# Neuro-Semiotic Reasoning: Computational Semiosis for Observational Causal Inference

**Martin Trajkow**

*Lead Researcher, o-machine*

martin@o-machine.com

https://o-machine.com

**Version:** 1.5 • May 8, 2026

**Initial Publication:** Ver. 1.0 • Feb 17, 2026

## Abstract

Current AI systems excel at statistical pattern matching and formal discovery but fail systematically at causal reasoning from real-world observational signals. While systems like AlphaEvolve demonstrate LLM-guided evolutionary search for synthesized innovation within formal spaces [17, 21], they remain blind to unstructured, hidden signals of physical and commercial reality. We identify this as the *Causal Reasoning Gap*: the inability to derive verified causation from observational signals that have not been textualized or modeled.

We present the **Neuro-Semiotic Reasoning Engine**, an architecture that constructs causal understanding through semiotic interpretation of differential relationships across temporal, spatial, conceptual, and absence dimensions. Unlike statistical models operating on tokens, our architecture introduces **Behavioral Signals** as the unit of computation. Drawing on Peirce's unlimited semiosis, we demonstrate that recursive interpretation enables emergent capabilities: inference of unobserved states, absence detection as a reasoning primitive, and complete auditability through traceable interpretive pathways.

We formalize the architecture through five pillars (temporal knowledge graphs, concept-mediated ontologies, behavioral modeling, adversarial verification, and evolutionary optimization) and propose that observational causal reasoning requires fundamentally different computational primitives than statistical correlation or formal search.

**Keywords:** Causal reasoning, Behavioral signals, Neuro-semiotic AI, temporal reasoning, absence detection, observational causal inference.

## 1. Introduction

Large language models have achieved remarkable fluency in text generation, yet leading AI researchers increasingly question their fundamental architecture as a path to human-level intelligence. As LeCun argues, LLMs lack a grounded model of the world because they are "limited to the discrete world of text" [16]. Sutton identifies deeper structural deficits: they lack ground truth for verification and mechanisms for surprise-driven learning [18].

Furthermore, even the most advanced evolutionary discovery systems, such as DeepMind's AlphaEvolve [17, 21], operate primarily within **formal problem spaces**—mathematical analysis, combinatorics, and code—where the environment is defined by rigid rules and automated evaluators. While these systems can discover novel algorithmic solutions that no human has yet written down, they are essentially performing **synthesized innovation**: exploring permutations of a known, modeled universe.

We identify a different, more fundamental structural limitation: the **Hidden Reality Blind Spot**. Current AI is blind to what has not yet manifested as a document or been modeled into a formal search space. The most valuable causal insights are often those still manifesting as raw, unrecorded behavioral signals. In highly dynamic environments, statistical correlation and formal search both face challenges because they lack access to the underlying causal mechanisms operating in real-time.

We define this as the **Causal Reasoning Gap**: the inability to derive causation from unstructured, temporal, multi-modal observational signals. This paper presents the **Neuro-Semiotic Reasoning Engine**, an architecture designed to address this gap by proposing computational semiosis – the recursive construction of causal understanding through differential interpretation of **Behavioral Signals**.

## 2. Background and Related Work

### 2.1 The Hierarchy of Knowledge Processing

We distinguish four levels of knowledge processing:

| Level | Definition | Example | Current AI Capability |
|---|---|---|---|
| **Information** | Raw facts | "Firm X hires Y" | High |
| **Knowledge** | Organized facts | "Firm X is building an engineering team" | High |
| **Understanding** | Causal relationships | "Firm X builds engineering team because…" | Limited |
| **Wisdom** | Actionable insight | "Firm X's hiring shift indicates strategic pivot" | Limited |

Current systems excel at Levels 1-2 but face severe challenges at Levels 3-4. We hypothesize this limitation is *architectural*, not merely computational.

### 2.2 Statistical Pattern Matching: Capabilities and Limits

LLMs predict text distributions based on training data. They manipulate language patterns without necessarily constructing causal models of underlying reality. As Sutton notes, "you can't have prior knowledge if you don't have ground truth." This can produce instances where LLMs optimize for token probability rather than correspondence with reality [7].

Even more fundamentally, LLMs lack explicit mechanisms for surprise-driven learning. They cannot readily detect when reality deviates from predictions or update their internal world models dynamically in real-time. This makes temporal reasoning from observational signals—such as detecting a strategic inflection point—structurally challenging.

### 2.3 Neuro-Symbolic and Discovery Systems

Recent work in neuro-symbolic AI seeks to combine neural learning with symbolic reasoning, but typically requires pre-defined, brittle ontologies [15]. Meanwhile, discovery systems like AlphaEvolve [17, 21] utilize LLMs to guide evolutionary search in formal spaces. While effective for algorithmic generation, these systems are constrained to environments where outcomes are formally verifiable. They are not explicitly designed to infer organizational strategy because such "source code" is not readily available; it must typically be inferred from observational signals.

### 2.4 Semiotic Foundations

We draw on Peirce's theory of unlimited semiosis [11]—the principle that signs generate interpretants which themselves become new signs. Meaning is dynamically constructed through the process of recursive interpretation. In our architecture, a fact gains meaning from its differential relationships across temporal, spatial, conceptual, and absence dimensions.

## 3. Theoretical Foundation: Causal Understanding Through Behavioral Signals

### 3.1 A Different Unit of Computation: Actions, Not Tokens

Every reasoning paradigm is defined by its fundamental unit of computation—what it "sees" of the world. - **LLMs see Tokens**: Statistical fragments of text, stripped of time and context. - **Symbolic AI see Propositions**: Rigid logical statements. - **Knowledge Graphs see Triples**: Static snapshots of entity-relation-entity.

**The proposed architecture processes Behavioral Signals**: observable actions by real-world entities—hiring, partnering, filing, releasing, expanding, contracting—anchored in time, linked to evidence, and analyzed in relation to their own history and the concurrent behaviors of adjacent actors.

Critically, a Behavioral Signal is **modality-independent**. The signal may be extracted from text documents, imagery, sensor data, or transactional records. The architecture reasons over the *fact of the behavior*, not solely over the medium that carried the signal. This decoupling aims to allow the system to ingest reality regardless of whether that reality has been comprehensively "textualized."

### 3.2 Semantic Atomicity as a Computational Constraint

Traditional Natural Language Processing (NLP) entity-relationship extraction typically produces compound, entangled relationships. When extracting logic from unstructured text, generative models naturally mirror the fuzzy, overlapping structure of human language. However, these compound structures cannot be subjected to strict tensor calculus.

The Neuro-Semiotic architecture resolves this limitation by enforcing a strict constraint we term **Semantic Atomicity**. Before any observational signal enters the temporal knowledge graph, the semiotic extraction layer is rigidly forced to decompose all raw text into indivisible $(s, p, o, t)$ tuples (subject, predicate, object, timestamp).

By strictly constraining the semiotic parser to output these atomic tuples, we ensure that the system does not ingest blurry heuristics. This constraint guarantees that the subsequent hyperdimensional binding operations (Track B) are mathematically commutative, reversible, and topologically pure, bridging the gap between linguistic ambiguity and formal logic.

### 3.3 The Four Dimensions of Differential Meaning

We propose that causal meaning emerges from analyzing how these signals relate across four dimensions: 1. **Temporal Dimension**: Causes precede effects; velocity and acceleration reveal intent. 2. **Spatial Dimension**: Entities connected through multiple independent pathways may indicate causal relationships. 3. **Conceptual Dimension**: Facts gain meaning when interpreted within concept phases. 4. **Absence Dimension**: The deviation between expected behavior and observed behavior serves as a diagnostic signal.

### 3.4 Unlimited Semiosis: The recursive construction of meaning

The central hypothesis is that **causal meaning can be approximated through recursive interpretation of differential relationships.** We model this through a semiotic chain where each interpretation generates a new sign (interpretant) for further analysis.

**Motivating Scenario 3.1: The Jony Ive Inference** To illustrate the proposed reasoning process, consider the following hypothetical scenario: *"What is Jony Ive building for OpenAI?"*—asked prior to official documentation.

| Step | Context Dimension | Interpretant |
|---|---|---|
| 1 | Atomic Signals | Ive's hiring of former Apple hardware leads; OpenAI's hardware team growth. |
| 2 | Spatial/Temporal | Satellite imagery shows facility buildouts at specific contract manufacturers. |
| 3 | Conceptual | Apple's concurrent shift away from specific device strategies; Google's ambient computing positioning. |
| 4 | Absence | No software-only patents for the new team; absence of traditional cloud-infra hiring. |
| 5 | Synthesis | Hypothesis: A specialized AI-integrated wearable device focused on ambient computing. |

The answer is constructed not from retrieval, but inferred from analyzing the convergence of heterogeneous behavioral signals.

## 4. System Architecture

The architecture utilizes five pillars to implement recursive interpretation.

### 4.1 Pillar 1: Temporal Knowledge Graph

We utilize a fact-first architecture where the fundamental unit is the timestamped Behavioral Signal with provenance tracking. This enables velocity and acceleration analysis (e.g., a hiring spike followed by deceleration) and domain-specific absence detection.

### 4.2 Pillar 2: Concept-Mediated Ontology

Entities connect through evolving concepts. Adding a new signal to a domain recontextualizes the concept itself, which recursively recontextualizes connected entities. Meaning is constructed dynamically rather than encoded entirely statically.

### 4.3 Pillar 3: Adversarial Falsification

Inferences undergo structured evaluation through a two-agent protocol. - **The Generator** proposes causal connections based on pattern matches. - **The Critic** searches for counter-evidence and contradictory signals across the four differential dimensions.

We reframe this validation step as adversarial falsification. The engine is designed such that an inference is only committed to the graph if it withstands critical assessment against available signals across all four dimensions. While this does not guarantee absolute ground truth, it produces **high-confidence inferences** that are logically defensible based on available observational data.

### 4.4 Pillar 4: Behavioral Modeling

Absence detection requires expectation. We maintain behavioral baselines for entities within specific concept phases. Significant deviations from these baselines are treated as diagnostic signals of structural shifts.

### 4.5 Pillar 5: Evolutionary Optimization

The system's reasoning strategies—including detection patterns and verification thresholds—are designed to evolve through selection pressure. Guided by LLM mutations (which leverage broad world knowledge), the system aims to dynamically adjust its reasoning procedures based on outcome validation, drawing inspiration from recent work in test-time computation and evolutionary meta-learning [14, 17, 19, 21].

## 5. Emergent Reasoning Capabilities

### 5.1 Inference of Unobserved States via Relational Clustering

The architecture enables inference of hidden structural states from weak, cross-domain signals (e.g., analyzing "talent clusters" to hypothesize new organizational developments prior to incorporation).

### 5.2 Rapid Falsification via Absence Signal Diagnostics

The system generates meaning from the non-occurrence of expected events to rapidly evaluate claims. If an entity claims "proprietary hardware breakthroughs" but exhibits zero evidence of specialized hiring or component procurement, the system generates a high-confidence anomaly signal.

### 5.3 Multi-Hop Causal Chain Tracing

The architecture traces multi-hop dependencies to map systemic exposures (e.g., identifying technological shifts or upstream supply constraints that render established structural advantages obsolete).

### 5.4 Complete Auditability

Conclusions are backed by a traceable interpretive chain leading back to atomic, evidence-linked Behavioral Signals.

## 6. Proposed Evaluation Methodology

### 6.1 Evaluation Philosophy

Evaluating causal reasoning systems presents distinct challenges. Precision and recall measure fact retrieval; they do not fully capture the quality of *causal inference*, the detection of *absence*, or the depth of *cross-domain reasoning*.

We propose a multi-tier framework with 12 distinct hypothesis-driven experiments organized by reasoning complexity. Principally, this involves implementing the system as outlined and empirically evaluating it against real-world observational datasets.

### 6.2 Research Questions and Hypotheses

**RQ1: Can causal inference of structured organizational activity exceed baseline LLM performance using purely observational, non-textualized behavioral signals?** *Hypothesis 1:* The Neuro-Semiotic Engine will demonstrate higher F1 scores in detecting unannounced organizational developments compared to base LLMs when restricted to identical sets of atomic behavioral signals.

**RQ2: Under what conditions does absence detection serve as a reliable falsifier of structural claims?** *Hypothesis 2:* The introduction of expected-behavior baselines (Pillar 4) will significantly reduce Type I errors (false positives) in claim validation compared to purely generative evaluation methods.

**RQ3: Does the Adversarial Falsification protocol incrementally improve the stability of multi-hop causal chains?** *Hypothesis 3:* Inferences subjected to multi-dimensional adversarial criticism (temporal, spatial, conceptual, absence) will exhibit lower variance and higher correspondence with eventual ground-truth revelation than initial Generator proposals.

### 6.3 Experimental Design

The planned evaluation involves instantiating the architecture on a bounded dataset of historical, timestamped behavioral signals (e.g., 2018-2022 hiring, patent, and facility data) and querying the system regarding strategic outcomes that were strictly "in the future" relative to the signal cutoff date.

Performance will be measured against standard LLM baselines and symbolic knowledge graph completion algorithms on three axes: 1. **Reasoning Depth**: Measured by the number of verifiable multi-hop dependencies successfully traced. 2. **Evidence Integrity**: Proportion of final inferences backed by unbroken, verifiable evidence chains. 3. **Absence Sensitivity**: The true positive rate of correctly falsifying claims based on diagnostic absence indicators.

### 6.4 Implementation Status

A working prototype implementing Pillars 1–4 has been developed, including timestamped behavioral signal storage with provenance tracking, concept-mediated entity relationships, a two-agent adversarial falsification protocol with mathematical confidence adjustment, and behavioral primitive detection with velocity analysis. The evolutionary optimization pillar (Pillar 5) exists as a framework but is not yet fully automated. Future work involves executing the proposed experiments across these operational components.

## 7. Discussion

The proposed transition to observational causal reasoning involves a shift from content-based frequency analysis to relationship-based structure. This structural shift is analogous to transitions in early information retrieval, where frequency-based metrics were augmented by topological relationship analysis (e.g., eigenvalue centrality measures applied to link structures).

| Dimension | Statistical Generation | Proposed Architecture |
|---|---|---|
| **Primary Input** | Tokens/Textual fragments | Modality-agnostic Behavioral Signals |
| **Logic Foundation** | Probabilistic Distribution | Differential interpretation |
| **Verification** | External Ground Truth | Adversarial Falsification Chains |
| **Primary Domain** | Synthesized formal discovery | Observational physical/organizational inference |

## 8. Conclusion

The Causal Reasoning Gap is an architectural gap. Closing it requires moving beyond the "discrete world of text" to a system that analyzes the world through modality-independent Behavioral Signals. By attempting computational semiosis through a fact-first temporal graph and an adversarial falsification protocol, the architecture aims to enable the transition from imitating human language patterns to reasoning about underlying structural and organizational reality.

This work represents a theoretical and architectural framework for observational causal inference. The architecture described here has been implemented as a working prototype. Future work focuses on the rigorous empirical evaluation of the hypotheses outlined in Section 6, comparing the architecture's performance against standard generative baselines in dynamic, unstructured environments.

## References

[1] Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.

[2] Garcez, A. d'A., et al. (2019). Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *arXiv preprint arXiv:1905.06088*.

[3] Goodfellow, I. J., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27.

[4] Hamilton, W. L. (2020). *Graph Representation Learning*. Morgan & Claypool Publishers.

[5] Hogan, A., et al. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1-37.

[6] Irving, G., Christiano, P., & Amodei, D. (2018). AI safety via debate. *arXiv preprint arXiv:1805.00899*.

[7] Ji, Z., et al. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.

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

[9] Madry, A., et al. (2018). Towards deep learning models resistant to adversarial attacks. *International Conference on Learning Representations*.

[10] Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

[11] Peirce, C. S. (1931-1958). *Collected Papers of Charles Sanders Peirce*, Vols. 1-8. Harvard University Press.

[12] Touvron, H., et al. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

[13] He, H., & Thinking Machines Lab. (2025). Defeating nondeterminism in LLM inference. *Thinking Machines Lab: Connectionism*. https://thinkingmachines.ai/blog/defeating-nondeterminism-in-llm-inference/

[14] Kamradt, G., & Chollet, F. (2025). ARC Prize 2025 results and analysis. *ARC Prize*. https://arcprize.org/blog/arc-prize-2025-results-analysis

[15] Mao, H., Chen, Z., Tang, W., Zhao, J., Ma, Y., Zhao, T., Shah, N., Galkin, M., & Tang, J. (2024). Position: Graph foundation models are already here. *arXiv preprint arXiv:2402.02216*.

[16] Reese, H. (2026). Yann LeCun's new venture is a contrarian bet against large language models. *MIT Technology Review*. https://www.technologyreview.com/2026/01/22/1131661/yann-lecuns-new-venture-ami-labs/

[17] Novikov, A., Vũ, N., Eisenberger, M., Dupont, E., Huang, P.-S., Wagner, A. Z., Shirobokov, S., Kozlovskii, B., Ruiz, F. J. R., Mehrabian, A., Kumar, M. P., See, A., Chaudhuri, S., Holland, G., Davies, A., Nowozin, S., Kohli, P., & Balog, M. (2024). AlphaEvolve: A coding agent for scientific and algorithmic discovery. *arXiv preprint arXiv:2506.13131*.

[18] Sutton, R. S. (2019). The Bitter Lesson. *Incomplete Ideas*. http://www.incompleteideas.net/IncIdeas/BitterLesson.html

[19] Darlow, L., Crowley, E. J., Storkey, A., & Hospedales, T. (2025). Continuous Thought Machines: Adaptive test-time computation for abstract reasoning. *arXiv preprint arXiv:2505.05522*.

[20] Silver, D., & Sutton, R. S. (2025). Welcome to the Era of Experience. In *Designing an Intelligence* (forthcoming). MIT Press. Available at: https://storage.googleapis.com/deepmind-media/Era-of-Experience%20/The%20Era%20of%20Experience%20Paper.pdf

[21] Georgiev, B., Gómez-Serrano, J., Tao, T., & Wagner, A. Z. (2025). Mathematical exploration and discovery at scale. *arXiv preprint arXiv:2511.02864*. https://doi.org/10.48550/arXiv.2511.02864

---

## Appendix A: Formal Definitions Summary

| Symbol | Definition |
|---|---|
| $f = (s, p, o, t, E)$ | Atomic Behavioral Signal (subject, predicate, object, timestamp, evidence set) |
| $M(f, C)$ | Meaning function: signal $\times$ context $\rightarrow$ interpretant |
| $\Delta_T(f, t_i, t_j)$ | Temporal differential of signal $f$ between times $t_i$ and $t_j$ |
| $\Delta_S(f)$ | Spatial differential: relational context set of signal $f$ |
| $c$ | Concept: higher-order abstraction mediating entity relationships |
| $\phi(c, t)$ | Phase of concept $c$ at time $t$ |
| $a = (f_{expected}, t, \epsilon)$ | Absence signal: expected behavior not observed within tolerance |
| $c_g$ | Generator confidence score |
| $\sigma$ | Convergence verification across four dimensions (temporal, spatial, conceptual, absence) |
| $s_d$ | Critic counter-evidence strength across dimensions |

| Symbol | Definition |
|---|---|
| $c_{final} = c_g \cdot \sigma - \max(s_d)$ | Committal Gate confidence with dimensional verification |
| $\theta_{high}, \theta_{low}$ | Empirically calibrated decision thresholds for committal gate |
| $S = \{i_0, i_1, ..., i_n\}$ | Semiotic chain: sequence of recursive interpretants |

## Appendix B: Domain-Specific Ontologies

Domain-specific ontologies and behavioral pattern libraries have been developed for multiple decision environments including autonomous driving, semiconductor supply chains, and energy infrastructure. These ontologies include:

- **Concept Types**: Technology categories, business models, strategic approaches, application domains, and risk factors
- **Behavioral Primitives**: Observable signals including partnerships, product releases, testing activity, leadership changes, technology development, deployment milestones, and workforce dynamics
- **Relationship Types**: Causal dependencies (ENABLES, REQUIRES, BLOCKS), competitive dynamics (COMPETES_WITH), behavioral signals (EXHIBITS, VALIDATES), and risk exposures (EXPOSES_TO)

The ontologies are designed to be extensible and transferable across domains while maintaining domain-specific precision. Each concept and relationship type includes temporal metadata, evidence requirements, and validation criteria.