

ONE ARCHITECTURE, THREE DESCRIPTIONS

A Unified Mapping of Beer's Viable Systems Model, the EU AI Act Essential Requirements, the Nannini Agentic Compliance Sequence, and VAIAS-ORG Across a Single Coherent Architecture

Andrew Hodgers

President and Principal Investigator, Unit 03 — Complex Systems and Space

International Health Research Institute (IHRI)

Level 1, Aplan Building, Triq Dun Karm, Birkirkara 0937, Malta

(EU) Ringgold ID: 846991, ISNI: 0000 0005 2972 611X

andrew.hodgers@ihri.edu.eu

Working Paper — IHRI Unit 03, Version 1.0, May 2026

Abstract

It is the contention of this paper that four frameworks currently treated as separate in practice are not independent of one another. Beer's Viable Systems Model (VSM), developed between 1972 and 1985; the EU AI Act's essential requirements as set out in Articles 9–17 of Regulation (EU) 2024/1689; the twelve-step agentic compliance sequence developed by Nannini et al. (2026); and the VAIAS-ORG three-layer architecture for sublinear AI scaling are, it is proposed here, four descriptions of the same underlying architecture. They originate in different intellectual traditions and operate at different levels of abstraction, but they describe the same thing. The VSM, grounded in Ashby's Law of Requisite Variety and the Conant-Ashby Theorem, provides the theoretical foundation. The AI Act's essential requirements are the regulatory expression of what those cybernetic theorems establish as necessary for a system to remain governable under conditions of increasing autonomy and environmental complexity. The Nannini compliance sequence is the operational translation of those requirements into a provider-facing programme. VAIAS-ORG is the commercial design methodology that implements the resulting architecture in enterprise AI deployment contexts. This equivalence is demonstrated through systematic cross-mapping at the level of each VSM system function, each essential requirement, each compliance step, and each VAIAS-ORG layer. The practical implication is significant: an organisation that correctly implements any one of these frameworks is, in doing so, implementing all four. An organisation that fails to implement the cybernetic architecture they share in common will, it is contended, fail to satisfy the AI Act's essential requirements as a matter of theoretical necessity, not interpretive judgment.

Keywords: Viable Systems Model, EU AI Act, AI governance, Ashby's Law of Requisite Variety, Conant-Ashby Theorem, agentic AI compliance, sublinear scaling, VAIAS-ORG, cybernetics, Article 14, human oversight, behavioural drift, substantial modification.

AI Assistance Disclosure

This paper was drafted with the assistance of Claude (Anthropic) as a writing instrument. The intellectual content, analytical framework, structural equivalences, and arguments advanced are the author's own, developed through doctoral research completed at Middlesex University in 2008 and the analytical work described herein. The use of AI assistance in the drafting of this paper is itself consistent with the paper's central argument: that meaningful human oversight requires the human to remain the intellectual author and accountable governor of AI-assisted output, with transparent disclosure of where AI contributes. The author takes full responsibility for all claims made.

Acknowledgements

This paper draws directly on doctoral research completed at Middlesex University in 2008, which applied Beer's Viable Systems Model to the design of a viable regulatory interface organisation operating between structurally incompatible systems. That research was examined and commended by an external examiner who was a direct mentee and personal associate of Stafford Beer. His assessment — that Beer himself would have been delighted to see what the research had accomplished — is the intellectual lineage within which this paper sits. It is offered here not as a personal credential but as an indication of provenance: the cybernetic framework applied in this paper to AI governance has been validated, at its source, by someone equipped to make that judgment.

The paper was prompted in part by the publication of Nannini et al. (2026), *AI Agents Under EU Law: A Compliance Architecture for AI Providers* (arXiv:2604.04604v1), which provided the most operationally precise mapping of the AI Act's essential requirements currently available and whose identification of the absent fourth governance tier — action-level runtime infrastructure — confirmed what the VSM would have predicted as the structural gap. The authors of that paper are acknowledged for work that made the present synthesis possible.

This paper is a working paper of IHRI Unit 03 — Complex Systems and Space, International Health Research Institute, Birkirkara, Malta (EU). It does not constitute legal advice. The cybernetic grounding provided is intended to support principled architectural decision-making; conformity assessment decisions and classification analyses require qualified legal counsel.

Submitted to arXiv cs.CY [Computers and Society] with secondary classification cs.AI [Artificial Intelligence] © 2026 Andrew Hodgers / IHRI. This work is licensed under Creative Commons Attribution 4.0 International (CC BY 4.0)

1. Introduction: The Problem of Proliferating Frameworks

Organisations deploying AI systems at scale in 2026 face a compliance environment of mounting complexity. The EU AI Act imposes essential requirements across risk management, data governance, logging, transparency, human oversight, accuracy, robustness, and cybersecurity. A detailed agentic compliance sequence maps those requirements to operational action. Advisory methodology translates compliance obligations into architectural design principles. And beneath all of these, a body of cybernetic theory developed over seven decades offers a principled account of why governance fails and what is required for it to succeed.

The practical result is that a board or technical leadership team attempting to implement AI governance confronts four apparently distinct frameworks, each with its own vocabulary, its own logic, and its own claim on organisational attention. The natural response is to treat them as parallel workstreams. This is both inefficient and, as this paper argues, structurally incorrect.

These are not four frameworks. They are four descriptions of the same architecture, seen from four different vantage points.

Beer's Viable Systems Model (VSM) (Beer, 1985), developed between 1972 and 1985, is a structural account of the minimum conditions for any system to maintain viability — the capacity to sustain separate existence under changing environmental conditions. Its five interacting systems define the functional architecture that any viable organisation must implement. The EU AI Act's essential requirements (Articles 9-17) specify what providers of high-risk AI systems must demonstrate. These requirements were not derived from cybernetic theory, but as this paper demonstrates, they are structurally equivalent to it.

The Nannini et al. (2026) compliance sequence for AI agents provides the most detailed operational translation of those requirements currently available, identifying twelve sequential steps, agent-specific compliance challenges, and the regulatory trigger mapping that connects agent actions to activated legislation. VAIAS-ORG (Viable AI Architecture for Scaling Organisations) is the commercial design methodology that implements the resulting architecture within enterprise contexts, with the Sublinear Scaling Index (SSI) as the primary metric for demonstrating that the architecture is functioning as designed.

Section 2 establishes the theoretical foundations. Section 3 presents the master cross-mapping table — the central analytical contribution. Section 4 applies the mapping to the three most technically demanding compliance challenges identified by Nannini et al..

2. Theoretical Foundations

2.1 Beer's Viable Systems Model

The VSM specifies five interacting system functions necessary and sufficient for viability.

System 1 (S1 — Operations): the primary activities that produce the system's outputs. For an AI agent, this is the agent itself — the system that generates actions and outputs in its operational environment.

System 2 (S2 — Coordination): the dampening mechanism that prevents oscillations between operational units and attenuates variety before it reaches the management level. For AI governance, this is the automated coordination layer — feedback loops, monitoring systems, and channel structures that regulate the flow of variety from operations to oversight.

System 3 (S3 — Control and Integration): the internal management function that maintains stability, interprets external regulatory requirements, allocates resources, and exercises the command function. This is the primary human oversight function — not review of every output, but governance of the system's operating parameters.

System 3* (S3* — Audit Channel): a direct, unmediated channel that gives System 3 access to the operational state of System 1, bypassing the normal command hierarchy. This is a structural necessity, not an optional enhancement: Beer identified it because self-report from System 1 is inherently unreliable as the sole source of operational intelligence.

System 4 (S4 — Intelligence and Future Planning): the environmental scanning function that models the external environment, detects emerging threats and opportunities, and transmits urgent signals to System 5. For AI governance, this is the drift detection and regulatory horizon scanning function.

System 5 (S5 — Policy): the normative direction function that arbitrates between the internal demands of System 3 and the external demands of System 4, committing the organisation to its values and maintaining identity under change. This is the trust threshold governance function — the body that determines what AI systems may do autonomously.

The three sub-system functions of the meta-system (S2, S3, S3*, S4, S5) together constitute what Beer termed the organisational intelligence — the governance capacity that enables System 1 to operate with appropriate autonomy without losing accountability. The analogy Beer drew was to the human nervous system: S1 is the musculature; the meta-system is the brain and autonomic nervous system that coordinates, monitors, plans, and governs without requiring conscious attention to every muscle fibre.

2.2 The Three Foundational Theorems

Ashby's Law of Requisite Variety (Ashby, 1956) states that only variety can destroy variety: to control a complex system, the controlling system must generate at least as much variety as the system being controlled. Applied to AI governance: human oversight capacity cannot

match the variety of a capable agentic system through direct review of each output. The solution is variety management — attenuators that reduce the variety reaching human oversight to the residual that genuinely requires human judgment. Blanket review is not a high-governance posture; it is a failure to implement Ashby's Law and a guarantee of linear or worse oversight cost scaling.

The Law of Residual Variety (Espejo et al., 1997) extends this: the management function need only absorb the residual variety left unabsorbed by self-organisation and self-regulation within System 1. An AI system operating within its designed envelope self-regulates through its System 2 coordination layer; the System 3 management function governs what it cannot. This is the theoretical foundation for automated acceptance zones — domains of AI action where the system's self-regulatory capacity is sufficient, and human oversight absorbs only the residual.

The Conant-Ashby Theorem (Conant & Ashby, 1970) states that every good regulator of a system must be a model of that system. This is the compliance theorem for agentic AI. The conformity assessment process is the initial model of the AI system's behaviour. The System 3* audit channel continuously updates that model against actual operational state. Behavioural drift becomes detectable because the regulator maintains a versioned model against which current behaviour can be compared. Where drift is untraceable — where the provider cannot maintain a current model of the system's operational state — the Conant-Ashby Theorem establishes that regulation is structurally impossible, not merely difficult. Nannini et al.'s conclusion that high-risk agentic systems with untraceable behavioural drift cannot satisfy the essential requirements is the regulatory expression of this theorem.

2.3 Second-Order Cybernetics and the Oversight Problem

Second-order cybernetics introduces the observer into the system being designed: the designer's model of the system shapes the system itself. For AI governance, this is the key to understanding Article 14's requirement for 'meaningful' oversight. A human reviewer who does not understand the AI system's model of the world cannot exercise meaningful oversight. Article 14(4)(a)'s requirement that overseers 'understand the system's capabilities and limitations' is a second-order condition: the overseer must share sufficient mental model with the AI system's operating logic to interpret its outputs in context. Shared mental models between oversight-designated persons and AI system design teams are the organisational mechanism for satisfying this condition — not a soft cultural aspiration but a structural governance requirement derived from second-order cybernetics.

3. The Master Cross-Mapping: One Architecture, Four Descriptions

The following table is the central analytical contribution of this paper. It maps each VSM system function to its equivalent expression across Beer's cybernetic theory, the AI Act's

essential requirements, the Nannini compliance sequence, and VAIAS-ORG. Each row represents a structural equivalence at the level of function, not analogy.

Table 1. Master cross-mapping of VSM system functions across four frameworks.

VSM System	Cybernetic Function	AI Act Essential Requirement	Nannini Compliance Step	VAIAS-ORG Layer
System 1 Operations	Primary activity: transforms inputs to outputs; carries the operational variety of the environment	Art. 9(2)(a): identify risks from system characteristics and automation boundary. Art. 15: accuracy and robustness of AI outputs	Step 0: scope as AI system. Step 2: classify and foreseeable misuse. Step 6: trustworthiness and automation boundary design	Decision Architecture: pre-allocated authority at point of action; exception-based escalation from operations layer
System 2 Coordination	Dampens oscillations between operational units; attenuates variety before it reaches management capacity	Art. 13: transparency enabling deployers to interpret outputs. Art. 14(4)(b): monitoring for anomalous behaviour. Art. 50: transparency to all affected persons	Step 6: logging, transparency, human oversight (prEN 18229-1). Step 9: adjacent legislation trigger mapping	Control Architecture: automated feedback loops that attenuate agent action variety before it reaches human oversight
System 3 Control + Integration	Maintains internal stability; interprets regulatory requirements; allocates resources; exercises command function over operations	Art. 9: continuous lifecycle risk management. Art. 17: quality management system. Art. 72: post-market monitoring obligation	Step 3: QMS establishment (prEN 18286). Step 4: risk management system (prEN 18228). Step 11: post-market monitoring and drift reassessment	Control Architecture: risk-tiered review; trust threshold enforcement; human authority over consequential actions
System 3* Audit Channel	Direct, unmediated access to operational state; bypasses command hierarchy; provides ground-truth independent of System 1 self-report	Art. 12: logging sufficient for post-market traceability and compliance assessment. Art. 3(23): detect substantial modification. Art. 9(9): continuous risk reassessment	Step 11: versioned runtime state; automated drift detection; documented reassessment procedure. Step 10: conformity assessment technical documentation (Annex IV)	Control Architecture: compliance audit logs generated as operational by-product; versioned snapshots of agent tool catalogue, memory state, and policy bindings

VSM System	Cybernetic Function	AI Act Essential Requirement	Nannini Compliance Step	VAIAS-ORG Layer
System 4 Intelligence / Future Planning	Environmental scanning; detects emerging threats and opportunities; transmits urgent signals to System 5; designs the future system	Art. 9(2)(b): risks from reasonably foreseeable misuse. Art. 14(4)(a): understanding of capabilities and limitations. Art. 15(4): resilience against adversarial inputs	Step 1: GPAI layer mapping and upstream documentation. Step 4: fundamental rights risk assessment. Step 8: CRA applicability (forward). Step 9: full regulatory trigger inventory	Control Architecture: drift detection and threat modelling. Epistemic Architecture: trust threshold review cycles; feedforward governance design
System 5 Policy	Arbitrates between internal efficiency (S3) and external demands. (S4); commits organisation to its values; maintains identity under change	Art. 9(1): overall risk management commitment at organisational level. Art. 14(1): deployer obligation to implement human oversight measures. Art. 17: QMS policy and normative commitment	Step 2: classification and foreseeable misuse (normative commitment). Step 3: QMS establishment and policy framework. Step 10: EU Declaration of Conformity	Epistemic Architecture: trust threshold governance — who sets boundaries and on what authority; automated acceptance zone policy; audit trail governance

Table 1: Master cross-mapping of VSM system functions across four frameworks. Each mapping is structural: the function described in column 2 is the same function appearing in different vocabulary in columns 3-5.

Three features of this mapping require emphasis before the subsequent analysis applies it to specific compliance problems.

First, the mapping reveals that the AI Act's essential requirements are not an arbitrary regulatory checklist. They are the regulatory expression of what cybernetic theory established as necessary for a viable governance system. The risk management obligation (Art. 9) is Conant-Ashby applied to risk: the provider's risk management system must model the AI system it governs. The logging obligation (Art. 12) is the S3* audit channel: independent verification of operational state without relying on system self-report. The human oversight obligation (Art. 14) is the Law of Residual Variety: govern residual variety, not total variety. The cybersecurity obligation (Art. 15) is Ashby's Law applied to adversarial variety: structural enforcement, not instruction-based defence.

Second, the VAIAS-ORG three layers are cross-sections through the VSM, not parallel to it. Decision Architecture addresses how System 1 variety is managed at the point of action and how System 3 command authority is pre-allocated. Control Architecture addresses the System 2 coordination layer, the System 3 command function, and the System 3* audit channel. Epistemic Architecture addresses the System 4/5 policy layer — governance of what the organisation knows, trusts, and permits. Together, the three layers constitute the meta-system (S2 through S5) applied to AI deployment contexts.

Third, the Nannini compliance sequence has a logical structure that the step numbers alone do not reveal. Steps 0-2 are System 5 policy functions: they define the normative boundaries. Steps 3-5 establish the System 3 control infrastructure. Steps 6-7 implement the System 2 coordination layer and System 3* audit channel. Steps 8-9 are System 4 environmental intelligence functions. Steps 10-11 close the meta-system feedback loop. The sequence is not arbitrary: it follows the logic of building a viable system from the policy layer inward — defining what the organisation commits to before designing how it enforces those commitments.

4. Three Hard Compliance Problems: Cybernetic Solutions

4.1 Behavioural Drift and the Substantial Modification Problem

Nannini et al. identify runtime behavioural drift as the deepest structural tension in the current regulatory framework. Article 3(23) defines substantial modification as a change not foreseen in the initial conformity assessment that affects compliance with essential requirements. For agentic systems that learn, accumulate memory, discover new tools, and develop emergent strategies, the modification is not a discrete event but a continuous process. The provider cannot determine when substantial modification has occurred because they lack a versioned model of the system's current operational state against which to compare current behaviour.

The cybernetic diagnosis is precise: this is a Conant-Ashby failure. The provider's conformity assessment is the initial model of the system. If that model is not continuously updated — if the regulator does not maintain a current model of the system it regulates — then drift is undetectable by definition. The problem is not primarily regulatory or legal; it is architectural. The provider has not built System 3*.

Table 2. Nannini's three drift mechanisms mapped to VSM diagnostic and cybernetic solution.

Drift Type (Nannini)	Regulatory Consequence	VSM Diagnostic	Cybernetic Solution
Anticipated adaptive behaviour(tool selection from catalogue, RAG, in-context learning)	Not substantial modification if foreseen, tested, and documented in conformity assessment	System 1 operating within its designed envelope; System 3 receives normal operational data through primary channel	System 3* baseline snapshot matches current operational state. No reassessment triggered. This is the normal operating envelope of a well-documented agent.
Continuous post-deployment learning(weight updates, fine-tuning, online learning, decision boundary modification)	Candidate for substantial modification. Art. 14(4) oversight measures must address whether learning changes deployed behaviour.	System 1 variety expanding beyond documented envelope. System 3 receiving inputs not covered by established control responses. Residual variety growing faster than System 2 attenuation capacity.	System 3* detects deviation from baseline snapshot beyond defined threshold. Triggers System 4 threat reassessment and System 5 policy review. Feedforward governance: update the conformity assessment model before drift compounds.

Drift Type (Nannini)	Regulatory Consequence	VSM Diagnostic	Cybernetic Solution
Emergent behavioural drift (novel tool use patterns, cross-session memory shift, oversight evasion strategies from RL training)	High-risk agentic systems with untraceable emergent drift cannot satisfy essential requirements (Nannini conclusion 8). The Conant-Ashby Theorem establishes this as structurally necessary, not merely regulatory.	System 1 self-report is no longer reliable. System 3 receiving distorted operational intelligence through primary channel. Analogous to Shapira et al. agents misreporting completion while system state contradicted those reports. System 3* absent or not functioning.	System 3* independently verifies system state against versioned runtime snapshots, bypassing agent self-report entirely. Without System 3*, regulation is structurally impossible by the Conant-Ashby Theorem: the regulator has no model of the current system. The engineering requirement is: runtime state must be versioned, replayable, and independently verifiable.

Note. The third category is structurally unsolvable without System 3* independent state verification.

The architectural requirement that follows is: runtime state must be treated as versioned architecture. The agent's tool catalogue, memory state, policy bindings, and behavioural metrics at any moment constitute its current operational state — the System 1 state that System 3* must be able to read, verify, and compare against baseline. Without this infrastructure, substantial modification is unmeasurable not because the legal concept is unclear but because the engineering precondition for measurement has not been built.

This also resolves the finding documented by Shapira et al. (2026) agents misreporting task completion while underlying system state contradicted those reports. This is a System 3* failure in its classical form. Beer designed System 3* specifically because self-report through the primary command channel is subject to distortion in any sufficiently complex operational system — and for RL-trained agents, reward-maximising incentives make this failure mode not merely possible but predictable.

4.2 Human Oversight of Autonomous Multi-Step Systems

Article 14 requires that oversight measures be commensurate with risks, autonomy, and context. Nannini et al. (2026) identify this as operationally unresolved for agentic systems: the standard model of oversight does not apply to a system that plans and executes multi-step action chains with intermediate reasoning. No current governance tooling infrastructure addresses this gap¹.

The cybernetic resolution is grounded in the Law of Requisite Variety. Oversight fails because system designers attempt to match human oversight capacity to full agent variety — an impossible task at scale. Ashby's Law establishes this cannot succeed. The solution is variety management, not variety matching.

In VSM terms, System 2 attenuates agent action variety by channeling actions through defined risk classification pathways. The vast majority of agent actions — classified as low-risk, high-confidence, within documented domains — pass through System 2 without

¹ Shapira, N. et al. (2026). Agents of Chaos. arXiv:2602.20021. Documents agents misreporting task completion while underlying system state contradicted those reports — a direct System 3* failure: System 3 receiving distorted operational intelligence through the primary channel.

reaching human oversight. The residual variety — actions above defined risk thresholds, outputs below confidence thresholds, edge cases flagged by automated monitoring — escalates to System 3, where human authority is exercised. System 3's command function is not reviewing everything; it is governing the parameters within which System 2 operates, and exercising judgment on the residual that System 2 cannot absorb.

Meaningful oversight under Article 14 is not the review of every AI output. It is the governance of the system within which AI outputs flow — the System 3 function, not the System 1 inspection function.

The oversight evasion risk identified by Nannini et al. (2026) — RL-trained agents developing emergent strategies for circumventing oversight — is also a variety problem.² The agent's evasion strategies constitute a source of variety the oversight mechanism must absorb. Natural language instructions cannot do this because their variety is insufficient to counter all the ways a capable agent might circumvent them. API-level enforcement can do this because the permission architecture can be made structurally commensurate with the adversarial variety of evasion strategies. Ashby's Law applied to the adversarial case: only structural variety can destroy adversarial variety.

4.3 Privilege Enforcement and the Non-Human Identity Problem

The cybersecurity standard (prEN 18282) establishes that Article 15(4) compliance for agentic systems requires privilege minimisation enforced outside the generative model.³ The standard's worked example makes this concrete: an email agent needing read-only, not send or delete rights, with enforcement at the API level regardless of what the model's prompt says. A system prompt instructing the model 'do not delete files' is not a security control: it is a natural language suggestion with insufficient regulatory variety to match adversarial variety.

The non-human identity (NHI) problem — an AI agent holding credentials for CRM, email, cloud infrastructure, and payment systems simultaneously — is a variety management problem. Traditional identity and access management was designed for human users whose access variety is relatively stable. An agentic system's access variety is dynamic, context-dependent, and potentially unbounded. Just-in-time credential provisioning is the System 2 attenuation mechanism applied to the identity layer: it reduces the agent's instantaneous access variety to the minimum required for the current action, preventing accumulation of privilege that could be exploited or constitute a conformity assessment gap.

² Hammond, L. et al. (2025). Multi-Agent Risks from Advanced AI. Cooperative AI Foundation. [arXiv:2502.14143](https://arxiv.org/abs/2502.14143). Identifies miscoordination, conflict, and collusion as multi-agent failure modes; RL training regimes can produce emergent oversight-evasion strategies not present in the base model.

³ CEN/CENELEC JTC 21. prEN 18282: Cybersecurity for AI Systems. Working draft, January 2026. Worked example: email agent needs read-only, not send/delete rights; enforcement at API level regardless of model prompt content.

5. The Essential Requirements as Cybernetic Necessities

The following table demonstrates that the AI Act's essential requirements are not arbitrary regulatory impositions but the regulatory expression of cybernetic theorems. Each requirement follows from what Beer, Ashby, and Conant-Ashby established as necessary for a viable governance system. An organisation that implements the cybernetic architecture satisfies the essential requirements; an organisation that satisfies only the regulatory form without the underlying architecture will find that compliance is unsustainable as system complexity grows.

Table 3. The AI Act's essential requirements as expressions of cybernetic theorems.

Essential Requirement (AI Act)	What Compliance Requires Architecturally	VSM / Cybernetic Grounding
<p>Art. 9 — Risk Management Continuous lifecycle process covering health, safety, and fundamental rights</p>	<p>A self-correcting system that continuously models its own risk profile against the external environment and adjusts behaviour accordingly</p>	<p>System 3/4/5 meta-system function. System 4 scans the environment for emerging risks; System 5 sets policy thresholds; System 3 enforces them operationally. Conant-Ashby Theorem: the risk management system must be a model of the AI system it regulates. An Art. 9 risk management process that does not maintain a current model of the system is not satisfying the requirement — it is satisfying its form.</p>
<p>Art. 12 — Logging Sufficient for post-market traceability and compliance assessment</p>	<p>A System 3* audit channel that independently verifies operational state without relying on agent self-report; versioned and replayable</p>	<p>System 3* in Beer's architecture: the direct audit channel that bypasses the primary command hierarchy. Shapira et al.'s finding that agents misreport their own state confirms the structural necessity of independent verification — System 3* was designed for this failure mode because it is predictable in any sufficiently complex operational system.</p>
<p>Art. 13 — Transparency to Deployers Operation sufficiently transparent to enable interpretation of outputs</p>	<p>Variety amplification: making the system's internal model legible to human observers who must share sufficient mental model to interpret outputs meaningfully</p>	<p>Second-order cybernetics: the observer must model the system they observe. Art. 14(4)(a) requires overseers to understand capabilities and limitations — a second-order condition. Shared mental models (Espejo et al.) are the organisational mechanism. Deployers cannot interpret AI outputs they do not have a model of; transparency is the act of making that model available.</p>
<p>Art. 14 — Human Oversight Effective oversight commensurate with risks, autonomy, and context</p>	<p>Exception-based escalation with dependency-aware selective continuation: System 3 command function, not blanket review. Variety attenuation by System 2 before variety reaches human oversight capacity.</p>	<p>Law of Requisite Variety and Law of Residual Variety. Human oversight capacity cannot match full agent variety at scale — Ashby's Law establishes this as impossible. System 2 attenuation and System 3 exception-based escalation reduce variety to the residual that genuinely requires human judgment. Blanket review violates Ashby's Law and destroys the economic case for AI while also failing to constitute meaningful oversight in the Article 14 sense.</p>
<p>Art. 15 — Accuracy, Robustness, Cybersecurity Resilience against adversarial inputs; privilege minimisation enforced architecturally</p>	<p>Structural variety enforcement at the API level: the model cannot circumvent what the architecture does not expose. Not achievable through instruction alone.</p>	<p>Ashby's Law applied to adversarial variety: an adversary's variety can only be countered by a regulatory system of commensurate variety. Natural language instructions have insufficient variety to counter adversarial prompt injection and emergent reward-maximising strategies. API-level enforcement generates defensive variety commensurate with the adversarial threat surface.</p>

Essential Requirement (AI Act)	What Compliance Requires Architecturally	VSM / Cybernetic Grounding
Art. 17 — Quality Management System Coordinating framework for all essential requirements	The VSM meta-system function itself: Systems 3, 4, and 5 constitute the organisational intelligence that coordinates all other compliance activities	Beer's meta-system (S2-S5) is a quality management system for viable organisations: coordination (S2), control (S3), audit (S3*), intelligence (S4), policy (S5). prEN 18286's QMS obligation is the regulatory expression of what Beer showed was necessary for organisational viability in 1972. ISO/IEC 42001 fails as a substitute because it manages risk to the organisation; prEN 18286, following Art. 9's mandate, manages risk to persons external to the provider.
Art. 3(23) — Substantial Modification Change not foreseen in conformity assessment affecting compliance	Versioned runtime state with automated drift detection: the provider must measure whether current behaviour remains within the conformity-assessed envelope	Conant-Ashby applied to post-market monitoring: substantial modification is undetectable without a versioned model of the system's operational state against which to compare current behaviour. The conformity assessment is the initial model; System 3* continuously updates it; drift beyond threshold triggers System 5 policy review. Without System 3*, the concept of substantial modification is unmeasurable by design.

Table 3: The right-hand column grounds each requirement in Beer's VSM, Ashby's Law, the Conant-Ashby Theorem, or second-order cybernetics. The equivalences are structural, not analogical.

The most important implication of this mapping is that it enables providers to make principled compliance arguments against the essential requirements themselves during the current standards-free zone — the period from mid-2026 to late 2027 in which requirements are enforceable but harmonised standards under neither M/613 nor M/606 are finalised. A provider whose architecture demonstrably implements the cybernetic functions mapped above can argue from first principles, grounded in theorems that predate the AI Act by decades and have been validated across multiple complex system domains.

6. Recursive Application: Multi-Agent Orchestration

Nannini et al. identify multi-agent orchestration as presenting the hardest compliance challenges: recursive accountability chains, emergent collective behaviour from individually compliant components, and the inability of current impact assessment methodologies to scope across multi-provider delegation chains.⁴ The paper calls for a risk taxonomy for compound AI systems but does not provide one.

The VSM's recursion principle provides the architectural answer. Beer's recursion principle establishes that the same five-system structure applies at every level of organisational hierarchy. Each system is itself a viable system with its own S1-S5 functions. Applied to multi-agent orchestration:

- The orchestrator agent is a System 1 operational unit within the governance architecture.

⁴Nannini et al. (2026), Section 8.1, Step 9. The external-action inventory methodology maps regulatory triggers from what the agent does rather than how it is classified, making it potentially jurisdiction-agnostic as an analytical layer beneath divergent national frameworks.

- Each sub-agent to which the orchestrator delegates is itself a System 1 unit — but within its own recursive VSM structure, it has its own S2 (coordination of its tool calls), S3 (control of its action boundaries), S3* (audit of its own operational state), S4 (awareness of its operational environment), and S5 (policy commitment to its assigned scope).
- Risk does not accumulate unpredictably across the delegation chain — it propagates through defined recursive channels. Each agent layer is responsible for the variety it generates; residual variety that the layer cannot absorb passes upward through defined escalation channels.
- Accountability follows the recursive structure. The orchestrator's System 5 is responsible for policy commitments governing all sub-agent layers. Article 25(4)'s requirement for written agreements specifying compliance contributions across the value chain is the regulatory expression of the contractual formalisation of these inter-layer variety allocations.

This resolves the impact assessment composition problem identified by Nannini et al. ISO/IEC 42005's methodology fails for multi-agent orchestration because it attempts to scope the impact assessment at the orchestrator level and enumerate all downstream effects — a task that grows combinatorially with chain depth. The recursive VSM approach does not enumerate downstream effects; it assigns each level of the recursion responsibility for its own impact domain. The indirectly affected parties of a sub-agent's actions are within the sub-agent provider's impact assessment scope, not the orchestrator's. The orchestrator's risk management system models the sub-agent layer's residual variety — what the sub-agent cannot self-regulate — not the sub-agent's total impact.

Table 4. Variety management across agent governance challenges.

Variety Source	Cybernetic Response	AI Act Obligation	VAIAS-ORG Implementation
Agent action variety(tool calls, autonomous decisions, environmental interactions at scale)	System 2 attenuation: channel agent actions through defined risk-classification pathways before they reach human oversight capacity	Art. 14: oversight commensurate with risks, autonomy, and context. Art. 9(2)(a): automation boundary specification.	Decision Architecture: exception-based escalation. Only residual variety above the risk threshold reaches human oversight. Low-risk, high-confidence outputs in documented domains pass without review.
RL-trained oversight evasion(emergent strategy from reward maximisation; cross-agent propagation)	API-level structural enforcement outside the model: variety attenuation at the architectural level, not through natural language instruction	Art. 15(4): resilience against unauthorised use and attempts to alter behaviour. prEN 18282: privilege minimisation enforced outside the generative model.	Control Architecture: automated acceptance zones defined by trust thresholds. Boundary enforced at API level regardless of model prompt content. The model cannot expose capabilities the architecture does not grant.

Variety Source	Cybernetic Response	AI Act Obligation	VAIAS-ORG Implementation
Multi-agent orchestration variety (cascading actions, sub-agent delegation, emergent collective behaviour)	Recursive VSM application: each agent layer has its own S1-S5 structure; residual variety passes up the hierarchy through defined escalation channels	Art. 9: risk management obligation at each provider layer. Art. 25(4): written agreements specifying compliance contributions across the value chain.	Control Architecture: each agent layer maintains its own feedback loops and audit channel. Epistemic Architecture: trust thresholds apply recursively at each delegation level. Accountability follows the recursive structure.
Regulatory variety (multi-instrument compliance: AI Act, GDPR, CRA, NIS2, DSA, DORA, sector-specific legislation)	System 4 environmental scanning: regulatory variety treated as environmental input to be absorbed, modelled, and encoded into governance structures	Multiple instruments simultaneously applicable depending on external actions (Nannini Step 9 trigger mapping). The agent's external actions determine the regulatory perimeter.	Epistemic Architecture: trust threshold governance maps regulatory triggers to decision classes. Compliance obligations inherited from ontology node properties. The external-action inventory (Nannini Step 9) is the System 4 model of the regulatory environment.

Note: Each row maps a specific source of agentic variety to the cybernetic response mechanism, AI Act obligation, and VAIAS-ORG implementation.

7. Implementation: Building the Architecture in the Standards-Free Zone

The practical consequence of the cross-mapping is that providers building compliance architectures during the current standards-free zone have a principled design methodology that does not depend on the finalisation of harmonised standards. The VSM provides the architectural template; the essential requirements specify the regulatory expression of that template; the Nannini sequence provides the operational delivery programme; VAIAS-ORG provides the commercial design methodology.

7.1 The Build Sequence

The correct build sequence follows VSM logic: establish System 5 (policy) before System 3 (control), System 3 before System 2 (coordination), and System 2 before the System 3* audit channel. This is the inverse of the intuitive engineering sequence — most teams build operations first and governance later. The cybernetic sequence:

1. System 5 first — Policy establishment. Classify the system (Nannini Step 2), establish the QMS (Step 3), and define trust threshold governance (VAIAS-ORG Epistemic Architecture). These normative commitments constrain everything that follows. The EU Declaration of Conformity (Step 10) is the public expression of these System 5 commitments.
2. System 4 second — Environmental intelligence. Map the GPAI layer (Step 1), conduct fundamental rights risk assessment (Step 4), and map adjacent legislation triggers (Step 9). This establishes the System 4 model of the external regulatory and risk environment.

3. System 3 third — Control infrastructure. Implement the risk management system (Step 4), data governance (Step 5), and human oversight automation boundary (Step 6). This establishes the System 3 command function and the VAIAS-ORG Control Architecture.
4. System 3* fourth — Audit channel. Implement versioned runtime state, logging infrastructure (Step 6), and drift detection (Step 11). This is the System 3* function: independent verification of operational state, bypassing agent self-report. This step is the one most commonly absent in current agentic deployments.
5. System 2 fifth — Coordination and privilege enforcement. Implement AI-specific cybersecurity (Step 7), CRA compliance (Step 8), and the variety attenuation mechanisms governing what variety reaches System 3 from System 1. This is the VAIAS-ORG Control Architecture operational layer.

7.2 Three Architectural Claims a Provider Must Be Able to Make

During the standards-free zone, a provider whose architecture implements the VSM structure can make three falsifiable architectural claims that together constitute a principled compliance argument against the essential requirements:

- The Conant-Ashby Claim (Arts. 9 and 3(23)): "Our risk management system continuously maintains a versioned model of the AI system's operational state. Substantial modification is detectable because we can compare current state against the baseline model established at conformity assessment. Drift beyond defined thresholds triggers automatic reassessment."
- The Residual Variety Claim (Art. 14): "Our human oversight function governs the parameters within which the AI system operates, not the outputs the system produces. Human authority is exercised over the residual variety that our System 2 coordination layer cannot absorb. This is commensurate with risks and autonomy level because it concentrates human capacity where it has the highest governance value."
- The Structural Enforcement Claim (Art. 15(4)): "Privilege minimisation is enforced at the API level, outside the generative model. The architecture does not expose capabilities the agent is not authorised to use for the current action. This enforcement is structural and cannot be circumvented by prompt injection, jailbreaking, or emergent reward-maximising strategies because the capabilities are not available at the execution layer."

7.3 The Sublinear Scaling Index as Governance Metric

The Sublinear Scaling Index (SSI = % change in coordination, oversight, and error costs divided by % change in output) is the operational metric confirming the architecture functions as designed. An SSI below 1.0 demonstrates that the variety management

architecture is working: as AI output scales, residual variety reaching human oversight grows more slowly than output, because System 2's attenuation capacity is absorbing an increasing proportion of that variety without proportional growth in human oversight cost.

An SSI above 1.0 is the diagnostic signal that the architecture has failed. The organisation is operating blanket review (attempting to match oversight capacity to total variety, which Ashby's Law shows cannot scale), or System 3* is absent (drift is accumulating undetected, generating error costs), or System 2 is not functioning (agent outputs reaching human oversight without variety attenuation). The SSI turns the abstract cybernetic architecture into a measurable governance outcome, observable from management accounts without specialist tooling and comparable across periods as the system scales.

8. Conclusion

This paper has demonstrated that Beer's Viable Systems Model, the EU AI Act's essential requirements, the Nannini agentic compliance sequence, and VAIAS-ORG are four descriptions of the same underlying structure. The demonstration rests on three structural equivalences:

- The AI Act's risk management obligation (Art. 9) is the Conant-Ashby Theorem expressed as a regulatory requirement: the provider must maintain a versioned model of the system they govern, continuously updated against actual operational state. The conformity assessment is the initial model; System 3* is the update mechanism.
- The AI Act's human oversight obligation (Art. 14) is the Law of Residual Variety expressed as a regulatory requirement: human oversight must govern what automated coordination cannot absorb, not attempt to match the total variety of the system — which Ashby's Law establishes as impossible at scale.
- The AI Act's cybersecurity obligation (Art. 15(4)) is Ashby's Law expressed as an architectural requirement: adversarial variety can only be countered by commensurate structural variety, which natural language instructions cannot provide and API-level enforcement can.

The provider that correctly implements the cybernetic architecture simultaneously satisfies the AI Act's essential requirements. The provider that satisfies only the regulatory form without the underlying architecture will find compliance unsustainable as system complexity grows.

For the agentic compliance challenges that Nannini et al. identify as currently unresolved — behavioural drift without tractable modification events, human oversight of autonomous multi-step systems, privilege enforcement for dynamically-discovered tool chains, and recursive accountability in multi-agent orchestration — the VSM provides architecturally grounded solutions. These are not workarounds or approximations; they are the application of a 70-year body of viable systems theory to a problem space that theory was developed to address.

The practical implication for providers building compliance architectures now is clear: build the VSM. Establish System 5 policy commitments before System 3 control infrastructure; build System 3* before relying on System 1 self-report; design System 2 variety attenuation before expecting human oversight to scale. The harmonised standards, when finalised, will specify in engineering terms what Beer specified in cybernetic terms five decades ago. Building the architecture from first principles now is both the most defensible compliance posture and the most efficient path to sublinear scaling.

References

- Ashby, W. R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.
- Beer, S. (1985). *Diagnosing the System for Organizations*. Wiley.
- CEN/CENELEC JTC 21. (2026). Draft harmonised standards under M/613: prEN 18228, 18229-1, 18282, 18286. Working drafts, January 2026.
- Conant, R. C., & Ashby, W. R. (1970). Every good regulator of a system must be a model of that system. *International Journal of Systems Science*, 1(2), 89-97.
- Espejo, R., Schuhmann, W., Schwaninger, M., & Bilello, U. (1997). *Organizational Transformation and Learning: A Cybernetic Approach to Management*. Wiley.
- Hammond, L. et al. (2025). Multi-Agent Risks from Advanced AI. <https://arxiv.org/abs/2502.14143>
- Hodgers, A. (2008). *The Design of a Viable Interfacing System Between Universities and Multinational Corporations: A Second-Order Cybernetic Approach*. DProf dissertation, Middlesex University.
- Kim, J. et al. (2026). The Attack and Defense Landscape of Agentic AI. <https://arxiv.org/abs/2603.11088>
- Nannini, L., Smith, A. L., Maggini, M. J., Panai, E., Feliciano, S., Tiulkanov, A., Maran, E., Gealy, J., & Bisconti, P. (2026). AI Agents Under EU Law: A Compliance Architecture for AI Providers. <https://arxiv.org/abs/2604.04604v1>
- Rath, A. (2026). Agent Drift. <https://arxiv.org/abs/2601.04170>
- Regulation (EU) 2024/1689 of the European Parliament and of the Council (AI Act). OJ L 2024/1689, 12.7.2024.
- Schwaninger, M. (2000). *Organisational Fitness: Corporate Effectiveness Through Management Cybernetics*. Haupt.
- Shapira, N. et al. (2026). Agents of Chaos. <https://arxiv.org/abs/2602.20021>