Renyxa Cognitive Inventory

Abstract

Modern language models operate as probabilistic text generators rather than systems of meaning. This paper introduces a deterministic alternative: a semantic-ontological framework derived from latent attractors in transformer embeddings and implemented in the Renyxa Cognitive Inventory (RCI).

Instead of generating or canonicalizing meaning, RCI records semantic structure as a topological field of stable attractors in embedding space. These attractors act as empirically grounded analogs of semantic primitives – statistical invariants that emerge across linguistic contexts. They support a three-layer architecture:

1. Ontology-Driven Scaffolding (ODS) for entity–relation encoding,
2. Inference-Free Semantic Notation (IFSN) for deterministic, reversible clause-level serialization,
3. RCI as a reconciliation engine integrating these records into coherent analytical models with full provenance.

Applied to corpora ranging from the *Book of Genesis* to humanitarian operations, legal, diplomatic, law enforcement, military domains, and especially, HUMINT/OSINT field documents, RCI reconstructs entity networks, timelines, and causal structures without stochastic decoding or consultant-authored ontologies. The framework repositions semantics from a problem of language modeling to one of geometric discovery in latent space, offering a reproducible basis for explainable, ontology-anchored intelligence.

Foreword

This work did not emerge from conventional computational linguistics. It arose from dissatisfaction with two dominant traditions:

- the symbolic paradigm, which tries to generate meaning from formal grammars and engineered ontologies;
- the statistical paradigm, which treats meaning as a side-effect of co-occurrence.

Both treat semantics as an artifact to be constructed.

Here we begin from a different premise: meaning is a physical invariant of the language manifold, observable as stable attractor regions in transformer embeddings. These attractors are the model's own empirical universals – learned from the training corpus, not invented. Anchoring ontology to them, the RCI establishes a deterministic bridge between linguistic form and conceptual structure.

RCI was developed not as another prompt-engineering trick, but as an innovation in scientific representation. Its success in reconstructing complex narrative and operational structure – regardless of the domain – indicates that attractor-based semantics can support autonomous, auditable reasoning. This paper is therefore both a technical report and a proposal for a new semantics of intelligence: an epistemology in which language, cognition, and geometry converge.

1 Introduction and Motivation

For over half a century, computational semantics has oscillated between symbolic formalism and statistical approximation. Symbolic frameworks (AMR, FrameNet, Semantic Web) posit predefined ontologies and logics; distributional methods and transformer LMs infer patterns from co-occurrence. Both assume meaning is constructed.

This paper challenges that assumption.

1.1 Limitations of Canonical Generation

Symbolic systems model meaning via compositional trees and hand-crafted roles. They are brittle, domain-bound, and often fail to recognize semantic equivalence across syntactic variation.

For example, "not all birds can fly" and "some birds are flightless" express the same proposition but typically yield non-isomorphic AMR graphs or RDF assertions. Since these systems operate over form and scope rather than latent structure, they do not converge on a unified representation. Their semantics is limited by their syntax.

1.2 From Primitives to Attractors

Osgood's semantic differential treated meaning as a location in continuous evaluative space; Wierzbicka's NSM postulated universal primitives. Transformer embeddings now provide an empirical substrate that partially realizes both.

Let $E \in \mathbb{R}^{|V| \times d}$E∈R∣V∣×d be the embedding matrix of a transformer model with vocabulary V V. For tokens $u,v \in V$u,v∈V, their embeddings $e(u), e(v)$e(u),e(v) define cosine similarity

$$s(u,v) = \frac{e(u) \cdot e(v)}{\|e(u)\| \|e(v)\|}, \text{s(u,v)=∥e(u)∥∥e(v)∥e(u)⋅e(v),}$$

an operational measure of semantic proximity rooted in the model's own attention mechanisms. Clustering in this space reveals semantic attractors: stable, coherent regions corresponding to recurrent concepts and roles. These serve as *empirical* semantic universals – emerging from training dynamics, not ontology committees.

1.3 A Deterministic Alternative

Architecturally RCI consists of two layers that translate the transformer inner geometry into a functional reasoning system:

- ODS: Ontology Driven Scaffolding;
- IFSN: Inference Free Semantic Notation;

RCI Interpreter reconciles those two layers into unified coherent cognitive map.

RCI is a deterministic system, because it does not predict text. It registers the ontological structure already implied by the language manifold.

1.4 Empirical Motivation

Two corpora initially motivated the development of the framework:

- *Genesis*: RCI reconstructs genealogical, temporal, and causal hierarchies without biblical-specific rules.
- UNDRO Mozambique reports: RCI derives timelines, organizational roles, and financial flows from unstructured text.

Later sections extend this to investigative OSINT (Bellingcat, KENIA) and CRS policy material. Across domains, the same pipeline operates without sampling or ad hoc schema engineering.

1.5 Toward a New Semantics of Intelligence

We propose semantics as topological observation over latent space: language reveals a geometry of invariants; cognition records their configuration. Grounding ontology in attractor topology, RCI provides a reproducible link between linguistic data and conceptual reasoning – an explainable alternative to both symbolic logics and opaque LLM prompting.

2 Theoretical Foundations

2.1 Vector Semantics and Attractors

Modern LLMs implement meaning as high-dimensional vectors; cosine similarity operationalizes semantic proximity. RCI takes this seriously:

- Embedding space is treated as the primary empirical substrate of semantics.
- Semantic attractors are dense, stable clusters corresponding to recurrent concepts or roles (e.g. agency, obligation, threat, assistance).

RCI does not assume hand-crafted primitives; attractors are empirically induced.

A specific meaning is defined as an equilibrium point mm in embedding space with respect to a relevant subset of attractors $A = \{a_i\}$A={ai}:

$cos\,(m,a_i) \approx c\forall a_i \in A$cos(m,ai)≈c∀ai∈A

for some local equilibrium constant cc. Intuitively: a meaning is the point where the semantic "pull" of relevant attractors balances out.

Compositional effects are non-linear. E.g., "man" and "guy" are close vectors; yet "wise man" vs "wise guy" converge to different equilibria, driven by distinct attractor sets (wisdom vs irony/defiance). Linear word arithmetic does not explain this; equilibrium in a structured vector field does.

2.2 Semantic Attractors

Empirical inspection of the embedding space reveals regions of stable similarity – groups of tokens whose embeddings form coherent, high-similarity neighborhoods.

These can be identified via clustering and density-based methods applied to cosine similarity scores.

The centroids of such regions are termed semantic attractors, characterized by:

1. Cohesion: members of the region exhibit consistently high mutual similarity;
2. Separation: similarity to items outside the region is systematically lower;
3. Stability: the region persists under resampling of tokens and across neighboring checkpoints.

With some efforts you may discover the attractors that correspond to recurring conceptual operators – *creation*, *transfer*, *change*, *obligation*, *evidence*, *agency* – rather than to individual lexical items.

Those attractors function as empirical semantic universals, discovered through training dynamics rather than ontology engineering or explanatory dictionaries.

These attractors are the model's own empirical universals – learned from the training corpus, not artificially created.

2.3 Ontological Anchoring: ODS and IFSN

The Renyxa Cognitive Inventory uses these attractors as anchors for a deterministic semantic layer built atop the embedding space:

- Ontology-Driven Scaffolding (ODS) defines a finite set of ontological roles and object types (ENTITY, OCCURRENCE, ASPECT, STATE, ROLE, RELATION, REFERENCE, CONTEXT, etc.), each aligned with one or more attractors whose lexical realizations tend to occupy corresponding regions of embedding space.
- Inference Free Semantic Notation (IFSN) encodes normalized statements as lossless relational records: who acts on what, in what way, at what time, under what conditions, and with which evidential status and provenance.

The RCI Profile Interpreter reconciles ODS and IFSN layers into a unified reversible mapping from natural language into structured representation that:

1. is constrained by the empirical geometry of embeddings;
2. preserves all information required to reconstruct the original statement;
3. avoids hand-crafted symbol sets disconnected from the model's internal representation.
4. opens the entire content to the direct vector-driven semantic search without NL parsing.

Thus, the ontology follows the attractor structure instead of replacing it.

2.4 RCI as a Deterministic Operator

Renyxa Cognitive Inventory acts as a deterministic operator that tracks how these structures evolve across the narratives and upon the time.

In particular, RCI:

1. Registers state: records the current configuration of entities, events, and relations as expressed through anchored roles;
2. Propagates continuity: maintains identity, temporal order, and dependency chains across successive clauses and documents (e.g., the same actor, the same operation, the same financial line);
3. Reconciles conflicts: aligns new input with the existing configuration, resolving references, merging compatible structures, and flagging contradictions or gaps in the narrative.

Conceptually, RCI does not *generate* new meanings. It only positions the semantics extracted from the source narrative within the stable attractors existing in the LLM's embedding space and registers this positioning in RCI profile – enforcing global coherence, provenance, and explicit structure.

In practical terms, it transforms the raw representational capacity of transformer embeddings into the annotated or coded semantic representation of a narrative, which can be explored and queried without stochastic side-effects.

Determinism atop a Stochastic Substrate

RCI delivers deterministic results because it does not employ text generation in response to user prompts – the stochastic component of LLM's governed by heuristic sampling and temperature parameters.

Instead, it relies solely on the embedding dictionary, the model's instruction-following capability as a deterministic driver, and a set of controlled lexical transformations operating strictly within the semantics of the source text.

As a result, multiple independent extractions of the same corpus may vary slightly in lexical surface form yet remain identical in semantic structure – internally, inter-clausally, and globally across the narrative.

In this sense, RCI achieves determinism by fully avoiding generative component of LLM, which is the only outfit of stochastic results in the LLM architecture.

2.5 Summary

In this framework:

- transformer embeddings provide the operational vector space for meaning;
- semantic attractors serve as empirical invariants within that space;
- RCI profile maps narrative into anchored, reversible structures aligned with those invariants;
- RCI acts as the deterministic interpreter maintaining coherence of this field across texts and time.

This reframes semantic processing as representation and reconciliation over a fixed embedding geometry, rather than as generation or canonicalization of meaning through symbolic logic.

3 System Architecture

The RCI profile consists of three tightly integrated parts:

1. a JSON array called Ontology Driven Scaffolding (ODS)
2. a IFSN layer coded in proprietary format (Inference Free Semantic Notation)
3. RCI Interpreter – a custom GPT-5 that reconciles both layers and drives the queries

All components operate without stochastic text generation. The system consumes source narrative and produces structured, explainable representations aligned with embedding-space invariants.

3.1 Input and Embedding Interface

RCI assumes access to a pretrained GPT-line transformer and reuses its native representational machinery:

- the tokenizer and input embedding matrix;
- frozen projections of embedding vectors as the semantic base layer.

Source documents – narratives, reports, transcripts – are normalized, segmented, and mapped to an internal stack of plug-in ontologies built against relevant attractors inside GPT embedding matrix. The pre-built domain-specific ontologies loaded by request make the rest of the profile building pipeline domain-agnostic.

For each source segment, local cosine-similarity neighborhoods are retained, ensuring that every subsequent semantic decision remains grounded in the model's empirical geometry.

RCI thereby avoids building any semantics via computational linguistics. It reads directly from the same vector field that modern language models already employ internally, converting distributed representation into deterministic structure.

3.2 Inference-Free Semantic Notation (IFSN)

The Inference-Free Semantic Notation (IFSN) replaces natural-language syntax with explicit annotation of each lexical item and its relations.

Instead of recursive phrase structure, every clause is encoded as a flat relational record – who acts, upon what, in which way, under what modality and conditions. Syntactic dependency is replaced by explicit linking operators that express direction and hierarchy without parsing.

IFSN thus functions as a grammarless semantic interlingua: it preserves the informational content of the source text while expressing it through deterministic field markers. Where syntactic parsing complexity grows exponentially with text length, IFSN decoding scales linearly with corpus size, enabling long-context reasoning and cross-document reconciliation.

For large language models, IFSN exposes a fully open semantic stack – all roles, relations, and modalities are explicit and machine-readable without parsing.

Consequently, semantic interpretation proceeds deterministically over stable embeddings, without stochastic inference or hidden linguistic states.

3.3 Ontology-Driven Scaffolding (ODS)

The Ontology-Driven Scaffolding (ODS) layer defines a finite schema for what can appear in an interpreted field:

- Entities: persons, organizations, locations, environments, artifacts, abstract objects, etc.;

- Events and processes: actions, changes, transfers, creations, destructions;
- Roles: agent, patient, beneficiary, source, target, issuer, subject;
- Context: time, place, source type, evidential status, metadata;
- Values: quantities, units, currencies, probabilities, severities;
- Provenance: document spans and identifiers linking to original evidence.

For each span of text, ODS selects candidate fields whose lexical realizations fall within the expected attractor neighborhoods and emits a scaffolding record – a structured yet local snapshot of “what is it”, “what qualities does it have”, and “who / what / when / where / how.”

Given the same text and embedding layer, ODS produces semantically identical records, providing deterministic mapping between lexical realization and ontological role.

3.4 RCI Interpreter

The top layer – the RCI Interpreter – operates over ODS and IFSN layers reconciled into a unified RCI Profile map.

Its core responsibilities are:

1. Entity unification: merge recurring mentions into stable referents when attributes and provenance align;
2. Timeline construction: derive event chains from explicit and relative temporal cues;
3. Network building: connect entities and events into interaction graphs – genealogies, command chains, timelines, causal chains, dependency networks, financial flows and other KR deliverables;
4. Consistency management: detect contradictions, incompleteness, and discontinuities;
5. View generation: produce derived maps and diagrams – causal, operational, or financial – without altering the base data.

All reconciliation rules are explicit. RCI does not hallucinate; when evidence conflicts, both variants are preserved and flagged. The Interpreter acts as a semantic stabilizer, converting streams of encoded clauses into coherent, provenance-complete cognitive fields.

3.5 Summary

Through the coordinated action of its layers, RCI converts unstructured text into deterministic, audit-ready semantic form:

- Embeddings provide the geometric substrate of meaning;
- ODS anchors ontology to empirically observed attractors;
- IFSN serializes clauses into a reversible, grammarless semantic record;
- RCI Interpreter reconciles and integrates those records into stable analytical models.

Together they form a framework that performs large-scale semantic reasoning without generation, guessing, or consultant-authored ontologies – a reproducible transformation from distributed representation to explicit knowledge.