraint legibility requires the agent to trace from constraint to rationale: "What happens to our compliance posture if this regulation changes?" is a causal question, not a retrieval question. Without causal reasoning, governance constraints remain opaque rules — satisfying Condition 2's boundary requirement but not its legibility requirement.

5.4 Metagraph-aware relationship modelling. Compliance chains (regulation → control → risk → asset) are relationships between relationships. Condition 2 requires the agent to trace constraints to their institutional rationale — but those rationale chains often cross levels of abstraction. First-class representation of meta-relationships (implemented via labelled property graphs with reification patterns) makes the constraint space navigable at the structural level. Without this, the agent can look up individual constraints but cannot trace the institutional reasoning that connects them.

The governance relocation substrate

5.5 Memory taxonomy underneath the lifecycle. Pavlyshyn's (2026) episodic/semantic/procedural/working memory taxonomy maps heuristically onto the lifecycle stages. Ingest corresponds to episodic acquisition. Consolidate corresponds to semantic structuring. Curate corresponds to memory consolidation. Expand corresponds to procedural and inferential extension. Apply corresponds to working memory at the point of decision.

This mapping is suggestive rather than rigorous — Pavlyshyn's taxonomy is practitioner-oriented rather than grounded in the established cognitive-science memory literature (Tulving, 1972; Squire, 2004), and the correspondence requires empirical validation that the lifecycle stages produce the same functional outcomes as the memory types they parallel. Its significance for Condition 3 is that governance relocation is not uniform. Different types of governance knowledge relocate through different substrate mechanisms. Procedural relocation (how to act within constraints) differs from semantic relocation (what the constraints protect). The memory taxonomy gives the architecture differentiated pathways for the migration from external governance enforcement to substrate-mediated reasoning.

5.6 Working memory for the Apply stage. The lifecycle governs the long-term knowledge base but is silent about what is in the agent's active processing context when a decision is being made. What claims are loaded? In what priority? At what freshness? Working memory design determines the quality of initiative at the point of action — specifically, it determines whether the agent's decision context includes enough causal structure for Condition 3 to hold at the moment of action selection. A well-designed working memory that loads governance rationale alongside operational context enables the agent to perceive constraints as structure rather than rules, even at the Apply moment.

---

6. The Domain Graph — The Third Tier

The world model literature operates with two knowledge tiers: the foundation model (broad, general prior) and the application context (narrow, specific posterior). The foundation model provides broad pattern density at low domain specificity. Application-layer context provides high specificity at low pattern density. Neither crosses the perceptual threshold E* for the kinds of domain-specific initiative that regulated industries require.

A third tier sits between them — the missing middle layer: the domain graph — accumulated, governed, reusable domain substrate that no single client owns but every client needs. Foundation models know too broadly; application context knows too locally. The domain graph supplies the governed middle layer where institutional initiative becomes possible.

The domain graph is a governed knowledge structure satisfying four properties:

Scope. Domain-specific knowledge between public domain (regulations, standards, published practices) and client-walled (engagement-specific operational knowledge). Examples: how SWIFT message types relate to settlement workflows, how Basel III capital requirements cascade through product hierarchies, how operational workarounds interact with control frameworks. This is knowledge that any practitioner in the domain carries but that no current AI system possesses as structured, validated, traversable intelligence.

Provenance. Every claim in the domain graph carries source, validation status, confidence tier, temporal validity, and named authority. It is governed L3 knowledge in the three-layer maturity model specified in Paper D — not accumulated reference material.

Compounding. Each engagement that operates on the domain graph enriches it. The graph deepens through use — gaps are identified, claims are validated against operational reality, cross-domain edges are discovered. This is Paper C's definition of fertility: "the organisational rate at which existing knowledge generates new knowledge, new capability, and new possibilities." The domain graph grows by generation — each use producing new connections, new pattern combinations, new adjacent possibilities (Kauffman, 2019) — rather than by accumulation alone.

Decay management. Different knowledge types within the domain graph decay at different rates. Regulatory interpretation has a half-life of weeks to months. Domain mechanics — how settlement actually works — decay over years. Operational workarounds decay in days to weeks. The Curate process in Paper D's lifecycle runs different maintenance cadences per decay class, and the epistemic immunity framework monitors for the six failure modes that would compromise the graph's integrity.

The domain graph matters for the three conditions in specific ways.

For Condition 1, the domain graph is the substrate that crosses E* — and its compounding property is fertility in action. Without it, the agent has either broad-but-shallow knowledge (foundation model, high density but zero fertility at the domain level) or narrow-but-deep knowledge (client context, moderate density but no cross-engagement compounding). Neither satisfies Condition 1's dual requirement: density and generativity at the domain level where initiative matters.

For Condition 2, the domain graph's governance structure makes the constraint space legible at the domain level — not just "this regulation exists" but "this regulation constrains this process in this way, as validated by these experts, with these known exceptions."

For Condition 3, the domain graph is the substrate over which governance relocates. As the graph deepens and generates new connections through operation, the agent's reasoning inherits its evolving causal structure, enabling the governance relocation mechanism described in Section 3.

The competitive implication is worth noting. Capability commoditises as foundation models converge. Autonomy commoditises as agentic frameworks proliferate. The domain graph does not commoditise because it compounds through governed operation — each engagement enriches the shared substrate, each validation deepens the causal structure, each contradiction resolved adds institutional reasoning the substrate did not previously contain. A competitor can replicate the schema. They cannot purchase the generative history.

---

7. From Active Inference to Initiative — and the Scale Objection

Active inference (Friston, 2010) provides a normative model for describing why agents operating on deep, governed institutional world models may move from reactive execution to anticipatory action selection — but only when the world model satisfies the three conditions. The paper does not require production LLM agents to literally implement variational free-energy minimisation. Rather, active inference is used as a formal vocabulary for modelling the transition from assigned-task execution to anticipatory policy selection under a governed world model. This section establishes the connection and confronts the strongest counterargument.

Part 1: The active inference model

The first objection is legitimate: an agent minimising variational free energy might minimise surprise by doing nothing or by looping trivial safe actions. Why would it produce initiative — proactive, organisation-aligned action?

Because the governed world model encodes organisational dynamics as ongoing processes. The model predicts that the organisation continues to operate — regulations change, claims decay, dependencies shift, workflows execute. Passivity produces mounting prediction error because the world model predicts change and the agent observes itself not responding to it. Active inference selects policies that minimise expected free energy, combining epistemic value (how much will I learn?) and pragmatic value (how close will I get to preferred states?). The governed world model defines preferred states as states where organisational processes function within governance constraints. An agent that does nothing watches its prediction error grow as the world moves. An agent that acts on the governed model reduces prediction error by maintaining alignment with the organisation's trajectory.

This raises the second objection: if preferences are encoded in the world model, the agent is following programmed goals — instruction by another name. But the preferences emerged from the substrate. The governed world model contains the causal structure — thousands of validated, connected claims — from which the agent's inference process derives that regulatory compliance is the prediction-error-minimising policy in situation X. The agent perceives regulatory compliance as the right move because its world model, shaped by governed exposure, makes the consequences of non-compliance salient at the point of decision. The difference surfaces in novel situations: an instructed agent fails when the rule set is incomplete — and in regulated environments, the rule set is always incomplete. An immersed agent generalises because its causal model extends beyond the boundary of any finite rule set, carrying the institutional reasoning that the rules were encoding.

The quality of initiative then scales with the depth of the governed world model. On a shallow world model, active inference produces shallow initiative — the agent can anticipate immediate consequences but cannot trace second- or third-order effects. On a deep, governed world model, it produces the kind of initiative the consulting partner exhibits: perception of consequences that span departments, timeframes, and regulatory domains. The agent's inference horizon is bounded by its world model's causal reach.

This completes the pyramid parallel. The partner who walks into the room and perceives the real problem is running active inference on a deep world model built through fifteen years of immersive exposure within a graduated control network. The governed intelligence lifecycle replicates this computationally: deep substrate (the lifecycle's five stages fulfilling Condition 1), legible constraints (the four governance authorities fulfilling Condition 2), progressive governance relocation (the enforcement-locus migration mechanism fulfilling Condition 3), and active inference as the normative model describing how all three convert into initiative.

Part 2: The scale objection

The strongest counterargument this paper must answer: initiative does not require governed architecture. Sufficiently large models, fine-tuned on organisational corpora, embedded in tool-rich workflows, may acquire enough latent pattern density to anticipate what matters, escalate appropriately, and propose aligned next actions. The governed substrate is one path to initiative — not the only one, and possibly not the primary one.

The response is not a dismissal but a narrowing. This paper does not claim that governed architecture is the only conceivable path to initiative-like behaviour. It claims that in regulated industries, governed architecture is the only path to initiative that is simultaneously:

Auditable. When the agent takes initiative, the reasoning chain is traceable through the governed graph: this action was selected because these claims, with this provenance and confidence, predicted these consequences. Scale-based initiative distributes the reasoning across billions of parameters — the action may be correct, but the audit trail is "the model predicted this." Regulators do not accept that. The EU AI Act's high-risk system requirements — phasing in across 2025-2027, with many obligations becoming applicable from August 2, 2026 and some embedded high-risk systems from August 2, 2027 — demand risk management systems, technical documentation, and human oversight that are architectural properties, not post-hoc documentation exercises.

Correctable. When initiative drifts — the agent begins pursuing action opportunities that diverge from organisational purpose — the governed substrate provides a diagnosis: which claims are stale, which edges have decayed, which governance constraints lack sufficient substrate depth for relocation. Scale-based initiative drifts opaquely. Correction requires retraining, not re-curation. The three-step governance relocation mechanism (Section 3) provides a diagnostic framework that scale-based approaches lack entirely.

Governable. The three conditions provide a formal framework for assessing whether initiative is developing appropriately. Condition 3's measurability criterion (governance intervention rate as a function of substrate depth) gives organisations an ongoing assurance mechanism. Scale-based initiative provides no equivalent — there is no substrate depth to measure, no governance migration to track, no per-action-class assessment of governance relocation.

The narrowed claim: in unregulated, low-consequence settings, scale alone may produce functionally adequate initiative. This paper's claim is scoped to environments where initiative must be institutionally legible — where the organisation needs to know not just that the agent acted well, but why it acted that way, and how to intervene when it does not. That describes every regulated industry. It also describes every enterprise deployment where agent actions carry material consequences.

The deeper point: the scale objection assumes initiative is a unitary phenomenon — either the agent has it or it does not. The three conditions suggest otherwise. Initiative exists on a spectrum defined by substrate depth, constraint legibility, and governance relocation degree. Scale may produce initiative-like behaviour at the shallow end — correct actions without traceable reasoning. Governed architecture produces initiative at the deep end — correct actions with institutional legibility. The claim is that the deep end is where regulated industries need to operate, and scale alone cannot get there.

---

8. Implications for Architecture and Theory

The three conditions for initiative have implications beyond the specific architecture proposed — they constitute a testable framework for evaluating any agentic system's capacity for organisationally aligned action.

If the three conditions are correct, they provide diagnostic criteria independent of implementation choice. Any architecture that satisfies Conditions 1-3 should produce initiative. Any architecture that fails one or more should not, regardless of capability level. The conditions are the contribution; the specific architecture is one implementation.

Theoretical implications

World model research has specified the substrate for perception but not the conditions under which perception aligns with organisational purpose. The most comprehensive mapping of this field — Chu et al.'s (2026) survey of agentic world modelling across physical, digital, social, and scientific domains — confirms the pattern: the entire research agenda is oriented toward environments where the governing laws are known or discoverable through experimentation. In regulated enterprise environments, the governing laws are institutional — created by human convention, enforced through compliance structures, and subject to political revision. This fifth regime is where the governed institutional world model operates, and where the three conditions for initiative must be satisfied. The governed institutional world model is the enterprise instantiation — world model depth embedded in a governance architecture that satisfies Conditions 2 and 3 alongside Condition 1.

The governance community faces the inverse problem. Compliance frameworks without world models partially satisfy Condition 2 but at the wrong level — opaque rules rather than causal models. Conditions 1 and 3 are architecturally impossible without a world model substrate. The EU AI Act's high-risk system requirements point toward a structural resolution — requiring risk management systems, technical documentation, and human oversight that are architectural properties, not documentation exercises. Governed world models satisfy these requirements by construction because the governance is the architecture.

Active inference, meanwhile, provides a normative model for what organisationally aligned anticipatory behaviour looks like, but lacks specification of the substrate it requires for enterprise-grade initiative. On shallow or ungoverned world models, it describes self-directed prediction error reduction — autonomy. The three conditions specify what the institutional world model must satisfy for the behaviour active inference describes to become organisationally aligned, connecting the mathematical framework (variational free energy, expected free energy, policy selection) to the institutional context it operates in. The gap between the normative model and production engineering remains — current LLM-based agents do not implement variational free-energy minimisation in any rigorous sense — but the framework identifies what the engineering must approximate.

Architectural implications

The domain graph (Section 6) represents the critical structural advantage. As Section 6 established, the domain graph compounds through governed operation — a competitor can replicate the schema but cannot purchase the generative history. The competitive moat is in the governed lifecycle that produces and maintains the substrate on which initiative depends, not in the world model technology itself.

Agent governance implications

The governed intelligence architecture (Paper D) specifies what the knowledge substrate must contain and how it is maintained. This paper adds the question of what agents may *do* with that substrate. Three governance layers are required, each with distinct controls.

First, **retrieval governance**: what claims may an agent access? Tier scoping and scope metadata (Paper D, Section 3) provide the mechanism — agents receive only the subgraph relevant to their task and authorization level.

Second, **reasoning governance**: what inferences may an agent draw? Neuro-symbolic reasoning constrains this — symbolic rules enforce consistency, prevent contradiction-crossing, and flag when reasoning chains include low-confidence premises. A conclusion cannot have higher confidence than its weakest necessary premise.

Third, **action governance**: what may an agent *do* based on what it knows? This is where confidence-to-action thresholds become critical. Agent action must be gated by claim confidence, source authority, scope match, and contradiction status:

- Low confidence or stale claim on critical path: block autonomous action, escalate to human review.

- Unresolved contradiction in reasoning chain: halt and surface the contradiction.
- Scope mismatch between claim and use context: refuse or flag.
- Foundational claim within declared scope: permit execution with audit trail.

These thresholds function as **epistemic circuit breakers** — halting reasoning or action when knowledge integrity conditions fall below threshold. In finance and engineering, circuit breakers are intuitive. Applied to epistemic infrastructure, they prevent agents from acting on defective knowledge with the same confidence they act on sound knowledge. In regulated workflows, insufficient epistemic confidence should block autonomous action rather than produce a best-effort answer — the system should fail closed, not fail open.

The more autonomous the consumer, the stronger the epistemic controls required. A human can compensate for ambiguity; an agent may not. The governance architecture must therefore distinguish what the system may *know* (what claims may enter or remain in the substrate), what the system may *say* (what claims may be presented to users, and with what caveats), and what the system may *do* (what claims may trigger autonomous action). A claim may be allowed in the graph for awareness but not allowed in an answer. It may be allowed in an answer but not allowed to trigger action. This three-layer distinction — knowing, saying, doing — is essential for regulated environments where the cost of misaligned action is high.

Limitations and scope

The three conditions are stated as necessary for organisationally aligned initiative. This paper provides one institutional analogue (the consulting pyramid, as representative of graduated immersion systems) and one architectural instantiation (the governed intelligence lifecycle). It does not claim these are the only analogue or the only instantiation. Other architectures that satisfy all three conditions should also produce initiative. The formal conditions are the contribution; the specific architecture is one implementation.

The paper also does not claim that all enterprise AI requires initiative. Many valuable applications operate well at the autonomy level — executing assigned tasks within defined parameters. Initiative is the target when the organisation needs agents that perceive what matters before being told, in domains where the cost of misaligned action is high enough to justify the substrate investment.

The governance relocation mechanism (Section 3) is specified at the architectural level. Empirical validation — demonstrating measurable governance intervention decline as substrate depth increases for specific action classes — is the necessary next step. The framework would be weakened if increasing substrate depth did not improve action-opportunity detection, if governance intervention rates declined only through action suppression rather than judgment improvement, or if the three conditions proved satisfiable without producing institutionally legible initiative. The formal conditions generate testable predictions, which is what working papers are for.

Future directions

Three research questions follow directly from this paper. First, empirical validation of the perceptual threshold E* and the governance relocation depth D*(A) across different action classes and domain types. These quantities are proposed as empirically estimable thresholds, not universal constants. Measurement would require controlled deployment of governed world models with instrumented governance layers — tracking intervention rates against substrate depth over time.

Second, the relationship between the domain graph's fertility rate and the quality of initiative. Paper C's six fertility dimensions provide the measurement framework. The hypothesis is that initiative quality correlates with fertility rate, not just substrate size — a prediction that distinguishes the governed architecture claim from the scale-alone alternative.

Third, the dual-mode attention architecture that this paper's analysis points toward but does not fully develop. If initiative requires both deep domain perception (focused analytical processing) and peripheral contextual awareness (ambient pattern detection across domain boundaries), the governed world model must support both modes simultaneously. This connects to McGilchrist's (2009, 2021) hemispheric analysis — the tenth tradition in Paper C's convergence — and to the observation that the consulting partner's initiative operates in both modes: deep analysis of the problem at hand and ambient awareness of what else is connected. Specifying the computational architecture for dual-mode attention on a governed knowledge graph is the subject of ongoing work.

References

Allen, J.F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11), 832-843.

Argyris, C. and Schön, D.A. (1978). *Organizational Learning: A Theory of Action Perspective*. Addison-Wesley.

Chu, M., Zhang, X.B., Lin, K.Q., Kong, L., Zhang, J., Tu, T., Ma, W., et al. (2026). Agentic world modeling: Foundations, capabilities, laws, and beyond. *arXiv preprint arXiv:2604.22748*.

Cohen, W.M. and Levinthal, D.A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128-152.

Frese, M., Kring, W., Soose, A. and Zempel, J. (1996). Personal initiative at work: Differences between East and West Germany. *Academy of Management Journal*, 39(1), 37-63.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127-138.

Gartner (2025). Predicts 2025: AI agents and their growing influence on enterprises.

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.

Ha, D. and Schmidhuber, J. (2018). World models. *arXiv preprint arXiv:1803.10122*.

Hutchins, E. (1995a). *Cognition in the Wild*. MIT Press.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265-288.

Kauffman, S. (2019). *A World Beyond Physics: The Emergence and Evolution of Life*. Oxford University Press.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K. and Kiverstein, J. (2018). The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface*, 15(138), 20170792.

Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

LeCun, Y. (2022). A path towards autonomous machine intelligence. *OpenReview*.

March, J.G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71-87.

Maturana, H.R. and Varela, F.J. (1980). *Autopoiesis and Cognition: The Realization of the Living*. Reidel.

McGilchrist, I. (2009). *The Master and His Emissary: The Divided Brain and the Making of the Western World*. Yale University Press.

McGilchrist, I. (2021). *The Matter with Things: Our Brains, Our Delusions, and the Unmaking of the World*. Perspectiva Press.

McKinsey (2025). The state of AI: How organizations are rewiring to capture value. *McKinsey Global Survey*.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pavlyshyn, A. (2026). *World Models for AI Agents*. Leanpub.

Squire, L.R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171-177.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.), *Organization of Memory* (pp. 381-403). Academic Press.

Precisely and Drexel University (2025). Data integrity trends and insights report.

Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences*. W.H. Freeman.

Prigogine, I. and Stengers, I. (1984). *Order out of Chaos: Man's New Dialogue with Nature*. Bantam Books.

Reichhart, W. and Gelas, A. (2026a). Dynamics Blindness: When AI Is Locally Correct and Globally Non-Compliant. SSRN Working Paper.

Reichhart, W. and Gelas, A. (2026c). Build the Medium: Why organizational intelligence is mechanism, not metaphor. SSRN Working Paper.

S&P Global (2025). AI and the enterprise: Market intelligence report.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

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