Epistemological Sequencing as an Architectural Principle: Structural Output Properties of Multi-Regime Inference over Large Language Models

Marcelo Manucci · ORCID: 0000-0002-9990-2316

CODHZ Research Laboratory · Open Repository Publication · May 2026

DOI 10.5281/zenodo.20121191

Abstract

Background. The dominant design pattern in applied large language model systems involves a single inferential regime: one prompt, one model, one analytical pass. This structural uniformity produces a convergence across systems, models, and analytical outputs that we term *epistemic monoculture* — a pattern that is not primarily a limitation of model capability but an architectural consequence of single-regime processing.

Objectives. To test whether *epistemological sequencing* — the controlled succession of distinct inferential regimes over the same analytical object — produces structural output properties not replicable by professional single-prompt engineering under controlled conditions (same generative model, same analytical case).

Methods. Six analytical frameworks instantiating epistemologically distinct regimes — spanning systems complexity theory, systemic constructivism, inductive empiricism, competitive morphogenesis, cognitive neuroscience, and third-order cybernetics — were each executed under sequenced (experimental) and single-prompt (control) conditions over the same organizational case. Output structure was measured across four metrics: variable diversity (M1), relational density (M2), configurational differentiation (M3), and inferential traceability (M4). A pre-specified success criterion required ≥3/4 metrics in ≥5/6 frameworks. Three independent blind assessors evaluated anonymized output pairs.

Results. The pre-specified success criterion was met in all six frameworks (≥3/4 metrics in 6/6): four frameworks achieved full 4/4 superiority, and two achieved 3/4 with a tie on M1. Relational density (M2) and inferential traceability (M4) were favorable to the sequenced condition in all six frameworks without exception — structural properties that depend on regime transitions rather than on the prescriptive content of any individual prompt. Blind assessment confirmed the structural advantage in five of six frameworks, with a majority in two and unanimity in one. One framework (NM), in which the single-prompt control was exceptionally strong, produced no majority verdict in the blind assessment and is documented and analyzed as an informative exception.

Conclusions. Epistemological sequencing produces architectural output properties — particularly relational density and inferential traceability — that were not observed in the strong single-prompt controls used in this study and are theoretically attributable to regime transitions. The consistency across six epistemologically heterogeneous regimes supports an architecture-level interpretation of the effect. Whether the principle generalizes to structurally distinct domains — that is, whether it is domain-agnostic — requires independent instantiation and remains an open research question. Verification requires instantiation outside the strategic analysis domain studied here. The complete measurement protocol, blind evaluation rubric, control prompt generation procedure with structural summaries, scoring tables, and an anonymized execution environment instantiating the six frameworks are openly deposited and operationally accessible. Current limitations — a single organizational case, a single domain, and a study design that does not fully isolate operator-level reproducibility from model-level variation — define the scope of the present evidence and the next stage of the research program.

Keywords: epistemological sequencing, inferential regime, large language models, structured analytical frameworks, relational density, inferential traceability, epistemic monoculture, prompt engineering comparison, blind evaluation

1. Introduction

Large language models have become the operational substrate of a rapidly expanding class of analytical systems: strategic diagnosis, organizational assessment, risk mapping, decision support, and complex problem-solving across a range of professional domains. Despite the diversity of applications, most of these systems converge on a structurally similar design pattern: a single prompt delivers a complex instruction to a model, the model produces a response, and that response constitutes the analysis. Variations in prompt sophistication, model selection, and output format account for the primary differentiation between competing systems.

This convergence on single-prompt architectures has a structural consequence that has received limited formal attention. A system operating under one inferential regime — however sophisticated its prompt — can only produce the kind of knowledge that regime makes visible. An *inferential regime*, as we use the term, is the set of rules, explicit or implicit, that determines what counts as valid evidence, which hypotheses are admissible, and what kinds of inference can be drawn from available data. When that regime is uniform across all processing steps, the system tends to produce structurally similar outputs regardless of the sophistication of its instructions or the capabilities of its model. We call this tendency *epistemic monoculture*.

Epistemic monoculture is not primarily a criticism of any particular system or model. It is a description of a structural consequence of single-regime design. A more capable model

operating under the same inferential regime may produce more fluent, contextually accurate, or volume-rich outputs, but it does not produce a structurally different *type* of knowledge. The outputs are improved instances of the same analytical category.

The epistemological intuition that distinct analytical frameworks reveal different aspects of a problem has a long history in philosophy of science (Kuhn, 1962; Longino, 1990; Chang, 2012) and in methodological practice across disciplines. What has not been formally investigated is whether this intuition can be operationalized as a computable architectural principle over language models: whether distinct inferential regimes can be made to succeed each other in a controlled way over the same analytical object, and whether that succession produces structural output properties that single-regime prompting does not produce under controlled conditions.

This paper proposes that it can, and reports empirical evidence in support of two claims. The first is that epistemologically sequenced analytical frameworks produce structural output properties — specifically, relational density and inferential traceability — that were not replicated by professional single-prompt engineering under the controlled conditions of this study. The second is that this advantage is consistent across six frameworks operating under epistemologically distinct regimes, which provides preliminary support for interpreting the result as a property of the sequencing architecture rather than of any particular framework's content.

The paper is organized as follows. Section 2 reviews related work on structured reasoning over LLMs, multi-agent systems, and knowledge representation. Section 3 describes the functional-level architecture of epistemological sequencing relevant to the study. Section 4 presents the experimental design, metrics, and control conditions. Section 5 reports results. Section 6 discusses the findings, their limitations, and the research program they open.

2. Related Work

Multi-agent and multi-step reasoning systems. A substantial body of work addresses the coordination of multiple reasoning steps or agents in LLM-based systems. Chain-of-Thought prompting (Wei et al., 2022) and its variants (Kojima et al., 2022; Wang et al., 2023) decompose reasoning into a sequence of steps within a single inferential regime. Tree-of-Thought (Yao et al., 2023) and related approaches explore branching paths in reasoning. Multi-agent frameworks including LangChain, AutoGPT, and related systems (Wooldridge, 2009; Park et al., 2023) distribute tasks across specialized agents or models. The key distinction between this prior work and epistemological sequencing is the nature of what varies across steps or agents. In chain-of-thought and tree-of-thought systems, the inferential regime is constant — the same validity criteria, evidence types, and hypothesis classes operate throughout. In multi-agent systems, agents are typically differentiated by

functional role (searching, summarizing, executing) rather than by epistemological regime. Epistemological sequencing varies the regime itself: each step operates under qualitatively distinct rules about what counts as valid inference, producing outputs governed by different validity criteria rather than merely functionally specialized roles.

Cognitive architectures for language agents. Sumers et al. (2024) survey cognitive architectures for language agents, identifying memory, action, and reasoning as the primary components. The present work is most closely related to the reasoning component, specifically the question of whether reasoning quality can be improved by controlling the epistemic framework under which inference occurs, rather than by improving the model or extending the context. Yao et al. (2023) and related work address what reasoning *does*; epistemological sequencing addresses under what *regime* reasoning occurs.

Knowledge representation and structured inference. The knowledge representation tradition (Russell & Norvig, 2020; Brachman & Levesque, 2004) addresses how to structure what a system knows. The present work addresses a different but related question: how to structure *how a system reasons* about what it knows. The distinction parallels the difference between a knowledge base and an inference engine. Epistemological sequencing operates at the level of the inference engine, not the knowledge base.

Prompt engineering and structured prompting. A large literature addresses the design of prompts to elicit specific behaviors from LLMs (Reynolds & McDonell, 2021; Liu et al., 2023; Sahoo et al., 2024). The present work does not compete with this literature but situates itself in relation to it: the experimental comparison is against the best available single-prompt engineering, not against naive prompting. The question is not whether structured prompts improve over unstructured prompts — they do — but whether epistemological sequencing produces properties that structured single-prompt engineering does not produce, given the same model and case.

Epistemic pluralism and multi-paradigm analysis. The philosophical literature on epistemic pluralism (Longino, 1990; Chang, 2012; Kellert et al., 2006) argues that distinct scientific paradigms can reveal aspects of phenomena unavailable to any single paradigm, and that their interaction can generate knowledge unavailable within either. The present work operationalizes this thesis computationally: it tests whether controlled succession of epistemologically distinct regimes over a language model produces the same kind of knowledge diversity that epistemic pluralism predicts in scientific practice.

Epistemological orchestration over LLMs. Most directly, this paper extends the functional architecture proposed in Manucci (2026), which demonstrated inter-framework emergence in a single documented case with full result traceability. The present study provides systematic empirical evidence across six frameworks and two analytical conditions, shifting the evidence base from existence demonstration to structured comparison.

3. Epistemological Sequencing: Functional Architecture

3.1 Core principle

Epistemological sequencing is an architectural design in which a complex analytical problem is processed through a succession of inferential regimes, each operating under distinct validity rules, before their outputs are integrated. The term *regime* refers to the configuration of rules that governs what counts as evidence, which hypotheses can be formulated, and what types of inference are admissible. Two regimes are epistemologically distinct when they do not share these criteria. Operationally, a regime is instantiated as a constrained step-level specification over four dimensions: evidence admissibility (what data types and sources qualify), hypothesis class (what kinds of claims can be generated), transformation rules (what operations convert inputs to outputs), and acceptance criteria (what conditions an output must satisfy to progress).

The core claim of the architecture is that the controlled succession of distinct regimes over the same analytical object produces structural properties in the integrated output that were not observed under strong single-regime controls in this study and are theoretically attributable to regime transitions. These properties emerge from the transitions between regimes — the points where the output of one regime becomes the input of the next — rather than from any individual step.

3.2 Key distinction from adjacent approaches

Epistemological sequencing differs from multi-agent task decomposition in the following respect: multi-agent systems divide work among agents differentiated by function (one agent searches, another summarizes, another executes). Epistemological sequencing divides the *reasoning regime* rather than the task. All agents in a multi-agent system may apply the same inferential standards; epistemologically sequenced frameworks deliberately apply different inferential standards at each step.

This distinction is not merely terminological. The structural properties documented in Section 5 — particularly relational density and inferential traceability — are properties of the regime transitions, not of any individual step. They were not observed in the stronger single-prompt controls used in this study and are theoretically attributable to the regime transitions themselves.

3.3 The six frameworks studied

The empirical study uses six analytical frameworks that instantiate epistemologically distinct reasoning regimes:

Framework CS (systems complexity analysis): operates under a regime derived from complex systems theory and bifurcation analysis (Prigogine, 1984; Lorenz, 1993). Its validity

criterion is structural coherence among interdependent variables; its primary output type is a set of qualitatively distinct system configurations with specified activation conditions.

Framework NM (niche and trend analysis): operates under a regime derived from competitive morphogenesis and ecological niche theory. Its validity criterion is competitive plausibility given observable market dynamics; its primary output type is a map of emergent opportunity territories.

Framework AP (adaptive planning): operates under a regime derived from the operational closure of third-order cybernetics and adaptive governance (Holling, 2001). Its validity criterion is strategic coherence across multiple organizational axes; its primary output type is an integrated project architecture with anticipatory indicators.

Framework 3D (experience design): operates under a regime derived from cognitive neuroscience and behavioral psychology (Damasio, 1994; Kahneman, 2011). Its validity criterion is behavioral plausibility given cognitive and emotional mechanisms; its primary output type is a triadic intervention architecture (cognitive, emotional, behavioral).

Framework SA (sentiment analysis): operates under a regime derived from inductive empiricism, emerging reality model, and fields of social meaning (Manucci, 2025), and social phenomenology (Schutz, 1967). Its validity criterion is empirical grounding in observable sentiment patterns; its primary output type is a diagnosis of collective cognitive and emotional climate.

Framework IM (issues management): operates under a regime derived from systemic constructivism and narrative epistemology (Luhmann, 1995; Bruner, 1990). Its validity criterion is narrative coherence across stakeholder fields; its primary output type is a structured risk map with stakeholder analysis and narrative field configuration.

The six regimes span the breadth from formal-structural (CS, NM) to phenomenological (SA), from behavioral (3D) to narrative (IM), and from emergent (CS, NM) to strategic-executable (AP). This span is intentional: if the structural advantage of sequencing were confined to frameworks with similar epistemological profiles, it would suggest a domain-specific effect rather than an architectural property.

3.4 Structure of a sequenced framework

Each framework is organized as a sequence of analytical steps. At each step, the system operates under the framework's regime to produce a structured output that becomes the input of the next step. The output of each step is constrained by the regime's validity criteria: outputs that violate the regime's rules of evidence or inference are rejected or reformulated before progressing.

The transition between steps — where one step's output becomes the next step's input — is the site of regime-induced transformation. At these transitions, information produced under one set of validity rules is processed under a different set, creating friction that forces

reformulation, selection, and reinterpretation. The structural properties we measure (relational density, inferential traceability) are cumulative products of these transitions.

The internal prompt sequences, validity criteria, and transition specifications of each framework are not released because they instantiate proprietary operational know-how developed over two decades of work in complex systems modeling. This paper, therefore, does not claim implementation-level reproducibility of the frameworks. Its reproducibility claim is narrower and explicit: the measurement protocol, the control prompt generation procedure, the consolidated scoring tables, the blind evaluation rubric, the frontend source code, and the live execution environment together allow independent inspection, re-measurement, and empirical challenge of the structural output properties reported here.

The reproducibility claim operates at the level of the architectural principle, not at the level of specific framework implementations. The consistency of structural effects across six epistemologically heterogeneous regimes supports an architecture-level interpretation: any independently constructed sequence of epistemologically distinct regimes should produce comparable structural properties when applied to a sufficiently complex analytical object. Whether this generalization holds across structurally distinct domains — the domain-agnostic conjecture — requires instantiation by independent researchers using their own frameworks in domains other than strategic organizational analysis. This is the natural next step of the research program described in Section 6.3.

Independent reproduction of the present study can occur at four levels: reproduction of the measurement protocol on the released materials; reproduction of the control condition using the deposited prompt-generation procedure and the structural summaries of the resulting prompts; direct verification through an anonymized execution environment that implements the full sequencing architecture for all six frameworks, allowing independent researchers to generate experimental outputs on their own cases under the same step-by-step protocol with operator validation at each transition, without exposing internal specifications; and conceptual reproduction of the architectural claim by instantiating independent multi-regime frameworks under the same evaluation protocol. The present paper does not make the proprietary framework implementations reproducible, but it does make the architectural claim empirically challengeable at multiple levels.

The measurement protocol (M1–M4), blind evaluation rubric, control prompt generation procedure, control prompt structural summaries, and consolidated scoring tables are openly deposited as companion materials, enabling independent evaluation of the structural claims. The anonymized execution environment is documented in Section 4.8 and operates at https://verification.codhz.com without registration or credentials.

4. Experimental Design

4.1 Research question and pre-specified criterion

Research question: Does epistemological sequencing produce structural output properties — specifically, variable diversity, relational density, configurational differentiation, and inferential traceability — that professional single-prompt engineering does not produce, under controlled conditions (same generative model, same analytical case)?

Pre-specified criterion: Sequenced frameworks outperform single-prompt controls on ≥3/4 metrics in ≥5/6 frameworks. This criterion was fixed before executing any condition.

4.2 Case

All six frameworks were executed over the same analytical case: the integration of diagnostic AI technology in a regional health network operating under simultaneous pressures from eight structural dimensions (technological, regulatory, demographic, financial, human capital, information systems, professional culture, and market). The case was selected to satisfy four criteria: sufficient complexity to activate the full depth of each framework's analytical sequence; simultaneous pressure from multiple structural dimensions requiring multi-regime analysis; documented evidence base to support inferential traceability measurement; and organizational specificity sufficient to distinguish between generic and case-grounded outputs.

The case is identified only at the level of organizational type and structural pressures throughout this paper. Domain-specific details are withheld to prevent case identification.

4.3 Generative model

Both conditions (experimental and control) were executed using Claude Sonnet 4 (claude-sonnet-4-20250514) in its standard configuration, without fine-tuning or system-level modification. The same model version was used for all twelve executions (six experimental, six control). The choice of a single fixed model is deliberate: the claim under evaluation concerns architectural contrast under a fixed model rather than benchmark ranking across models. The current verification environment (see Section 4.8) operates on Claude Sonnet 4.6 — the current commercial model in the same family — and therefore supports structural verification and inspection of new cases rather than an exact replication of the original study's model version.

4.4 Experimental condition

The experimental condition consists of the full epistemological sequencing of each framework: six analytical steps (five for one framework) executed in sequence, with the output of each step constrained by the framework's regime and reformulated before serving as input to the next. Each step was executed as a separate model call. Human operator

validation was applied at each transition to verify that outputs met the regime's validity criteria before progressing. Operator intervention was constrained to validity checking — verifying that outputs satisfied the regime's formal criteria before serving as input to the next step — and did not include content rewriting, substantive augmentation, or post hoc selection among semantically distinct alternatives.

4.5 Control condition

The control condition was designed to represent the highest-quality single-prompt engineering achievable without knowledge of the sequencing architecture. Control prompts were generated using Google AI Studio's professional prompt optimization feature, given a briefing that described the analytical objective of each framework without revealing its internal structure, sequencing logic, or epistemological regime. The resulting prompts are sophisticated, professional-quality instructions that request structured analysis, multi-dimensional thinking, explicit justification, and organized output — representing the state of the art in single-prompt engineering for complex analytical tasks.

Using AI Studio to design control signals ensures that the control condition reflects what a competent signal engineer would design, providing a solid benchmark for comparison.

The control prompt generation procedure and structural summaries by framework are openly deposited (see Data and materials availability).

The comparison in this study is therefore between two system-level analytical conditions: a multi-step sequenced condition with regime-constrained transitions, and a single-pass professional prompting condition using the same model and analytical case. It is not a comparison of isolated prompt quality, but of analytical architectures operating under controlled conditions.

4.6 Metrics

Standard evaluation metrics for LLM outputs — including BLEU, ROUGE, BERTScore, and human preference ratings — measure properties of individual responses: fluency, factual accuracy, relevance, or user satisfaction. The structural properties relevant to epistemological sequencing are different in kind: they concern the internal relational architecture of an analytical document (how many cross-domain connections exist, how inferential chains link evidence to recommendations) rather than the quality of any individual statement. No existing benchmark captures inter-domain relational density or multi-link inferential traceability, because these properties are architectural — they describe the structure of the analysis as a whole, not the quality of any single analytical claim. The four metrics defined below were designed specifically to measure these architectural properties. They are study-specific but not case-specific: they are defined over structural properties of analytical documents and are therefore portable across domains. Their definitions, scoring procedures, and inter-coder reliability are fully specified

to enable independent replication. Domain coding for M1 and edge typing for M2 were performed independently by two coders following a pre-specified protocol; agreement statistics and coding instructions are openly deposited in the reproducibility materials (see Data and materials availability).

Output structure was measured across four metrics. Measurement was applied to both the experimental and control outputs for each framework, producing twelve measurement sets (six experimental, six control).

M1 — Variable diversity (δ). Counts the number of variables identified in the analysis and their disciplinary distribution across a pre-specified ontology of eight analytical domains derived from the case's structural dimensions. The index $\delta = D^2 / (V \times D_{max})$, where D is the number of domains with at least one variable, V is the total variable count, and D_{max} is 8. The ontology was fixed before executing any condition; variables were classified by two independent coders.

M2 — Relational density (ρ, π). Measures the number of typed relational edges between identified variables, the proportion of inter-domain edges (π), and the proportion of non-trivially typed edges (τ_a, defined as edges with explicit causal, tensional, or cascade typing). Higher values on all three sub-indices indicate greater relational richness.

M3 — Configurational differentiation (Δ). Measures the qualitative distinctness of the output's configurations, scenarios, or strategic spaces. Δ is calculated as 1 minus the mean proportion of dominant variables shared across configuration pairs. A value approaching 1 indicates that each configuration is dominated by a distinct set of variables; a value approaching 0 indicates that all configurations share the same variable dominance profile.

M4 — Inferential traceability (τ). A random sample of 20% of the output's operative elements (recommendations, actions, interventions) was drawn from three distributional thirds of each document. For each sampled element, the evaluator attempted to reconstruct the complete inferential chain: evidence → variable → pattern or relationship → configuration → operative element. τ is the proportion of sampled elements for which a complete five-link chain could be reconstructed. τ_p (partial traceability) records the proportion of elements where 3–4 contiguous links were reconstructable. Inferential chain reconstruction was performed using a fixed coding protocol with pre-specified criteria for each link in the five-step chain. Partial and complete traceability were coded separately in order to reduce all-or-nothing judgment effects and to preserve analytically meaningful distinctions between incomplete and fully reconstructable inferential paths.

4.7 Blind evaluation

Three independent assessors, none of whom participated in framework design or execution, evaluated anonymized output pairs (experimental and control) for each framework. Documents were randomized as A and B with the assignment sealed until data

consolidation. Assessors evaluated four dimensions using a structured rubric: variable breadth (corresponding to M1), relational density (M2), configurational differentiation (M3), and traceability (M4). Scores used a 0–1 scale with three anchor points defined by observable signals. Justification text of maximum 100 words per dimension was required.

Concordance across assessors was measured. No divergence exceeded 20 percentage points on any dimension, eliminating the need for third-evaluator resolution.

The identity of the specific framework, its analytical tradition, and the sequencing architecture were not disclosed to assessors. Assessors were given only the anonymized output documents and the evaluation rubric.

The blind evaluation was designed as triangulation rather than as a replacement for structural measurement: it tests whether the measured differences are also recognizable to independent readers without access to the production process.

4.8 Open infrastructure for verification and instantiation

The methodological choices in this study — the M1–M4 measurement protocol, the blind evaluation rubric, the control prompt generation procedure, and the experimental execution itself — are documented at a level of operational specificity intended to enable independent use beyond reading the paper. Five companion materials are openly deposited: the M1–M4 measurement protocol with codebook rules, framework-specific adaptations, and the case ontology procedure; the blind evaluation rubric with scoring dimensions, assessment forms, and adjudication protocol; the control prompt generation procedure, including the briefing structure provided to the independent generator (Google AI Studio) and structural summaries of the resulting prompts by framework; the consolidated scoring tables, including raw scores by assessor and the inter-rater reliability summary; and the frontend source code of the execution environment under MIT license, with framework prompts retained on the server.

An anonymized execution environment instantiating the six frameworks operates at https://verification.codhz.com. It accepts a case description and executes the analytical sequence on Claude Sonnet 4.6 (claude-sonnet-4-6, the current generative model; the experimental data reported in this paper were generated on claude-sonnet-4-20250514). The environment produces step-by-step output in the same protocol as the experimental condition, with operator validation between transitions.

The infrastructure supports four use modes:

- **Output structure inspection**. Execution of any framework on a new case to inspect the architectural properties reported in this study (relational density, configurational differentiation, inferential traceability) on outputs the user can examine directly.

- **Inter-framework comparison**. Execution of multiple frameworks on a single case to observe how distinct epistemological regimes treat the same analytical object — the structural foundation of the inter-framework operations described in Section 6.3.
- **Replication on the published case**. Execution of any framework on the case used in this study, with independent comparison against the deposited scoring tables.
- **Application of the M1–M4 protocol**. Independent generation of sequenced and non-sequenced output pairs and application of the protocol to verify whether the structural properties reported here generalize to the user's own analytical contexts.

The framework prompts are not disclosed. The deposited materials and the execution environment enable verification, comparison, and instantiation of output structure under the documented protocol. They do not disclose the internal mechanism by which the structural properties are produced.

5. Results

5.1 Structural measurement (M1–M4)

Table 1 presents the comparative measurement results for all six frameworks across the four metrics.

Table 1. Structural measurement results: experimental (EXP) vs. control (CTRL) conditions across six frameworks and four metrics.

FRAMEWORK	EPISTEMOLOGICAL REGIME	M1	M2	M3	M4	SCORE
CS	Systems complexity / bifurcation	EXP ✓✓	EXP ✓✓	EXP ✓✓	EXP ✓✓	4/4
NM	Competitive morphogenesis	EXP ✓	EXP ✓	EXP ✓	EXP ✓	4/4
AP	Strategic foresight / adaptive governance	EXP ✓	EXP ✓	EXP ✓✓	EXP ✓	4/4
3D	Cognitive neuroscience / behavioral	EXP ✓✓	EXP ✓✓	EXP ✓✓	EXP ✓✓	4/4
SA	Inductive empiricism / social phenomenology	Tie	EXP ✓✓	EXP ✓✓	EXP ✓✓	3/4
IM	Systemic constructivism / narrative	Tie	EXP ✓✓	EXP ✓✓	EXP ✓✓	3/4

✓ = difference clear and consistent; ✓✓ = difference of nature, not only degree.

Pre-specified criterion: ≥3/4 metrics in ≥5/6 frameworks. Result: criterion met in 6/6 frameworks. Four frameworks (CS, NM, AP, 3D) achieved full 4/4 superiority; two frameworks (SA, IM) achieved 3/4 with a tie on M1, where the strong single-prompt control matched the experimental condition on variable diversity but not on the other three metrics.

5.2 Metric profiles

M2 — Relational density. The most consistent advantage. Favorable to the experimental condition in all six frameworks without exception. The structural explanation is that each regime transition in the sequencing produces a layer of relational structure that single-regime processing did not generate in this study: a two-layer relational structure (variable→variable before variable→configuration) in NM; a three-layer structure (actors→narratives→tensions) in IM; a quantitative-qualitative decoupled structure with inter-layer relationships in SA. These layers are properties of the sequential process — they emerge from the transitions between regimes — and were not observed under strong single-prompt controls regardless of prompt sophistication.

M4 — Inferential traceability. The second most consistent advantage, and the most surprising from a prompt engineering perspective. The sequential structure of the experimental condition produces traceability as an architectural property: each step's output is constrained to reference prior steps, creating a structured chain of inferential warrant by design. The control condition achieves operational traceability (recommendations connected to stated objectives) but not process traceability (evolution of hypotheses, documentation of contradictions, activation-conditioned rather than calendar-conditioned recommendations). Experimental τ values did not fall below approximately 78%; control values for the strongest controls (AP and IM) reached approximately 55–60%.

M1 — Variable diversity. Favorable to the experimental condition in four frameworks (CS, NM, AP, 3D); a tie in two (SA, IM), where the strong single-prompt control matched the experimental output on variable count and disciplinary distribution. The advantage, where present, reflects what we term *the value of the question*: the epistemological regimes embed disciplinary distinctions that professional prompt engineering does not spontaneously generate. A prompt engineer does not request a 5/5/5 stratification of variables by temporal stability, a triadic C-E-C objective structure, or a five-field narrative configuration — not because of insufficient sophistication, but because these categories belong to specific disciplinary traditions rather than to general prompt engineering practice. The ties on M1 in SA and IM indicate that sufficiently prescriptive single-prompt engineering can approach the experimental condition with variable inventory; the consistent advantage on M2 and M4 in the same frameworks shows that the architectural difference is concentrated in process-dependent properties rather than in variable inventory alone.

M3 — Configurational differentiation. Favorable in all six frameworks, with intensity depending on the presence of regime-specific operators. The advantage is highest where the framework includes an operator that structurally forces differentiation: a formal contradiction operator in IM (Δ ≈ 0.84 vs. 0.55 for the control), a triadic architecture constraint in 3D (Δ ≈ 0.76 vs. 0.40), a free concept selection in AP (Δ ≈ 0.68 vs. 0.45). It is more modest where differentiation derives from depth rather than from a structural operator: NM (Δ ≈ 0.70 vs. 0.55).

5.3 The NM exception

The NM control is the strongest control in the study. The AI Studio prompt for NM requested 15 variables stratified across stability dimensions, 3–5 opportunity spaces with strategic unit profiles, and 20 prioritized actions with KPIs — a structure superficially similar to the experimental output. On structural measurement (M1–M4) the experimental advantage was present in all four metrics. In blind assessment, however, NM produced no majority verdict: one assessor favored the control, one favored the experimental output, and one assessed the pair as tied. The combination — structural advantage on M1–M4 with no majority in blind assessment — warrants closer analysis.

The analysis of this exception is informative. The NM control matches the experimental condition on surface structure but not on three properties that are products of the sequential process: (1) the two-layer relational structure (variable-to-variable relationships before variable-to-opportunity mapping), (2) the non-obvious variable — a variable that does not appear in standard competitive analysis but emerges from the framework's regime and has structural leverage in the system — and (3) the evolution of hypotheses across steps, where early analytical outputs constrain and refine later ones in ways that a single-pass system did not replicate in this study.

The NM exception demonstrates that when a single-prompt control is sophisticated enough to request the right categories, the advantage of sequencing concentrates in process properties (M2, M4) rather than content properties (M1, M3). This is consistent with the interpretation that M2 and M4 are structural products of the sequencing architecture, while M1 and M3 are partially replicable by sufficiently informed prompting.

5.4 Blind evaluation

Table 2 presents the results of the blind evaluation.

Table 2. Blind evaluation results: experimental (EXP) vs. control (CTRL) per framework and assessor.

FRAMEWORK	ASSESSOR 1	ASSESSOR 2	ASSESSOR 3	MAJORITY
CS	EXP (4/4)	Tied	EXP (3/4)	**EXP**
NM	Tied	CTRL (2/4)	EXP (3/4)	**No majority**
AP	EXP (4/4)	Tied	EXP (4/4)	**EXP**
3D	EXP (4/4)	EXP (3/4)	EXP (3/4)	**EXP (unanimous)**
SA	EXP (4/4)	EXP (3/4)	Tied	**EXP**
IM	EXP (3/4)	EXP (3/4)	EXP (4/4)	**EXP (unanimous)**

Summary: EXP wins by majority in 5/6 frameworks; 3D and IM by unanimity. NM produced no majority verdict, with one assessor favoring CTRL, one favoring EXP, and one tie. The structural M1–M4 measurement still favors the experimental condition in NM (4/4); the absence of a blind majority is consistent with the interpretation that strong single-prompt controls can approach the experimental condition on surface features visible to readers, while the structural advantage concentrates in process-dependent properties.

Assessors' dimensional scores show that D2 (relational density / connections) was the dimension most consistently favorable to the experimental condition: 83% of individual assessments (15/18) rated the experimental output at least equal to the control on this dimension, with 11/18 rating it higher. D4 (traceability) was the least consistently favorable dimension in blind assessment (61%), despite being highly favorable in structural measurement. Analysis of assessor justifications suggests that blind assessors recognized operational traceability in both conditions but could not distinguish process traceability (the experimental advantage) from operational traceability without awareness of how the outputs were produced.

One assessor correctly identified the experimental document in all six pairs without prior knowledge of the framework's structure or the sequencing architecture, citing consistent structural patterns in analytical depth, relational explicitness, and hypothesis evolution. This finding suggests that the structural signature of epistemological sequencing may be detectable from output structure alone.

5.5 Trade-off: systemic depth versus operational concreteness

A consistent pattern in assessor justifications warrants specific documentation. Assessor 2 favored the control condition in 3 of 6 frameworks on Dimension 1 (variable breadth), and in 2 of 6 on Dimension 4 (traceability), citing the control outputs' greater specificity and closer connection to actionable recommendations. This assessor's profile suggests a preference for operational concreteness over systemic depth — a genuine evaluative dimension that the experimental condition trades off against analytical scope.

The trade-off is real and is not an artifact of measurement. The epistemological sequencing in several frameworks (most clearly CS and SA) produces analytical outputs with higher

systemic depth and broader variable scope, but the connection between that analysis and immediate operational recommendations requires additional interpretive work from the user. A different assessor's preferences, or a different organizational context emphasizing immediate actionability, might close the gap between conditions on these dimensions.

This trade-off does not invalidate the structural advantage of sequencing; it contextualizes it. The framework produces a richer analytical structure that requires more interpretive effort to operationalize. Whether that trade-off is favorable depends on the analytical purpose.

6. Discussion

6.1 Two sources of structural advantage

The results support decomposing the structural advantage of epistemological sequencing into two functionally distinct sources.

The first is *value of question*: the epistemological regime embeds disciplinary distinctions — categories, classifications, and analytical structures — that prompt engineering does not spontaneously generate. The frameworks studied here draw on complex systems theory, cognitive neuroscience, social phenomenology, narrative epistemology, and strategic foresight. The categories these disciplines provide — temporal stability stratification, triadic C-E-C architecture, quantitative-qualitative decoupling, narrative field configuration — are not produced by asking a general analytical question, however sophisticated. They require knowing which disciplinary tradition to invoke. This source of advantage is primarily reflected in M1 and M3: variable diversity and configurational differentiation.

The second is *value of process*: the sequential structure of regime transitions produces relational layers and inferential chains that single-pass systems did not produce in this study, independent of what categories each step requests. The double-layer relational structure, the cross-regime desynchronization, and the hypothesis evolution documented across all six frameworks are properties of the transitions between regimes, not of any individual step. This source of advantage is primarily reflected in M2 and M4: relational density and inferential traceability.

The NM exception clarifies the relationship between these two sources. A sufficiently sophisticated control prompt can replicate the *value of question* by requesting comparable categories; but it did not replicate the *value of process* in this study, because the regime transitions that generate relational layers and hypothesis evolution are architectural properties absent from the single-pass controls examined here. The NM results show this pattern directly: the advantage concentrates in M2 and M4 precisely because the control captures M1 and M3 through prescriptive prompting.

The NM case thus functions as a discriminative stress test of the theoretical model. If the advantage of sequencing were merely a matter of asking better questions — that is, if value of question exhausted the explanation — then a control prompt that successfully replicates the categories should eliminate the advantage entirely. The fact that the advantage persists, and persists specifically in the process-dependent metrics (M2, M4) rather than the content-dependent metrics (M1, M3), is the pattern the theory predicts and the strongest control fails to eliminate.

6.2 Relational density as a diagnostic signature

The consistency of the M2 advantage — present in all six frameworks, the largest single source of blind assessor recognition, not reproduced by any single-prompt control in this study — warrants interpretation as a candidate diagnostic signature of epistemological sequencing. We propose that relational density, specifically the proportion of inter-domain non-trivial edges, functions as a structural marker for the presence of regime transitions in the analytical process.

This interpretation has methodological implications. If relational density is consistently higher in sequenced outputs, then measurement of relational density in LLM-generated analytical outputs could serve as a provisional proxy for the presence of epistemological sequencing. Conversely, analytical outputs that achieve high relational density without sequencing would be evidence against the proposed mechanism and would require an alternative explanation.

Testing this implication requires comparing epistemologically sequenced systems against non-sequenced systems that have been specifically optimized to produce relational density — a comparison that the present study, designed against best-available prompting practice, does not fully address.

The architectural interpretation would be weakened if non-sequenced systems optimized specifically for cross-domain relational structure were shown to match sequenced systems on M2 and M4 under the same measurement protocol. This constitutes a direct falsification condition for the claim that regime transitions are the primary source of the observed structural advantage.

6.3 Toward a domain-agnostic grammar

The present study is conducted within a single domain (organizational strategic analysis) using six frameworks developed for that domain. The question of whether epistemological sequencing is a general architectural principle — applicable across domains wherever distinct inferential regimes can be instantiated — is not answerable from this evidence alone.

However, the evidence provides preliminary support for the generalization. The six frameworks studied span epistemological traditions whose origins are entirely independent

of strategic organizational analysis: complex systems theory, cognitive neuroscience, social phenomenology, narrative epistemology, competitive morphogenesis, and adaptive governance. The consistency of the structural advantage across these traditions suggests that the results are not artifacts of strategic analysis as a domain but reflect properties of the sequencing architecture that are domain-indifferent.

If this interpretation is correct, epistemological sequencing would imply a domain-agnostic architectural grammar: any controlled succession of epistemologically distinct inferential regimes over a common analytical object would produce comparable structural properties — relational density and inferential traceability in particular — independent of the specific content of the regimes and of the specific domain of application. The frameworks studied here would then be instances of a general architecture rather than a proprietary system. The empirical evidence in this paper does not establish this generalization; it identifies an architecture-level effect within one domain and six frameworks, and renders the domain-agnostic conjecture testable under the conditions specified below.

Verifying this requires instantiating the same architectural principles in frameworks drawn from epistemologically distinct domains outside strategic analysis: clinical diagnostic reasoning, regulatory policy design, interdisciplinary research synthesis, legal argumentation. The test is whether the same structural properties emerge under comparable measurement conditions when the regimes are drawn from these domains. That program of research is the natural extension of the present findings.

To qualify as evidence of domain-agnostic generalization, an independent instantiation would need to satisfy four minimal conditions: (1) the inferential regimes must be drawn from epistemological traditions not used in the present study; (2) the analytical domain must be structurally distinct from organizational strategic analysis; (3) the measurement protocol (M1–M4) must be applied without modification, to enable direct comparison; and (4) the frameworks must be designed and executed by researchers who did not participate in the present study's design, to control for operator effects. If M2 and M4 advantages of comparable magnitude emerge under these conditions, the architectural interpretation would be substantially strengthened. If they do not, the boundary conditions of the principle will have been empirically identified — an equally valuable scientific outcome.

6.4 Relationship to epistemic pluralism

The results reported here provide computational evidence for a thesis that epistemic pluralism (Longino, 1990; Chang, 2012; Kellert et al., 2006) has defended on philosophical grounds: that distinct epistemic frameworks reveal aspects of phenomena unavailable to any single framework, and that their controlled interaction can generate knowledge unavailable within any one. The present study operationalizes this thesis as a measurable architectural property: the relational layers and inferential chains produced by regime

transitions are the computational analog of the interaction effects that epistemic pluralism predicts in scientific practice.

This connection has implications for both communities. For AI researchers, it suggests that epistemological diversity in analytical systems is not merely a user-experience preference but a structural property with measurable output consequences. For epistemologists, it provides empirical evidence — from a controlled computational experiment rather than from theoretical argument — that the benefits of paradigm interaction are real and measurable.

6.5 Boundary conditions and operational frontiers

The findings reported here are conditional on a set of choices that the present study makes explicit. Each of these conditions defines a direction in which the architectural interpretation requires further test, and each maps to an operational frontier where the architecture is being instantiated or could be instantiated. They are presented not as caveats but as the structural shape of the work this paper opens — research, applied instantiation, and integration alike.

Single case. All six frameworks were executed over the same organizational case. Generalization requires replication across cases with distinct structural profiles. Two factors partially mitigate this limitation. First, the prior study (Manucci, 2026) documented convergent results using three distinct generative models over the same case, suggesting that the structural advantage is model-independent. Second, the present study's six frameworks — spanning epistemological traditions as diverse as complex systems theory and social phenomenology — constitute six partially independent tests of the architectural principle on the same case. The consistency of M2 and M4 across all six provides stronger evidence than a single framework tested on six cases would, because it demonstrates that the structural advantage survives variation in the reasoning regime, not only variation in the analytical object. Nevertheless, multi-case validation remains essential and is the next empirical requirement for strengthening the architectural interpretation.

Single domain. All six frameworks are designed for strategic organizational analysis. Empirical validation in epistemologically distinct domains — clinical diagnostic reasoning, public health analysis, urban policy design, regulatory impact assessment, educational design, institutional risk anticipation — is the next stage both for the architectural interpretation advanced here and for the program of applied research the architecture supports.

Operator-model confound. Different human operators used different generative models in the multi-operator multi-model validation documented in related work. The present study's single-case design uses a single model, reducing but not eliminating this confound within the study. A fully controlled multi-operator experiment with a fixed model would isolate the operator contribution.

Operator effect at transitions. Although operator intervention was constrained to validity checking at transitions — not to content rewriting or substantive augmentation — the present design does not fully isolate the operator's contribution to the structural advantage. A validity check that rejects an output and triggers regeneration may influence the quality of the resulting chain in ways that differ from a fully automated pipeline. Determining the magnitude of this effect requires comparing identical frameworks executed by multiple independent operators under identical model and case conditions, which is planned as part of the next-stage replication design.

Process traceability in blind evaluation. Blind assessors recognized the experimental advantage in M2 (relational density) with high consistency but not in M4 (inferential traceability). Analysis of assessor justifications suggests that process traceability — the experimental condition's specific advantage — is not visible without knowledge of the production process. This raises the question of whether inferential traceability, as measured by the M4 protocol, is a property that adds analytical value or primarily serves audit purposes. We consider this an open question worthy of further investigation.

Single-pass controls. The control condition uses a single-pass prompt. An iterative prompt engineer who refined outputs through multiple turns might narrow the gap on some dimensions. The comparison is against single-pass professional prompting, not against iterative expert prompting.

7. Conclusion

This paper reports empirical evidence that epistemological sequencing — the controlled succession of distinct inferential regimes over the same analytical object — produces structural output properties not replicated by professional single-prompt engineering under controlled conditions. The evidence is based on a comparison across six frameworks operating under epistemologically distinct regimes, using four structural metrics and blind assessor evaluation.

The two most consistent structural advantages are relational density and inferential traceability. Both depend on the sequential process itself — on the transitions between regimes — rather than on the prescriptive content of any individual prompt or the capability of the underlying generative model. They are architectural properties, not prompt properties.

The consistency of the advantage across six epistemologically heterogeneous frameworks — spanning traditions as distinct as complex systems theory, cognitive neuroscience, and narrative epistemology — supports an architecture-level interpretation of the effect within the domain studied. Whether the principle generalizes to structurally distinct domains is

left as a research horizon: the systematic test requires instantiation outside the strategic analysis domain studied here, on the conditions specified in Section 6.3.

More broadly, the results suggest that the structural convergence of current LLM analytical systems — what we term epistemic monoculture — is not primarily a limitation of model capability but a consequence of single-regime architectural design. Addressing it does not require more capable models; it requires architectures that allow distinct inferential regimes to interact in a controlled way. More generally, variation in inferential regime may be an overlooked axis of architectural design in LLM-based analytical systems, alongside model choice, prompting strategy, and task decomposition. Epistemological sequencing is one candidate architecture of this kind. Whether it is the only one, or the optimal one, is an open question.

The work this paper opens has three trajectories that share the same empirical foundation. The first is independent verification and instantiation of the architectural principle in epistemologically distinct domains, supported by the open infrastructure documented in Section 4.8. The second is the formal characterization of the regime transitions where the structural properties emerge, including the inter-framework operators required to compose multiple sequenced analyses on a single object. The third is applied work in domains where structured reasoning under uncertainty is the operative challenge — clinical, regulatory, urban, educational, institutional — for which the architecture provides a domain-indifferent grammar that can be instantiated rather than imitated. The architectural interpretation advanced here is open to test on the explicit condition identified in Section 6.2: non-sequenced systems specifically optimized for cross-domain relational structure that match sequenced systems on M2 and M4 under the same protocol would falsify the central claim. The infrastructure required to test that condition, and to extend the architecture to the trajectories above, is publicly available.

References

Brachman, R. J., & Levesque, H. J. (2004). *Knowledge Representation and Reasoning*. Morgan Kaufmann.

Bruner, J. (1990). *Acts of Meaning*. Harvard University Press.

Chang, H. (2012). *Is Water H2O? Evidence, Realism and Pluralism*. Springer.

Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. Putnam.

Holling, C. S. (2001). Understanding the complexity of economic, ecological, and social systems. *Ecosystems*, 4(5), 390–405.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kellert, S. H., Longino, H. E., & Waters, C. K. (Eds.). (2006). *Scientific Pluralism*. University of Minnesota Press.

Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large language models are zero-shot reasoners. *Advances in Neural Information Processing Systems*, 35.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35.

Longino, H. E. (1990). *Science as Social Knowledge*. Princeton University Press.

Lorenz, E. N. (1993). *The Essence of Chaos*. University of Washington Press.

Luhmann, N. (1995). *Social Systems*. Stanford University Press.

Manucci, M. (2025). El Algoritmo de la Confianza: La transformación digital de la IA aplicada a la estrategia de vínculos. Unknown Frontier LLC.

Manucci, M. (2026). Epistemological orchestration over language models: A functional architecture with preliminary evidence. *Zenodo*. https://doi.org/10.5281/zenodo.19266911

Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences*, 120(13), e2215907120.

Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. *Proceedings of UIST 2023*.

Prigogine, I. (1984). *Order Out of Chaos*. Bantam Books.

Reynolds, L., & McDonell, K. (2021). Prompt programming for large language models: Beyond the few-shot paradigm. *Extended Abstracts of CHI 2021*.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S., & Chadha, A. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv:2402.07927*.

Schutz, A. (1967). *The Phenomenology of the Social World*. Northwestern University Press.

Sumers, T. R., Yao, S., Narasimhan, K., & Griffiths, T. L. (2024). Cognitive architectures for language agents. *Transactions on Machine Learning Research*.

Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Narang, S., ... & Zhou, D. (2023). Self-consistency improves chain of thought reasoning in language models. *International Conference on Learning Representations*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35.

Wooldridge, M. (2009). *An Introduction to MultiAgent Systems* (2nd ed.). Wiley.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T., Cao, Y., & Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36.

Correspondence: Marcelo Manucci, CODHZ Research Laboratory. *Competing interests:* The author is the founder of CODHZ Research Laboratory, where the analytical frameworks examined in this study were developed. The experimental design, the M1–M4 measurement protocol, the blind evaluation rubric, and the procedure for generating control prompts via an architecturally independent system (Google AI Studio) were structured to enable comparability and auditability independently of this relationship. All companion materials are openly deposited and the verification environment operates without intermediation. *Data and materials availability:* All companion materials are openly deposited at Zenodo and accessible via the paper's DOI. The deposit comprises the M1–M4 measurement protocol with codebook (Version 2.0); the blind evaluation rubric and assessment forms (Version 2.0); the control prompt generation procedure with structural summaries by framework (Version 2.0); the consolidated scoring tables with raw scores by assessor (Version 2.0); the frontend source code of the verification environment under MIT license; and the documentation of the verification environment itself. The execution environment operates at https://verification.codhz.com without registration or credentials. The framework prompts and the unmodified experimental output documents are not disclosed because they contain proprietary specifications; the deposited materials enable verification of output structure, measurement protocol, and experimental design without disclosing the internal mechanism. Researchers and operators interested in the instantiation of the architectural principle in domains beyond strategic organizational analysis may contact the laboratory through codhz.com. *Repository deposit:* This paper is openly deposited at Zenodo DOI: 10.5281/zenodo.20121191. Prior preprint by the author: Manucci (2026), Zenodo DOI 10.5281/zenodo.19266911.