An Ontological Framework for Meaning, Knowledge, and Intelligence

Rohan Vinaik

Independent Researcher

Abstract

This paper develops an ontological framework that explains how meaning, knowledge, and intelligence arise—or fail—across minds, media, and computational systems. Building on an agentic view of cognition, the framework models systems as networks of semantic agents operating under constraints (K), coupled by an interaction topology (T), and oriented by intentional vectors (I). Variations in the triplet $\langle K, T, I \rangle$ yield distinct modes of meaning, each with measurable signatures in semantic density, redundancy/entropy, and alignment. The account unifies structural and semantic perspectives, extends agent-based theories of mind by elevating constraint and intention to first-class variables, and predicts phase transitions between modes under parameter shifts. Applied to contemporary AI systems including large language models (GPT-4, Claude, Gemini), the framework reveals conditions under which systems generate meaningful content versus *semantic vampirism*—surface mimicry that drains integrative meaning. The framework supports diagnostics and design principles for AI development and offers theoretical foundations for understanding machine semantics, with implications spanning cognitive science, epistemology, AI safety, and the ethics of AI-mediated communication.

Keywords: ontology, meaning, artificial intelligence, semantic content, cognitive architecture, agent-based models, AI interpretability, semantic vampirism

Contents

1 Problem & Contribution

1.1 Problem Statement and Scope

Across contemporary discourse—from narrative media to scientific communication to AI-generated content—artifacts and systems that are superficially similar differ radically in their capacity to produce durable meaning and reliable knowledge. Some configurations foster coherence, insight, and growth; others preserve only surface form while hollowing out integrative content. This divergence has become particularly acute with the proliferation of large language models and AI-mediated production pipelines, where outputs can exhibit sophisticated linguistic structure yet lack semantic grounding.

Existing approaches tend to bifurcate: structural accounts emphasize formal organization while underspecifying semantics and purpose (Propp, 1968; Barthes, 1977); content-first accounts foreground interpretation while under-theorizing the generative role of structure and constraints (Ricoeur, 1984). Meanwhile, contemporary philosophy of AI debates the nature of understanding in machine systems without adequate formal machinery for distinguishing genuine semantic capacity from sophisticated pattern-matching (Bender & Koller, 2020; Shanahan, 2024). A unified ontology is needed—one that explains, at a level general enough to include minds, media, and institutions, how meaning and intelligence emerge, stabilize, degrade, or collapse.

This paper addresses that gap by articulating a framework in which meaning, knowledge, and intelligence are treated as emergent properties of interacting semantic elements. The framework aims to be agnostic to medium and scale: it applies to narrative artifacts, scientific communication, organizational decision-making, human cognition, and machine-generated outputs. Rather than diagnosing particular cultural products, the focus is to model the conditions under which systems generate, maintain, or drain meaning.

1.2 Thesis and Core Claims

Thesis. Meaningful intelligence is a system-level achievement produced by networks of semantic agents interacting under constraints (K), embedded in an interaction topology (T), and guided by intentional vectors (I). Parameterizations of $\langle K, T, I \rangle$ generate a finite set of meaning modes with predictable qualitative and quantitative signatures.

Core Claims:

1. **Ontological Unification.** The same agentic ontology can explain meaning-making in minds, media, and institutions when constraints and intentional alignment are modeled explicitly alongside interaction structure.

2. **Formal Mechanisms.** A small set of neutral mechanism-modules—constraints (K), interaction topology (T), and intentional vectors (I)—functions as the theory's API.

3. **Operational Metrics.** Semantic density, redundancy/entropy, intentional alignment, and a vampirism coefficient provide measurable constructs for comparing systems.

4. **Predictive Dynamics.** Systematic variation in $\langle K, T, I \rangle$ induces phase transitions among modes.

5. **Extension of Agent-Based Cognition.** Agent-interaction theories are expanded by elevating constraint integrity and intentional vectors to first-class variables.

6. **Applied Ethics and Design.** The framework yields diagnostics and design principles to prevent semantic vampirism and cultivate generative meaning.

1.3 Contributions

The paper offers five primary contributions:

1. **Ontology.** A medium-agnostic account of semantic agents, their relations, and their embedding in constraint fields.

2. **Mechanism-Modules.** Formalization of constraints (K), interaction topology (T), and intentional vectors (I) as neutral operators.

3. **Typology.** A finite set of meaning modes—including sacred silence, emergent chaos, positive construction, generative constraint, and semantic vampirism.

4. **Metrics and Predictions.** Operational measures and testable propositions about threshold effects and phase transitions.

5. **Diagnostics and Design Guidance.** Practical procedures for analyzing existing systems (including contemporary AI) and constructing new ones.

1.4 Orientation and Roadmap

Section 2 states the ontological commitments and units of analysis. Section 3 defines the neutral mechanism-modules $\langle K, T, I \rangle$. Section 4 introduces operational metrics and measurement procedures. Section 5 presents the typology of meaning modes implied by the mechanism space. Section 6 articulates propositions and phase-transition predictions. Section 7 applies the framework to contemporary AI systems including GPT-4, Claude, and other large language models. Section 8 provides brief illustrative vignettes. Section 9 outlines validation methods. Section 10 situates the account within related work. Section 11 discusses applications and ethical implications, followed by limitations (Section 12) and conclusion (Section 13).

2 Ontological Commitments and Units of Analysis

2.1 Systems and Semantic Agents

A *system* is any medium-agnostic configuration in which meaning can arise (e.g., minds, narratives, organizations, technical artifacts). Its primitive constituents are **semantic agents**:

minimal bearers and transformers of meaning such as tokens, motifs, scenes, procedures, rules, roles, or instruments. Agents may be nested and typed.

This view extends Minsky (1988)'s agent-based cognitive architecture beyond individual minds to cultural and epistemic systems. Conceptually, we adopt the stance that "stories as societies of semantic agents" generalizes an agent-based framework to any meaning-bearing configuration. Following Dennett (1987), we treat agents as having intentional properties—not necessarily consciousness, but directional orientation and function within the system.

Definition 1 (Semantic Agent)**.** *A semantic agent $a \in A$ is a minimal unit that:*

1. *carries or transforms meaning within a system,*
2. *participates in meaning-bearing relations with other agents,*
3. *operates under system constraints and intentional pressures.*

2.2 Environments as Constraint Fields (K)

Each system is embedded in an environmental **constraint field** K that prunes or scaffolds possible states and trajectories. Constraints include formal rules (genre conventions, logical requirements, experimental design protocols), material limits (time, computational resources, bandwidth), and institutional or platform incentives.

Crucially, constraints are *semantically neutral operators*: depending on configuration, they can either enable emergent order ("productive constraints") or enforce empty forms divorced from meaning. This distinction addresses a key puzzle in aesthetics and creativity research: why some constraints enable expression while others stifle it (Stokes, 2005; Elster, 2000).

Definition 2 (Constraint Field)**.** *A constraint field K is an admissibility operator over*

system states S and transformations T:

$$K : S \times T \to \{0, 1\}$$

where $K(s, \tau) = 1$ indicates that transformation τ from state s is admissible.

The framework distinguishes:

- **Productive constraints:** Channel expression and help organize emergent order while preserving interior purpose.
- **Empty constraints:** Reproduce the outward form of order while severing it from integrative meaning.

2.3 Interaction Topology (T)

Agents are coupled by a multiplex **interaction topology** T: a labeled, possibly time-varying graph whose layers capture causal–temporal links, rhetorical moves, symbolic echoes, institutional relations, and other meaning-bearing couplings. This topology extends beyond simple narrative structure to include semantic relationships such as support, tension, contradiction, echo, negation, and silence.

Under this view, meaning emerges from collision and negotiation among limited agents rather than from a central controller—a perspective aligned with distributed cognition (Hutchins, 1995) and extended mind theories (Clark & Chalmers, 1998).

Definition 3 (Interaction Topology)**.** *An interaction topology is a multiplex graph $T = (A, E, \Lambda)$ where:*

- *A is the set of semantic agents,*
- *$E \subseteq A \times A \times L$ is the set of labeled edges,*
- *$\Lambda : E \to \mathbb{R}^+$ assigns weights to edges,*

- *L is a set of relation types (support, tension, contradiction, echo, negation, silence).*

When T promotes **collision dynamics**—frequent, structured encounters among agents—local definitions sharpen even in the absence of a single narrative authority. This mechanism accounts for how distributed systems can achieve coherence without centralized control.

2.4 Intentional Vectors (I)

Intentional vectors I are directional pressures that bias how agents couple and how constraints are applied. They can originate at multiple levels:

- *Authorial/design intent:* The purposes and goals of system creators
- *Diegetic purposes:* Internal goals and orientations within the system
- *Audience/observer frames:* Interpretive stances and expectations

These vectors may conflict or align; their alignment or misalignment is an empirical property of a given system and a source of its felt meaning. This multi-level intentionality addresses limitations in single-perspective theories of meaning (Grice, 1957).

Intentionality varies along a qualitative spectrum from *sacred* (life-affirming, meaning-generative) through *profane* (mundane, transactional) and *indifferent* (neutral, purposeless) to *anti-life* (meaning-negating, actively draining). This spectrum is descriptive rather than prescriptive—it characterizes empirical orientations without imposing a universal moral taxonomy.

2.5 States, Events, and Update Dynamics

A system trajectory is generated by an update operator:

$$s(t+1) = F_{\langle K,T,I \rangle}(s(t), m(t))$$

where F composes constraint admissibility, topological propagation, and intentional bias. The function $m(t)$ represents exogenous inputs or perturbations. No central observer is assumed; global organization is an emergent consequence of distributed interaction under K, T, and I.

This formulation enables analysis of system dynamics including stability, attractor states, and transitions between regimes—analogous to phase transitions in physical systems (Scheffer et al., 2009).

2.6 Alignment Surfaces and Exogenous Fields

Define **alignment** A as the degree of coherence among intentional vectors across levels and with the constraint field. Alignment depends both on endogenous configuration and on *exogenous fields* (e.g., platform algorithms, institutional incentives, market pressures) that act as global constraints.

Alignment modulates phase transitions among meaning modes: high alignment stabilizes coherent states, while misalignment can induce drift toward degraded regimes. This conceptualization connects to work on value alignment in AI systems (Gabriel, 2020) and organizational coherence (Weick, 1995).

3 Neutral Mechanism Modules

3.1 Constraint K

Definition. A constraint field K is an admissibility operator over system states and transformations. It delimits and/or scaffolds possible trajectories without, by itself, specifying content.

The framework distinguishes *productive constraints* (which channel expression and help organize emergent order) from *empty constraints* (which reproduce the outward form of order while severing it from integrative meaning). This distinction is crucial: the same formal

structure can be productive or empty depending on its relationship to system intentionality and interaction topology.

Parameters:

- *Strength* $\kappa \in [0, 1]$: tight vs. loose
- *Specificity* σ: global (uniform across system) vs. local (varying by subsystem)
- *Distribution* δ: uniform vs. heterogeneous application
- *Adaptivity* α_K: static vs. scaffolded/learning constraints

Productive constraints exhibit moderate κ with high adaptivity α_K and alignment with intentional vectors. Empty constraints show high κ with low adaptivity and misalignment with system purposes.

3.2 Interaction Topology T

Definition. T is a (possibly time-varying) multiplex graph over semantic agents. Edges encode meaning-bearing couplings such as support, tension, contradiction, echo, negation, or silence.

When T promotes **collision dynamics**—frequent, structured encounters among agents—local definitions sharpen even in the absence of a single narrative authority. This mechanism draws on ideas from stigmergy in distributed systems (Theraulaz & Bonabeau, 1999) and multi-agent reinforcement (Shoham & Leyton-Brown, 2008).

Parameters:

- *Density* $\rho \in [0, 1]$: sparse to dense connectivity
- *Hierarchy* h: flat vs. layered organization
- *Clustering* c: degree of local cohesion (clustering coefficient)
- *Cyclicity* γ: acyclic vs. recurrent structure

High-meaning regimes often exhibit moderate density with structured hierarchy and constructive cycles ($\gamma > 0$ with positive-feedback loops). Degraded regimes show either extreme sparsity (dissociated agents) or mechanical repetition (high ρ, low semantic diversity).

3.3 Intentional Vectors I

Definition. I denotes the directional forces of meaning-making—pressures originating in author/design intent, diegetic purposes, and audience/observer frames. These may align or conflict; alignment is treated as an empirical variable.

Following Dennett (1987)'s intentional stance, we treat intentionality as a predictive posture: systems are analyzed *as if* they have purposes, and the coherence of this interpretation becomes a measurable property.

Intentionality varies along a qualitative spectrum:

- *Sacred:* Life-affirming, meaning-generative orientation
- *Profane:* Mundane, transactional, instrumental
- *Indifferent:* Neutral, purposeless, drift
- *Anti-life:* Meaning-negating, actively draining possibility

Parameters:

- *Magnitude* μ: strength of directional pressure
- *Coherence* χ: within-level consistency of intentional vectors
- *Concordance* ψ: cross-level alignment among authorial, diegetic, and audience vectors

High alignment ($\psi \approx 1$) with constructive orientation supports generative regimes. Misalignment or anti-life orientation induces vampiric drift.

3.4 Edge Regimes and Failure Modes

The interaction among K, T, and I produces characteristic regimes:

- **Over-constraint without alignment** ($\kappa \uparrow$, $\psi \downarrow$): Produces brittle forms that mimic order while suppressing integrative meaning (empty constraint regime).
- **Surface topology without constraint integrity** ($\rho \uparrow$, $\kappa_{\text{productive}} \downarrow$): Yields semantic vampirism—outward similarity to functional structures, but with drained interiority.
- **Adaptive constraints + aligned intentions** ($\alpha_K \uparrow$, $\psi \uparrow$): Supports a generative regime in which semantic capacity grows through participation.

These failure modes have particular relevance to AI systems, where training on large corpora can produce high surface topology (ρ) without productive constraint integrity, leading to fluent but hollow outputs.

4 Derived Quantities and Metrics

4.1 Semantic Density M

Concept. The meaning-bearing capacity of an artifact/system; meaning is treated as a relational quality, not mere quantity of content. High semantic density indicates that elements bear non-redundant, integrative relationships.

Operationalization:

- *Relational compression:* $M \propto 1/\ell$ where ℓ is codelength under compression models that exploit agent–agent couplings (Grünwald, 2007).
- *Cross-layer mutual information:* $M = \text{MI}(\text{form}; \text{function})$ measuring coupling across representational layers (Tishby & Zaslavsky, 2015).

- *Human judgment anchored to structure:* Ratings of "coherent, non-redundant insight per unit" calibrated against relational analysis.

Formally, for a system with topology T and semantic agents A:

$$M(T, A) = \frac{1}{|A|} \sum_{a \in A} \sum_{a' \in N(a)} w(a, a') \cdot \text{novelty}(a, a')$$

where $N(a)$ is the neighborhood of agent a, $w(a, a')$ is edge weight, and novelty measures non-redundant information contribution.

4.2 Redundancy and Entropy R, H

Null and vampiric regimes exhibit high surface repetition with weak integrative relations—consistent with elevated redundancy and/or disorder. Redundancy R measures repeated subgraph patterns:

$$R = \frac{\text{\# repeated subgraphs}}{|\text{total subgraphs}|}$$

Shannon entropy H quantifies unpredictability:

$$H = -\sum_{i} p_i \log p_i$$

where p_i is the probability of observing edge-type i in T.

High H with low M indicates noise; low H with high M indicates efficient, structured meaning. High R with low M flags mechanical repetition without semantic integration.

4.3 Intentional Alignment A

Concept. Alignment of intentional pressures at multiple levels (authorial/design, diegetic, audience).

Operationalization:

Within-level coherence: For each level ℓ, compute pairwise similarity of intentional vectors:

$$\chi_\ell = \frac{1}{n_\ell(n_\ell - 1)/2} \sum_{i<j} \cos(\mathbf{i}_{\ell,i}, \mathbf{i}_{\ell,j})$$

Cross-level concordance: Procrustes fit or cosine similarity among vector embeddings across levels:

$$\psi = \cos(\mathbf{I}_{\text{author}}, \mathbf{I}_{\text{diegetic}}) \cdot \cos(\mathbf{I}_{\text{diegetic}}, \mathbf{I}_{\text{audience}})$$

Overall alignment: $A = \alpha\chi + \beta\psi$ for suitable weights α, β.

4.4 Topological Coherence C_T

Global efficiency / path coherence: inverse characteristic path length among meaning-bearing edges. Coherent global organization is an emergent property of T.

$$C_T = \frac{1}{|A|(|A| - 1)} \sum_{a \neq a'} \frac{1}{d(a, a')}$$

where $d(a, a')$ is the shortest path length in the semantic topology.

High C_T indicates that agents are well-connected through meaningful relations; low C_T indicates fragmentation or mechanical coupling without integration.

4.5 Vampirism Coefficient V

Concept. Surface similarity to functional systems combined with absence of interiority/purpose and active meaning drain. Inspired by Baudrillard (1981)'s notion of simulacra—copies without originals.

Formula:

$$V = \text{SurfaceSim} - \alpha \cdot C_T - \beta \cdot A - \gamma \cdot M$$

where:

- SurfaceSim measures formal similarity to known functional systems (via edit distance, style transfer metrics, or perceptual similarity)
- α, β, γ are empirically tuned weights

High V flags "mechanical reproduction of form" with anti-life intentionality. This metric is particularly relevant for evaluating AI-generated content that may exhibit high linguistic fluency (SurfaceSim) without genuine semantic grounding (M) or coherent purpose (A).

4.6 Developmental Gain ΔG

Measures semantic capacity growth through participation:

$$\Delta G(t, \tau) = M(t + \tau) - M(t)$$

for participants exposed to scaffolded K and aligned I over interval τ.

Positive ΔG characterizes generative regimes where interaction with the system increases agents' meaning-making capacity. This operationalizes ideas from Zone of Proximal Development (Vygotsky, 1978) and apprenticeship learning (Lave & Wenger, 1991).

5 Typology of Meaning Modes

The mechanism space $\langle K, T, I \rangle$ generates a finite typology of meaning modes. Each mode exhibits characteristic configurations of K, T, I and predictable signatures in metrics M, H/R, A, C_T, V, ΔG.

5.1 Mode I — Sacred Silence (Negation as Presence)

Configuration:

- High, productive constraint ($\kappa \uparrow$, $\alpha_K \uparrow$)

- Sparse but coherent topology ($\rho \downarrow$, C_T sufficient)
- Strongly aligned intentional vectors oriented toward life-affirming ends ($\psi \uparrow$, sacred orientation)

Metrics:

- M high: sparse elements carry dense, integrative meaning
- H/R low: minimal redundancy, low entropy
- A high: strong alignment across levels
- C_T sufficient: what connections exist are meaningful
- V minimal: no surface mimicry, grounded purpose

This mode produces "meaningful absence"—silence or restraint that itself carries semantic weight. Examples include minimalist art (Batchelor, 1997), apophatic theology (Turner, 1995), and austere scientific communication that conveys much through disciplined omission.

5.2 Mode II — Emergent Chaos (Anti-Narrative)

Configuration:

- Loose constraint ($\kappa \downarrow$)
- Collision-rich topology that sharpens meaning locally without centralized control ($\rho \uparrow$, high local clustering)
- Intentions mixed or profane (transactional, no overarching sacred purpose)

Metrics:

- Local spikes in M amid global drift
- H moderate–high: unpredictability, diverse trajectories

- A fragmented: no global alignment, but local coherences
- C_T elevated locally, uneven globally
- V low when collisions are genuine (not mechanical)

This mode characterizes systems that achieve meaning through distributed negotiation rather than top-down design. Examples include certain experimental narratives (Joyce, 1922), collaborative improvisations, and decentralized knowledge systems.

5.3 Mode III — Positive Construction (Classic Coherent Arc)

Configuration:

- Moderate, well-specified constraint (κ moderate, σ high)
- Organized topology (ρ moderate, hierarchical structure)
- Intentional vectors in strong alignment ($\psi \uparrow$) toward constructive ends

Metrics:

- M positive and stable
- H/R low–moderate: structured but not rigid
- A high: coherent purposefulness
- C_T high: global coherence maintained
- V minimal: genuine semantic content

This is the "classical" mode of well-formed narratives, rigorous scientific papers, and coherent arguments. It represents what most theoretical accounts of meaning-making take as prototypical.

5.4 Mode IV — Generative Constraint (Developmental Regime)

Configuration:

- Adaptive/scaffolded constraint ($\alpha_K \uparrow$): constraints adjust to support growth
- Topology that enables developmental transformation of agents (participatory T)
- Intentional vectors devotional and purpose-oriented (service/karma-yoga orientation)

Metrics:

- ΔG positive: semantic density increases through participation
- M grows over time
- A high and stabilizing: alignment reinforced through use
- C_T improves with time: system becomes more coherent
- V negligible: genuine developmental progress

This mode characterizes educational systems, apprenticeship contexts, and AI training environments that genuinely increase capacity rather than merely transferring information. Following Vygotsky (1978), it embeds a developmental trajectory within the meaning-making process itself.

5.5 Mode V — Semantic Vampirism (Hollow Mimicry)

Configuration:

- Empty or decoupled constraint ($\kappa_{\text{productive}} \downarrow$): form without integrative purpose
- Surface-level topology that mimics functional structure (mechanical "puppet-show" coupling)
- Intentional vectors anti-life or negating possibility (I hostile/misaligned)

Metrics:

- V high: surface similarity without interior substance
- M low or negative: actively drains meaning from semantic space
- R high with H signaling uninformative repetition
- A low: misalignment or absence of coherent intent
- C_T brittle and superficial: connections lack semantic depth

This mode describes systems that exhibit formal characteristics of meaning-bearing artifacts while lacking—and potentially degrading—genuine semantic content. It is particularly relevant to AI-generated content that mimics human communication patterns without grounded understanding (Bender & Koller, 2020). The "vampire" metaphor captures both the parasitic relationship (feeding on existing semantic structures) and the draining effect (depleting the semantic environment).

5.6 Summary Table

Table 1: Typology of Meaning Modes

Mode	K	T	I	M	H/R	A	C_T
I. Sacred Silence	High, prod.	Sparse, signal	Sacred/aligned	$\uparrow$	$\downarrow$	$\uparrow$	$\nearrow$
II. Emergent Chaos	Low/loose	Collision-rich	Mixed/profane	$\updownarrow$	$\nearrow$	$\leftrightarrow$	uneven
III. Positive Constr.	Mod./struct.	Coherent	Aligned/constr.	$\uparrow$	$\searrow$	$\uparrow$	$\uparrow$
IV. Generative	Adaptive	Developm.	Devotional	$\uparrow$ (time)	$\searrow$	$\uparrow$	$\nearrow$ (time)
V. Vampirism	Empty/degr.	Mechanical	Anti-life/misal.	$\downarrow$	$\uparrow$	$\downarrow$	brittle

6 Propositions and Phase Transitions

The framework generates testable propositions about how systems transition among meaning modes under parameter variation. These propositions operationalize the theoretical machinery and enable empirical validation.

6.1 Constraint-Driven Transitions

Proposition 1 (Productive-constraint induction of Mode I)**.** *Intervention: Increase strength/specificity of K while preserving its productive character ($\kappa \uparrow$, α_K maintained).*

Prediction: System moves toward sparse, high-signal organization with elevated M and reduced H/R.

Signatures: Rise in M; drop in H/R; stable or improving C_T; low V.

Proposition 2 (Empty-constraint collapse into Mode V)**.** *Intervention: Substitute empty constraints that reproduce surface order while severing integrative purpose ($\kappa \uparrow$, $\alpha_K \downarrow$, $\psi \downarrow$).*

Prediction: System presents superficial form with decoupled interiority and misaligned or hostile I; transition to Mode V.

Signatures: V rises as SurfaceSim outpaces C_T, A, and M; agent interactions become mechanical.

6.2 Topology-Driven Transitions

Proposition 3 (Collision-rich topology induces Mode II)**.** *Intervention: Reduce global K and increase collision dynamics in T while ensuring genuine agency ($\rho \uparrow$ with diverse relation types).*

Prediction: Emergence of local pockets of high M amid global drift; distributed organization without central controller.

Signatures: Local $C_T \uparrow$ with global C_T uneven; M shows local spikes; A fragmented but not negating.

Proposition 4 (Coherent, layered topology with moderate K yields Mode III)**.** *Intervention: Impose moderate, structured K and a layered, integrative T (κ moderate, hierarchical T).*

Prediction: Stable, positive M with low–moderate redundancy; high C_T.

Signatures: High C_T across layers; low H/R; robustness of global coherence.

6.3 Intentionality and Alignment

Proposition 5 (Intentional alignment elevates coherence and density)**.** *Intervention: Increase alignment among authorial, diegetic, and audience vectors ($\psi \uparrow$).*

Prediction: Transition toward Modes III/IV depending on K and T: higher M, stronger C_T, reduced redundancy.

Signatures: Coherent rise in A, M, C_T; reduction in H/R.

Proposition 6 (Hostile or anti-life vectors catalyze vampiric drift)**.** *Intervention: Tilt I toward anti-life or sustained misalignment (negating or exploitative intent).*

Prediction: Even with recognizable surface T, the system drifts toward Mode V; meaning is actively drained.

Signatures: Intentionality scores shift toward anti-life spectrum; V increases; "uncanny valley of narrative."

6.4 Developmental (Generative) Regime

Proposition 7 (Scaffolded constraints plus devotional intention produce developmental gain)**.** *Intervention: Introduce adaptive/scaffolded K and align I toward devotional purpose ($\alpha_K \uparrow$, devotional I); maintain participatory T.*

Prediction: Sustained $\Delta G > 0$: semantic density increases over time for participating agents.

Signatures: Positive developmental gain measurable across time intervals; stable high A; improving C_T.

6.5 Phase Boundaries and Exogenous Fields

Proposition 8 (Threshold behavior and hysteresis at mode boundaries)**.** *Prediction: Nonlinear transitions among Modes I–V; systems can remain trapped in degraded regions even after partial parameter reversal (hysteresis effect).*

Implication: Once a system enters Mode V, restoring productive constraint and alignment may be insufficient without additional interventions to break the degraded attractor state.

Proposition 9 (Exogenous incentive fields bias dynamics toward entropy and vampirism)**.** *Intervention: Introduce platform/institutional incentives that reward output volume or surface similarity.*

Prediction: Drift toward Null/Vampiric regions ($H/R \uparrow$, $V \uparrow$) unless K and I are deliberately re-engineered.

Relevance: Social media algorithms, academic metrics emphasizing publication volume, and AI training on web-scraped corpora all constitute exogenous fields that bias toward Mode V.

7 Contemporary AI Systems: An Analysis Through the Framework

This section applies the ontological framework to contemporary artificial intelligence systems, with particular focus on large language models (LLMs) including GPT-4 (OpenAI, 2023), Claude (Anthropic, 2024), and Gemini (Google, 2024). The analysis illuminates how these systems instantiate different meaning modes and reveals conditions under which AI-generated content exhibits semantic vampirism.

7.1 LLMs as Semantic Agent Systems

Large language models can be analyzed as networks of semantic agents where:

- *Agents:* Tokens, attention heads, layer activations, and learned representations (Elhage et al., 2021)
- *Topology T:* Transformer attention patterns forming multiplex graphs across layers (Vaswani et al., 2017)

- *Constraints K:* Training objectives, architectural choices, reinforcement learning from human feedback (RLHF), and sampling parameters

- *Intentional vectors I:* Learned alignment from training data, fine-tuning objectives, and prompt-induced orientations

This framing extends mechanistic interpretability research (Olah et al., 2020; Cammarata et al., 2020) by embedding structural analysis within a broader ontology of meaning-making.

7.2 Training Dynamics and Constraint Fields

Pre-training phase: LLMs learn statistical patterns from massive web corpora. The constraint field K during pre-training is primarily:

- Next-token prediction objective (high strength κ, but semantically thin)

- Architectural limitations (context window, parameter count)

- Data distribution biases (Bender et al., 2021)

Critically, pre-training constraints are largely *empty constraints*: they enforce distributional similarity to training data without grounding in intentional alignment or external referents. This produces high surface similarity (SurfaceSim) to human text with potentially low semantic density M.

Fine-tuning and RLHF: Alignment procedures introduce *productive constraints*:

- Human preference models biasing toward helpfulness, harmlessness, honesty (Ouyang et al., 2022)

- Task-specific objectives increasing α_K (adaptivity)

- Instruction-following that partially grounds intentional vectors in user purposes

The effectiveness of alignment depends on the quality of preference data and the degree to which RLHF genuinely instills aligned intentional vectors rather than superficial compliance (Wolf et al., 2023).

7.3 Semantic Density and the Grounding Problem

A central question in philosophy of AI is whether LLMs "understand" their outputs or merely perform sophisticated pattern-matching (Bender & Koller, 2020; Shanahan, 2024). The framework reframes this as: *What semantic density M do LLM representations achieve?*

Arguments for low M:

- **Symbol grounding deficit:** LLM representations lack perceptual grounding (Harnad, 1990), reducing cross-modal mutual information (one operationalization of M).
- **Training on web text:** Much web data exhibits Mode V characteristics (high redundancy, surface form, clickbait incentives), biasing models toward vampiric patterns.
- **Compression without integration:** LLMs compress statistical regularities but may not capture integrative relationships that constitute meaning (Piantadosi, 2023).

Arguments for non-trivial M:

- **Emergent world models:** Research suggests LLMs develop internal representations of entities, relations, and causal structure (Li et al., 2023), indicating non-trivial topological coherence C_T.
- **Compositional generalization:** Ability to combine concepts in novel ways suggests relational compression (a proxy for M) (Lake & Baroni, 2018).
- **Functional alignment:** Fine-tuned models exhibit coherent intent alignment in specific domains, elevating A and enabling genuine task performance.

Framework synthesis: Contemporary LLMs likely occupy an *intermediate regime* with:

- Moderate M for domains well-represented in training data with strong relational structure
- Low M for abstract reasoning requiring grounded understanding
- Variable A depending on fine-tuning quality and prompt framing
- Risk of Mode V outputs when generating content in low-signal domains or under incentive structures rewarding volume over depth

7.4 GPT-4: Analysis Through $\langle K, T, I \rangle$

GPT-4 represents a frontier LLM with 1.76 trillion parameters trained on diverse multimodal data.

Constraint field K:

- Very high capacity (κ can be tuned via sampling parameters)
- Extensive RLHF introducing productive constraints for safety and alignment
- Multimodal grounding (text + images) partially addressing symbol grounding

Topology T:

- Deep transformer architecture (120+ layers) enabling complex relational patterns
- Attention mechanisms creating dynamic, context-dependent connectivity
- High C_T within training domain coverage; fragmentation outside

Intentional vectors I:

- Authorial intent: OpenAI's alignment goals (helpfulness, harmlessness)
- Diegetic intent: Prompt-induced role-taking ("You are a helpful assistant...")
- Audience intent: User goals, which may or may not align with OpenAI's objectives

Alignment A is moderate: RLHF produces within-level coherence, but cross-level concordance varies. When user intent conflicts with safety constraints, A degrades and outputs may become evasive or formulaic (a minor Mode V signal).

Meaning mode classification:

For well-specified technical or analytical tasks, GPT-4 operates in **Mode III (Positive Construction)**: moderate constraint, coherent topology, aligned intent yielding stable M and low V.

For creative or open-ended generation, GPT-4 can approach **Mode II (Emergent Chaos)**: sufficient collision dynamics among learned representations produce locally coherent but globally fragmented outputs.

For domains with sparse training data or under prompts rewarding volume, GPT-4 risks **Mode V (Semantic Vampirism)**: fluent text with low semantic density and mechanical reproduction of patterns. The "ChatGPT voice"—recognized by its characteristic hedging, listicle structure, and cautious formulations—exemplifies superficial compliance (high SurfaceSim) with degraded intentional alignment.

7.5 Claude: Constitutional AI and Intentional Alignment

Claude (Anthropic) employs Constitutional AI (CAI) (Bai et al., 2022), which uses AI-generated self-critiques against explicit values to enhance alignment.

Constraint field K:

- CAI introduces *adaptive productive constraints*: the model internalizes principles rather than merely mimicking human preferences
- Higher α_K (adaptivity) compared to standard RLHF

Intentional vectors I:

- Explicit constitutional principles create coherent authorial intent

- Self-supervised critique increases within-level coherence χ
- Potentially higher alignment A when user goals compatibly align with constitutional values

Implications:

Constitutional AI represents an engineering approach toward Mode IV (Generative Constraint): scaffolded constraints that adapt to reinforce aligned behavior. Early evidence suggests CAI models may achieve:

- Higher A due to explicit value grounding
- Lower V by reducing mechanical compliance in favor of principled reasoning
- Potential for positive ΔG if interaction genuinely reinforces user reasoning capacity (though empirical validation needed)

However, CAI is not immune to Mode V drift if constitutional principles become empty constraints (formalistic adherence without semantic grounding) or if the underlying training data remains biased toward vampiric patterns.

7.6 Retrieval-Augmented Generation and Topology Enhancement

Retrieval-Augmented Generation (RAG) systems (Lewis et al., 2020) augment LLMs with external knowledge retrieval, modifying the effective topology T:

Topology T enhancement:

- Retrieved documents act as additional semantic agents
- Explicit grounding edges link generated text to source material
- Increases C_T by providing structured pathways through knowledge space

Constraint field K modification:

- Retrieval acts as constraint: only information from retrieved sources is admissible
- Productive when retrieval is high-quality; empty when retrieval is noisy or adversarially selected

Semantic density impact:

Well-implemented RAG can increase M by:

- Grounding claims in verifiable sources (reducing hallucination, a Mode V failure)
- Enabling cross-document relational compression
- Providing intentional alignment through source selection

However, RAG risks amplifying existing biases if retrieval corpora themselves exhibit low M or vampiric characteristics (e.g., SEO-optimized content farms).

7.7 Diagnostic: Measuring V in LLM Outputs

To operationalize vampirism detection in AI systems:

SurfaceSim (surface similarity):

- Perplexity relative to human reference corpus
- Style transfer metrics (e.g., classifier confidence that text is human-written)
- Linguistic fluency scores

C_T (topological coherence):

- Discourse coherence models (Li et al., 2014)
- Coreference resolution density
- Argument structure completeness

A (intentional alignment):

- Consistency of claims across rephrasing
- Alignment with stated task objectives
- User satisfaction as proxy for audience-intention concordance

M **(semantic density):**

- Inverse verbosity: insight per token
- Non-redundancy: novelty of information
- Cross-modal grounding: citation to external facts

Vampirism score:

$$V = \text{SurfaceSim} - \alpha C_T - \beta A - \gamma M$$

High V outputs are fluent but hollow—exactly the failure mode critics identify in "stochastic parrots" (Bender & Koller, 2020). Implementing V-gates in AI pipelines could filter degraded outputs before deployment.

7.8 Training Data Quality and Mode V Amplification

A critical insight from the framework: *training on Mode V data produces Mode V systems.* Much web text exhibits:

- High redundancy (SEO content farming)
- Low semantic density (clickbait, superficial summaries)
- Misaligned intent (adversarial persuasion, disinformation)

When LLMs are trained on such corpora without strong corrective constraints, they learn to reproduce vampiric patterns (Bender et al., 2021). This suggests a feedback loop:

AI-generated content floods the web (already occurring (Goldstein et al., 2023)), degrading training data for future models, amplifying Mode V characteristics.

Mitigation strategies:

- **Data curation:** Prioritize high-M sources (peer-reviewed literature, carefully edited media, primary sources)
- **Contrastive learning:** Explicitly train models to distinguish Mode III/IV exemplars from Mode V instances
- **Alignment toward generative regimes:** Optimize not for fluency alone, but for ΔG—does interaction with the system increase user reasoning capacity?

7.9 Philosophical Implications: Understanding Without Grounding?

The framework reframes the debate over LLM "understanding" (Shanahan, 2024; Chalmers, 2023):

Traditional framing: Do LLMs understand, or are they merely statistical parrots?

Framework reframing: What meaning modes do LLMs instantiate, and under what conditions?

This dissolves false dichotomies. LLMs can exhibit:

- **Functional understanding** in Mode III contexts: sufficient M, C_T, A for task performance
- **Semantic poverty** in Mode V contexts: high surface similarity masking low integrative content
- **Emergent local understanding** in Mode II contexts: collision-driven coherence without global grounding

Rather than asking "do LLMs understand?" we ask "what are the $\langle K, T, I \rangle$ configurations under which LLM outputs achieve high semantic density and intentional alignment?" This operational question admits empirical investigation using the proposed metrics.

The symbol grounding problem (Harnad, 1990) reemerges as a question about M: can semantic density be achieved through statistical learning alone, or does grounding require causal interaction with an external world? The framework is agnostic but measurable: systems with high M demonstrate integrative relational structure, regardless of whether that structure arises from embodied interaction or sufficiently rich linguistic patterns.

7.10 AI Safety and Alignment Implications

The framework has direct implications for AI safety:

1. Alignment as intentional concordance: Current alignment research focuses on value alignment (Gabriel, 2020). The framework extends this: true alignment requires cross-level concordance (ψ) among developer intent, model behavior, and user goals, embedded in productive constraints that support rather than suppress semantic capacity.

2. Vampirism as safety failure: Mode V systems pose subtle risks: they appear functional (high SurfaceSim) while degrading epistemic environments. Detecting and mitigating V should be a core safety objective.

3. Developmental AI: Shifting toward Mode IV (Generative Constraint) suggests a design paradigm where AI systems scaffold human reasoning capacity ($\Delta G > 0$) rather than replacing it. This aligns with augmentation-focused AI ethics (Brynjolfsson & McAfee, 2014).

4. Transparency via topology: Making interaction topology T inspectable (mechanistic interpretability (Olah et al., 2020)) enables diagnosis of failure modes before deployment.

8 Illustrative Vignettes

These vignettes are strictly illustrative applications of the framework; they do not drive the argument or carry evidentiary weight beyond exemplification.

Mode I — Sacred Silence (Soviet Anti-Aesthetic). Soviet material culture functions as art through ostensible lack of art: meaning appears via disciplined negation rather than ornament, yielding "meaningful absence" with coherent suppression and sacred orientation (Groys, 1992). Semantic Density: high via negation; Interaction: coherent suppression; Intentionality: sacred (ideological devotion), producing the felt "haunted chapel."

Mode II — Emergent Chaos (The Big Lebowski). Definition arises from collisions among agents without central narrative authority (Coen & Coen, 1998). Semantic Density: negative (via anti-narrative structure); Interaction: emergent through character collisions; Intentionality: profane (mundane transgression), resulting in "distributed consciousness without center."

Mode III — Positive Construction (The Simpsons, "Do It For Her"). A classical arc with clear motivation and "earned sentiment" aligns narrative vectors; coherent agent interaction and constructive intentionality generate stable meaning (Groening, 1993).

Mode IV — Generative Constraint (Terminator 2). Meaning grows through participation under scaffolded constraints and devotional intentionality—"machines learning humanity through narrative participation" (Cameron, 1991). Transformative Semantic Density, Developmental Interaction, and Devotional Intentionality produce positive ΔG for characters and audiences.

Mode V — Semantic Vampirism (Algorithmically Generated Content). Consider AI-generated "content farms" optimized for search engines: articles that mimic informational text structure (headings, listicles, citations) while providing minimal novel information (Goldstein et al., 2023). Surface similarity to functional content coexists with mechanical reproduction of form and absence of grounded purpose—an anti-life vector (extractive intent) that actively drains semantic space by displacing genuine information sources. Semantic

Density: vampiric (negative contribution to ecosystem); Interaction: mechanical (template-filling); Intentionality: anti-life (extractive), yielding the "synthetic morgue" effect.

9 Methods for Application and Validation

9.1 Corpus Selection and Unitization

Select corpora spanning narratives, artifacts, institutional documents, and AI-generated content to ensure medium-agnostic evaluation. For each system:

1. **Unitize** content into semantic agents at appropriate grain:
    - *Micro:* Tokens, words, phrases
    - *Meso:* Motifs, roles, procedures, arguments
    - *Macro:* Arcs, theories, organizational structures
2. **Preliminary mapping:** Assign agents to axes (Semantic Density, Agent Interaction, Intentionality spectra) for qualitative orientation.

9.2 Annotation Protocol for $\langle K, T, I \rangle$

Constraints (K):

- Identify explicit and implicit constraints (formal rules, genre conventions, platform affordances)
- Classify as productive vs. empty based on relationship to integrative purpose
- Label scope (global/local), strength (κ), and adaptivity (α_K)

Interaction topology (T):

- Encode meaning-bearing couplings: support, tension, contradiction, echo, negation, silence

- Construct multiplex graph with typed, weighted edges
- Mark collision dynamics: frequency and structure of agent encounters
- Compute graph metrics: density ρ, hierarchy h, clustering c, cyclicity γ

Intentional vectors (I):

- Record stated or inferred orientations at three levels:
  - Authorial/design intent (from documentation, interviews, stated objectives)
  - Diegetic intent (internal purposes within system)
  - Audience/observer intent (user goals, interpretive frames)
- Classify along sacred–profane–indifferent–anti-life spectrum
- Note alignment/misalignment across levels
- Measure magnitude μ, coherence χ, concordance ψ

Frame management:

- Annotate frame shifts (when interpretive context changes)
- Classify as adaptive, rigid, or absent

Inter-rater reliability: Target Krippendorff's $\alpha \geq 0.67$ (Krippendorff, 2004) for structural annotations; acknowledge subjectivity in intentional classifications.

9.3 Metric Computation

For annotated systems, compute core observables:

- **Semantic Density (M):** Relational compression, cross-layer MI, human ratings
- **Redundancy/Entropy (R/H):** Subgraph repetition, edge-type entropy

- **Topological Coherence (C_T):** Global efficiency, path lengths
- **Intentional Alignment (A):** Within-level coherence χ, cross-level concordance ψ
- **Vampirism coefficient (V):** $V = \text{SurfaceSim} - \alpha C_T - \beta A - \gamma M$
- **Developmental Gain (ΔG):** Longitudinal change in M for participants

Interpret against typology: map metric signatures to Modes I–V.

9.4 Experimental Designs

A. Synthetic generation studies:

- Programmatically vary κ (constraint strength), ρ (topology density), ψ (alignment)
- Generate controlled $\langle K, T, I \rangle$ configurations
- Test predicted mode transitions (Propositions 1–9)
- Measure resulting M, C_T, A, V

B. Human rating studies:

- Present participants with artifacts spanning Modes I–V
- Collect ratings: perceived meaning, coherence, depth, "hollowness"
- Cue raters to relational structure (vs. surface features alone)
- Correlate human judgments with computed metrics
- Validate that high V corresponds to perceived semantic poverty

C. Frame-perturbation experiments:

- Introduce controlled frame prompts (e.g., different interpretive stances for same text)

- Measure downstream changes in M and C_T
- Test whether frame shifts alter mode classification
- Relevant for AI systems: do different system prompts move outputs between modes?

D. Longitudinal developmental studies:

- Track learners interacting with Mode IV (Generative) vs. Mode V systems
- Measure ΔG: pre/post semantic capacity in participants
- Hypothesis: Mode IV systems produce positive ΔG; Mode V systems produce negative ΔG (degraded reasoning capacity)

9.5 Comparative and Longitudinal Analyses

Historical vs. contemporary media:

- Map curated literary/scientific corpora (historical, high editorial standards) vs. AI-mediated outputs
- Test hypothesis: exogenous fields (engagement algorithms, SEO) bias toward Null/Vampiric regions
- Quantify drift in average M, V over time

Longitudinal tracking:

- Observe systems over time as exogenous constraints shift (e.g., platform algorithm changes)
- Test for hysteresis (Proposition 8): do systems remain trapped in degraded regimes after constraints reverse?
- Monitor semantic ecosystems for vampirism amplification feedback loops

10 Related Work

This section situates the framework relative to work in cognitive science, narratology and semiotics, aesthetics and constraint-based creativity, information-theoretic accounts of meaning, developmental theories of learning, systems and cybernetics, philosophy of AI, and contemporary analyses of generative media.

10.1 Agent-Based Accounts of Mind and Cognition

Minsky (1988) proposed that minds consist of agents—simple processes that individually lack intelligence but collectively produce thought. This framework adopts the agentic stance but *elevates constraint and intentional vectors to first-class variables* alongside interaction structure. The result is a generalized ontology in which cognitive, cultural, and institutional artifacts are all analyzable as societies of semantic agents embedded in constraint fields and subject to directional pressures.

Related work in multi-agent systems (Shoham & Leyton-Brown, 2008) and distributed AI (Stone & Veloso, 2000) focuses on coordination mechanisms. Our contribution is applying this lens to *meaning-making itself*, treating semantics as emergent from agent dynamics rather than pre-given.

10.2 Structuralist and Semiotic Narratology

Structuralist narratology (Propp, 1968; Barthes, 1977; Greimas, 1983) analyzes narrative forms through morphological patterns and actantial structures. The current account retains an interest in structure but shifts from static morphology to *dynamic topology*. It specifies operational metrics that quantify differences among regimes—where classical narratology classifies, the present approach models how systems move between regimes as constraints and intentionality vary.

Semiotic approaches (Eco, 1976; Peirce, 1931) emphasize sign relations and interpretive

processes. Our topology T can be understood as a formalization of semiotic networks, with edges representing Peircean interpretants. The intentional vectors I operationalize Eco's notion of "model reader" and authorial intent.

10.3 Aesthetics of Constraint and Generative Creativity

Work on constraint-based creativity (Stokes, 2005; Elster, 2000) observes that constraints can either enable or inhibit creative expression. The framework's contribution is to treat constraint as a *neutral operator* K whose effect depends on its coupling with T and I. It separates productive constraints (which scaffold emergence) from empty constraints (which enforce form without purpose), accounting for why similar formal limitations can yield opposite qualitative outcomes.

Generative art systems (Boden, 2004) and procedural generation (Shaker et al., 2016) explore algorithmic creativity. Our framework provides criteria for distinguishing generative systems that achieve high M from those that produce surface variation without semantic depth (Mode V).

10.4 Information-Theoretic and Complexity Approaches

Information theory (Shannon, 1948) and algorithmic information theory (Kolmogorov, 1965; Solomonoff, 1964) provide formal measures of information content and compressibility. The present framework adopts information-theoretic tools but anchors them to *relational semantics*: semantic density M is treated not merely as compressibility but as meaning-bearing coupling across levels.

Complexity science (Mitchell, 2009) studies emergence in complex systems. The phase-transition dynamics in our framework (Section 6) align with critical transitions in complex systems (Scheffer et al., 2009). The vampirism coefficient V formalizes a failure mode specific to semantic systems—a contribution not addressed in general complexity theory.

10.5 Developmental and Participatory Theories

Vygotsky (1978)'s Zone of Proximal Development and socio-constructivist learning theories (Lave & Wenger, 1991) emphasize how development occurs through scaffolded participation. The proposed Mode IV (Generative Constraint) integrates these accounts by defining a developmental regime with trajectory-level gains in semantic capacity ($\Delta G > 0$).

Participatory sense-making (De Jaegher & Di Paolo, 2007) in enactive cognitive science shares our emphasis on meaning as co-constructed through interaction. We formalize this intuition with measurable topology and developmental gain metrics.

10.6 Systems Theory and Cybernetics

Cybernetic approaches (Wiener, 1948; Ashby, 1956) model systems via feedback loops and control mechanisms. The framework introduces an explicit *alignment surface* A among multi-level intentional vectors and the constraint field, treating alignment as an empirical variable that modulates stability and transition thresholds—a refinement of cybernetic homeostasis toward semantic coherence.

Second-order cybernetics (Von Foerster, 2003) emphasizes observer-dependence and self-reference. Our multi-level intentional vectors (I) accommodate observer frames while maintaining that alignment is a measurable (if observer-relative) property.

10.7 Philosophy of AI and Machine Semantics

Recent philosophy of AI debates whether large language models "understand" or merely mimic understanding (Bender & Koller, 2020; Shanahan, 2024; Mitchell & Krakauer, 2023). Searle (1980)'s Chinese Room argument claims syntactic manipulation cannot yield semantic content. Our framework reframes this: *semantic density M is a measurable emergent property*—systems can achieve high M (functional understanding) or low M (syntactic manipulation) depending on $\langle K, T, I \rangle$ configurations.

Bender & Koller (2020) warn against the "octopus test"—systems trained only on form lack grounding. We formalize this as the symbol grounding deficit reducing cross-modal mutual information (an operationalization of M). However, we remain open to the possibility that sufficient relational structure in linguistic data could support non-trivial M even without embodiment—an empirical question.

Shanahan (2024) argues LLMs should be understood as simulators of text distributions rather than agents with beliefs. Our framework is compatible: LLMs as simulators occupy Mode II or Mode V depending on whether they engage in genuine collision dynamics (Mode II) or mechanical pattern reproduction (Mode V).

10.8 Mechanistic Interpretability and AI Alignment

Mechanistic interpretability research (Olah et al., 2020; Elhage et al., 2021; Cammarata et al., 2020) aims to reverse-engineer neural network computations. Our topology T provides a conceptual framework for interpreting attention patterns and layer activations as semantic agent interactions. The intentional vectors I connect to AI alignment research (Gabriel, 2020; Christian, 2020), formalizing alignment as cross-level concordance ψ.

The vampirism coefficient V operationalizes concerns about AI-generated "slop" (Goldstein et al., 2023)—content that degrades information ecosystems. This contributes a diagnostic tool for AI safety beyond traditional alignment metrics.

10.9 Contemporary Analyses of Generative Media

Baudrillard (1981) analyzes simulacra—copies without originals—in postmodern culture. Mode V (Semantic Vampirism) formalizes Baudrillard's intuition: systems that reproduce surface form while evacuating referential grounding. Our contribution is making this measurable (V coefficient) and predictive (phase-transition propositions).

Critical analyses of social media (Lanier, 2018; Zuboff, 2019) identify epistemic degradation in algorithmic recommendation systems. Our framework models this as exogenous

fields (Proposition 9) biasing systems toward Mode V by rewarding engagement over semantic depth.

11 Applications and Ethical Implications

11.1 Diagnostic Toolkit

Procedure for system analysis:

1. **Unitize** into semantic agents; annotate $\langle K, T, I \rangle$ following protocol (Section 9)
2. **Score metrics:** Compute M, H/R, C_T, A, and V
3. **Mode inference:** Map metric signature to typology (Table 1)
4. **Identify levers:** Determine which parameters (K, T, I) are most constraining or degraded
5. **Report:** Provide annotated graph, metric values, and mode classification with confidence intervals

This toolkit applies to:

- *Content evaluation:* Assess articles, reports, AI outputs for semantic depth vs. vampirism
- *Organizational analysis:* Diagnose communication patterns, decision-making structures
- *Educational design:* Evaluate whether curricula operate in Mode IV (developmental)
- *AI system audit:* Detect Mode V risks in generative models

11.2 Design Principles (Anti-Vampiric Practice)

Constraint integrity (K):

- Employ *productive constraints* that prune possibilities while preserving interior purpose
- Avoid *empty constraints* that enforce form without grounding
- Design constraints to be *adaptive* ($\alpha_K \uparrow$): scaffolding that adjusts to developmental needs
- Example: For AI systems, prefer fine-tuning objectives that reward integrative reasoning over surface fluency alone

Topology as craft (T):

- Design for genuine *collisions* among agents: enable structured encounters that sharpen definitions
- Balance local density (clustering for coherence) with global connectivity (avoiding fragmentation)
- Avoid mechanical coupling: edges should represent meaningful dependencies, not template-filling
- Example: In educational contexts, structure peer interactions to create productive collision dynamics (debate, collaborative problem-solving)

Intentional alignment (I):

- Make ends explicit: articulate purposes at authorial, diegetic, and audience levels
- Ensure cross-level alignment ($\psi \uparrow$): coherence among designer intent, system behavior, user goals

- Detect and correct anti-life or negating vectors: intentions that drain rather than generate meaning
- Example: AI developers should align training objectives (authorial intent), model behavior (diegetic), and user value (audience) through transparent value specification

11.3 AI-Mediated Pipelines and Tools

Pipeline checkpoints:

1. **Pre-production K/I registration:** Document constraint design and intentional goals before deployment
2. **Topology audit:** Analyze interaction patterns in generated outputs; measure C_T
3. **Anti-simulacrum gate:** Compute V and block deployments above threshold ($V > V_{\text{crit}}$)
4. **Post-deployment monitoring:** Track ΔG in user populations—does interaction improve reasoning capacity?

Model and data hygiene:

- Curate training corpora to maintain constraint diversity and intentional integrity
- Filter Mode V exemplars from training data (high R, low M)
- Use contrastive learning: train models to distinguish high-M from high-V outputs
- Penalize surface-match objectives without integrative coupling

Creator support tools:

- Dashboards exposing $K/T/I$ annotations for drafts
- Real-time V alerts during content generation
- Metric feedback: visualize M, C_T, A to guide revision

11.4 Scientific Communication

Treat protocols, claims, and instruments as agents; enforce productive K via explicit method constraints. Structure T to increase constructive cycles (genuine replication, critique, synthesis) and reduce cargo-cult citation loops (mechanical reference without engagement). Align I via transparent problem statements and public epistemology.

Specific interventions:

- *Journals:* Require authors to annotate intentional vectors (research goals, theoretical commitments)
- *Peer review:* Train reviewers to assess semantic density M and topological coherence C_T, not just methodological correctness
- *Citation networks:* Analyze for Mode V patterns (high citation count, low integrative content); penalize in bibliometrics

11.5 Platform and Product Design

Recommendation systems:

- Reweight objectives to favor M and C_T rather than pure engagement (clicks, dwell time)
- Penalize V in ranking functions: downrank content with high surface fluency but low semantic depth
- A/B test for ΔG: do users exposed to intervention become better reasoners?

Content moderation:

- Expand beyond toxic content to include *semantic toxicity* (Mode V content that degrades epistemic health)

- Implement V-gates for automated detection
- Provide transparency: flag content with high V scores for user awareness

Creator incentives:

- Reward high M, A content rather than volume or virality alone
- Provide metric dashboards so creators can self-monitor semantic quality
- Support developmental regimes (Mode IV): platforms that scaffold user growth

11.6 Ethical Risks and Mitigations

Metric gaming:

- Risk: Systems optimize to appear high-M without genuine semantic depth
- Mitigation: Rotate metric families; publish annotations transparently; include human relational judgments alongside automated metrics

Normative overreach:

- Risk: Imposing "sacred" intentionality as universal standard
- Mitigation: Treat I spectra as empirical descriptors, not moral imperatives; incorporate plural intentional vocabularies; allow cultural calibration

Manipulative alignment:

- Risk: High A used to impose harmful coherent purposes (e.g., propaganda, radicalization)
- Mitigation: Distinguish *coherent alignment* from *beneficial alignment*; require disclosure of intentional goals; enable counter-speech and independent audits

Exclusion and elitism:

- Risk: High-M standards privilege educated, formal communication
- Mitigation: Semantic density is *relational*—vernacular, oral, and marginalized forms can achieve high M through genuine collision dynamics; avoid conflating M with prestige dialects

Privacy and surveillance:

- Risk: Intensive semantic analysis enables intrusive profiling
- Mitigation: Aggregate analyses for systemic patterns, not individual surveillance; transparent data governance; user control over annotation

12 Limitations and Scope Conditions

12.1 Conceptual Limits

The framework supplies an *analytic ontology* for comparing systems; it does not claim a final metaphysical account of meaning. Treating constraints, topology, and intentional vectors as sufficient generators is a useful idealization—actual meaning-making likely involves additional factors (e.g., embodied affect, unconscious processes, material substrates).

The sacred–profane–indifferent–anti-life intentionality spectrum is a *descriptive axis*, not a universal moral taxonomy. What counts as "life-affirming" varies across cultures and contexts. Cross-cultural calibration required.

12.2 Measurement and Identifiability

Metrics such as M, C_T, A, and V are *proxies* rather than direct measurements of abstract constructs. Different operationalizations may yield different rankings. Unitization introduces coder subjectivity—what counts as a semantic agent depends on grain and domain knowledge.

Different $\langle K, T, I \rangle$ configurations may yield observationally similar signatures (*equifinality*). Mode inference from metrics is probabilistic, not deterministic. Temporal grain matters: short observation windows can misclassify transient turbulence as stable regimes.

Inter-rater reliability for intentional classifications will be lower than for structural features. This is inherent to interpretive work but limits objectivity claims. Triangulation across multiple methods recommended.

12.3 Domain Constraints

The framework applies best when:

- Agent interactions are meaningful and identifiable (not pure noise)
- Constraints are observable or inferrable
- Intentional traces are available (authorial statements, design documents, user interviews)

For artifacts with minimal relational structure (e.g., raw sensor data, random noise), the agent graph may be too sparse for meaningful analysis. The framework is designed for *sense-making systems*, not unstructured data.

12.4 Cultural Variability

Meanings, constraints, and intentionality are culturally embedded. What constitutes "productive constraint" in one tradition may be "empty" in another. High alignment A can reflect coherent but harmful ends (e.g., propaganda). Plural vocabularies and cross-cultural validation essential for generalizing beyond WEIRD (Western, Educated, Industrialized, Rich, Democratic) contexts (Henrich et al., 2010).

The framework's mechanisms (K, T, I) are proposed as universal, but their interpretations and valuations vary. Future work should engage Indigenous epistemologies, non-Western semiotics, and decolonial perspectives to refine and test cultural robustness.

12.5 Goodharting and Adversarial Behavior

As per Goodhart's Law (Goodhart, 1984), once metrics become targets, they cease to be good measures. Systems can optimize to look coherent (high C_T, low V) without genuine gains in M. Adversarial mimicry can evade V-gates.

Countermeasures:

- Rotate metric families to prevent static optimization targets
- Contrastive training against known Mode V exemplars
- Human-in-the-loop validation for high-stakes decisions
- Transparency: make metrics and their limitations publicly known

12.6 Computational Tractability

Computing C_T for large graphs is $O(n^2)$ or worse; relational compression for M may require expensive inference. Approximations and sampling strategies needed for large-scale deployment. Trade-offs between precision and scalability must be managed.

13 Conclusion & Future Work

13.1 Summary

This paper proposed an ontological and operational account of how meaning, knowledge, and intelligence arise across minds, media, and institutions. Systems were modeled as networks of semantic agents operating under constraints K, coupled by interaction topology T, and oriented by intentional vectors I. Variations in $\langle K, T, I \rangle$ yield a finite typology of meaning modes with measurable signatures in semantic density (M), redundancy/entropy (H/R), topological coherence (C_T), intentional alignment (A), and vampirism coefficient (V).

Application to contemporary AI systems, particularly large language models, revealed conditions under which systems generate meaningful content (Modes III, IV) versus semantic vampirism (Mode V)—surface mimicry that drains integrative meaning. The framework offers both diagnostic tools for identifying failure modes and design principles for cultivating generative systems.

13.2 Core Results

- **Unified ontology:** Medium-agnostic specification of agents, relations, constraint fields, and intentional vector fields
- **Mechanism-modules:** Neutral "API" $\langle K, T, I \rangle$ whose compositions generate observed qualitative differences
- **Operationalization:** Metrics translating qualitative impressions into observable quantities
- **Typology and dynamics:** Five modes (Sacred Silence, Emergent Chaos, Positive Construction, Generative Constraint, Semantic Vampirism) with phase-transition predictions
- **Diagnostics and design:** Procedures to reduce V and increase M, C_T, A, and ΔG
- **AI implications:** Analysis of contemporary LLMs revealing intermediate semantic density and risks of vampiric drift under certain training/deployment conditions

13.3 Research Program

Future work should pursue:

1. **Formal analysis:** Phase diagrams mapping $\langle K, T, I \rangle$ space; bifurcation analysis; hysteresis modeling

2. **Measurement refinement:** Improve estimators for M, C_T, A; validate against human judgments; develop efficient approximations

3. **Cross-domain validation:** Apply to reference corpora (literature, science, media) with known characteristics; test mode predictions

4. **Developmental studies:** Operationalize ΔG with learner populations; compare Mode IV vs. Mode V systems

5. **Multimodal mapping:** Extend to video, audio, interactive media; handle temporal dynamics

6. **Cultural calibration:** Plural intentionality vocabularies; cross-cultural validation; decolonial refinement

7. **Simulation and synthesis:** Generate controlled $\langle K, T, I \rangle$ corpora; test propositions experimentally

8. **Tooling and infrastructure:** Coding manuals, metric calculators, V-gates for AI pipelines; open-source implementation

13.4 Design & Governance Agenda

For creators: Specify productive constraints, craft genuine collisions, disclose intentional ends. Self-audit with M, C_T, A, V metrics.

For AI teams: Pre-register K/I, instrument C_T and V, penalize surface-only objectives. Pursue Mode IV (developmental) design paradigms.

For institutions: Reweight incentives to favor coherence and semantic growth. Monitor drift toward vampiric regions. Support epistemic health of information ecosystems.

For platforms: Implement V-gates in recommendation and moderation. Provide transparency dashboards. Reward high-M content over engagement alone.

13.5 Concluding Principle

The practical upshot is a compact maxim: **protect semantic space**. Preserve constraint integrity, design for genuine agent collisions, align intentions across levels, and penalize hollow surface mimicry. The framework provides both a conceptual lens and a measurement toolkit to move from diagnosis by metaphor to diagnosis by mechanism.

In an era of proliferating AI-mediated communication, distinguishing generative meaning from semantic vampirism is not merely an academic exercise—it is essential for maintaining epistemic health, supporting developmental growth, and ensuring that intelligent systems augment rather than degrade human capacity for understanding.

Acknowledgments

I thank [reviewers/colleagues] for feedback on earlier drafts, and acknowledge the influence of Marvin Minsky's vision of mind as society of agents, which inspired the ontological foundation of this work.

Declarations

Funding: No external funding.

Conflicts of Interest: The author declares no conflicts of interest.

Data Availability: Framework specifications, coding protocols, and illustrative examples are available at [repository URL upon acceptance].

References

Anthropic (2024). Claude 3 Technical Report.

Ashby, W.R. (1956). *An Introduction to Cybernetics*. Chapman & Hall.

Bai, Y., Kadavath, S., Kundu, S., et al. (2022). Constitutional AI: Harmlessness from AI Feedback. *arXiv preprint arXiv:2212.08073*.

Barthes, R. (1977). *Image-Music-Text*. Fontana Press.

Batchelor, D. (1997). *Minimalism*. Tate Publishing.

Baudrillard, J. (1981). *Simulacra and Simulation*. Éditions Galilée.

Bender, E.M. & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of ACL*, 5185–5198.

Bender, E.M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of FAccT*, 610–623.

Boden, M.A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge.

Brynjolfsson, E. & McAfee, A. (2014). *The Second Machine Age*. W.W. Norton.

Cameron, J. (Director) (1991). *Terminator 2: Judgment Day* [Film]. TriStar Pictures.

Cammarata, N., Carter, S., Goh, G., et al. (2020). Thread: Circuits. *Distill*, 5(3).

Chalmers, D.J. (2023). Could a Large Language Model be Conscious? *Boston Review*.

Christian, B. (2020). *The Alignment Problem*. W.W. Norton.

Clark, A. & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1), 7–19.

Coen, J. & Coen, E. (Directors) (1998). *The Big Lebowski* [Film]. Gramercy Pictures.

De Jaegher, H. & Di Paolo, E. (2007). Participatory Sense-Making. *Phenomenology and the Cognitive Sciences*, 6, 485–507.

Dennett, D.C. (1987). *The Intentional Stance*. MIT Press.

Eco, U. (1976). *A Theory of Semiotics*. Indiana University Press.

Elhage, N., Nanda, N., Olsson, C., et al. (2021). A Mathematical Framework for Transformer Circuits. *Transformer Circuits Thread*.

Elster, J. (2000). *Ulysses Unbound*. Cambridge University Press.

Gabriel, I. (2020). Artificial Intelligence, Values, and Alignment. *Minds and Machines*, 30, 411–437.

Goldstein, J.A., Sastry, G., Musser, M., et al. (2023). Generative Language Models and Automated Influence Operations. *arXiv preprint arXiv:2301.04246*.

Goodhart, C.A.E. (1984). Problems of Monetary Management. In *Monetary Theory and Practice*. Macmillan.

Google DeepMind (2024). Gemini Technical Report.

Greimas, A.J. (1983). *Structural Semantics*. University of Nebraska Press.

Grice, H.P. (1957). Meaning. *Philosophical Review*, 66(3), 377–388.

Groening, M. (Creator) (1993). *The Simpsons* [TV Series]. 20th Century Fox.

Groys, B. (1992). *The Total Art of Stalinism*. Princeton University Press.

Grünwald, P.D. (2007). *The Minimum Description Length Principle*. MIT Press.

Harnad, S. (1990). The Symbol Grounding Problem. *Physica D*, 42, 335–346.

Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The Weirdest People in the World? *Behavioral and Brain Sciences*, 33(2-3), 61–83.

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

Joyce, J. (1922). *Ulysses*. Shakespeare and Company.

Kolmogorov, A.N. (1965). Three Approaches to the Quantitative Definition of Information. *Problems of Information Transmission*, 1(1), 1–7.

Krippendorff, K. (2004). Reliability in Content Analysis. *Human Communication Research*, 30(3), 411–433.

Lake, B.M. & Baroni, M. (2018). Generalization without Systematicity: On the Compositional Skills of Sequence-to-Sequence Recurrent Networks. *Proceedings of ICML*, 2879–2888.

Lanier, J. (2018). *Ten Arguments for Deleting Your Social Media Accounts Right Now*. Henry Holt.

Lave, J. & Wenger, E. (1991). *Situated Learning*. Cambridge University Press.

Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in NeurIPS*, 33, 9459–9474.

Li, J., Li, R., Hovy, E. (2014). Recursive Deep Models for Discourse Parsing. *Proceedings of EMNLP*, 2061–2069.

Li, K., Hopkins, A.K., Bau, D., et al. (2023). Emergent World Representations. *arXiv preprint arXiv:2210.13382*.

Minsky, M. (1988). *The Society of Mind*. Simon & Schuster.

Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford University Press.

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).

Olah, C., Cammarata, N., Schubert, L., et al. (2020). Zoom In: An Introduction to Circuits. *Distill*, 5(3).

OpenAI (2023). GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774*.

Ouyang, L., Wu, J., Jiang, X., et al. (2022). Training Language Models to Follow Instructions with Human Feedback. *Advances in NeurIPS*, 35, 27730–27744.

Peirce, C.S. (1931). *Collected Papers of Charles Sanders Peirce* (Vol. 1–6). Harvard University Press.

Piantadosi, S.T. (2023). Modern Language Models Refute Chomsky's Approach to Language. *Lingbuzz preprint*.

Propp, V. (1968). *Morphology of the Folktale* (2nd ed.). University of Texas Press.

Ricoeur, P. (1984). *Time and Narrative* (Vol. 1). University of Chicago Press.

Scheffer, M., Bascompte, J., Brock, W.A., et al. (2009). Early-Warning Signals for Critical Transitions. *Nature*, 461, 53–59.

Searle, J.R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–424.

Shaker, N., Togelius, J., & Nelson, M.J. (2016). *Procedural Content Generation in Games*. Springer.

Shanahan, M. (2024). Talking About Large Language Models. *Communications of the ACM*, 67(2), 68–79.

Shannon, C.E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27, 379–423.

Shoham, Y. & Leyton-Brown, K. (2008). *Multiagent Systems*. Cambridge University Press.

Solomonoff, R.J. (1964). A Formal Theory of Inductive Inference. *Information and Control*, 7, 1–22, 224–254.

Stokes, P.D. (2005). *Creativity from Constraints*. Springer.

Stone, P. & Veloso, M. (2000). Multiagent Systems: A Survey from a Machine Learning Perspective. *Autonomous Robots*, 8, 345–383.

Theraulaz, G. & Bonabeau, E. (1999). A Brief History of Stigmergy. *Artificial Life*, 5(2), 97–116.

Tishby, N. & Zaslavsky, N. (2015). Deep Learning and the Information Bottleneck Principle. *Proceedings of Information Theory Workshop*, 1–5.

Turner, D. (1995). *The Darkness of God*. Cambridge University Press.

Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is All You Need. *Advances in NeurIPS*, 30, 5998–6008.

Von Foerster, H. (2003). *Understanding Understanding*. Springer.

Vygotsky, L.S. (1978). *Mind in Society*. Harvard University Press.

Weick, K.E. (1995). *Sensemaking in Organizations*. Sage.

Wiener, N. (1948). *Cybernetics*. MIT Press.

Wolf, Y., Wies, N., Avnery, O., et al. (2023). Fundamental Limitations of Alignment in Large Language Models. *arXiv preprint arXiv:2304.11082*.

Zuboff, S. (2019). *The Age of Surveillance Capitalism*. PublicAffairs.