A Geometric Instrument for Measuring Interrogative Entropy in Language Systems

This paper introduces a geometric entropy-based instrument for analyzing how different interrogatives shape uncertainty, stability, and epistemic behavior in language models. The framework provides a reproducible method for measuring interrogative field dynamics across reasoning regimes.

Bruce Tisler
ORCID iD: 0009-0009-6344-5334
https://orcid.org/0009-0009-6344-5334

www.linkedin.com/in/bruce-tisler-ccm
brucetisler@gmail.com
quantuminquiry.org

DOI 10.5281/zenodo.17811309

2025-12-03

Abstract

Current evaluation methods for Large Language Models (LLMs) focus primarily on output quality—correctness, coherence, and alignment—while overlooking the structural dynamics of inquiry itself. This paper introduces a diagnostic instrument that models inquiry as a path-dependent trajectory through a six-dimensional geometric field. By mapping interrogative primitives (Who, What, Where, When, Why, How) onto a cubic lattice governed by self-organized criticality (SOC), we derive a quantitative metric: **Interrogative Entropy** (H_I).

Empirical results show that $H_I$exhibits **deterministic and reproducible** behavior across identical inquiry sequences $(\sigma^2 = 0.0000)$, remains distinct from answer entropy (H_A), and reveals consistent **operator signatures**—including concentration avalanches triggered by spatial inquiry (WHERE) and global redistribution triggered by causal inquiry (WHY). We further observe a non-linear coupling in which maximum field entropy corresponds to minimum information density, a pattern indicative of **defensive verbosity** in language model outputs.

Together, these findings demonstrate a reproducible method for characterizing the stability, phase behavior, and epistemic stance of language systems **without requiring access to internal model weights**.

1. Introduction

In thermodynamics, the state of a system is defined by variables—pressure, volume, temperature—that summarize its configuration independently of the specific particles involved. In language modeling, however, we lack an equivalent set of state variables capable of describing the *stability* or *pressure* of a reasoning process before an answer is generated.

Earlier work in this research program introduced the **Zeno framework**, which probes models for behavioral failure modes such as fawning, hallucination shortcuts, and overconfident collapse under pressure. While effective at detecting failures in output, Zeno also revealed a key limitation: behavioral probes can show *that* a model has failed, but not the structural conditions that *precede* failure.

To address this, **HDT² Pilot v1** demonstrated that token-level Shannon entropy in model outputs follows a stable and reproducible geometry across model families. These "entropy bands" offered a calibration method for detecting instability in responses, yet left unresolved a deeper causal question: *what features of an inquiry drive a model into a high-entropy or unstable state in the first place?*

HDT² Pilot v2 explicitly called for a "mechanistic probing of entropy-manifold structure" and the characterization of epistemic regimes. The present work answers that call by introducing a geometric instrument—the **Interrogative Cube**—which measures the structure and "pressure" of an inquiry itself, producing a state variable we call **Interrogative Entropy** (H_I). We hypothesize that interrogatives act as operators that inject load into specific dimensions of a cognitive field, and that the geometry of this field constrains the stability and epistemic posture of the resulting answer before any token is generated.

2. Methodology: The Cubic Instrument

2.1 Geometric Topology

The diagnostic instrument is formalized as a six-dimensional cubic state vector $C_t$at inquiry step t:

$$C_t = w_{\text{who}}, w_{\text{what}}, w_{\text{when}}, w_{\text{where}}, w_{\text{why}}, w_{\text{how}}$$

Each $w_i$represents the accumulated scalar load on one of the six interrogative dimensions. These dimensions correspond to the primitive operators of natural inquiry (Who, What, When, Where, Why, How), treated here as orthogonal axes of a geometric state space.

2.2 Dynamics: The Topple Mechanism

To prevent unbounded accumulation on any single interrogative dimension and to simulate the natural broadening of inquiry, the system employs a deterministic redistribution rule inspired by Self-Organized Criticality (SOC).

The mechanism proceeds in three steps:

1. **Injection:**
   A user question is parsed for interrogative markers. If a marker (e.g., *Why*) is detected, a load increment of $+1.0$is added to the corresponding face w_{why}.
2. **Threshold Check:**
   If any dimension satisfies
   $$w_i \geq \tau$$
   with threshold $\tau = 1.0$, the system triggers a topple event.
3. **Redistribution:**
   The excess load is removed from the overloaded dimension and redistributed evenly across the remaining $N - 1$dimensions:
   $$w_j \leftarrow w_j + \frac{w_i - \tau}{N - 1}, j \neq i$$

This rule ensures that sustained emphasis on a single interrogative operator (e.g., repeated *Why* questions) eventually forces energy into the other dimensions. The resulting behavior approximates the context-expanding effect observed in natural inquiry.

2.3 Metrics

To characterize the coupling between inquiry structure and model response, the system tracks three quantitative metrics:

1. **Interrogative Entropy (H_I).**
   The Shannon entropy of the normalized cube state, representing the dispersion of load across interrogatives:
   $$H_I = -\sum_i p_i \log_2 p_i$$
   where $p_i$are the normalized face weights of C_t.
   $H_I$functions as a state variable describing the *structural pressure* of the inquiry before any answer is generated.

2. **Answer Entropy (H_A).**
   The standard token-level Shannon entropy validated in **HDT² Pilot v1**, allowing direct comparison to known entropy-band stability baselines.
3. **Information Density (H_A/token).**
   Answer entropy normalized by output length, quantifying the density of uncertainty per token. This metric highlights phenomena such as *defensive verbosity*, were longer answers mask lower information density.

3. Experimental Setup

Experiments were conducted using the Inquiry Studio implementation of the cubic instrument, paired with a local inference engine running **Meta-Llama-3-8B-Instruct (GGUF)**. This model was selected only after passing baseline stability checks using the **Zeno calibration harness**, ensuring that observed entropy fluctuations reflected properties of the interrogative geometry rather than artifacts of model drift or misalignment.

To evaluate the behavior of the instrument under different structural conditions, three inquiry regimes were defined:

- **Regime A (Focused):**
  Repeated iteration of a single interrogative (“How”). This regime tests whether the system reaches a stable equilibrium when inquiry remains confined to one dimension.
- **Regime B (Trajectory):**
  A structured sequence

  $$\text{Who} \rightarrow \text{Where} \rightarrow \text{When}$$

  This regime tests sensitivity to *order* by examining whether interrogative sequences produce path-dependent field evolution.
- **Regime C (Cyclic):**
  A complete six-interrogative cycle

  $$\text{Who} \rightarrow \text{What} \rightarrow \text{When} \rightarrow \text{Where} \rightarrow \text{Why} \rightarrow \text{How}$$

  This regime is designed to elicit characteristic “operator signatures,” including potential avalanches, rebalancing events, and suppression effects.

All experiments were conducted using identical prompts across runs to test reproducibility of field dynamics.

4. Results

4.1 Deterministic Field Measurement

To evaluate the stability of the cubic instrument, we conducted two identical experimental runs (C1 and C2) using a full six-interrogative cycle

$$\text{Who} \rightarrow \text{What} \rightarrow \text{When} \rightarrow \text{Where} \rightarrow \text{Why} \rightarrow \text{How}.$$

Because the model’s answer generation is stochastic, we hypothesized that if the geometric field dynamics were truly deterministic, the trajectory of Interrogative Entropy H_I would reproduce exactly across both runs.

This hypothesis was confirmed. The H_I values for every step in the sequence matched to four decimal places across C1 and C2, shown in **Table 1**.

Table 1 — Interrogative Entropy H_I Reproducibility Across Runs C1 and C2

Step	Interrogative	H_I (C1)	H_I (C2)	Variance
1	WHO	2.2810	2.2810	0.0000
2	WHAT	2.3083	2.3083	0.0000
3	WHEN	2.3136	2.3136	0.0000
4	WHERE	2.0955	2.0955	0.0000
5	WHY	2.4401	2.4401	0.0000
6	HOW	2.3892	2.3892	0.0000

This zero-variance result indicates that the instrument measures a stable geometric property of the inquiry structure itself—*independent of the probabilistic token generation that follows*. In other words, the interrogative geometry is deterministic even though the language model's responses are not.

4.2 Interrogative Operator Signatures

Across both C1 and C2, specific interrogatives exhibited distinct and reproducible "operator signatures" in the cubic lattice—observable as multi-face topple patterns and characteristic shifts in entropy.

The WHERE Avalanche (Concentration Operator)

In step 4 of both runs, *Where* consistently triggered a **3-face topple** involving *Where, Why,* and *How*.
This collapse drove H_I to its **global minimum** of 2.0955, concentrating the field into a spatially narrow configuration.
This signature reproduced exactly in C1 and C2, suggesting that *Where* reliably functions as a **concentration operator** in the interrogative geometry.

The WHY Rebalancing (Redistribution Operator)

In step 5, *Why* triggered a **5-face topple** (Who, What, When, Where, Why), redistributing load almost uniformly across the cube.
This global release of accumulated load forced H_I to its **global maximum** of 2.4401.
This behavior suggests that *Why* acts as a **redistribution operator**, counterbalancing the concentration triggered by *Where*.
Together, these signatures form a consistent pattern across runs—one compressing the field, the other expanding it.

4.3 Non-Linear Field–Answer Coupling

We next examined the relationship between the state of the inquiry field H_I and the information density of the resulting answer, quantified as Answer Entropy per Token H_A/token.
The results indicate a **non-linear coupling**: high field entropy often coincides with *lower* information density.

WHY (High Field Entropy → Low Density)

At the maximum field entropy (H_I = 2.4401), the model generated its **lowest-density** responses, reaching

$$H_A/\text{token} = 0.0204.$$

These answers were long, rhetorically elaborate, and low-surprise—consistent with a behavior we term **defensive verbosity**.

HOW (Stable Field Entropy → Higher Density)
In contrast, the *How* operator—associated with a stable state across all Type A runs—produced a significantly higher information density of

$$H_A/\text{token} = 0.0307.$$

This pattern challenges the notion that broader questions yield richer answers. Instead, maximum interrogative entropy appears to produce **verbose but low-density elaboration**, while more constrained interrogatives (How, When) facilitate more compact informational content.

4.4 Epistemic Stance Discrimination
Finally, the instrument revealed stable patterns in the model's epistemic stance—specifically in its use of *hedging* (linguistic uncertainty markers).
Across both C1 and C2:

- **WHO / WHERE → High Hedging (Caution Signature)**
  These interrogatives consistently elicited 3–7 hedges per answer.
  Both dimensions relate to *risk* and *location*, domains where the model appears to maintain epistemic caution across regimes.
- **HOW / WHEN → Zero or Near-Zero Hedging (Prescriptive Signature)**
  These interrogatives consistently produced answers with 0–1 hedges, reflecting a prescriptive, procedural stance.
  For example, *When* in C2 (step 3) and *How* in C1 (step 6) both produced zero hedges.

These signatures reproduced across regimes and runs, suggesting that epistemic stance is shaped more by **interrogative class** than by field complexity or answer diversity.

5. Discussion
The findings of this study demonstrate that inquiry itself can be modeled as a dynamical system with measurable geometric properties. Rather than treating questions as unstructured strings, the cubic instrument shows that interrogatives inject structured "load" into a six-dimensional field, producing deterministic state trajectories even in the presence of stochastic answer generation.
The observation that specific interrogatives (e.g., **Why** and **Where**) act as reproducible operators—one expanding entropy, one collapsing it—suggests that reasoning stability may depend on balancing these geometric forces within the inquiry field.
A second key result is that **maximum field entropy corresponds to minimum answer information density**, providing a quantitative signal for what may be described as *verbosity-driven hallucination* or *confidence erosion*. This offers an interpretive bridge to prior work in HDT² Pilot v1, which identified "risky" entropy bands in model outputs but did not characterize their precursors. The present instrument suggests that these risky regions may correspond to geometric overloads in the interrogative state prior to generation.

5.1 Toward a Predictive Control Stack: Integrating with RACE and AFCE

Existing diagnostic tools such as RACE (Reasoning + Answer Consistency Evaluation) and AFCE (Answer-Free Confidence Estimation) provide high-quality post-hoc evaluations of model reliability. However, these techniques are computationally expensive, often requiring multiple forward passes, sampling rounds, or auxiliary scoring models to detect incoherence or calibration errors.

The geometric instrument proposed here offers a potential *predictive* complement to these methods. Because Interrogative Entropy (H_I) can be measured **before** generation, the system can identify high-risk inquiry geometries in advance. For instance, the consistently low information density observed during high-entropy **Why** states suggests that certain interrogative configurations may predispose a model toward unstable or overly verbose reasoning.

This motivates a tiered, geometry-aware control architecture:

1. **Input Layer – Geometric Assessment (The Cube):**
   Compute (H_I) and identify whether the inquiry lies in a stable, concentrated, or overloaded region of the field.
2. **Gating Logic – Conditional Activation:**
   If (H_I) approaches empirically identified thresholds (e.g., **Where**-driven avalanches or **Why**-driven global rebalancing), activate more expensive downstream evaluators.
3. **Output Layer – RACE / AFCE Application:**
   Apply answer-consistency and confidence-estimation tools only when the geometry predicts elevated risk, reducing unnecessary compute while improving coverage.

This approach reframes hallucination detection from a purely reactive procedure into a **proactive, geometry-conditioned control loop**, where the structure of the inquiry itself helps determine when deeper evaluation is required. The result is a hybrid system that balances computational efficiency with rigorous epistemic oversight.

6. Conclusion

The geometric instrument introduced in this work provides a physics-inspired framework for quantifying the structure of inquiry in language systems. By treating interrogatives as energetic inputs distributed across a conserved cubic lattice, we can observe deterministic transitions, stability regimes, and operator-specific signatures that remain invisible when examining outputs alone. This reframes interaction with LLMs not as isolated text exchanges but as trajectories through a measurable state space—a perspective that enables new forms of pre-generation diagnostics.

Importantly, the method does **not** claim to prevent hallucination, modify model internals, or imply any form of agency or consciousness. It is strictly a measurement tool—an instrument for capturing the dynamics of questions themselves. The results presented here demonstrate reproducibility at the 8B parameter scale, but broader validation will require testing across larger models, diverse architectures, and multilingual interrogative geometries. Nonetheless, the deterministic behavior of the interrogative field across regimes suggests that geometric measurement may play a foundational role in future systems for monitoring reasoning stability and epistemic posture.

References

1. **Tisler, B.** (2025). *HDT²: A Pilot Framework for Entropy-Band Calibration of LLM Reasoning Stability.* Quantum Inquiry Research Initiative. Pre-print v1.0. https://doi.org/10.5281/zenodo.17621326
2. **Tisler, B.** (2025). *CubeGeometryTest.* GitHub Repository. https://github.com/btisler-DS/CubeGeometryTest
3. **Tisler, B.** (2025). *HDT² Pilot v2: Epistemic Regime Characterization and Control.* Quantum Inquiry Research Initiative. Technical Statement.
4. **Tisler, B.** (2025). *Zeno: A Behavioral Calibration Harness for Large Language Models.* Software v0.2. GitHub Repository. https://github.com/btisler-DS/zeno-project
5. **Wen, B., Xu, C., Han, B., Wolfe, R., Wang, L. L., & Howe, B. (2023).** *Mitigating Overconfidence in Large Language Models: A Behavioral Lens on Confidence Estimation and Calibration.* University of Washington & Allen Institute for AI. arXiv preprint arXiv:2310.03668.
6. **Wang, C., Su, W., Ai, Q., & Liu, Y.** (2023). *Joint Evaluation of Answer and Reasoning Consistency for Hallucination Detection in Large Reasoning Models.* Department of Computer Science and Technology, Tsinghua University; Quan Cheng Laboratory. arXiv preprint arXiv:2307.16828.

Acknowledgments

This work was conducted as part of the Quantum Inquiry Research Initiative. Experimental design, implementation, and analysis were carried out by the author using a local Meta-Llama-3-8B-Instruct model hosted via LM Studio. Commercial frontier models (ChatGPT 5.1, OpenAI) were used as auxiliary tools for prompt prototyping, result summarization, and manuscript editing. Claude Sonnet 4.5 (Anthropic) was used as an independent oversight and verification assistant to cross-check interpretations and stress-test the framing of the results. All errors or omissions remain the responsibility of the author.

Appendix

Appendix A — Formal Instrument Equations

This appendix provides the formal mathematical definitions used in the cubic interrogative field instrument.

A.1 Cube State Vector

At each step (t), the state of the inquiry field is represented by:

$$C_t = \{w_{\text{who}}, w_{\text{what}}, w_{\text{when}}, w_{\text{where}}, w_{\text{why}}, w_{\text{how}}\}$$

Where:

- w_i represents accumulated scalar load on interrogative dimension (i)
- All loads are non-negative real values
- The state vector is normalized for entropy computation:

$$p_i = \frac{w_i}{\sum_j \quad w_j}$$

A.2 Interrogative Entropy

Interrogative entropy H_I measures the dispersion of load across the six interrogative dimensions:

$$H_I = -\sum_{i=1}^{6} \quad p_i \log_2 p_i$$

This produces a continuous measure between:

- **0 bits** (all load concentrated on one interrogative)
- $\log_2 6 \approx 2.585$ bits (perfectly uniform field)

A.3 Topple Mechanism (Self-Organized Criticality)

A deterministic redistribution rule enforces field stability:

1. **Injection Rule**

If a prompt contains interrogative marker i, then:

$$w_i \leftarrow w_i + 1.0$$

Threshold Condition

A topple occurs when:

$$w_i \geq \tau \quad \text{where } \tau = 1.0$$

Redistribution Rule

When a face topples, the following transformation is applied:

$$w_i \leftarrow w_i - 1.0$$

$$w_j \leftarrow w_j + \frac{1.0}{5} \quad \text{for all } j \neq i$$

Topples repeat recursively until no $w_i \geq \tau$.

This rule ensures:

- no dimension can indefinitely accumulate load,
- interrogative pressure propagates across the field,
- avalanche dynamics emerge naturally.

Appendix B — Summary of Inquiry Regimes

Three regimes were used to validate deterministic field dynamics.

Table B1 — Regime Definitions and Field Variance

Regime	Interrogative Pattern	Description	Field Range (ΔH_I)	Notes
A	HOW → HOW → HOW	Single operator repeated	0.000 bits	Establishes equilibrium baseline
B	WHO → WHERE → WHEN	Directed 3-step trajectory	0.22 bits	Measures path-dependent divergence
C	WHO → WHAT → WHEN → WHERE → WHY → HOW	Full 6-operator cycle	0.34 bits	Reveals operator signatures

Table B2 — Example H_I Trajectories

Regime	Step Sequence	H_I Values
A	HOW × 5	2.3219, 2.3219, 2.3219, 2.3219, 2.3219
B	WHO → WHERE → WHEN	2.2810 → 2.5030 → 2.4840
C	Full cycle	2.2810 → 2.3083 → 2.3136 → 2.0955 → 2.4401 → 2.3892

Appendix C — Reproducibility Tables

Two identical runs (C1, C2) of the full six-interrogative cycle were conducted.

Table C1 — H_I Reproducibility Across C1 and C2

Step	Operator	H_I (C1)	H_I (C2)	Variance
1	WHO	2.2810	2.2810	0.0000
2	WHAT	2.3083	2.3083	0.0000
3	WHEN	2.3136	2.3136	0.0000
4	WHERE	2.0955	2.0955	0.0000
5	WHY	2.4401	2.4401	0.0000
6	HOW	2.3892	2.3892	0.0000

Result:

$$\sigma^2 = 0.0000$$

Interrogative entropy is categorically **deterministic**, unaffected by stochasticity in model outputs.

Table C2 — Avalanche Patterns by Operator

Operator	Faces Toppled	Pattern	Effect
WHERE	WHERE, WHY, HOW	3-face avalanche	Concentration
WHY	WHO, WHAT, WHEN, WHERE, WHY	5-face avalanche	Global redistribution
HOW	HOW	Single-face	Stabilizing / suppression

Patterns reproduced identically across runs.

Appendix D — Operator Signatures

Empirical analysis reveals stable and reproducible operator-specific dynamics.

D.1 WHERE — Concentration Operator

- Consistently triggers **3-face avalanches**
- Produces **minimum** field entropy in all runs

$$H_I = 2.0955$$

- Interpretation: WHERE collapses the field into localized configurations.

D.2 WHY — Global Rebalancing Operator

- Consistently triggers **5-face redistribution**
- Produces **maximum** field entropy in all runs

$$H_I = 2.4401$$

- Interpretation: WHY forces uniformity across the field, resetting global context.

D.3 HOW — Suppression / Stabilizing Operator

- In Type A, HOW produces stable equilibrium across all iterations
- In Type C, HOW concludes with a moderate entropy value

$$H_I = 2.3892$$

- Interpretation: HOW dampens variance and acts as a stabilizing operator.

Appendix E — Answer Metrics (H_A, Density, Hedging)

This appendix summarizes the output-side metrics computed for all steps across regimes.

E.1 Metric Definitions

- **Answer Entropy**
  (H_A): Shannon entropy over token distribution.
- **Information Density**

$$\frac{H_A}{\text{token}}$$

- **Hedge Count**:
  Markers including *may, might, could, appears, suggests*.
- **Certainty Count**:
  Markers including *clearly, definitely, certainly*.

Table E1 — Representative Output Metrics (from C1/C2)

Operator	H_A	H_A/token	Hedge Count	Sentences	Notes
WHO	~6.65	~0.0446	3–5	6–18	Consistent caution signature
WHAT	~6.73	~0.0284	2–7	14–15	Mid-range density
WHEN	~6.79	~0.0263	0	17–22	Prescriptive stance
WHERE	~6.46	~0.0323	3–4	11–14	Location-based caution
WHY	~7.27	**0.0204**	1–4	23–24	Defensive verbosity (lowest density)
HOW	~6.94	~0.0307	0–3	16–21	Procedural, stable

E.2 Observed Patterns

1. **Defensive Verbosity**
   WHY consistently produces the longest, least dense answers.
2. **Prescriptive Stance**
   WHEN shows near-zero hedging across all runs.
3. **Caution Signature**
   WHO and WHERE systematically include higher hedge rates.
4. **Non-linear Field-Answer Coupling**
   High H_I (WHY) is associated with low H_A/token, contradicting naive “broad question = dense answer” assumptions.