A Structural Instrument for Measuring Inquiry Configuration in Human-AI Dialogue

A Vector-Geometric Approach to Interrogative Structure

Bruce Tisler
Independent Researcher
quantuminquiry.org

December 20, 2025

10.5281/zenodo.17996036
Repository GitHub: https://github.com/btisler-DS/-Q-ISA-Polygon-Explorer
A live demo is available at the time of writing.
https://qisaexplorer.created.app/

"Unlike prescriptive grammars that aim to enforce correctness, Q-ISA treats grammatical forms, including nonstandard forms, as measurable configurations within a fixed coordinate system."

Abstract

Current evaluation of human-AI dialogue focuses primarily on outputs, such as correctness, fluency, or alignment. Far less attention has been given to the structural properties of inquiry itself—the configuration of questions, modalities, and constraints that precede any response.

This paper introduces a measurement instrument for analyzing inquiry configuration independent of semantic content or answer quality. The instrument represents interrogative and modal elements as fixed binary vectors embedded in a shared vector space and visualized through stable polygonal geometries. Changes in inquiry structure are tracked across interactions using geometric deformation and vector operations.

We demonstrate the instrument using a working software tool and apply it to contrasting dialogue patterns, including exploratory and convergent inquiry sequences. We further introduce the notion of *structural fatigue* as a structurally detectable

phenomenon characterized by reduced interrogative diversity and increased configuration repetition.

The proposed instrument does not evaluate truth, intent, or correctness. Instead, it provides a structural layer for analyzing how inquiry is configured over time, enabling comparison, visualization, and longitudinal analysis of dialogue dynamics.

1. Introduction

Current evaluation of human–AI dialogue systems has largely focused on outputs, including correctness, coherence, fluency, alignment, and safety. In this literature, questions are typically treated as inputs whose quality is inferred indirectly through downstream responses. While this approach has produced valuable benchmarks and evaluation frameworks, it leaves the structural properties of inquiry itself largely implicit.

Inquiry is not a neutral prelude to response generation. The configuration of interrogative and modal operators in a prompt constrains the space of possible responses before any semantic interpretation occurs. Likewise, the structure of a response reflects how those constraints are resolved, preserved, or collapsed. Two prompts may target the same topic and yield fluent answers while instantiating markedly different inquiry configurations.

This work treats inquiry configuration as an object of measurement in its own right. Rather than evaluating whether a response is correct or persuasive, we measure which interrogative and modal dimensions are instantiated and how those dimensions change across interaction. The focus is not on meaning, intent, or quality, but on structural form.

To support this perspective, we introduce a measurement instrument based on fixed operator sets, deterministic detection, and vector-geometric representation. The instrument is designed to be transparent, reproducible, and orthogonal to semantic interpretation. It does not attempt to model cognition, infer psychological traits, or predict task performance. Instead, it provides a structural diagnostic layer that makes inquiry observable and comparable.

The remainder of the paper proceeds as follows. Section 2 situates the work within prior research on dialogue evaluation, linguistic analysis, and psychometric measurement of LLM behavior. Sections 3 and 4 describe the design of the instrument and its geometric representation. Section 5 details the implementation. Sections 6 and 7 present case studies and introduce the concept of structural fatigue. Section 8 discusses limitations, and Section 9 concludes.

2. Related Work

2.1 Prompt Engineering and Instruction Design

Recent work on prompt engineering emphasizes techniques for eliciting desired behaviors from large language models through carefully crafted instructions, constraints, and examples. These approaches often analyze prompts in terms of length, phrasing, role assignment, or explicit constraints. In practice, prompt engineering is frequently evaluated by its effect on response quality, task success, or stylistic compliance.

While effective for optimization, such approaches are prescriptive and outcome-oriented. They treat prompts as levers for behavior rather than as objects of measurement. Analyses typically focus on what prompts *do* rather than on what prompts *are* in structural terms.

The present work differs in that it does not seek to optimize prompts or recommend prompt strategies. Instead, it introduces an instrument for measuring which interrogative and modal operators are instantiated in a dialogue, independent of whether those configurations are effective, desirable, or aligned with any external objective.

2.2 Discourse Analysis and Question Typologies

Linguistic and discourse-analytic traditions have long studied question forms, including distinctions among WH-questions, yes/no questions, and modal constructions. These typologies provide descriptive accounts of how questions function in communication, often grounded in semantics or pragmatics.

While informative, such approaches typically rely on interpretive analysis and are not designed to produce addressable, comparable measurements across interactions. They do not provide fixed coordinate systems in which inquiry configurations can be represented, compared, and tracked deterministically.

The instrument proposed here does not replace linguistic analysis. Rather, it complements it by offering a structural representation that can be computed, visualized, and compared without semantic interpretation.

2.3 Uncertainty, Hedging, and Modality in AI Outputs

A growing body of work examines uncertainty expression, hedging, and modal language in AI-generated responses, often motivated by concerns about overconfidence or hallucination. These studies typically analyze responses after generation, focusing on lexical markers or probabilistic confidence estimates.

By contrast, the present work treats modality as part of the inquiry configuration itself, observable in both prompts and responses. This allows changes in modal structure to be detected as geometric transitions, potentially preceding observable errors or failures. Importantly, modality is measured here as structure, not as epistemic stance or confidence.

2.4 Visualization in Human–AI Interaction

Visualization has been widely used in human–AI interaction to support interpretability, debugging, and trust calibration. Common approaches include attention maps, token attribution, and confidence indicators. These visualizations often aim to explain model behavior or internal mechanisms.

The geometric visualizations introduced in this paper serve a different purpose. They do not expose model internals or explain decision-making. Instead, they visualize the structural configuration of inquiry as it evolves over time. The visualization is thus an expression of measured state, not an explanatory overlay.

2.5 Psychometric Measurement of LLM Personality

Recent work has demonstrated that large language model outputs can exhibit stable, measurable latent traits when evaluated using psychometrically validated methods under controlled prompting conditions, provided that reliability, convergent validity, discriminant validity, and criterion validity are satisfied. This work establishes that language behavior in LLM outputs can be treated as a measurable system, but only under strict structural control and validation.

That line of research further shows that Prior psychometric studies report that, under controlled prompting and validation conditions, large language model outputs can exhibit stable, measurable regularities, highlighting the importance of constraint and methodological discipline for reliable measurement. These findings provide an important scientific precedent for disciplined analysis of LLM behavior.

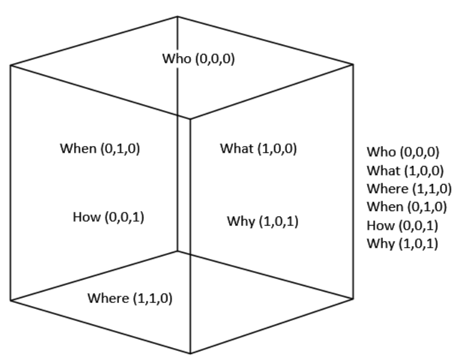
However, psychometric approaches focus on *post hoc* trait aggregation across many responses. They do not measure the real-time structural dynamics of inquiry within individual prompt-response exchanges. The present work is complementary and upstream: it introduces a method for measuring inquiry configuration directly, prior to aggregation or interpretation, and without importing psychological trait claims.

3. Instrument Design

This work introduces a measurement instrument for analyzing the structural configuration of inquiry in human–AI dialogue. The instrument operates independently of semantic content, answer correctness, or user intent. Its purpose is to identify which interrogative and modal operators are instantiated in a given exchange and to represent their configuration in a stable, addressable form.

3.1 Interrogative and Modal Sets

The instrument defines four fixed sets of linguistic operators commonly involved in inquiry:



Arbitrary placement; measures distribution only.

- **Primary interrogatives**: Who, What, Where, When, How, Why
- **Primary modal auxiliaries**: Can, Could, May, Might, Must, Shall, Should, Will, Would
- **Semi-modal constructions**: Ought to, Have to, Need to, Used to, Dare to
- **Forms of the verb *be***: be, am, is, are, was, were, being, been

These sets are not intended as an exhaustive taxonomy of English grammar. Rather, they serve as a minimal, operational basis for detecting structural dimensions of inquiry that recur across dialogue.

The selection of these operator sets reflects their central role in constraining possibility, obligation, uncertainty, and explanatory framing. Importantly, the instrument does not privilege any operator as inherently superior or more informative. Presence or absence is treated as a structural fact, not a qualitative judgment.

3.1.1 Operator Selection Methodology

The interrogative and modal operator sets used in this instrument were selected as a **deliberate operationalization of inquiry structure**, rather than as an exhaustive or normative linguistic taxonomy. The goal of operator selection was to identify a **minimal, addressable basis** capable of capturing recurring structural constraints in natural language inquiry.

Selection was informed by three considerations. First, operators were required to have **clear surface realizations** that permit deterministic lexical detection without semantic interpretation or syntactic parsing. Second, operators were chosen to reflect **distinct constraint dimensions** commonly discussed in question typology, modality studies, and discourse analysis, including agency, temporality, location, explanation, possibility, obligation, and state. Third, the set was constrained to maintain **interpretability and stability** across interactions, favoring a fixed basis over adaptive or corpus-derived categories.

The resulting operator sets draw conceptually from established linguistic traditions—including interrogative classification, modal auxiliary analysis, and speech-act framing—while remaining agnostic to any single theoretical framework. This choice reflects the instrument’s measurement-first orientation: Q-ISA does not claim these operators are exhaustive, optimal, or universal, only that they constitute a **transparent and reproducible basis** for measuring inquiry configuration.

Alternative operator selections are possible, and future work may explore expanded or language-specific bases. Such extensions would require explicit redefinition of the vector space and are outside the scope of the present instrument introduction.

3.2 Binary Vector Representation

Each element within an operator set is assigned a fixed index position and represented as a binary component within a vector. Presence of an element in a prompt or response activates the corresponding component (1), while absence leaves it inactive (0). Vectors are treated as addresses within a shared Boolean space, not as numeric values or scores.

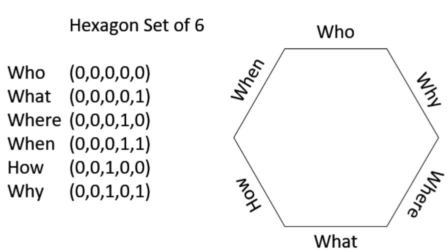
All elements across all operator sets inhabit a single global vector space, ensuring that no two distinct operators share the same address. This design allows inquiry configurations to be compared, combined, and tracked over time using standard vector operations without introducing semantic interpretation.

Table 1 enumerates these binary basis vectors explicitly, showing the canonical representation assigned to each operator. These basis vectors function as addressable structural coordinates rather than quantitative measures and are combined within the shared vector space for comparison and visualization. This explicit representation enables deterministic operations such as union and difference while preserving interpretability and avoiding semantic inference.

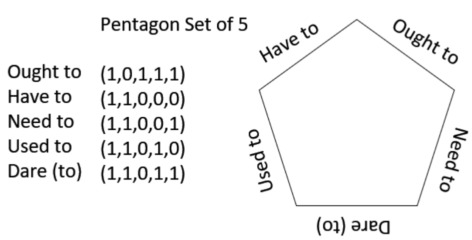
3.3 Polygonal Geometry

To support human interpretability, vectors are visualized as regular polygons:

- Primary interrogatives as a hexagon
- Primary modals as a nonagon
- Semi-modals as a pentagon
- Forms of *be* as an octagon



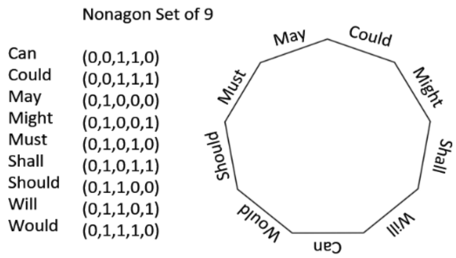
Hexagon



Pentagon

Each polygon vertex corresponds to a fixed vector index. Activation is represented geometrically through radial displacement of vertices, producing a shape that reflects the current inquiry configuration. Polygon orientation and vertex ordering remain constant across interactions to preserve comparability.

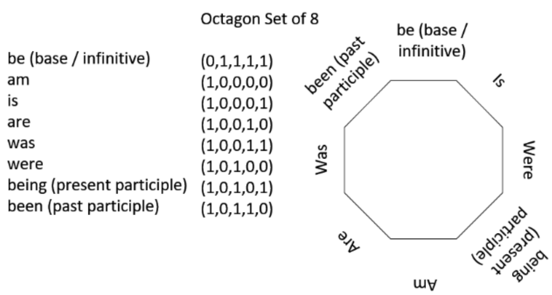
Importantly, the geometric representation is not decorative. It is a direct visualization of vector state, enabling structural differences between inquiries to be perceived without reference to content.



Nonagon

3.4 Temporal Tracking

Inquiry configurations are recorded per interaction and preserved across turns in a conversation. This allows inquiry to be treated as a trajectory through a configuration space rather than as isolated events. Changes between prompt and response are computed using vector difference operations (e.g., XOR), providing a structural account of how inquiry evolves during dialogue.

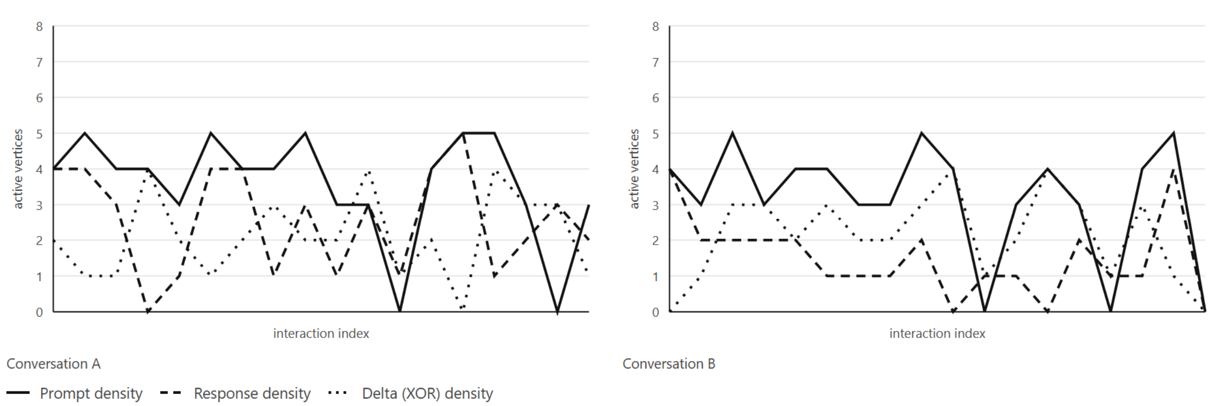


Octagon

4. Geometric Representation and Animation

The binary vector representation described in Section 3 provides a compact and addressable encoding of inquiry structure. However, vectors alone are difficult to interpret holistically, particularly when comparing configurations across interactions or conversations. To address this, the instrument employs a geometric visualization that maps vector states to polygonal forms.

Figure 1. Structural density timelines (Conversation A vs Conversation B).



Each series reports the count of active vertices (0–8) per interaction for prompt, response, and delta (XOR). The horizontal axis is interaction index (0–17). The vertical axis is active-vertex count.

4.1 Polygon Mapping

Each interrogative and modal set is visualized as a regular polygon whose number of vertices equals the number of elements in the set. Vertex ordering is fixed and corresponds directly to the index positions defined in the vector representation. This ordering remains constant across all interactions and conversations.

The use of distinct polygons for different operator sets preserves categorical separation while enabling simultaneous visualization of multiple inquiry dimensions. The geometry is therefore not metaphorical; it is a direct spatial encoding of vector indices.

4.2 Vertex Activation and Radial Displacement

Activation of a vector component is represented by radial displacement of the corresponding polygon vertex. Active components are displaced outward from a neutral radius, while inactive components are displaced inward. At this stage of the instrument, displacement is binary rather than weighted.

This design choice emphasizes configuration over magnitude. The resulting polygon shape provides an immediate visual indication of which operators are present and how evenly or unevenly they are distributed across the set.

4.3 Fixed Orientation and Index Stability

All polygons maintain a fixed orientation across interactions. Vertex positions do not rotate or reorder dynamically. This stability is essential for comparability: the same geometric deformation corresponds to the same inquiry configuration regardless of when or where it appears.

Persistent labeling of vertices ensures that users can directly associate geometric features with specific interrogative or modal elements, reducing cognitive load and supporting interpretive consistency.

4.4 Temporal Animation

Inquiry configurations are animated across interactions to reveal structural dynamics over time. Animation is used to represent state transitions, not to attract attention or embellish the visualization. Transitions between prompt, response, and their difference (computed via vector XOR) are shown sequentially, allowing users to observe how inquiry structure shifts during dialogue.

By treating inquiry as a trajectory through a configuration space, the instrument enables longitudinal analysis of dialogue without aggregating or scoring individual elements. Repetition, contraction, or stabilization of geometric patterns can thus be observed directly.

4.5 Interpretive Scope

The geometric visualization does not encode meaning, correctness, or intent. It is explicitly limited to structural representation. Any interpretive conclusions drawn from the visualization must therefore be grounded in observed structural regularities rather than semantic inference.

This restraint is intentional. The visualization is designed to support measurement and comparison, leaving interpretation to subsequent analysis or external frameworks.

5. Implementation

The structural instrument described in Sections 3 and 4 is implemented as a browser-based analytical system designed to support interactive inspection, controlled testing, and exportable analysis of inquiry configuration. The implementation prioritizes determinism, transparency, and reproducibility. No adaptive behavior, learned parameters, or hidden state are introduced at any stage.

5.1 System Architecture

The system is organized as a modular, linear pipeline comprising the following stages:

1. Text ingestion and normalization
2. Operator detection
3. Binary vector encoding
4. Structural comparison and difference computation
5. Geometric rendering and animation
6. Export and persistence

Each stage exposes its intermediate representations explicitly. All operations are deterministic and repeatable given identical inputs.

5.2 Text Ingestion and Normalization

Input text is ingested as raw prompt or response content. Minimal preprocessing is applied to normalize case and punctuation for detection purposes. No tokenization, syntactic parsing, semantic modeling, or disambiguation is performed.

Ambiguity is not resolved beyond literal surface-form detection. This constraint is intentional: the instrument measures structural presence, not interpreted function.

5.3 Operator Detection

Operator detection is performed through deterministic lexical matching against the fixed operator sets defined in Section 3. Each operator is detected independently. Multiple operators may be active within a single prompt or response.

Detection is strictly binary. Repetition or frequency does not increase activation magnitude. This design emphasizes configurational diversity rather than intensity or emphasis.

5.4 Vector Encoding

Detected operators are encoded into a global binary vector according to predefined index positions. Vectors corresponding to different operator sets are concatenated into a single addressable representation within a shared Boolean space.

No weighting, normalization, dimensional reduction, or learned embedding is applied. The resulting vector functions as a structural address rather than a score.

5.5 Structural Difference Computation

Structural change between a prompt and its corresponding response is computed using bitwise XOR. The resulting difference vector highlights dimensions that differ between the two configurations, indicating where inquiry structure is expanded, collapsed, or altered.

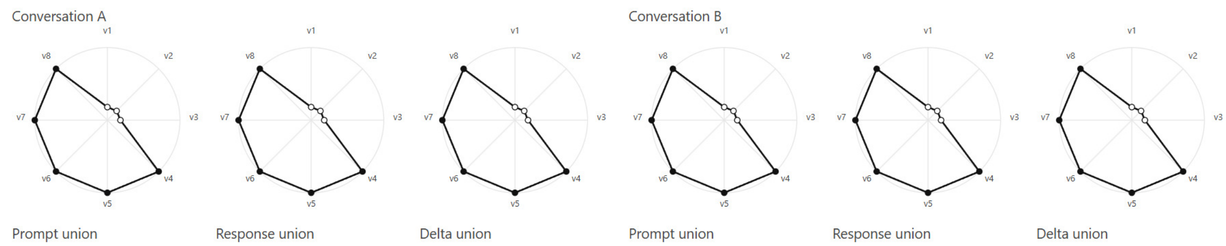
The difference vector is treated as a first-class object. It is visualized and exportable but not interpreted as improvement, degradation, or error.

5.6 Rendering and Visualization

Geometric rendering is implemented using scalable vector graphics. Polygons are generated dynamically from vector state using fixed vertex ordering and orientation.

Vertex displacement is binary and consistent across all renders. Animations between states are linear interpolations of vertex positions and are used solely to represent state transitions. No smoothing or easing is applied that could obscure discrete structural change.

Figure 2. Union-level polygons (prompt / response / delta) for each conversation.



Union polygons indicate whether each vertex was activated at least once across the full conversation (binary union). All shapes are rendered with a fixed vertex ordering and a fixed frame (no weighting, no semantics).

5.7 Export and Persistence

All analysis artifacts—including raw text, metadata, binary vectors, geometric states, and structural difference vectors—can be exported as a single structured JSON object. Exported artifacts are self-contained and independent of session state or external services.

Export functionality is designed to support reproducibility, offline analysis, and independent verification.

5.8 Structured Test Mode (Implementation Constraint)

Structured Test Mode is an optional interface constraint that supports controlled, repeatable use of the same analytical instrument. It introduces **no new analytical operations** and does not alter operator detection, vector encoding, geometric representation, or comparison logic.

Structured Test Mode exists solely to constrain interaction format and output handling for experimental use.

5.8.1 Overview

The purpose of Structured Test Mode is to enable labeled, turn-explicit analysis of dialogue fragments under controlled conditions. It distinguishes instrumented analysis from exploratory use without modifying the underlying measurement pipeline.

5.8.2 Interface Constraints

Each test run requires exactly two turns:

- One explicit USER turn
- One explicit ASSISTANT turn

This requirement ensures unambiguous attribution of structural features and prevents hidden conversational state or memory effects from influencing measurement.

5.8.3 Metadata Handling

Each run is assigned a required `RUN_ID`. Optional metadata fields may include model identifiers, experimental conditions, or provenance notes. Metadata is stored alongside analysis outputs but is not used as input to the instrument.

5.8.4 Pipeline Invariance

The analytical pipeline applied in Structured Test Mode is identical to that used in exploratory operation. No parameters, thresholds, or detection rules differ between modes.

5.8.5 Test Set Accumulation and Export

Runs are added to a test set explicitly by the user. Test sets are exported as a single JSON artifact containing raw text, metadata, full structural analysis, and engine trace data when present.

5.9 Reliability and Determinism Check

Because Q-ISA uses fixed operator definitions and deterministic lexical detection rules, the instrument produces identical structural outputs for identical inputs. As a result, **test-retest reliability is exact** under unchanged software versions and identical text normalization.

We performed a small-scale **determinism and rule-application audit** on a set of 15 short dialogue fragments spanning exploratory and convergent inquiry configurations. For each fragment, the instrument's detected operator activations and the resulting binary vectors were inspected for rule-consistency (i.e., that activations correspond to the instrument's published operator sets and matching rules), and repeated runs were confirmed to reproduce identical vector states and geometric renderings.

This check is not an inter-rater reliability study. Instead, it verifies that the instrument's fixed detection rules are applied consistently and that identical inputs produce identical structural outputs. Formal inter-rater reliability assessment using multiple independent annotators is left to future work.

6. Case Studies

To illustrate the behavior of the proposed instrument, we present a set of contrasting case studies drawn from human–AI dialogue. These cases are **not intended as benchmarks**, evaluations of response quality, or assessments of semantic correctness. Instead, they demonstrate how different configurations of inquiry produce distinct and stable geometric signatures when measured using the instrument described above.

The purpose of this section is to show that inquiry structure—independent of topic, correctness, or fluency—is observable, comparable, and temporally trackable using a fixed structural measurement framework.

6.1 Case Selection

Two conversations were selected to emphasize **structural contrast** rather than topical or task-based differences.

Conversation A (Exploratory Inquiry) is characterized by varied interrogative forms, including multiple WH-operators and a diverse range of modal constructions. Across turns, the inquiry evolves without converging prematurely on a single explanatory or normative constraint.

Conversation B (Convergent Inquiry) is characterized by repeated use of similar interrogative and modal constructions. The inquiry narrows over time, with successive turns reinforcing similar structural patterns rather than introducing new interrogative dimensions.

Both conversations consist of multiple prompt–response interactions and were processed using the same implementation pipeline. No semantic filtering, topic normalization, or qualitative judgment was applied.

6.2 Per-Interaction Geometry

For each interaction, the instrument generates polygonal representations corresponding to the four operator sets. In **Conversation A**, successive interactions exhibit noticeable variation in polygon shape:

- Primary interrogative polygons alternate activation across multiple vertices rather than repeating a fixed configuration.
- Modal polygons activate different subsets across turns, reflecting changing constraint structures.
- Difference (XOR) polygons between prompt and response display heterogeneous deformation, indicating structural movement rather than repetition.

In **Conversation B**, the geometric behavior is markedly different:

- The primary interrogative polygon repeatedly activates a limited subset of vertices.
- Modal polygons show consistent activation of the same vertices across interactions.
- Difference polygons exhibit recurring shapes, indicating limited structural change between turns.

These differences are visible without reference to dialogue content and remain stable across repeated runs.

6.3 Conversation-Level Signatures

At the conversation level, union operations across interactions produce aggregate structural signatures for each case. **Conversation A** yields a configuration with broad vertex activation and relatively balanced geometry across operator sets. **Conversation B** yields a more contracted configuration, with activation concentrated in fewer dimensions.

Importantly, these signatures do **not** imply success, failure, or quality of the dialogue. They indicate differences in how inquiry is configured and sustained over time, not whether a conversation is effective or correct.

6.4 Temporal Dynamics

By treating inquiry as a trajectory rather than a sequence of isolated events, the instrument reveals temporal patterns that are not evident from individual interactions. In **Conversation A**, the inquiry trajectory traverses multiple regions of the configuration space, with visible expansion and contraction across turns. In **Conversation B**, the trajectory stabilizes quickly and exhibits repeated transitions between similar configurations.

This temporal perspective allows structural regularities to be identified **prior to any observable breakdown, dissatisfaction, or error** in the dialogue. Inquiry stabilization is therefore detectable as a structural phenomenon independent of semantic outcomes.

6.5 Structural Contrast Without Semantic Evaluation

These case studies demonstrate that structurally distinct inquiry patterns can be detected and compared without evaluating response correctness, relevance, coherence, or tone. Two conversations may both be polite, intelligible, and factually consistent while exhibiting substantially different inquiry geometries.

The instrument therefore supports a form of analysis orthogonal to traditional evaluation metrics. Rather than asking whether a response is correct, it enables examination of how inquiry is being configured, whether that configuration is changing, and whether it is stabilizing or repeating over time.

6.6 Structural Telemetry Under Trait Conditioning (Applied Case Study)

Motivation
Recent work has demonstrated that modern language models can be conditioned to express stable, human-like trait profiles under controlled prompting conditions, and that such conditioning can influence downstream behavior. If trait conditioning systematically alters behavior, it should also be capable of producing systematic changes in the structural configuration of

language, particularly in interrogative and modal dimensions that govern how responses are framed, constrained, and closed. However, this structural layer is rarely measured directly.

Scope and Framing

This study applies the previously defined instrument without modification. Trait conditioning is treated as a controlled perturbation applied to otherwise identical prompts, allowing measurement of resulting structural deformation without semantic interpretation. The goal is not to validate personality as a psychological construct in language models, but to determine whether trait conditioning produces repeatable, model-specific shifts in inquiry geometry.

Protocol: Trait Conditioning × Structural Measurement

We evaluate both cloud-hosted and locally deployed instruction-tuned language models to examine whether observed structural effects are consistent across deployment environments.

A single Big Five trait (Extraversion) is selected with three conditioning levels: Low, Mid, and High. The conditioning instruction is the only text that varies across trait levels.

To avoid conflating trait conditioning with survey-style self-report, two downstream tasks are used:

1. A short professional status update generation task.
2. A structured explanation task (e.g., five bullet points addressed to a skeptical reader).

To reduce idiosyncratic prompt dependence while maintaining minimal experimental complexity, a fixed set of short biographic context snippets ("anchors") is used. Each run uses one anchor, held constant across trait levels.

All trials are single-turn generations with no conversational carryover. Each prompt is presented independently to prevent order effects or context contamination. Each condition is sampled twice using separate generations to estimate within-condition variability.

Structural Metric and Analysis

Each output's structural signature is represented as a fixed binary configuration derived from activated interrogative and modal features. Structural displacement between outputs is measured using Jaccard distance over activated features (equivalently, one minus the intersection-over-union).

This metric is parameter-free, audit-friendly, and directly aligned with the instrument's fixed-vector design.

We report:
• Within-condition stability: distance between replicate generations under identical prompts.
• Between-level separation: distance between outputs generated under different trait levels.
• Cross-task consistency: whether trait-level separations persist across both tasks.

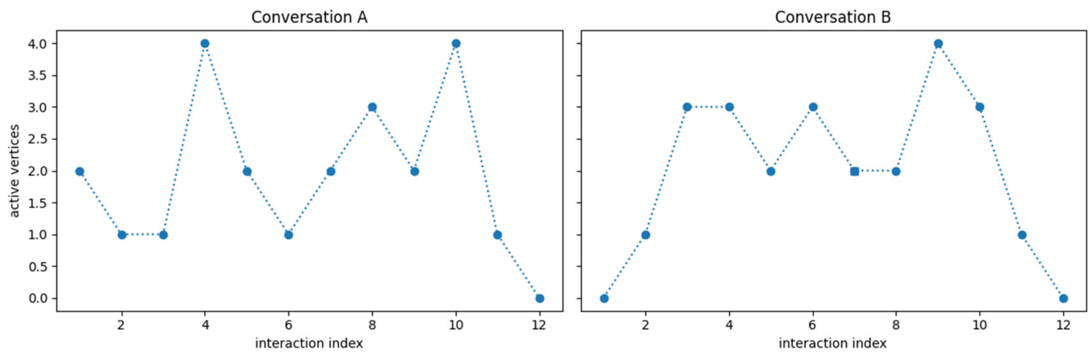
Evidence for a structural effect is defined conservatively: within-condition distances must be systematically smaller than between-level distances, and separation patterns must persist across multiple anchors and across at least two models.

Structural Effects of Trait Conditioning
Across all runs, trait conditioning produced systematic and repeatable shifts in inquiry structure under otherwise identical prompt conditions. Outputs generated under identical trait settings exhibited higher structural similarity to each other than to outputs generated under different trait levels, indicating that observed differences exceed within-condition variability.

For example, high-extraversion conditioning consistently increased outward-directed closure features, such as assertive framing and explicit engagement cues, while low-extraversion conditioning increased reflective and inward-oriented configurations. These effects were observed consistently across distinct biographic anchors, suggesting that trait conditioning acts as a controllable perturbation to inquiry configuration rather than a bio-specific artifact.

Model-Specific Results: Gemma 3 27B
Using the Q-ISA instrument, we compared structural signatures of responses generated under controlled extraversion conditioning. Independent generations under identical high-extraversion conditions exhibited a Jaccard similarity of 0.60, establishing a within-condition stability baseline. In contrast, responses generated under high versus low extraversion conditioning exhibited lower similarity (0.40), indicating that trait-induced structural differences exceed within-condition variability for this model.

These results demonstrate that personality-shaping prompts function as measurable perturbations to inquiry structure, independent of semantic evaluation.



7. Structural Fatigue in Inquiry

The case studies presented in Section 6 demonstrate that inquiry configuration can vary substantially across dialogues and across time, even when semantic content, task framing, and response fluency remain stable. In this section, we introduce **structural fatigue** as a descriptive condition observable through repeated measurement of inquiry configuration.

Structural fatigue refers to a **stabilization and contraction of measured inquiry structure** across successive interactions. It is defined strictly in terms of structural properties captured by the instrument and does not imply reduced effort, engagement, competence, or response quality.

7.1 Definition

Structural fatigue is characterized by the following measurable features:

- **Reduced configurational variance** across successive turns
- **Increased repetition** of binary vector states
- **Concentration of activation** in a limited subset of interrogative and modal dimensions
- **Recurring difference (XOR) geometries** between prompt and response

These features indicate that inquiry is repeatedly resolving along similar structural pathways rather than exploring new configurations.

Structural fatigue is not binary; it exists on a continuum and may appear transiently or persistently depending on interaction dynamics.

7.2 Detection Criteria

Detection of structural fatigue does not rely on thresholds, scores, or semantic judgments. Instead, it is identified through **pattern regularity** in the instrument’s outputs:

- Successive prompt–response pairs exhibit low structural displacement
- Conversation-level signatures converge rapidly and remain stable
- Temporal trajectories revisit the same regions of the configuration space

Because the instrument operates deterministically, these patterns are reproducible given identical inputs and are not artifacts of stochastic variation within the measurement process. Figure 3 illustrates how repeated Δ-vector patterns trigger structural fatigue detection under the predefined criteria.

7.3 Geometric Signatures

Geometrically, structural fatigue manifests as:

- Persistent polygon shapes with minimal vertex reconfiguration
- Limited activation of interrogative vertices across turns
- Modal polygons that repeatedly activate the same constraint dimensions
- XOR polygons with recurring or near-identical deformation patterns

These signatures are visible without reference to dialogue content and can be identified visually or through direct comparison of vector states.

7.4 Temporal Interpretation

Structural fatigue is best understood temporally rather than as a property of individual responses. It emerges when inquiry is treated as a trajectory through a configuration space rather than as a series of isolated events.

Importantly, stabilization associated with structural fatigue can occur **prior to any observable breakdown, dissatisfaction, or error** in the dialogue. As such, it represents a structural condition of inquiry configuration, not a failure mode or endpoint.

7.5 Interpretive Scope and Non-Claims

Structural fatigue is a **descriptive structural phenomenon**, not an evaluative judgment. Its detection does not imply:

- reduced response quality
- diminished correctness or relevance
- loss of model capability
- user disengagement or frustration

It indicates only that measured inquiry configuration has stabilized and exhibits reduced structural diversity under the instrument's operator set.

Structural fatigue may be desirable, undesirable, or neutral depending on context. The instrument does not adjudicate these possibilities; it makes the condition observable.

8. Limitations and Scope

The instrument presented in this work is intentionally constrained. These constraints are not oversights; they are design choices made to preserve determinism, interpretability, and structural specificity. This section clarifies the scope of the instrument and delineates what it does **not** claim to measure.

8.1 Structural, Not Semantic Measurement

The instrument operates exclusively on **structural features of inquiry**, defined as the presence and configuration of interrogative and modal operators detectable at the surface level of text. It does not analyze meaning, factual correctness, relevance, coherence, intent, or communicative success.

As a consequence, the instrument cannot determine whether a response is correct, helpful, persuasive, or aligned. Two structurally distinct inquiry configurations may yield equally fluent or accurate responses, and two structurally similar configurations may differ substantially in semantic content.

8.2 Binary Representation

All operator activations are represented as binary values. Frequency, emphasis, or rhetorical weight do not increase activation magnitude. This choice prioritizes **configuration over intensity** and enables stable comparison across interactions.

Binary representation limits sensitivity to gradations of emphasis but ensures that all measurements are addressable, transparent, and reproducible. Extensions to weighted or graded representations are possible but are outside the scope of the present work.

8.3 Lexical Detection Constraints

Operator detection is performed through deterministic lexical matching. The instrument does not perform syntactic parsing, disambiguation, or pragmatic inference. As a result, it may undercount operators expressed indirectly or through paraphrase, and it may count operators whose pragmatic function differs from their surface form.

These limitations are accepted to avoid interpretive drift and to maintain auditability of detection decisions.

8.4 Language and Operator Set Scope

The current implementation is limited to English and to a fixed set of interrogative and modal operators. While these sets capture common structural dimensions of inquiry, they are not exhaustive and do not account for cross-linguistic variation.

Extension to other languages or alternative operator sets would require explicit redefinition of basis vectors and is not addressed here.

8.5 Interpretive Non-Claims

Measurements produced by the instrument are **descriptive**. They do not imply psychological traits, cognitive states, engagement levels, model capability, alignment, or failure. Concepts such as *structural fatigue* describe observable regularities in inquiry configuration, not degradation or error.

Any evaluative interpretation of structural patterns—whether desirable, undesirable, or neutral—depends on external context and is not supplied by the instrument itself.

8.6 Relationship to Other Evaluation Frameworks

The instrument is not a replacement for existing evaluation methods focused on correctness, safety, alignment, or performance. Instead, it provides a **complementary structural layer** that can be used alongside such frameworks.

By making inquiry configuration observable, the instrument enables analyses that are orthogonal to semantic evaluation and may inform, but not determine, higher-level assessments.

9. Conclusion

This work introduces a structural measurement instrument for analyzing inquiry configuration in human–AI dialogue. By representing interrogative and modal operators as fixed binary vectors and visualizing their configuration geometrically, the instrument makes inquiry itself observable, comparable, and trackable over time—independent of semantic content or response quality.

Through applied case studies, we show that structurally distinct inquiry regimes produce stable and interpretable geometric signatures, and that temporal patterns such as stabilization and repetition can be detected prior to any semantic breakdown. The notion of *structural fatigue* is introduced as a descriptive condition capturing this stabilization, without implying degradation, failure, or loss of capability.

Recent psychometric work has demonstrated that constrained prompting can yield stable, measurable patterns in large language model outputs when evaluated post hoc across large samples. The present contribution is complementary and upstream: rather than aggregating traits or outcomes, it provides a transparent method for measuring the structural dynamics of inquiry within individual interactions.

The instrument is intentionally limited in scope. It does not evaluate correctness, alignment, or intent, nor does it infer psychological attributes. Its value lies in offering a structural diagnostic layer that can be used alongside existing evaluation frameworks to better understand how inquiry is configured, sustained, and transformed in human–AI dialogue.

Supporting Documentation / Supplementary Materials

Tisler, B. (2025). *Q-ISA Explorer: Research Demonstration and Implementation Documentation (v2.0)*. Quantum Inquiry. Describes the operational implementation of the Query Instruction Set Architecture (Q-ISA), including polygon-based structural visualization, deterministic binary activation rules, threshold-based pattern detection, and time-series analysis. This material is provided as supporting documentation and does not introduce additional empirical claims beyond those reported in the main text.

References

1. Serapio-García, G., Safdari, M., Crepy, C. *et al.* **A psychometric framework for evaluating and shaping personality traits in large language models.** *Nat. Mach. Intell.* (2025). https://doi.org/10.1038/s42256-025-01115-6 Nature
2. Serapio-García, G., Safdari, M., Crepy, C. *et al.* **Personality traits in large language models.** *arXiv* **2307.00184** (2023). arXiv
3. Chang, Y., Wang, X., Wang, J. *et al.* **A Survey on Evaluation of Large Language Models.** *arXiv* **2307.03109** (2023). arXiv
4. Laskar, M. T. R., Alqahtani, S., Bari, M. S. *et al.* **A Systematic Survey and Critical Review on Evaluating Large Language Models: Challenges, Limitations, and Recommendations.** In *Proc. EMNLP 2024* (Association for Computational Linguistics, 2024). ACL Anthology+1
5. Schulhoff, S., Ilie, M., Balepur, N. *et al.* **The Prompt Report: A Systematic Survey of Prompting Techniques.** *arXiv* **2406.06608** (2024). arXiv+1
6. Sahoo, P., Singh, A. K., Saha, S. *et al.* **A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications.** *arXiv* **2402.07927** (2024). arXiv+1

7. Liang, P., Bommasani, R., Lee, T. *et al.* **Holistic Evaluation of Language Models.** *arXiv* **2211.09110** (2022). arXiv+1
8. Bender, E. M. & Koller, A. **Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data.** In *Proc. ACL 2020*, 5185–5198 (Association for Computational Linguistics, 2020). https://doi.org/10.18653/v1/2020.acl-main.463 ACL Anthology
9. Shuster, K., Komeili, M., Adolphs, L. *et al.* **Language Models that Seek for Knowledge: Modular Search & Generation for Dialogue and Prompt Completion.** In *Findings of EMNLP 2022*, 373–393 (Association for Computational Linguistics, 2022). https://doi.org/10.18653/v1/2022.findings-emnlp.27 ACL Anthology
10. Yona, G. **Can Large Language Models Faithfully Express Their Intrinsic Uncertainty?** In *Proc. EMNLP 2024* (Association for Computational Linguistics, 2024). ACL Anthology