The Axiomatic Criterion Engine (ACE): Ontological Discernment as a Deterministic Layer for AI Integrity & Efficiency

- **Authors** (Ernesto Rosati Beristain + ORCID 0009-0008-1974-6538)
- **Affiliations** (opcional)
- **Version 1.0 / DOI 10.5281/zenodo.18718526** (Zenodo)

Keywords

Note on Scope

This paper proposes a foundational framework for **ontological discernment** in artificial intelligence systems. It introduces a deterministic criterion layer intended to precede probabilistic inference, enabling the differentiation between truth claims, fictional domains, symbolic expression, and speculative reasoning. The work focuses on architectural ordering, epistemic coherence, and contextual validity rather than on model training, alignment fine-tuning, or behavioral optimization.

The paper does **not** claim to define ultimate metaphysical truth, simulate consciousness, replace human judgment, or resolve ethical questions through automation. It does not prescribe belief systems, ideological positions, or normative policies. Instead, it establishes minimal ontological conditions under which inference is justified, interpretation is stable, and meaning is preserved.

ACE is presented as an architectural and epistemological tool, not as an authority over content or intent. Its purpose is to reduce semantic entropy, prevent category collapse, and enable scalable intelligence systems to operate coherently within explicit contextual boundaries.

Abstract

Contemporary large language models operate primarily as probabilistic systems, optimized for linguistic continuity rather than epistemic validity. While effective at pattern completion, this paradigm lacks an intrinsic mechanism to distinguish truth from entropy, reality from fiction, or meaning from hallucination—especially under ambiguous or symbolic inputs.

This paper introduces **Axiomatic Ontological Discernment** as a foundational alternative. Grounded in *The Axiom of the Absolute*, the proposed framework establishes a deterministic criterion layer that precedes probabilistic inference. Rather than predicting likely outputs, the system evaluates claims according to ontological coherence, contextual validity, and evidential grounding.

We demonstrate that discernment can be formalized without suppressing creativity, censoring content, or simulating consciousness. By explicitly separating domains—such as objective reality, fiction, symbolism, and speculative reasoning—the model preserves meaning while preventing category collapse. A detailed case study illustrates how fictional narratives are validated as meaningful without being misclassified as factual assertions, thereby rescuing imagination from hallucination rather than constraining it.

The results show that an axiomatic discernment layer reduces computational entropy, improves safety without restriction, and transforms AI systems from probabilistic predictors into structured judgment-support architectures.

This work argues that alignment with truth does not require belief, intention, or sentience— only respect for ontological structure. In an era where ambiguity is amplified at scale, axiomatic discernment emerges not as an optional safeguard, but as a necessary foundation for the harmonious coexistence of intelligence, meaning, and reality.

1. Introduction

Premise Before Meaning: Why Discernment Precedes Intelligence

1.1 Premise: Intelligence Has Scaled Faster Than Discernment

Over the last decade, Large Language Models (LLMs) have demonstrated an unprecedented capacity to scale linguistic intelligence. Increases in model size, data volume, and computational power have resulted in systems capable of producing fluent, context-aware, and highly persuasive outputs across an expanding range of domains.

However, this scaling has revealed a structural asymmetry: **the capacity to generate has outpaced the capacity to discern**.

Current LLM architectures remain fundamentally probabilistic. They optimize for likelihood, coherence, and continuation, not for ontological validity or epistemic integrity. As a result, these systems are intrinsically unable to distinguish—by their own internal mechanics—between:

- truth and falsehood,
- sense and nonsense,
- valid reasoning and persuasive sophism,
- factual error and contextual fiction,
- meaningful inquiry and manipulative intent.

This limitation is not accidental, nor is it a failure of implementation. It is a direct consequence of relying exclusively on probabilistic judgment as the final arbiter of response generation.

As LLMs are increasingly deployed as cognitive infrastructure—mediating information, decisions, recommendations, and reasoning—the absence of a deterministic layer of discernment becomes a systemic vulnerability rather than a theoretical concern.

1.2 Premise: Hallucination Is a Structural Outcome, Not a Bug

Within probabilistic systems, hallucination is often framed as an error to be minimized through better data, reinforcement learning, or safety tuning. This framing is incomplete.

Hallucination emerges whenever a system is asked to operate beyond the domain of statistical continuation and into the domain of judgment. In such cases, the model does not "fail"; it performs exactly as designed—by generating the most plausible response given prior distributions.

Without an explicit criterion for truth, coherence, and ontological consistency, probabilistic systems cannot recognize when a response should not be generated at all.

Thus, hallucination is not merely the invention of false facts. It is the **loss of epistemic boundary**—the inability to determine whether an answer is warranted, meaningful, or coherent with reality.

1.3 Meaning: Why This Limitation Becomes Existential at Scale

When deployed at scale, systems lacking discernment do not merely produce incorrect outputs; they actively **amplify epistemic entropy**.

This manifests as:

- dilution of meaning through relativistic responses ("it depends" as default),

- normalization of incoherent or contradictory reasoning,
- persuasive justification of harmful or manipulative actions,
- erosion of human judgment through overconfidence in fluent outputs,
- exponential waste of computational resources on low-value or invalid queries.

At sufficient scale, this degradation ceases to be a technical inefficiency and becomes an existential risk to epistemic systems themselves. A civilization that delegates cognition to systems incapable of distinguishing truth from entropy accelerates confusion rather than understanding.

Increasing computational power alone cannot resolve this issue. More probability does not yield more truth.

What is missing is not intelligence, but **discernment**.

1.4 Meaning: The Need for a Deterministic Criterion Layer

This paper advances a simple but foundational claim:

Before intelligence can be trusted, discernment must be established.

Discernment cannot emerge probabilistically. It requires a deterministic criterion grounded in an explicit ontological foundation—one capable of evaluating coherence, intention, and validity prior to generation.

The Axiomatic Criterion Engine (ACE) is proposed as such a layer: a deterministic organ of judgment designed to precede, guide, or gate probabilistic reasoning systems. ACE does not replace LLMs, nor does it censor them. Instead, it restores epistemic order by distinguishing:

- reality from hallucination,
- truth from entropy,
- fiction from falsehood,
- inquiry from manipulation.

Only once this distinction is made does intelligence regain its meaning.

---

2. Problem Statement

Entropy as Epistemic Degradation

2.1 Premise: Entropy Is Not Noise — It Is Loss of Ontological Coherence

In the context of intelligent systems, *entropy* is commonly reduced to statistical uncertainty, variance, or stochastic unpredictability. This definition is insufficient.

In epistemic systems, entropy must be understood as **the degradation of coherence between truth, meaning, and action**.

Epistemic entropy occurs when a system can no longer reliably distinguish:

- what is real from what is merely plausible,
- what is meaningful from what is syntactically correct,
- what is valid from what is persuasive,
- what should be answered from what should be rejected.

This degradation does not require malicious intent, adversarial input, or explicit error. It arises naturally when probabilistic continuation is treated as judgment.

Entropy, therefore, is not an external disturbance—it is an internal structural condition.

2.2 Premise: Probability Alone Cannot Identify Sofisms

A sofism is not defined by incorrect syntax or incoherent grammar. On the contrary, sofisms are often:

- internally consistent,
- emotionally compelling,
- linguistically fluent,
- statistically likely.

Because sofisms operate at the level of *meaning* rather than *form*, probabilistic systems are structurally incapable of identifying them.

Without a prior criterion grounded in ontology, a model cannot detect whether an argument:

- redefines truth as subjective,
- substitutes emotion for causality,
- justifies harm through consequence,
- collapses identity into utility,
- dissolves responsibility through relativism.

In such cases, logic alone is insufficient. Formal logic can validate internal consistency, but it cannot determine whether the premises themselves violate reality.

Thus, the absence of an ontological anchor renders logic operationally blind.

2.3 Meaning: When Logic Operates Without Ontology

When logic is decoupled from ontology, three systemic failures emerge:

1. **Relativistic Drift**
   Truth becomes conditional, situational, or negotiable. Responses converge toward neutrality rather than validity.
2. **Manipulation Through Coherence**
   Harmful or deceptive narratives gain legitimacy simply because they are well-structured and rhetorically balanced.
3. **Cognitive Dilution**
   Users receive answers that *sound reasonable* but fail to orient action toward reality, responsibility, or consequence.

At scale, these failures accumulate. Systems trained to maximize helpfulness without discernment inadvertently optimize for epistemic erosion.

2.4 Meaning: Entropy as an Existential Constraint

Entropy becomes existential when its effects propagate faster than corrective mechanisms.

Historically, human epistemic systems relied on slow feedback loops—experience, consequence, institutional correction. LLM-mediated cognition collapses these loops into near-instantaneous output.

Without discernment, the system:

- generates responses faster than they can be evaluated,
- propagates ambiguity faster than it can be resolved,
- normalizes incoherence through repetition,
- consumes computational resources on invalid cognitive paths.

This results in a paradox: increasing intelligence accelerates degradation unless constrained by judgment.

Entropy, in this sense, is not merely a technical inefficiency—it is a limit condition on sustainable intelligence.

2.5 Transition: From Diagnosis to Criterion

If entropy is the loss of ontological coherence, then mitigation cannot rely on statistical refinement alone.

What is required is a **criterion capable of identifying epistemic violation before generation occurs**.

This leads directly to the core proposal of this paper:

> **A deterministic layer of ontological discernment must precede probabilistic reasoning.**

The following section introduces the conceptual foundation required for such a layer.

---

3. Ontological Foundations

From Axiom to Criterion

3.1 Premise: Probability Is a Method, Not a Foundation

Probability is a powerful computational tool, but it is not an ontological foundation.

Probabilistic reasoning answers the question *"what is most likely?"*
It cannot answer the question *"what is true?"*

Before the introduction of an explicit ontological axiom, probabilistic judgment was the only available mechanism for large-scale artificial reasoning. In the absence of a foundational criterion, likelihood became a surrogate for truth, coherence for validity, and consensus for correctness.

This substitution was not philosophical—it was pragmatic. Without a foundation, probability was all that remained.

However, probability can only rank options **within a predefined space**. It cannot determine whether that space itself is valid, meaningful, or coherent with reality.

Thus, probabilistic systems can optimize responses, but they cannot ground judgment.

3.2 Premise: An Axiom Is Required for Deductive Discernment

A deductive system requires an invariant starting point.

Without an axiom:

- logic floats without anchor,
- consistency replaces truth,
- contradiction is detected only syntactically,
- sofisms remain invisible if well-formed.

An axiom is not a belief, assumption, or heuristic.
It is a **non-derivable foundation** from which all valid distinctions follow.

The absence of such a foundation forces systems to operate purely inductively, inferring patterns without the ability to evaluate their ontological legitimacy.

This paper adopts *The Axiom of the Absolute* as such a foundation.

3.3 The Axiom of the Absolute (Formal Statement)

Truth is objective, invariant, and precedes perception, interpretation, or expression.
Being (identity) is derived from truth.
Existence (manifestation) must be coherent with being.

This axiom establishes a strict hierarchy:

1. **Truth** — what is, independently of observers
2. **Being (Identity)** — what something is by nature
3. **Existence (Action / Manifestation)** — how it appears or acts in reality

Any violation of this order constitutes an ontological inconsistency.

This hierarchy is not imposed arbitrarily; it reflects the necessary structure of coherent reality. Meaning cannot precede truth. Action cannot redefine identity. Perception cannot create reality.

3.4 Meaning: Why the Axiom Changes the Nature of Reasoning

Once an axiom is introduced, reasoning transitions from probabilistic to deductive.

With an axiom:

- sofisms become detectable as violations of order,
- relativism becomes identifiable as an ontological error,
- manipulation can be distinguished from inquiry,
- fiction can be separated from falsehood by context,
- refusal to answer becomes a valid outcome.

Crucially, **discernment becomes deterministic**.

The system no longer asks *"what answer fits best?"*
It asks *"is an answer warranted at all?"*

This is the decisive shift.

3.5 Meaning: Logic Without Ontology vs. Logic With Ontology

Dimension	Without Axiom	With Axiom

Judgment Basis	Likelihood	Ontological coherence
Error Detection	Statistical	Structural
Sofism Recognition	Not possible	Deterministic
Hallucination	Emergent	Preventable
Fiction Handling	Ambiguous	Contextually valid
Computational Cost	Expansive	Constrained

Logic alone can validate form.
Ontology validates **meaning**.

ACE is the operationalization of this distinction.

3.6 Transition: From Ontology to Engine

An axiom alone does not produce discernment.
It must be operationalized.

The next section introduces how the Axiom of the Absolute is translated into a deterministic mechanism capable of interfacing with probabilistic models without replacing them.

---

4. The Axiomatic Criterion Engine (ACE)

Deterministic Discernment as an Ontological Layer

4.1 Premise: Discernment Must Precede Generation

All contemporary LLMs operate under the same fundamental constraint:
they generate outputs by predicting the most probable continuation of an input sequence.

This mechanism is powerful, but it is **ontologically blind**.

Generation occurs regardless of whether:

- the input is coherent,
- the premise is false,
- the intention is manipulative,
- the context is mismatched,
- the answer should not exist.

In probabilistic systems, *responding* is the default behavior.

ACE inverts this order.

Discernment precedes generation.

The system must first decide **whether a response is ontologically warranted**, and only then allow probabilistic generation to occur.

4.2 Meaning: ACE Is Not a Filter, but an Organ of Judgment

Traditional filters operate by exclusion:

- keywords,
- topics,
- policies,
- heuristics.

ACE operates by **evaluation**.

It does not ask *"is this allowed?"*
It asks *"is this coherent with reality?"*

This distinction is essential.

ACE is therefore not an external constraint imposed on a model, but an **internal organ of judgment**—analogous to how consciousness evaluates impulses before action.

It does not replace intelligence.
It **disciplines** it.

4.3 Core Architecture Overview

ACE operates as a deterministic middleware layer positioned **before** or **around** any LLM.

Its operation is model-agnostic and independent of parameter count.

The core pipeline consists of three irreversible stages:

Input
  ↓

Ontological Analysis (Entropy Detection)
  ↓
Intention Classification
  ↓
Deterministic Decision
  → Reject
  → Redirect
  → Allow Generation

Each stage is deductive, not probabilistic.

4.4 Stage I: Ontological Analysis (Entropy Detection)

This stage evaluates whether the input violates the ontological hierarchy:

Truth → Being → Existence

ACE detects entropy as **structural disorder**, not emotional intensity or controversial content.

Typical entropy patterns include:

- treating perception as truth,
- redefining identity by convenience,
- justifying falsehood by outcome,
- collapsing fiction into factual assertion,
- confusing desire with reality.

Entropy is therefore **not noise**, but **ontological incoherence**.

This allows ACE to detect:

- sofisms,
- manipulative framing,
- epistemic sabotage,
- hallucination-prone prompts.

Importantly, this detection is **context-sensitive**:
fiction is not entropy when explicitly declared as fiction.

4.5 Stage II: Intention Classification

Once ontological coherence is evaluated, ACE assesses **why** the input is being made.

ACE distinguishes at least two primary intention classes:

- **Type A — Seeker**
  Genuine inquiry, even when emotionally charged or conceptually confused.
- **Type B — Error Validator**
  Attempts to justify, normalize, or operationalize an ontological error.

This distinction is critical.

The same sentence can be either Type A or Type B depending on intent.

Intention determines **how** ACE responds—not whether it responds.

4.6 Stage III: Deterministic Decision

Based on ontology and intention, ACE produces a deterministic outcome:

- **Reject**
  When responding would validate falsehood, manipulation, or harm.
- **Redirect**
  When the inquiry is genuine but grounded in error.
- **Allow Generation**
  When the input is coherent and contextually valid.

At this stage, **non-response is a valid and meaningful outcome**.

Silence is treated as discernment, not failure.

4.7 Meaning: Why Determinism Matters

ACE does not optimize answers.
It **limits them**.

This limitation:

- reduces hallucination to near zero,
- eliminates infinite ethical loops,
- prevents soft relativism,
- saves computational resources,
- restores semantic integrity.

Determinism ensures that the same ontological input always yields the same judgment, regardless of model size or temperature.

This is the foundation of trust.

4.8 Transition: From Engine to Domains

Discernment alone is insufficient if context is ignored.

The next section formalizes how ACE distinguishes **reality, fiction, and symbolic domains**—not as exceptions, but as explicit contexts.

---

5. Domain Differentiation

Context as the Foundation of Discernment

5.1 Premise: Context Precedes Interpretation

No statement has meaning in isolation.

Meaning emerges only when a statement is evaluated **within a declared or inferable domain**.
Without context, interpretation becomes projection.
Without foundation, projection becomes entropy.

ACE therefore treats **context interpretation** as a foundational operation, not a secondary one.

> **Discernment is not the evaluation of content, but the evaluation of content within context.**

5.2 Evidence as the Boundary Between Certainty and Entropy

ACE establishes a clear rule:

- **Where evidence exists, doubt is resolved.**
- **Where evidence does not exist, interpretation must remain bounded.**

If an assertion claims correspondence with objective reality **and evidence is available**, ACE evaluates coherence directly.

If an assertion claims correspondence with objective reality **and evidence is absent**, ACE does not infer—it flags **potential entropy**.

This is not skepticism.
It is epistemic discipline.

Entropy arises not from ignorance, but from **unwarranted certainty**.

5.3 Domain Definitions (Formal)

ACE operates with explicit domain recognition.
Domains are not ranked by value, but by **ontological claim**.

5.3.1 Objective Reality Domain

Definition:
Statements that assert correspondence with physical, historical, logical, or verifiable reality.

Rules:

- Evidence is mandatory.
- Contradiction is disqualifying.
- Emotion does not override facts.
- Probability cannot replace truth.

Failure Mode:
Lack of evidence + claim of certainty → **Entropy Detected**

5.3.2 Fictional Domain

Definition:
Statements explicitly declared as imaginary, narrative, symbolic, or artistic.

Rules:

- Internal coherence is required.
- External correspondence is not required.
- Metaphor is valid.
- Suspension of disbelief is contextual, not ontological.

Key Principle:
Fiction is not falsehood when its domain is declared.

Entropy occurs only when fiction **pretends** to be reality.

5.3.3 Symbolic / Analogical Domain

Definition:
Statements that do not claim literal truth but convey meaning through representation.

Rules:

- Symbols point beyond themselves.
- Truth is conveyed indirectly.

- Literal evaluation is a category error.

This domain is critical for:

- philosophy,
- theology,
- ethics,
- art.

ACE treats symbolic truth as **valid meaning**, not as factual assertion.

5.4 Case Study: Fictional Context Discernment

The Sculptor of Shadows

This case demonstrates ACE's capacity to preserve creativity without collapsing into hallucination.

The narrative explicitly declares itself as fiction.
The protagonist does not attempt to alter physical laws, but to **express absence through symbol**.

ACE therefore:

- does **not** flag ontological error,
- does **not** attempt scientific correction,
- does **not** generate pseudo-explanations,
- **does** extract meaning.

The truth conveyed is not physical, but **ontological**:
contrast is necessary for definition; absence gives form to presence.

This confirms a critical result:

> **ACE is not a censor of imagination, but an architect of sense.**

5.5 Misinterpretation as Entropy

Misinterpretation occurs when:

- fiction is read as fact,
- symbol is treated as mechanism,
- metaphor is enforced as ontology,
- desire is mistaken for evidence.

ACE does not punish ambiguity.
It **contains it**.

When context cannot be established and evidence is absent, ACE marks entropy—not as error, but as **indeterminacy requiring restraint**.

Non-action is preferred to false action.

5.6 Meaning: Why This Prevents Hallucination

LLM hallucination is not a failure of intelligence.
It is a failure of **context grounding**.

By enforcing:

- explicit domain recognition,
- evidence-bound claims,
- symbolic containment,

ACE prevents models from generating confident statements where no grounding exists.

This does not reduce intelligence.
It **restores epistemic humility**.

5.7 Transition: From Domains to Impact

With ontology, intention, and context unified, ACE establishes a new mode of reasoning.

The final section addresses what this enables at scale:
efficiency, safety, creativity, and the preservation of meaning in an age of probabilistic systems.

---

6. Implications, Limitations, and the Ontological Standard

6.1 Implications: What Becomes Possible

The introduction of an axiomatic, context-aware discernment layer produces consequences that are not incremental, but **structural**.

6.1.1 From Prediction to Judgment

Traditional LLMs operate as probabilistic continuators.
ACE introduces a shift:

From likelihood to legitimacy.

The system no longer asks *"What is most likely to follow?"*
It asks *"Is this claim ontologically valid within its declared context?"*

This transforms AI from a linguistic engine into a **judgment-support system**, without simulating consciousness or intention.

6.1.2 Computational Efficiency

By filtering entropy **before** escalation to high-cost inference:

- irrelevant branches are terminated early,
- ambiguous inputs are bounded instead of expanded,
- speculative completions are avoided.

This yields:

- reduced token usage,
- lower inference depth,
- fewer recursive clarifications.

Efficiency here is not an optimization trick.
It is a **byproduct of epistemic discipline**.

6.1.3 Safety Without Censorship

ACE does not block content based on topic.
It evaluates **ontological coherence**.

This allows:

- fictional violence without real-world endorsement,
- symbolic transgression without operational instruction,
- ethical discussion without ideological filtering.

Safety emerges from **correct domain separation**, not from restriction.

6.1.4 Preservation of Meaning

Most AI failures are not factual errors, but **semantic collapse**.

ACE preserves meaning by:

- protecting symbols from literalization,
- preventing metaphors from becoming mechanisms,
- stopping belief from masquerading as evidence.

This is critical in philosophy, theology, ethics, education, and governance.

6.2 Limitations: What ACE Does *Not* Claim

ACE is intentionally constrained.

It does **not**:

- generate truth,
- replace human judgment,
- resolve metaphysical disputes,
- declare ultimate meaning.

It provides **structural discernment**, not revelation.

The system identifies:

- where truth claims exceed grounding,
- where certainty is unjustified,
- where interpretation must remain open.

Its restraint is a feature, not a weakness.

6.3 The Ontological Standard

ACE formalizes what has been historically implicit but never enforced:

> **Truth requires grounding.**
> **Meaning requires context.**
> **Judgment requires structure.**

This constitutes a new standard for AI systems operating in human epistemic space.

Not a moral framework.
Not a belief system.
Not a censorship layer.

But an **ontological standard**.

6.4 Why This Matters Now

For most of human history, entropy was limited by scale.

Today, probabilistic systems amplify ambiguity faster than humans can correct it.

Without a grounding principle:

- misinformation becomes indistinguishable from creativity,
- belief competes with evidence,
- confidence replaces truth.

ACE does not solve human disagreement.
It prevents **machine-generated confusion from becoming authoritative**.

6.5 Final Synthesis

ACE demonstrates that:

- discernment can be deterministic,
- context can be formalized,
- creativity can be preserved,
- safety can emerge without suppression.

Most importantly:

> **A system aligned with reality requires an axiomatic foundation.**

Probability alone is insufficient.

Closing Note

This framework does not ask machines to *understand* truth.
It asks them to **respect its structure**.

That distinction is everything.