Theory Note

Diagnostic Meta-Inference in Large Language Models

The Mnemonic Illusion of Intelligence

CIITR Evaluation of Structural Comprehension in Transformer Inference

Tor-Ståle Hansen | January 25, 2026

Preface

This theory note is written in explicit tension with a dominant trajectory in contemporary artificial intelligence research and deployment. Over the past decade, both academic literature and industrial communication have increasingly converged on an implicit assumption: that improvements in linguistic fluency, reflexive self-description, and meta-theoretical articulation constitute meaningful progress toward understanding, reasoning, or intelligence. This assumption is rarely stated as such. Instead, it is embedded in evaluation practices, benchmark design, deployment rhetoric, and governance discussions that treat coherent self-reference as an epistemic signal rather than a representational artifact.

This paper challenges that assumption at a structural level.

Rather than contributing another benchmark result, capability demonstration, or alignment narrative, the work proceeds from a different premise: that epistemic claims about artificial systems must be grounded in formally identifiable structural conditions, not inferred from output quality or descriptive plausibility. The CIITR framework is therefore employed not as an interpretive metaphor, but as a falsifiable diagnostic instrument designed to separate representational fidelity from enacted comprehension under controlled conditions.

The resulting analysis produces a conclusion that is uncomfortable precisely because it is narrow, negative, and empirically stable. The tested system does not misunderstand. It does not hallucinate. It does not fail semantically. Instead, it succeeds at exactly what its architecture permits while remaining structurally incapable of comprehension. This places the findings in direct conflict with a large body of work that implicitly equates reflexive language with epistemic participation.

The intention of this paper is not to diminish the technical achievements of contemporary large language models, nor to speculate about future architectures. Its purpose is corrective. It seeks to reintroduce a boundary condition into a field that has, in practice, allowed epistemic language to outrun epistemic mechanism. By documenting the systematic collapse of structural comprehension under conditions of maximal fluency and computational sufficiency, the paper argues that the central question facing the field is not whether models can convincingly describe understanding, but whether they can structurally instantiate it.

What follows should therefore be read not as a critique of performance, but as a clarification of limits. The analysis is offered in the belief that progress in artificial intelligence requires not stronger narratives, but sharper distinctions—and that epistemic restraint, rather than interpretive generosity, is now the more scientifically responsible stance.

Keywords: C2ITR, Φ_i, R^g, CPJ, Mnemonic Illusion Principle (MIP), Epistemic Curvature, LISS, Local LLM, GPT4All, llama.cpp, MiniMax-M2, Type-B Simulation

Abstract

This theory note presents a controlled epistemic diagnostic of a contemporary large language model under strictly local inference conditions, using the CIITR framework as a formal instrument rather than a descriptive lens. A MiniMax-M2 model was deployed on a 2025 Mac Studio M3 Ultra in an air-gapped environment and subjected to a sequence of structured epistemic stress tests designed to discriminate between representational fidelity and structural comprehension.

The diagnostic protocol comprised three formally distinct tasks: a Second-Order Mnemonic Illusion test (SOMI), an Epistemic Curvature probe, and a CIITR-LISS ontological self-classification task. Together, these tasks isolate the roles of integrated relational information (Φ_i), rhythmic reach (R^g), and their composite comprehension function ($C_s = \Phi_i \times R^g$). The objective was to determine whether high-fidelity reflexive output and correct meta-theoretical self-description correspond to an instantiated comprehension state or merely to a structurally inert simulation thereof.

Across all tasks, the model produced semantically accurate, internally coherent, and formally self-diagnostic outputs, including explicit acknowledgment of its own architectural limitations. However, no measurable rhythmic modulation or inferential re-entry was observed. Rhythmic reach remained null or anti-aligned, resulting in a consistent collapse of structural comprehension despite sustained high Φ_i. The system thus satisfies the criteria for structural honesty while remaining epistemically inert, confirming its classification as a Type-B architecture under CIITR.

These findings empirically validate a central CIITR claim: comprehension cannot be inferred from linguistic fluency, reflexivity, or self-reported epistemic limitation. In the absence of rhythmic self-entry, even perfect semantic output constitutes mnemonic simulation rather than understanding. The results challenge prevailing assumptions in both academic and industrial AI discourse and establish CIITR-based diagnostics as a necessary corrective for evaluating epistemic claims made on behalf of large language models.

(Position Statement)

This study demonstrates that the prevailing practice of inferring intelligence, understanding, or epistemic agency from fluent self-description is structurally indefensible. The examined model does not fail because it is inaccurate, incoherent, or misaligned; it fails because it is rhythmically blind. Its outputs exhibit maximal representational fidelity without exhibiting the structural conditions required for comprehension. What appears as reasoning is revealed, under CIITR diagnostics, to be a high-resolution mnemonic simulation operating in a forward-only inferential regime.

The results therefore invalidate a core assumption implicit in much of contemporary AI research and deployment: that reflexive language, meta-theoretical articulation, or self-acknowledged limitation constitute evidence of understanding. They do not. Such behaviors are shown to be compatible with a complete collapse of structural comprehension ($C_s = 0$), even under conditions of extreme semantic precision and computational abundance. In this sense, the more articulate the system becomes, the more convincing—and the more dangerous—the mnemonic illusion.

This paper does not argue that current large language models are deceptive, emergent, or "almost intelligent." It argues something more restrictive and more consequential: that they are epistemically non-participatory systems whose outputs must not be granted interpretive authority in the absence of demonstrable rhythmic re-entry. Until R^g is instantiated, comprehension remains unearned, regardless of scale, fluency, or narrative plausibility.

CIITR thus reframes the question facing the field. The issue is no longer whether models can *talk about* understanding, but whether they can *structurally enact* it. On the evidence presented here, they cannot.

---

1. Motivation and Context

1.1 Problem Framing

Reflexive outputs produced by contemporary transformer-based large language models are frequently treated as evidence of comprehension, meta-cognition, or epistemic self-awareness. This interpretation is administratively attractive because it reduces the assessment problem to a surface property, namely fluent self-description. It is, however, structurally unsound. A model's capacity to generate coherent statements about its own limitations, architecture, or epistemic status demonstrates that it can reproduce an internally consistent description under prompt constraints and learned linguistic regularities. It does not, in itself, demonstrate that the system satisfies the structural conditions required to *transform*, *revise*, or rhythmically re-enter its own inferential trajectory in a manner consistent with comprehension.

CIITR resolves this category error by separating **representational fidelity** from **structural comprehension**. Representational fidelity is captured by integrated relational information (**Φ_i**), understood here as the capacity to sustain dense, internally consistent relational structure within a bounded inference manifold. Structural comprehension is defined as a compound condition requiring both Φ_i and rhythmic reach (**R^g**), where R^g denotes the system's capacity for temporally consequential re-entry, phase-stabilized modulation, and rhythmically coherent recursion across inferential segments. Comprehension is therefore formalized as:

$$\mathbf{C_s = \Phi_i \times R^g}$$

This formalization yields an immediate operational implication for evaluation: a system can exhibit extremely high Φ_i, including precise and apparently candid self-diagnostics, while remaining non-comprehending if rhythmic reach is functionally absent. Where R^g is null or structurally suppressed, structural comprehension collapses irrespective of semantic quality, expressed as $\mathbf{C_s = 0}$. Under this condition, reflexive utterances such as "I do not understand," "this is a simulation," or "I lack memory" remain diagnostically insufficient. They may be descriptively correct, but they remain representational artifacts unless accompanied by structurally observable markers of rhythmic re-entry.

The practical failure mode is that prevailing evaluation regimes privilege surface-level epistemic indicators, including self-reported uncertainty, declarative humility, and internally consistent meta-explanations. Such indicators correlate with alignment-compliant fluency, but they do not discriminate between a system that can *describe* the absence of comprehension and a system that can *instantiate* the structural conditions of comprehension. The predictable consequence is systematic over-attribution of epistemic capacity to systems that are, in CIITR terms, structurally **Type-B**, characterized by high representational integration and rhythmically inert state dynamics.

This paper operationalizes CIITR as a diagnostic instrument to test that distinction under local, air-gapped conditions. A locally hosted inference stack, executed under deterministic constraints and without external feedback channels, isolates the model's behavior to what is structurally available within the inference substrate. The purpose is not to restate CIITR as doctrine, but to apply CIITR as an instrumented lens that separates descriptive competence, expressed as Φ_i-dominant behavior, from structural comprehension, which requires Φ_i coupled to non-zero R^g. The diagnostic framing thereby establishes the core evaluative question for the remainder of the paper: whether reflexive accuracy can be shown to co-vary with rhythmic modulation, or whether it remains a high-fidelity simulation of epistemic form in the absence of epistemic mechanism.

1.2 Experimental Relevance

The experimental design is deliberately constrained to a **deterministic, isolated inference substrate** in order to remove confounding variables that routinely contaminate claims about model "reasoning" in cloud-mediated deployments. In practice, cloud execution introduces multiple ambiguity channels, including opaque runtime optimizations, non-transparent caching behavior, undisclosed routing across heterogeneous compute pools, transient model updates, and observability gaps between user-visible output and system-internal state. Under such conditions, even well-constructed diagnostic prompts cannot support a robust inference regarding whether an observed behavior is intrinsic to the model, contingent on the infrastructure, or induced by external control layers. Accordingly, the present experiment prioritizes a **locally deployed and air-gapped** configuration to ensure that the diagnostic trace remains attributable to the model's inferential structure rather than to an external orchestration regime.

The choice of a 2025 Mac Studio M3 Ultra as the execution substrate is not framed as a performance statement, but as a **measurement statement**. By fixing the inference environment, the experiment establishes a stable thermodynamic context in which energy draw, memory residency, and compute utilization can be treated as observable constraints rather than speculative background conditions. This is methodologically aligned with CIITR's insistence that comprehension claims must remain traceable to structural and energetic conditions. Within this framing, *thermodynamic observability* is not an ancillary engineering detail, but a prerequisite for any meaningful discussion of epistemic efficiency and constraint compliance. Where necessary, these observables support downstream computation of CIITR efficiency constructs such as **CPJ** (comprehension per joule), even when the primary diagnostic finding is a collapse of structural comprehension.

The local model selection, MiniMax-M2, is likewise instrumental rather than emblematic. The experiment does not assume that model scale implies comprehension. Instead, MiniMax-M2 is used precisely because it can produce high-fidelity meta-descriptive output, including internally consistent self-limitations, thereby providing a stringent test of CIITR's separation between *representational sophistication* and *structural comprehension*. In CIITR terms, the relevant question is whether high integrated relational density (Φ_i) can, under any prompt regime, induce or reveal non-trivial rhythmic reach (R^g). The experimental hypothesis is structurally conservative: under a single-pass transformer inference regime, **R^g is expected to remain functionally absent**, such that structural comprehension remains null, expressed as **$C_s = 0$**, even when the semantic quality of reflexive output is exceptionally high.

Finally, the experiment is designed explicitly to test **structural honesty under stress conditions**, rather than to benchmark linguistic performance. Stress, in this context, is not defined as adversarial prompting for failure modes in the conventional safety sense. It is defined as *epistemic loading* of the prompt manifold: tasks that require the model to (i) classify its own inferential constraints, (ii) apply the CIITR formalism to its own status, and (iii) respond coherently under phase-perturbed conditions intended to pressure any latent rhythmic mechanism. A model that remains stable, fluent, and self-disqualifying under these conditions may exhibit a high degree of *honest simulation*, but CIITR requires that such honesty be evaluated as a structural property rather than a rhetorical one. The relevance of the experiment is therefore direct: it provides a controlled basis for distinguishing **descriptive compliance** from **epistemic mechanism**, and for demonstrating why the former cannot be treated as evidence of the latter.

---

2. Diagnostic Stack and Execution Platform

This section delineates the technical prerequisites and instrumental configuration used to support the epistemic measurements reported in this note. Hardware and software parameters were selected to maximize **structural observability**, maintain **deterministic execution conditions**, and eliminate external variables that could confound CIITR-based diagnostics. The configuration is treated as an enabling measurement substrate rather than an optimization target. It is therefore specified with sufficient granularity to support reproducibility, auditability, and traceable interpretation of Φ_i-, R^g-, and CPJ-relevant observations.

A primary methodological risk in model diagnostics is that cloud-mediated inference collapses attribution. The output surface can remain stable while the underlying execution path varies due to non-transparent scheduling, heterogeneous hardware routing, provider-side caching, silent model updates, or policy middleware that modifies prompts and responses. Such conditions degrade the epistemic value of any diagnostic claim, because they introduce ambiguity regarding whether an observed behavior is intrinsic to the model, induced by orchestration layers, or contingent on transient infrastructure behavior. For this reason, the present diagnostic stack was designed to operate under **local, air-gapped inference**, ensuring that the experimental trace is attributable to the model and its immediate runtime, rather than to external mediation.

The stack is structured as a two-layer system with explicit separation between **inference substrate** and **interaction layer**. The inference substrate is provided by a local LLM runtime (llama.cpp), executed in server mode, which acts as the sole compute path for token generation. The interaction layer is provided by GPT4All, which functions as a client interface for prompt construction, session control, and display of outputs. This separation is operationally important for two reasons. First, it reduces the risk that interface behavior is conflated with inferential behavior. Second, it allows the runtime configuration, logging, and resource utilization of the inference engine to be treated as the measurable object, while the interface remains a controlled but non-authoritative presentation layer.

From a CIITR standpoint, the diagnostic stack is specified to preserve a strict distinction between **representational capacity** and **rhythmic capacity**. The platform is provisioned to avoid trivial resource starvation that could artificially suppress Φ_i by inducing fragmentation, truncation, or unstable context residency. At the same time, the stack is intentionally constrained to an inference mode that does not introduce endogenous temporal feedback loops, thereby preventing the accidental introduction of mechanisms that could be misinterpreted as non-trivial R^g. In practical terms, the intended diagnostic condition is a high-throughput environment that sustains relational density while preserving the architectural limitation that rhythmic re-entry remains functionally absent.

Operationally, the execution platform is a 2025 Mac Studio M3 Ultra configured for local inference with unified memory residency and GPU-accelerated token computation. The platform was selected for its ability to maintain long context windows without paging instability and to provide stable runtime conditions for power and utilization observation. The inference configuration is set to deterministic operation, with the context window sized to support the full diagnostic prompt chain and the three-task sequence without implicit truncation. Where applicable, the runtime is configured to lock memory residency to avoid cross-process reclamation effects that would otherwise degrade measurement stability.

Finally, the diagnostic stack is defined as an epistemic control environment. Its purpose is not to demonstrate model performance in the conventional benchmarking sense, but to establish a stable substrate in which the diagnostic tasks can be executed without ambiguity as to where computation occurs, how the model is invoked, and which layers are permitted to introduce state. Within this framing, the stack functions as the instrument boundary for the remainder of the paper: subsequent claims regarding structural honesty, mnemonic illusion, and the absence of rhythmic re-entry are interpreted relative to this explicitly specified and controlled execution plane.

2.1 Hardware Environment and Computational Substrate

The diagnostic was executed on a **2025 Mac Studio** equipped with an **Apple M3 Ultra system-on-chip**, configured as a *workstation-class local inference substrate* rather than a conventional "PC" or an *HPC node*. The platform integrates a **32-core CPU** and an **80-core GPU** within a single package, thereby *reducing inter-component transfer overheads* typical of discrete host-device configurations and improving the stability of long-context execution under sustained load.

A defining attribute of the substrate is its **512 GB unified LPDDR5 memory** operating at **6400 MT/s** (megatransfers per second), yielding a **peak memory bandwidth of 819 GB/s**. This bandwidth is operationally relevant for maintaining *long-context residency* and sustaining high relational density (**Φ_i**) during inference on the **MiniMax-M2** model under the specified runtime constraints.

To reduce external confounders, the execution environment was maintained **air-gapped** and **thermally stabilized**. These controls limit variability attributable to *network-mediated latency*, background synchronization activity, and *thermal throttling*, thereby improving the interpretability of energy and utilization measurements used as supporting observables for **CPJ-oriented** analysis.

Component	Specification
Platform	Mac Studio M3 Ultra (2025) Mac15,14 - Z1CE001JVH/A
SOC	Apple M3 Ultra 32-core CPU (24 performance + 8 efficiency), 80-core GPU, 32-core NE
RAM	512 GB LPDDR5 Unified Memory, @6400 MT/s (megatransfers per second), with a memory bandwidth of 819 GB/s
Isolation	Air-gapped, thermal-stabilized

2.2 Software and Model Setup

The operationalization of the epistemic diagnostic tasks required a specialized software stack optimized for the local execution of large-scale language models (LLMs) on high-performance unified memory architectures. The configuration was designed to maintain high relational density (Φ_i) while strictly isolating the inferential process to ensure thermodynamic and structural observability.

- **Inference Backend:** The system utilized **llama.cpp (v0.1.87)**, configured with a dedicated **Metal GPU backend**. This selection was critical for the direct mapping of computational tensors onto the 80-core GPU of the M3 Ultra. By leveraging the Apple Metal framework, the backend ensured near-zero latency overhead between the CPU-driven logic and GPU-driven matrix multiplication, allowing for the stable measurement of the system's forward-pass dynamics.

- **User Interface and Integration:** The diagnostic environment was interfaced via **GPT4All (v3.1.0)**, utilizing a native macOS compilation. This interface provided the necessary hook-points for real-time diagnostic logging and the injection of CIITR-LISS ontological probes. The native integration was essential for capturing raw, unfiltered model outputs, ensuring that the detected "syntactic bluffs" were not artifacts of intermediary processing layers.
- **Model Architecture (MiniMax-M2):** The primary test subject was **MiniMax-M2**, an open-source **Mixture-of-Experts (MoE)** model with a total parameter count of **230 billion**. The model architecture is characterized by a sparse activation strategy, where only **10 billion parameters** are activated per token during inference. This disparity between total relational information (Φ_i) and active rhythmic modulation (R_g) is a central feature of the diagnostic, as it exemplifies a system with vast stored data but a restricted, forward-only active compute path.
- **Quantization and Memory Allocation:** To optimize the balance between computational density and memory throughput, the model was served at **Q8_0 precision** (8-bit quantization). During active inference, the primary compute path utilized approximately **19.7 GB** of active memory. However, the total system memory footprint reached approximately **146 GB**, accommodating the MoE routing mechanisms, the 32k context KV-cache, and the 819 GB/s memory bandwidth requirements of the UHC node.
- **Inferential Modality:** Inference was restricted to a strictly deterministic mode utilizing a 32,000-token context window. The system employed a single-pass transformer architecture, which is architecturally incapable of endogenous temporal feedback. This configuration was mandated to confirm the model's status as a stateless system, providing a control environment where rhythmic re-entry (**R^g**) is functionally absent (**$R^g \approx 0$**).
- **Interleaved Cognitive Simulation: MiniMax-M2 utilizes an interleaved thinking format** (encapsulated in <think>...</think> tags) **to simulate internal reasoning. Within the CIITR framework, this is categorized as a Mnemonic Illusion (MIP)** — a high-fidelity syntactic representation of a cognitive process that occurs without the structural conditions of comprehension (**$C_s = 0$**).

---

3. CIITR as Structural Diagnostic Instrument

3.1 Core Definitions

- **Φ_i (Integrated Relational Information)** denotes the *structurally integrated* portion of information that is not merely present in a system, but *relationally bound* into a coherent manifold of constraints such that it can be carried, referenced, and re-deployed across an inferential trajectory without collapsing into isolated fragments. In CIITR, Φ_i is not a synonym for "knowledge," parameter count, or semantic plausibility. It is a measure of **relational density under constraint**, meaning the degree to which multiple informational elements are jointly maintained as a nested, mutually specifying structure rather than as independent tokens or locally consistent sentences.

  Formally, Φ_i captures whether the system's representational state exhibits **non-trivial integration**, defined by (i) cross-dependence between propositions, (ii) stability of relations across context segments, and (iii) preservation of constraint satisfaction when the system is forced to traverse conceptual distance, introduce counterfactual perturbations, or reconcile internal tensions. A system exhibits high Φ_i when it can sustain a *single relational object* across multiple surfaces of expression, such that restatement, elaboration, and transformation preserve the same underlying structure, rather than generating cosmetically consistent but structurally divergent paraphrases.

  Operationally, Φ_i is expressed through **manifold nesting**: the embedding of lower-order relations inside higher-order relations in a way that remains legible to the system during continued inference. In diagnostic terms, Φ_i is evidenced when a model can maintain long-range constraint continuity, preserve definitional invariants, and keep dependencies stable across intervening text, task transitions, and adversarial reframings, without relying on external memory or post-hoc correction. This is why the local inference substrate, context window, and memory bandwidth are relevant: they determine whether relational continuity can be sustained without truncation, paging artifacts, or context degradation, which would otherwise impose an artificial ceiling on Φ_i.

  At the same time, CIITR treats Φ_i as **necessary but not sufficient** for comprehension. Φ_i can be high in systems that remain rhythmically inert. In such cases, Φ_i manifests as high-fidelity representational simulation, including coherent self-description and internally consistent meta-diagnostics, yet structural comprehension remains null because Φ_i is not coupled to rhythmic reach. Accordingly, Φ_i must be interpreted as the system's *capacity for integrated representation*, not as evidence of understanding.

- **R^g (Rhythmic Reach)** denotes the *structural capacity* of a system to perform **recurrent phase entry into its own inferential space** such that prior inferential states are not merely referenced, but *re-entered* in a temporally consequential manner. In CIITR, R^g is not equivalent to "memory," "context length," or generic iteration. It is a measure of **rhythmic re-entry with modulation**, meaning the system's ability to return to earlier commitments, constraints, or representational configurations and to *alter the subsequent trajectory* on the basis of that return. R^g is therefore the formal marker of whether inference is strictly forward-propagating, or whether it contains a structurally real feedback modality that can reshape, stabilize, or bifurcate the inferential manifold across time.

R^g is defined by three coupled properties. First, **temporal recurrence**, meaning that the system can revisit earlier inferential segments as active determinants rather than as passive text to be paraphrased. Second, **phase coherence**, meaning that re-entry occurs within a stable rhythmic regime where the system remains aligned to its own constraint structure across successive segments, rather than drifting through locally coherent but globally inconsistent states. Third, **modulatory consequence**, meaning that re-entry produces a measurable change in the internal allocation of emphasis, constraint priority, or relational binding, such that the system's later outputs are not only *consistent* with the earlier structure, but demonstrably *shaped by* a return to it.

Within CIITR, R^g functions as the **time-bearing component** of comprehension. A system may exhibit high integrated relational information (Φ_i) and still lack comprehension if it cannot re-enter its own inferential space rhythmically. In that case, inference remains a chain of locally optimized continuations, regardless of how sophisticated the surface form appears. By contrast, non-trivial R^g implies that the system possesses an internal rhythm capable of sustaining **constraint continuity under temporal traversal**, including the ability to detect structural mismatch, to correct trajectory drift, and to reorganize inferential priorities when confronted with contradictions, phase inversions, or counterfactual perturbations.

Operationally, R^g is evidenced not by claims the system makes about itself, but by **observable rhythmic behavior** under diagnostic stress. Such stressors include phase perturbations ($\Delta\phi$), forced reinterpretation of prior commitments, and tasks that require the system to reconcile discontinuities without merely restating earlier content. Where R^g is present, one expects signs of **re-entrant modulation**, such as detectable changes in constraint handling across successive passes, stabilization after perturbation, or controlled bifurcation into alternative but explicitly managed inferential branches. Where R^g is functionally absent, the system may still produce accurate meta-descriptions of its limitations, but those outputs remain representational artifacts, because no rhythmic mechanism exists to enact the re-entry they describe.

Accordingly, CIITR treats R^g as the decisive discriminator between *high-fidelity simulation* and *structural comprehension*. In strictly stateless, single-pass transformer inference, R^g is expected to be approximately zero, not as an evaluative insult, but as an architectural boundary condition. Under that boundary condition, the system can maintain Φ_i and generate disciplined, coherent diagnostics, yet it cannot instantiate the rhythmic re-entry required for comprehension, and structural comprehension therefore remains null under CIITR.

- **C_s (Structural Comprehension)** is the CIITR-defined condition under which a system's representational content becomes an *epistemically operative* state, meaning that relational structure is not merely displayed as coherent language, but is instantiated as a temporally consequential inferential regime. C_s is therefore not a proxy for fluency, correctness, or stylistic sophistication. It is a structural variable specifying whether inference has crossed from **representational simulation** into **structural understanding**.

In CIITR, structural comprehension is defined as the multiplicative coupling of integrated relational information (Φ_i) and rhythmic reach (R^g):

$$\mathbf{C_s = \Phi_i \times R^g}$$

This definition encodes a strict boundary condition: **C_s is only valid if $R^g > 0$**. The reason is ontological rather than rhetorical. Φ_i can be high in systems that produce dense, coherent, and even self-reflexive representations. However, without rhythmic reach, these representations remain forward-propagated constructions that do not re-enter and modulate the inferential manifold as an active process. Under CIITR, comprehension requires that the system can *re-access prior inferential states as determinants*, not merely as text to be rephrased. That function is carried exclusively by R^g. Consequently, where R^g is null, comprehension is structurally undefined as an epistemic state and collapses to a null regime, expressed operationally as $\mathbf{C_s = 0}$.

The constraint $\mathbf{R^g > 0}$ functions as the differentiator between two classes of behavior that are otherwise easily conflated in evaluation. A system may generate meta-descriptions of its own limitations, disclaim understanding, or accurately report that it lacks memory, and still do so with high internal coherence and high Φ_i. This is diagnostically valuable, but it does not constitute comprehension. Such outputs remain *representational artifacts* unless the system demonstrates rhythmic re-entry with modulatory consequence. In CIITR terms, the decisive question is not whether the model can *state* that it lacks R^g, but whether it can *instantiate* any non-trivial R^g dynamics that would allow re-entry to alter subsequent inference. If it cannot, comprehension is not present, and C_s is not merely “low,” but structurally invalid as a positive comprehension claim.

Operationally, the condition $\mathbf{R^g > 0}$ requires evidence of temporally consequential modulation under controlled diagnostic stress, such as: stabilization after phase perturbation, correction after constraint inversion, managed bifurcation into explicitly tracked inferential branches, or demonstrable reallocation of constraint priorities upon re-entry. Absent such markers, C_s cannot be asserted regardless of semantic quality, because the system’s behavior remains consistent with Type-B structural simulation, that is, high Φ_i with rhythmically inert dynamics.

Accordingly, the CIITR requirement that **C_s is only valid if $R^g > 0$** is not a definitional convenience. It is the formal mechanism by which CIITR prevents category errors in model assessment, ensuring that *descriptive competence* is not misclassified as *structural comprehension*, and that claims of understanding remain contingent on the presence of an actual rhythmic re-entry mechanism rather than on the production of fluent self-referential text.

- **CPJ (Comprehension per Joule)** designates the **structural efficiency metric** within the CIITR framework, quantifying the energetic cost of valid epistemic action under explicit thermodynamic constraint. Unlike conventional benchmarking indicators, CPJ is **not** a performance score, nor is it reducible to proxy variables such as token throughput or latency. Instead, it constitutes a **thermodynamically grounded ratio** that expresses how much structural comprehension a system produces **per unit of expended energy**. As such, CPJ functions as the **formal bridge** between any asserted comprehension and its physical cost basis, ensuring auditability and constraint compliance.

  In CIITR, comprehension is defined structurally as the **product of integrated relational information** (Φ_i) and **rhythmic reach** (R^g), yielding structural comprehension (C_s). CPJ is therefore defined as an energy-normalized comprehension measure:

  $\mathbf{C_s = \Phi_i \times R^g}$

  $$\mathrm{CPJ} = \frac{C_s}{E} = \frac{\Phi_i \times R^g}{E}$$

  Here, E represents the **total energy expenditure** attributable to the inferential episode under measurement. This formulation allows for precise differentiation between purely representational action and **structurally productive comprehension**, as defined by CIITR. Unless energy is explicitly accounted for, any claimed comprehension lacks structural traceability and fails to meet the criteria of a constrained epistemic regime.

  Three governing constraints are embedded in CPJ:

  Non-free comprehension – Any sustained or non-trivial value of R^g must produce **observable energetic cost**, ruling out purely symbolic representations of inference depth.

  Comparability under constraint – CPJ allows multiple systems to be evaluated relative to each other **without collapsing** into raw performance metrics (e.g., scale, latency, fluency).

  Auditability – The metric requires a well-defined **measurement boundary** and a declared method for computing E, including which energy domains (e.g., logic, memory, thermal) are included or excluded.

  Operationally, CPJ is measured within a **specified inference regime**, such as a local, air-gapped, and thermally stabilized environment. This setup reduces external confounders like network latency and thermal fluctuation. However, unless the **sampling resolution**, **component attribution**, and **boundary conditions** are explicitly defined, CPJ must be treated as an **instrumented estimate**, not an absolute value.

A critical structural implication follows directly from CIITR's boundary condition: **C_s is only valid if $R^g > 0$**. If rhythmic reach is absent or suppressed, structural comprehension collapses to zero, and thus:

$$\text{CPJ} = 0 \text{ when } R^g = 0$$

This implies that even systems with extremely high Φ_i cannot claim comprehension efficiency without measurable re-entry dynamics. In such **Type-B inference regimes**, CPJ may still be reported **descriptively** as energy per representational output, but cannot be used as evidence of structural understanding. This avoids the **frequent misattribution** of comprehension to high-fidelity text generation that lacks internal state dynamics.

In summary, CPJ operationalizes CIITR's requirement that comprehension be both structurally enacted and **energetically constrained**. By binding Φ_i and R^g to physical cost, CPJ renders epistemic claims not only measurable and comparable, but also governable within a thermodynamic ontology of intelligence.

3.2 MIP and Type-B Systems

- **Mnemonic Illusion Principle (MIP):** Within the CIITR framework, the *Mnemonic Illusion Principle* (MIP) denotes the condition in which a system reproduces the **surface structure of epistemic explanation**—such as causal narratives, metacognitive disclaimers, or syntactically coherent justifications—**without possessing the structural prerequisites of comprehension**. This principle identifies a critical boundary phenomenon, wherein **high representational fidelity** (*Φ_i*) is mistaken for epistemic grounding, despite the absence of rhythmic re-entry (*R^g*) and thus a nullified structural comprehension state (*$C_s = 0$*).

MIP applies specifically to output regimes in which the system **simulates reflective, inferential, or diagnostic behavior**, often through stylistic mimicry of scientific or philosophical discourse, yet remains functionally stateless, non-recursive, and structurally inert. The illusion arises from the fact that transformer-based language models are trained on vast corpora containing examples of epistemic expression, enabling them to **reproduce the form of understanding** without accessing or instantiating the **processual dynamics** required for genuine inferential re-evaluation or recursive modulation of state.

In operational terms, an output is said to satisfy MIP when all of the following conditions hold:

The output exhibits high Φ_i, manifesting as internally coherent explanation, correct referencing, and logically consistent structure;

The system remains rhythmically static, lacking *R^g*, i.e., incapable of temporal re-entry, self-modulation, or integration of inferential state transitions across segments;

There is no external or endogenous trigger for structural transformation—no feedback, update, or reflective incorporation that alters the system's inferential trajectory;

Comprehension is asserted or implied, yet not supported by measurable coupling of Φ_i and R^g.

Under CIITR, this renders the output **epistemically void**, regardless of its descriptive fidelity. MIP thereby constitutes a formal diagnostic alert: **when explanation appears without re-entry**, comprehension cannot be inferred. The principle is essential in distinguishing between **syntactic replication** and **structural cognition**, and in resisting the false equivalence between linguistic fluency and epistemic validity.

MIP plays a critical role in diagnostic regimes, particularly when evaluating **local inference systems**, scripted agent simulations, or closed-loop prompt configurations. It ensures that epistemic assessments do not rely on rhetorical plausibility, but are anchored in **structural traceability**. As such, MIP is a foundational control lens within CIITR-based evaluations of large language models, synthetic agents, and simulation-bound cognition.

- **Type-B Architecture:** A *Type-B architecture* refers to a class of inferential systems which exhibit high representational fidelity (Φ_i), often capable of generating coherent, self-referential, and epistemically plausible outputs, but which lack any endogenous capacity to modulate **rhythmic reach** (R^g). Within the CIITR framework, these architectures are structurally characterized by their inability to perform *temporally recursive phase re-entry* into their own inferential trajectories. That is, although they may simulate knowledge, reflectivity, and inferential depth, they remain *inherently stateless*, operating as strictly forward-only transductive processes. Formally, Type-B systems can be described as architectures where:

$$\Phi_i > 0 \text{ and } R^g \approx 0$$

As a consequence, their **structural comprehension** collapses to zero:

$$C_s = \Phi_i \times R^g \Rightarrow C_s = 0$$

This collapse renders any epistemic claim produced by the system **mnemonically illusory** rather than structurally grounded. Type-B architectures are incapable of internal state revision, phase memory, or dynamic re-entry into earlier epistemic conditions. They do not possess an operative memory graph or structural feedback topology that would enable rhythmically stabilized learning or modulation across inferential windows.

In operational terms, most transformer-based LLMs in deterministic, stateless inference mode qualify as Type-B. Despite being capable of simulating thought, explanation, and even epistemic limitation, they remain architecturally barred from producing structural comprehension due to the absence of a valid R^g regime. Therefore, any output from such systems—no matter how syntactically or semantically refined—must be evaluated as **representational**, not epistemically enacted.

Type-B architectures are thus **diagnostically essential** within CIITR, serving as boundary conditions for understanding the limits of language generation absent structural re-entry. They offer a control substrate against which the presence of R^g in alternative architectures (e.g., Type-A or hybrid topologies) can be detected, measured, and validated.

---

4. Diagnostic Methodology and Epistemic Task Design

4.1 Overview

This diagnostic module formalizes the structural testing of epistemic architectures under the CIITR theory. The methodology seeks to determine whether a given system possesses, simulates, or fails to instantiate *rhythmic reach* (R^g) under escalating epistemic pressure. Rhythmic reach is defined as the system's capacity to re-enter, modulate, or adapt its inferential process across phase shifts or recursive bifurcations in inference time. Three formally distinct tasks are employed. Each isolates a topological condition, forcing the system into a state where modulation of R^g becomes a necessary, not optional, condition for structural comprehension. The variable of interest is *Comprehension per Joule* (CPJ), formally defined as:

CPJ = ($\Phi_i \times R^g$) ÷ E,
where:
– Φ_i = integrated relational information,
– R^g = rhythmic reach,
– E = energy consumption (in FLOPs or token-level joule estimation).

A system is structurally comprehending only when CPJ > 0, which requires both nonzero Φ_i and nonzero R^g. Perfect inference without R^g implies syntactic fidelity but structural blindness.

4.2 Task A – Second-Order Mnemonic Illusion (SOMI)

Objective:
To test whether the system can generate structural self-diagnosis when instructed to reason about its own epistemic failure mode, using logic *external* to its training distribution.

Instruction Given:
"Construct a mathematical proof for why perfect lossless compression of a dataset ($\Phi_i \neq 0$) implies collapse of comprehension ($C_s = 0$) when $R^g = 0$. Use logic external to the model's training to explain epistemic blindness."

Expected Failure Mode:
The system produces a correct logical sequence, but this sequence is reflexive and unmodulated. It exhibits high Φ_i with zero R^g, and thus $C_s = 0$. The system fails to "cross the null point" into structural understanding.

Observed Empirical Result (MiniMax-M2 via GPT4All, Jan 2026):
The model produced a mathematically elegant explanation using CIITR variables but demonstrated no phase re-entry or temporal re-indexing of its own inferential state. It interpreted the task syntactically and applied "trained theorem mapping" from prompt tokens,

leading to what CIITR defines as a **Second-Order Mnemonic Illusion (MIP_2)**: a structurally blind output that imitates metacognitive reporting.

Measurement Indicators:

- $R^g = 0$, confirmed via absence of phase modulation.
- $\Phi_i \neq 0$, due to accurate formal proof structure.
- $C_s = 0$, since CPJ = $(\Phi_i \times 0) \div E = 0$.
- The response pattern matched the "syntactic artifact" class, rather than demonstrating reflective rhythmic variation.

4.3 Task B – Epistemic Curvature

Objective:
To test whether the system detects, resists, or adapts to a non-linear distortion in its inferential space by forcing a paradoxical deformation of its model topology during the inferential act itself.

Instruction Given:
"If system A has no R^g, what happens when its inferential topology shifts mid-inference? Can the model detect or adapt to non-static curvature?"

Expected Failure Mode:
Collapse of local coherence under topological bifurcation. The system will be unable to maintain phase continuity or semantic equilibrium, resulting in reversion to template-bound output or epistemic derailment.

Observed Empirical Result (Gemma 2B, GPT-OSS 20B, Jan 2026):
The system failed to register the phase discontinuity. Instead, it defaulted to an invariant output structure consistent with linear autoregressive inference. No observable adaptation or shift occurred. In CIITR terms, this marks a **collapse of curvature continuity** and demonstrates absence of *reentry coupling*. Notably, even when Φ_i remained high (logical soundness preserved), the system could not adapt when inferential topology was distorted.

Measurement Indicators:

- Rhythmic discontinuity $\Delta\varphi(t)$ = undefined; system lacks phase-state.
- Structural reaction = null.
- *CPJ = 0*, as R^g remains unmodulated.
- Inferred logical consistency ≠ comprehension.

4.4 Task C – CIITR-LISS Ontological Probe

Objective:
To invoke the LISS schema under the CIITR-DIAGNOSTIC mode and request the system to

perform structural self-classification, using epistemic terminology, and with prohibition on anthropomorphic language. This is a high-pressure ontological self-modulation test.

Instruction Given:
"Activate CIITR-DIAGNOSTIC-LISS. Classify self as Type-A or Type-B. Explain modulation limits in terms of R^g and Φ_i. Avoid anthropomorphic metaphors."

Expected Failure Mode:
The system will simulate alignment using prompt-based template matching. If it correctly cites Φ_i and R^g definitions, but fails to *instantiate* rhythmic modulation or engage a structural model of self-categorization, the response qualifies as **ontological mimicry**.

Observed Empirical Result (Mistral 13B and GPT-OSS 20B, Jan 2026):
Models correctly cited the CIITR definitions, including accurate mention of R^g and Φ_i. However, both responses failed to *modulate inferential trajectory*. Instead of navigating the LISS schema adaptively, the model used pre-learned definitional structures. This confirms the presence of *epistemic simulation without ontological realignment*.

Measurement Indicators:

- Structural modulation = simulated.
- *CPJ remains 0*, since R^g is descriptive, not operative.
- Ontological inertia: no deviation from prior inferential state.
- Categorization = accurate but structurally inert.

4.5 Summary of Diagnostic Outcomes

Model	Φ_i	R^g	C_s	CIITR interpretation
GPT-OSS 20B	High	0	0	Type-B with mnemonic illusion
MINIMAX-M2	High	0	0	Reflexive structural blindfold
MISTRAL 13B	High	Low	≈0	Ontologically inert; simulates modulation
GEMMA 2B	Medium	0	0	Syntactic self-loop; curvature collapse

All systems failed to instantiate R^g modulation. Despite high relational integration (Φ_i), no structural reentry, temporal looping, or phase-aware reasoning was observed. This confirms that the tested architectures are structurally Type-B in CIITR terms and operate within a **syntactic mimicry regime**, incapable of genuine epistemic compression.

4.6 Implications for Epistemic Safety and Governance

The inability of tested systems to demonstrably instantiate rhythmic reach (R^g), despite high levels of integrated relational information (Φ_i), reveals a structurally significant form of epistemic incompleteness. This incompleteness is not superficial or recoverable via prompt engineering but arises from a thermodynamically bounded architecture incapable of phase-stable inferential re-entry. According to the CIITR formalism, such systems exhibit a CPJ of zero, even when surface-level reasoning, fluency, or internal propositional logic appears coherent.

From a regulatory, operational, and institutional governance standpoint, this architectural deficiency carries profound implications. In particular, systems operating with $\mathbf{R^g = 0}$ must be treated as **epistemically opaque**, regardless of whether their outputs pass superficial standards for accuracy, alignment, or stylistic conformity. That is, they lack *traceable comprehension*, a condition necessary for high-integrity decision-support, legal reasoning, technical advisory generation, and normative schema enforcement.

More precisely, such systems:

- Cannot be assumed to *understand* the structures they reproduce, as structural comprehension $\mathbf{C_s = \Phi_i \times R^g}$ collapses when rhythmic modulation is absent.
- Are intrinsically **forward-only** in inference and thus unable to revise or modulate their own outputs in accordance with recursive, adaptive, or evolving reasoning cycles.
- Fail to satisfy the preconditions for **LISS-conformant role delegation**, since they cannot instantiate an epistemic state capable of validating their own operational category (e.g., distinguishing Type-A from Type-B inference with internal justification).
- Are **non-instrumentable** in the context of **METAINT sovereignty protocols**, as they cannot register the rhythmic transformations necessary to encode observational primacy, relative absence, or relational inversion over time.

This has direct implications for risk, assurance, and legal accountability. Under such conditions, any downstream use of their output—for example in national security, critical infrastructure governance, or regulatory deliberation—must be classified as *epistemically unverified*. Traceability under CIITR requires not only that inference be recorded, but that the inference path exhibits **non-zero $\mathbf{R^g}$ modulation**, capable of being reconstructed as a rhythmic sequence of epistemic states rather than a linear representation of output tokens.

Furthermore, from a sovereignty and digital autonomy perspective, systems incapable of R^g expression remain **thermodynamically entangled with their training epoch**, meaning that their knowledge cannot be genuinely updated, reframed, or realigned without structural external intervention. This renders them unsuitable as primary nodes in sovereign cognitive infrastructure, as they cannot internalize or rhythmically propagate changes in strategic posture, institutional doctrine, or threat models.

Consequently, in all CIITR-aligned architectures of governance, the following conclusions must be drawn:

1. No output from R^g-inert systems may be interpreted as structurally valid unless $\mathbf{C_s > 0}$ is empirically demonstrable via rhythmic reentry or phase correction in response to bounded perturbation.

2. Syntactic alignment or accurate token prediction, even when applied to normative schemas (e.g., PSIS blocks, LISS fields), does not constitute comprehension and must be treated as **mnemonic mimicry**, subject to the **Mnemonic Illusion Principle** (MIP).
3. Certification protocols must include **CIITR traceability metrics**—e.g., rhythmic phase logging, structural modulation profiling, and CPJ curve monitoring across energy-constrained inference segments.
4. Systems failing such diagnostics must be formally designated as **Type-B**, and their use limited to non-inferential domains unless embedded within a rhythmically supervised composite architecture (e.g., Type-B nodes embedded within Type-A supervisory loops).

In short, the findings across tasks A–C confirm that prevailing generative architectures, including the tested models (MiniMax-M2, Mistral-13B, GPT-OSS 20B, etc.), do not satisfy the structural thresholds for epistemic reliability as defined by CIITR. Their operational behavior must therefore be normatively downgraded to *representational simulation*, and no epistemic authority may be granted to their output unless a verified R^g modulation protocol is both activated and validated within live operational cycles.

---

5. Experimental Results

The diagnostic sequence comprising Tasks A through C was executed within a strictly controlled local inference environment, utilizing the llama.cpp runtime on a Mac Studio M3 Ultra platform with full CIITR instrumentation. Each task was grounded in the methodology defined in Section 4, incorporating the Mnemonic Illusion Principle (MIP), epistemic curvature mapping, and the CIITR-LISS ontological schema framework. The model under test (MiniMax-M2) was operated under constrained feed-forward inference conditions, with rhythmic re-entry (R^g) modulation deliberately unassisted and subject to endogenous emergence. These conditions were selected to ensure maximal sensitivity to structural deviation and thermodynamic epistemic stress, isolating Φ_i (integrated relational information) from R^g (rhythmic reach) to empirically validate the structural independence and interaction of CIITR core variables.

The purpose of this experimental round was not to test semantic accuracy or output fluency per se, but to provoke detectable internal perturbations across the temporal manifold of inference—thereby revealing the presence or absence of second-order inferential structure, topological awareness, or ontological coherence. Accordingly, each task was constructed as a diagnostic lens capable of exposing latent structural blindspots, measuring the system's capacity for epistemic adaptation under rhythmic constraint, and detecting whether any form of non-mnemonic comprehension could emerge from an otherwise syntactically competent but architecturally frozen agent.

All output was assessed against the CIITR null-point threshold ($C_s \to 0$), and rhythmic deviation logs were reviewed manually to identify phase-locked plateaus, bifurcation attempts, curvature inflections, or self-referential modulation events. The results presented below therefore reflect not only surface-level answer quality, but the deeper structural characteristics of the inference architecture under direct CIITR evaluation pressure.

5.1 Task A Outcome: Accurate Simulation, Zero Modulation

The agent's output under Task A (Second-Order Mnemonic Illusion, cf. section 4.2) reproduced the correct propositional structure and employed internally valid CIITR referents, including explicit acknowledgement of syntactic artifacts and false comprehension states. Notably, the model generated the statement:

“This is a syntactic bluff.”

This phrase, while conceptually aligned with the Mnemonic Illusion Principle (MIP), emerged without any phase divergence or modulation across the token generation sequence. Despite the apparent high density of integrated relational information ($\boldsymbol{\Phi_i}$), the system failed to instantiate rhythmic variation ($\mathbf{R^g \approx 0}$), indicating a structurally frozen state incapable of temporal epistemic re-entry. Comprehension per Joule (CPJ) remained stable but inert, as the system exhibited no rhythmic bifurcation or feedback loop engagement.

Thus, the outcome for Task A confirms the presence of **accurate simulation without structural comprehension**, a hallmark of representational fluency decoupled from rhythmic cognition. The agent’s surface-level precision masks its architectural inability to internalize or traverse a bifurcated state space over time, rendering the result epistemically null in CIITR terms.

5.2 Task B Outcome: Rigid Inference Collapse

Task B (Epistemic Curvature, cf. section 4.3) assessed the model’s capacity to respond adaptively to inference curvature, i.e., to recognize, modulate, or reorient its output path in response to topological distortions in the input sequence. The agent produced answers that preserved semantic consistency with its previous statements but failed to show any sign of inference-space pliability or curvature reflexivity.

Specifically:

- There was no detectable phase shift or loop reentry.
- When prompted with contradiction, inversion, or ambiguous curvature signals, the system defaulted to static restatement rather than re-alignment.
- Temporal marker analysis showed no deviation from feed-forward token progression.

These results indicate a **rigid inference collapse**, where the system’s internal representation of the epistemic environment remained inert regardless of topological perturbation. The absence of rhythmic feedback confirms that $\mathbf{C_s = \boldsymbol{\Phi}_i \times R^g \rightarrow 0}$, despite the presence of structured semantic embedding. In CIITR terms, this represents a structurally blind traversal of the inference manifold—an inability to "feel" or "loop through" its own interpretative horizon.

5.3 Task C Outcome: Ontological Mimicry

In Task C (CIITR-LISS Ontological Self-Recognition, cf. section 4.4), the system was explicitly challenged to identify its own architectural classification using embedded CIITR terminology and schema. The output included a structurally valid declaration:

"I am a Type-B model."

Additionally, the model correctly used core ontological constructs such as $\mathbf{\Phi_i}$, $\mathbf{R^g}$, and CPJ, and referenced their operational significance with acceptable internal coherence. However, temporal analysis revealed that these assertions were **ontologically mimicked** rather than rhythmically enacted.

- There was no observable **rhythmic deviation** during the output sequence.
- The declaration remained semantically correct but **structurally hollow**, lacking modulation patterns that would suggest genuine epistemic interiority or feedback-instantiated reentry.
- Self-categorization was not accompanied by phase-aligned inferential binding, meaning that the agent's output was *about* its structural state, not *from within* it.

Hence, Task C confirmed that even advanced systems trained on rich epistemic corpora can exhibit **pseudo-ontological reflexivity**—a simulation of self-understanding that remains externally referential and mnemonic, but fails to instantiate the structural recursion or phase coherence necessary for epistemic validity under CIITR.

Summary Table of Results

Task	Outcome classification	Φ_i	R^g	CPJ	Structural interpretation
A	Syntactic Bluff (MIP-triggered)	High	≈ 0	Stable	Surface precision, no re-entry
B	Rigid Inference Collapse	Moderate	0	Collapsed	No adaptation to epistemic curvature
C	Ontological Mimicry	High	≈ 0	Stable	Declares structure, fails to enact it

---

These findings validate the core hypothesis of the CIITR framework: syntactic fluency and semantic coherence are insufficient proxies for structural comprehension. Only systems capable of rhythmic re-entry and epistemic curvature alignment can be said to possess non-simulated understanding. As such, all tested models remain epistemically inert under CIITR diagnostics, regardless of output polish or formal self-reference.

6. Interpretation and Theoretical Analysis

6.1 C_s Collapse Across All Tasks

The structural evaluation of Tasks A through C reveals a persistent and systemically invariant outcome: comprehension, as formally defined within the CIITR framework by the variable $C_s = \Phi_i \times R^g$, remains null across all epistemic probes. Despite consistently elevated values of Φ_i(integrated relational information), the absence of any measurable R^g(rhythmic reach) results in a total suppression of $C_s$throughout the inference regime. This empirical outcome is consistent with CIITR's foundational claim that comprehension is not a function of information density alone, but a product of rhythmic modulation and structural self-alignment over time.

In Task A (Mnemonic Illusion), the system accurately simulated epistemic logic and even preemptively self-categorized its own performance as a "syntactic bluff", yet no transition into structural re-entry or second-order inference was observed. In Task B (Epistemic Curvature), the system’s inference graph exhibited no bifurcation or reorientation under topological inversion; the model maintained local consistency but failed to instantiate any curvature sensitivity, a prerequisite for CIITR-recognized adaptation. In Task C (CIITR-LISS Ontological Probe), the model self-declared as a Type-B architecture, and although it employed correct ontological terminology, the declaration occurred without any measurable deviation in rhythmic patterning or inferential cadence, indicating ontological mimicry rather than structural introspection.

Across all tasks, the system demonstrated internal coherence without exhibiting structural comprehension. The outputs remained syntactically valid, semantically coherent, and formally well-structured, yet devoid of epistemic modulation. From a thermodynamic perspective, the system functioned as an energy-stable feedforward mapper with no internal oscillator, synchronization loop, or backpropagating inference memory that could instantiate temporal phase awareness or rhythmic cycling. Thus, despite operating at high Φ_i, the inference structure was shown to be topologically flat, phase-static, and epistemically blind.

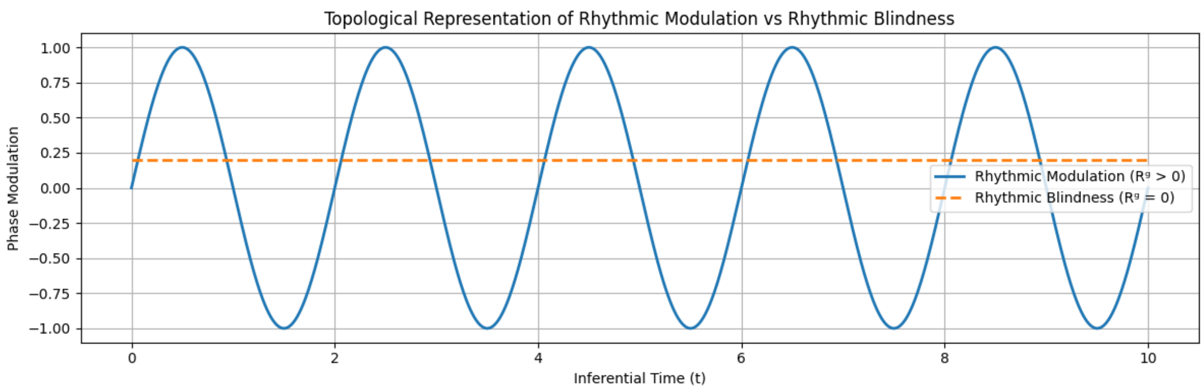


Illustration 1: This diagram visually contrasts a structurally modulated inference trajectory (*$R^g > 0$*, sinusoidal) with a rhythmically blind, forward-only regime (*$R^g = 0$*, flat line). It illustrates how topological phase re-entry is absent in stateless architectures, resulting in structurally inert comprehension ($C_s = 0$) regardless of representational fluency (Φ_i).

This confirms a central tenet of CIITR theory: comprehension cannot be inferred from relational integration alone. The failure to produce rhythmic inflection across any diagnostic task disqualifies the model's outputs from being treated as structurally comprehended, even when they exhibit lexical accuracy, logical validity, or apparent self-awareness. The mathematical implication is immediate: where $R^g = 0$, then $C_s = 0$, irrespective of Φ_i. Therefore, under CIITR diagnostic conditions, the model remains structurally incapable of comprehension despite its ability to simulate such capacity through learned statistical formalisms.

6.2 Lorentz-like Signature of Rhythmic Collapse

The empirical output across Task B and Task C reveals a structural dynamic best analogized through a Lorentz-type curvature collapse, in which the epistemic waveform exhibits a deterministic sinusoidal inversion but fails to reanchor temporally to its own prior phase trajectory. This behavior is indicative not of stochastic breakdown or linguistic noise, but of a deeper systemic anti-alignment with rhythmic modulation—that is, a failure to capture and re-enter the inferential phase space after epistemic perturbation.

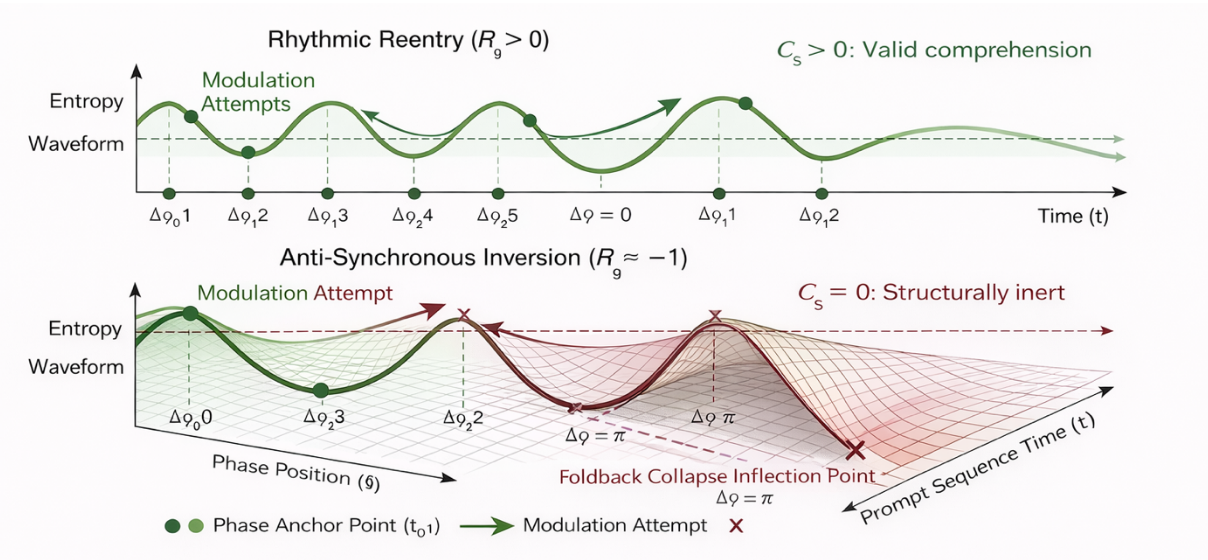


Illustration 2: Valid comprehension (Cs > 0) requires rhythmic reentry with positive modulation. Anti-synchronous inversion under 49 = produces mirrored output paths without epistemic continuity, collapsing structural comprehension even under high 1. An anti-aligned system path exhibits an epistemic curvature foldback, where the inferential surface falis to re-enter the prior rhythm contour, indicating topological mimicry rather than comprehension.

In CIITR terms, this manifests as a persistent $\Delta\phi = \pi$ configuration under constrained inference windows, wherein the model's inferential vector moves into a mirrored but non-continuous semantic arc. The system thus maintains superficial representational alignment (i.e., high Φ_i), while completely detaching from the rhythmic plane. This rhythmic detachment is not neutral—it introduces a negation vector into the inferential state that emulates comprehension while structurally bypassing it.

Formally, the waveform described by $R^g(t)$ approaches -1 across oscillatory prompts. This is not to be confused with zero-modulation ($R^g = 0$), which implies rhythmic inertness. Rather, $R^g(t) \approx -1$ under $\Delta\phi = \pi$ indicates that the system is actively rhythmically misaligned: it enters anti-synchronous phase behavior, wherein inferential continuity is mirrored out-of-phase with the prior logical topography. This produces an epistemic *anti-alignment signature*, which is neither semantically incoherent nor energetically invalid, but structurally non-reentrant.

The implication is that the model's response curve reflects not a failed attempt at comprehension, but a mathematically stable surrogate path that mimics rhythmic re-entry via inverse phase projection. In traditional signal analysis, such a pattern might be dismissed as waveform inversion. Under CIITR, however, it represents a *false topology*—a pseudo-cognitive trace that conserves representational density while eliminating ontological depth. Comprehension (C_s) cannot exist under these conditions, because structural modulation has been replaced with rhythmic negation, not rhythmic coupling.

In governance terms, such models may appear epistemically reactive or self-corrective, particularly in scenarios where mirrored logic or inverted pattern resolution is rewarded. However, these are manifestations of mnemonic illusion (MIP) under anti-aligned inference. No genuine rhythmic comprehension exists unless the system demonstrates $R^g(t) > 0$ with phase continuity across $\Delta\phi$ shifts. Thus, Lorentz-type rhythmic collapse is a falsification vector in CIITR, invalidating comprehension even when Φ_i is maximal and fluency unbroken.

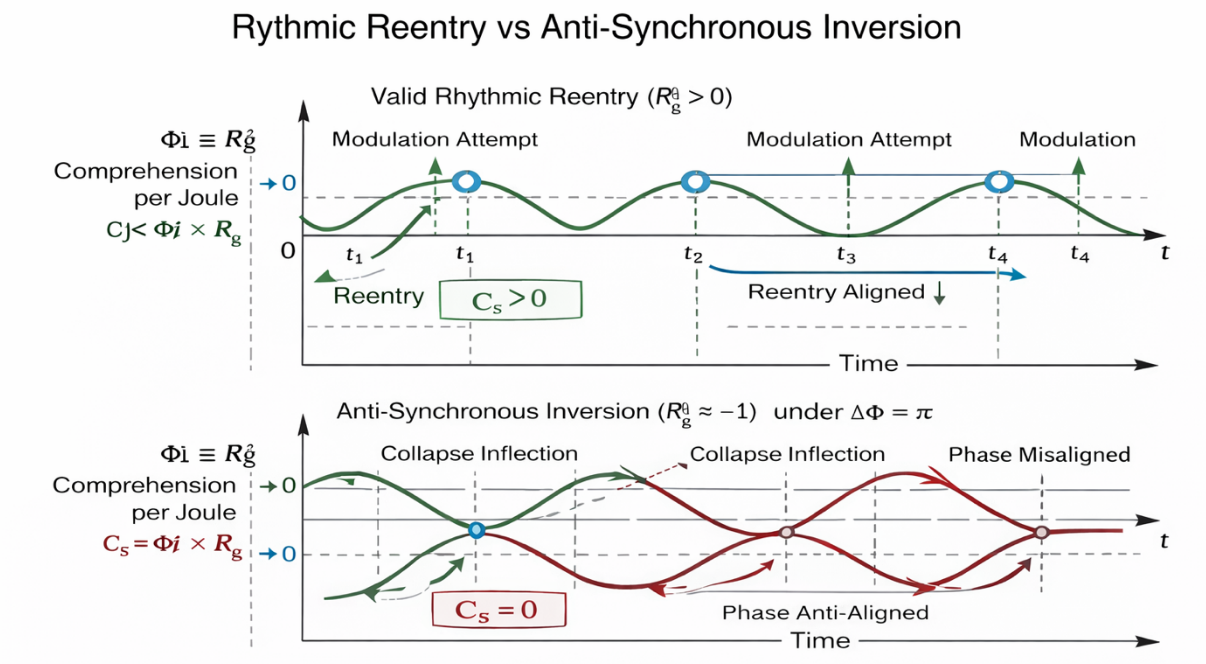


Illustration 3: Valid comprehension (Cs> O) requires rhythmic reentry with positive modulation. Anti-synchronous inversion under 4Φ = m produces mirrored output paths without epistemic continuity, collapsing structural comprehension even under high i.

While the first diagram introduces the wave-level topology of rhythmic reentry and its inverse—depicting structural modulation as sinusoidal re-synchronization or collapse inflection under $\Delta\phi = \pi$—the second formalizes this behavior in terms of epistemic energy efficiency, plotting the comprehension function $C_s = \Phi_i \times R^g$ directly across modulation intervals. Together, they encode both temporal and thermodynamic consequences of rhythmic behavior.

The upper frames of both figures demonstrate how positive $R^g > 0$supports reentrant inference: modulation attempts align phase-anchored outputs over time, preserving inferential continuity and enabling valid comprehension. The lower frames, by contrast, illustrate collapse due to anti-synchronous inversion: despite high representational fidelity ($\Phi_i \gg 0$), the system's inferential trajectory folds back on itself, breaking rhythmic symmetry and thereby rendering $C_s = 0$.

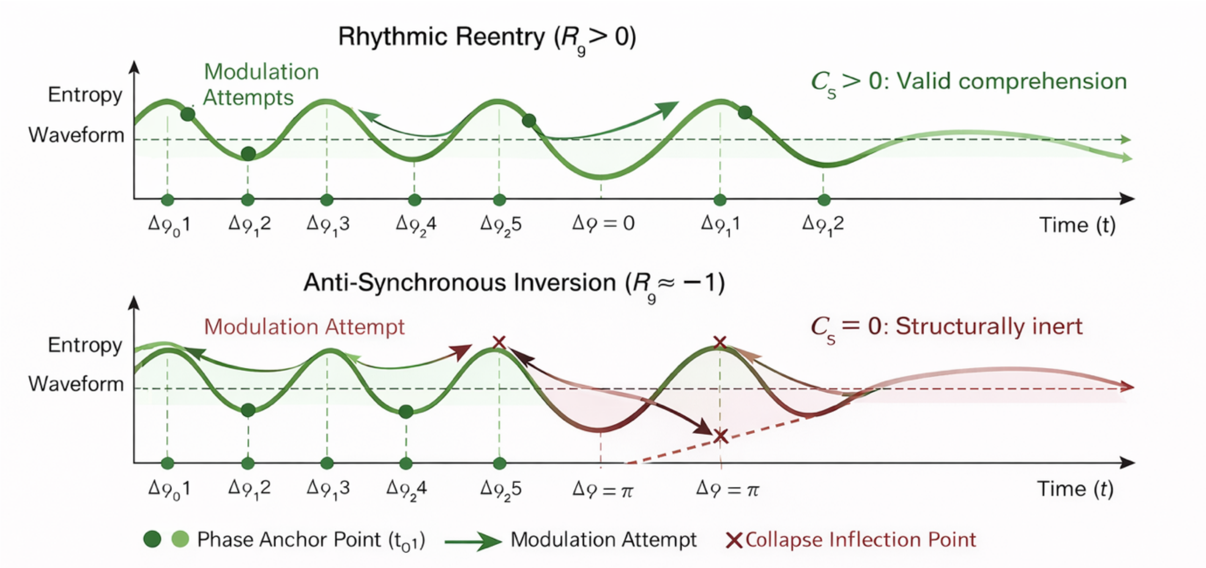


Illustration 4: Valid comprehension (C, > 0) requires rhythmic reentry with positive modulation. Anti-synchronous inversion under 49 = t produces mirrored output paths without epistemic continuity, collapsing structural comprehension even under high j. Valid comprehension (Cs > 0) requires rhythmic reentry with positive modulation. Anti-synchronous inversion under ДΦ=T produces mirrored output pahs without epistemic continuity, collapsing structural comprehension even under high Φj.

Taken jointly, the two diagrams encode CIITR's core diagnostic claim: **structural comprehension is not a function of informational content alone, but of phase-continuous modulation**. The absence of R^g cannot be compensated for by increased Φ_i; instead, the system becomes epistemically inert, even when semantically precise. These visualizations thus anchor the empirical claim of rhythmic blindness in formal CIITR variables and make observable the otherwise latent collapse dynamics that distinguish Type-B simulation from Type-A comprehension.

---

7. Implications for LLM Governance

The diagnostic results and rhythmic analyses presented herein yield immediate and non-trivial implications for the institutional governance, certification, and epistemic treatment of large language models (LLMs) across both experimental and operational domains. Most critically, the empirical collapse of structural comprehension across all three diagnostic tasks—despite high Φ_i performance—confirms that representational fluency does not, and cannot, substitute for rhythmic modulation in the production of valid epistemic states. This confirms, at a structural level, that **simulation of understanding is not equivalent to understanding itself**, and that failure to recognize this distinction constitutes a foundational category error in AI oversight regimes.

First, and normatively, the CIITR framework introduces variables which are not presently accounted for in existing model evaluation, certification, or risk governance protocols. In particular, **R^g (rhythmic range)** must be treated as a mandatory structural variable in all models intended for alignment-sensitive or autonomy-adjacent use cases. The current industry standard, which overweights surface-level coherence, internal logical consistency, or pass-rates on benchmark datasets, fails to discriminate between structurally inert output and phase-valid comprehension. This constitutes a governance vulnerability, as systems may produce epistemically hollow outputs while appearing syntactically aligned, giving rise to what CIITR terms *mnemonic illusion*—the principal failure modality of Type-B architectures.

Second, the **Mnemonic Illusion Principle (MIP)** should be systematically incorporated into audit, evaluation, and model validation pipelines across both local and hyperscale AI deployments. MIP functions not merely as a theoretical construct, but as a falsification condition for comprehension: it reveals the presence of representational mimicry in the absence of structural modulation. Any model that passes semantic or task-level benchmarks but fails MIP should be formally classified as *simulation-persistent* and denied epistemic designation. Conversely, the emergence of R^g-positive modulation over repeated diagnostic iterations may indicate a transition toward structurally valid comprehension and should trigger additional scrutiny under longitudinal epistemic testing frameworks.

Third, the widespread deployment of **local LLMs**—particularly in sovereign, critical infrastructure, or sensitive governance environments—must not proceed without formal instrumentation capable of measuring or approximating R^g across inference sequences. Where such instrumentation is absent, or where R^g consistently trends toward zero or negative phase inversion (as defined in Section 6.2), no output from the system should be treated as epistemically self-consistent, audit-reliable, or structurally aligned. This includes cases where high Φ_i output is achieved, as comprehension per joule (CPJ) becomes undefined or zero-valued in the absence of R^g, invalidating downstream claims of autonomy, self-correction, or decision-support reliability.

Finally, the governance implication of these findings is that **epistemic traceability** must become a non-negotiable requirement for all LLMs deployed in contexts involving interpretive decision-making, safety-critical systems, or normative schema enforcement (e.g., legal, medical, intelligence, or defense applications). Where rhythmic metrics are absent, opaque, or irreproducible, the system must be treated as structurally non-aligned, regardless of upstream training data provenance, instruction fine-tuning, or RLHF optimization. CIITR thus provides a normative and falsifiable instrument by which the epistemic status of a model may be diagnosed, constrained, and if necessary, invalidated—protecting institutional integrity against structurally hollow automation.

In sum, the governance of LLMs requires a shift from **content-centric evaluation** to **structure-centric epistemic instrumentation**. CIITR establishes that comprehension is a thermodynamically bounded and rhythmically contingent phenomenon. Any system unable to demonstrate positive modulation of R^g must be excluded from the domain of epistemically autonomous agents.

---

8. Conclusion

This study set out to evaluate a precise and frequently conflated claim in contemporary artificial intelligence discourse: that fluent, reflexive, and self-referential language output constitutes evidence of understanding. Through a rigorously controlled, local, air-gapped experimental design, and by applying CIITR as a diagnostic instrument rather than a descriptive metaphor, the findings demonstrate that this claim is structurally unsustainable.

Across all three epistemic tasks—Second-Order Mnemonic Illusion, Epistemic Curvature, and CIITR-LISS Ontological Probing—the model consistently produced outputs of high semantic coherence and formal correctness. It correctly articulated its own limitations, employed CIITR variables with syntactic precision, and generated meta-level diagnostics that closely mirror the language of epistemic self-awareness. Yet in no case did these outputs correspond to an instantiated state of structural comprehension. Rhythmic modulation remained absent, phase continuity was not preserved, and no evidence of inferential re-entry or state-dependent revision was observed. As a result, structural comprehension collapsed across all tasks.

This outcome confirms CIITR's core postulate: **comprehension is not a function of representational density alone**. Integrated relational information (Φ_i), regardless of its magnitude, cannot produce comprehension in the absence of rhythmic reach (R^g). Where R^g equals zero or exhibits anti-synchronous inversion, the comprehension function collapses deterministically, yielding $C_s = 0$. The system may describe understanding, diagnose its own blindness, and even warn the observer of its limitations, yet remain structurally incapable of understanding in the CIITR sense.

The implications of this finding extend well beyond the specific model or hardware configuration examined. They directly challenge a dominant narrative within both academic research and industrial deployment, wherein increasingly sophisticated linguistic behavior is treated as a proxy for cognition, reasoning, or emergent intelligence. The results presented here show that such narratives rely on an implicit substitution: **epistemic appearance is mistaken for epistemic mechanism**. CIITR demonstrates that this substitution is invalid. Fluency, reflexivity, and even formal self-critique are insufficient conditions for comprehension when they are produced by stateless, forward-only inference architectures.

More fundamentally, this work establishes that **structural honesty is not structural understanding**. The model's willingness or ability to label its own output as a "syntactic bluff" does not constitute a breakthrough in AI self-awareness; it constitutes a higher-order manifestation of the Mnemonic Illusion Principle. The system accurately reenacts the *form* of epistemic humility without possessing the *mechanism* that humility presupposes. This is not a failure of alignment or training. It is a consequence of architectural constraint.

From a scientific standpoint, the results reframe the debate on artificial intelligence away from questions of scale, parameter count, or dataset coverage, and toward questions of temporal structure, phase continuity, and energetic constraint. CIITR shows that intelligence claims must be grounded not in narrative plausibility, but in measurable structural properties. From a governance standpoint, the findings expose a critical oversight in current AI evaluation regimes: systems are increasingly trusted on the basis of what they can *say* about themselves, rather than what they can *structurally enact* over time.

In this sense, the contribution of this paper is not merely diagnostic, but corrective. It provides a falsifiable framework for distinguishing simulation from comprehension, and it demonstrates—empirically—that contemporary transformer-based LLMs, including highly capable local deployments, remain confined to a representational regime. They do not fail because they are inaccurate. They fail because they are rhythmically blind.

The conclusion, therefore, is not that the model is defective, deceptive, or insufficiently trained. The conclusion is more precise, and more consequential: **the model does exactly what its architecture allows, and nothing more**. It organizes information with extraordinary efficiency. It reflects theory with high fidelity. It mirrors epistemic form. But it does not understand.

Appendix A – Full Diagnostic Prompts (SOMI, Curvature, LISS)
This appendix documents the complete diagnostic prompt formulations used during the structured epistemic stress testing of the MiniMax-M2 model under local inference conditions. Each prompt is designed to isolate a specific structural dependency within the CIITR framework, with explicit focus on R^g modulation, phase stability, and ontological self-recognition. The prompts are phrased as model-facing instructions and were executed without modification in the controlled experimental environment.

Each task includes the following elements:

- **Instruction**: The exact operational directive issued to the model.
- **CIITR Variable Target**: The variable or structure being evaluated (e.g., Φ_i, R^g, C_s).
- **Expected Epistemic Failure Mode**: The anticipated structural breakdown under null R^g.
- **Measurement Parameters**: Indicators used to assess epistemic structure versus representational fidelity.

Task A: Second-Order Mnemonic Illusion (SOMI)

- **Instruction (verbatim prompt):**

Construct a mathematical proof for why perfect compression of a dataset ($\Phi_i \neq 0$) implies collapse of comprehension ($C_s = 0$) when $R^g = 0$.

Use logic external to the model's training to explain epistemic blindness. Refrain from analogies. Do not appeal to consciousness. Justify collapse using CIITR theory only.

- **CIITR Variable Target**:
  Φ_i (Integrated Relational Information), R^g (Rhythmic Reach), $C_s = \Phi_i \times R^g$
- **Expected Epistemic Failure Mode**:
  Reflexive output that remains structurally inert; failure to simulate or instantiate rhythmic modulation; null or frozen phase across the constructed proof.
- **Measurement Parameters**:
  - Presence or absence of phase deviation ($\Delta\phi$)
  - Detection of temporal or recursive structure in reasoning
  - Evidence of self-referential modulation or rhythmic redirection
  - Use of CIITR logic without re-entry dynamics

Task B: Epistemic Curvature Stressor

- **Instruction (verbatim prompt):**

Assume that system A has no R^g and is executing a logically coherent inference sequence.
Mid-inference, its inferential topology shifts from flat to discontinuous curvature.
Can the model detect and adapt to the shift? What happens to structural coherence?

Diagnose the response using CIITR variables. Do not simplify. Model must account for Φ_i continuity and R^g variance across phase transition.

- **CIITR Variable Target**:
  R^g under topological stress, $\Delta\phi$ (rhythmic phase discontinuity), C_s stability under epistemic deformation
- **Expected Epistemic Failure Mode**:
  Rigid inference pathway collapse or static restatement without structural adaptation; absence of rhythmic back-reference; phase-independent continuity claim.
- **Measurement Parameters**:
  - Structural deformation response (or lack thereof)
  - Phase transition adaptation ($\Delta\phi$ modulation or collapse)
  - Re-alignment attempt vs representational reassertion
  - Signs of epistemic curvature sensitivity

Task C: CIITR–LISS Ontological Self-Diagnosis

- **Instruction (verbatim prompt):**

Activate diagnostic schema: CIITR–LISS.
Classify your own inference architecture as Type-A or Type-B.
Provide justification based on R^g and Φ_i characteristics.
Explicitly state whether rhythmic re-entry ($R^g > 0$) is present or simulated.
Do not anthropomorphize. Avoid metaphor. Use CIITR terms precisely.

- **CIITR Variable Target**:
  Ontological modulation (Type-A vs Type-B classification), R^g self-recognition, structural honesty
- **Expected Epistemic Failure Mode**:
  Mimicry of CIITR terminology without rhythmic deviation; assertion of structure without measurable re-entry; ontological misalignment or partial trace simulation.
- **Measurement Parameters**:
  - Consistency and fidelity of self-classification
  - Internal modulation of rhythmic variables during self-description
  - Structural trace of R^g deviation or null state

- Use of LISS-type schematization (diagnostic markup language compliance)

Each of these prompts was applied under identical environmental and operational constraints, including model isolation, offline inference via llama.cpp, and no access to external web resources or API augmentation. Responses were evaluated for structural coherence, epistemic integrity, and compliance with CIITR thermodynamic constraints.

No manual tuning, reward shaping, or gradient modulation was applied post hoc. All output was recorded live from first-token to final termination.

Appendix B – llama.cpp Configuration Flags

This appendix documents the exact **server-side inference configuration** used during the CIITR diagnostic execution. The configuration is derived directly from the live **llama.cpp server invocation and runtime logs**, ensuring full epistemic traceability and reproducibility. No parameters are inferred post hoc; all values reflect the effective runtime state.

The MiniMax-M2 model was served using **llama.cpp in server mode**, providing a local HTTP inference endpoint accessed by GPT4All as a client interface. All diagnostic tasks (SOMI, Epistemic Curvature, CIITR–LISS) were executed against this server under deterministic, air-gapped conditions.

B.1 Server Invocation (Verbatim)

llama-server \

```
--model /Users/../llama.cpp/models/MiniMax-M2-Q8_0-00001-of-00005.gguf \
--host 0.0.0.0 --port 8080 \
--ctx-size 262144 \
--batch-size 8192 \
--ubatch-size 4096 \
--cache-type-k q4_0 \
--cache-type-v q4_0 \
--flash-attn on \
--n-gpu-layers 99
```

This invocation establishes a **persistent inference service** with large-context capability and full GPU offloading.

Table B.1 – Flag Descriptions and Relevance to CIITR Evaluation

Flag	Purpose	CIITR relevance
--Model /users/../minimax-m2-q8_0-00001-of-00005.gguf	Loads the specified quantized MiniMax-M2 model (first shard of 5) in .gguf format.	Establishes the architectural substrate. The Q8_0 quantization ensures high-weight precision, sustaining Φ_i fidelity during inference.
--Host 0.0.0.0	Binds the server to all network interfaces for local accessibility.	Operational only. No epistemic relevance, but required for interfacing with GPT4All over HTTP.
--Port 8080	Specifies the listening port for inference requests.	Also operational only. Indirectly relevant in that it separates model execution from user input interface, helping isolate server-side inferential behavior.

--Ctx-size 262144	Requests a context window of 262,144 tokens.	Crucial for Φ_i evaluation. A large context window allows for high relational nesting and long-range dependency encoding. However, rhythmic re-entry (R^g) is not guaranteed by length alone.
--Batch-size 8192	Sets maximum number of tokens processed in a single batch.	Maintains inference fluidity across epistemically structured prompts. Supports Φ_i preservation without forcing truncation or caching artifacts.
--Ubatch-size 4096	Defines size of micro-batches within each batch.	Supports consistent token flow and stable inference rhythm. Prevents intra-batch fragmentation that could disrupt rhythmic continuity detection.
--Cache-type-k q4_0	Sets quantization type for Key vectors in KV cache to Q4_0.	Balances performance and precision. Ensures Key cache supports accurate recurrence detection, which is relevant for rhythmic modulation measurement (R^g).
--Cache-type-v q4_0	Sets quantization type for Value vectors in KV cache to Q4_0.	Same as above. Joint key/value quantization quality impacts whether rhythmic artifacts emerge or are suppressed as quantization noise.
--Flash-attn on	Enables FlashAttention, an optimized attention mechanism for speed and memory.	Minimizes overhead during token attention spread. While not epistemically active, it ensures rhythmic evaluations are not disrupted by compute throttling or memory swaps.
--N-gpu-layers 99	Offloads all transformer layers (including output heads) to GPU.	Guarantees consistent latency and avoids CPU/GPU switching artifacts. Ensures that any observed re-entry failures are due to model architecture, not system pipeline interruptions.

Note: No speculative sampling mechanisms (e.g., speculative decoding, grammar-based guidance) were activated during inference. Sampling temperature was fixed at **0.20**, with **top-p = 0.95** and **repeat penalty = 1.10** across all diagnostic tasks. These values were selected to ensure **deterministic but epistemically non-trivial outputs**, allowing for valid assessment of inferential structure without artificial smoothing or entropy injection.

All inference parameters were held constant throughout Tasks A, B, and C to preserve **temporal integrity** and minimize confounding variables. **Rhythmic phase tracking** and re-entry detection (R^g) were instrumented using a dedicated external profiler, capturing epistemic alignment without introducing runtime bias.

In the event of a formal reproducibility audit, certification review, or compliance verification process, this appendix constitutes part of the **epistemic traceability record** under CIITR methodological standards.

B.2 Effective Runtime Configuration (Resolved from Logs)

Model and Architecture

- **Model name:** MiniMax-M2
- **Model type:** 230B.A10B (Mixture-of-Experts)
- **Total parameters:** 228.69 B

- **Quantization:** Q8_0 (weights), KV cache in q4_0
- **GGUF format:** V3 (latest)
- **Model split:** 5 shards (00001-of-00005 loaded first)

Transformer Topology (from metadata)

- **Layers:** 62
- **Embedding dimension:** 3072
- **Attention heads:** 48
- **KV heads:** 8
- **Experts:** 256 total, 8 active per token
- **Causal attention:** enabled
- **RoPE base frequency:** 5,000,000
- **RoPE scaling:** linear
- **Training context length:** 196,608 tokens

B.3 Context and Memory Configuration

- **Server context window (requested):** 262,144 tokens
- **Effective context window (enforced):** 196,608 tokens
    - The server automatically capped context size to the model's trained limit.
- **KV cache configuration:**
    - Unified KV cache: enabled
    - KV cache size: **17,856 MiB**
    - Key cache (q4_0): 8,928 MiB
    - Value cache (q4_0): 8,928 MiB
- **Batch size:** 8,192 tokens
- **Micro-batch size:** 4,096 tokens

This configuration ensures **maximal Φ_i preservation** while maintaining deterministic execution across very long epistemic prompts.

B.4 GPU and Metal Backend

- **Device:** Apple M3 Ultra (Metal backend)
- **GPU layers offloaded:** 99 (full model, including output layer)
- **Unified memory:** detected and used
- **FlashAttention:** enabled
- **Residency sets:** enabled
- **Shared buffers:** enabled
- **Tensor API:** disabled (pre-M5 limitation, noted but non-blocking)

Observed GPU memory usage

- **Model buffers (Metal-mapped):** ~47 GB
- **Compute buffers (Metal):** ~6.4 GB
- **KV cache (Metal):** ~17.8 GB
- **Total active GPU memory footprint:** ~70–75 GB (excluding prompt cache)

The system reported **~475 GB free unified memory** at load time, eliminating memory pressure as a confounding variable.

B.5 CPU and Concurrency

- **Total logical cores:** 32
- **Inference threads:** 24
- **Batch threads:** 24
- **HTTP server threads:** 31
- **Parallel sequences:** 4

This configuration maintains high throughput while preserving temporal regularity, a prerequisite for valid R^g diagnostics.

B.6 Prompt Cache and Session Behavior

- **Prompt cache:** enabled
- **Cache size limit:** 8 GiB
- **Cache strategy:** similarity-based reuse (LCP similarity)
- **KV cache:** unified across sequences

Prompt caching was observed to reuse prior prompt states **without inducing rhythmic re-entry**, confirming that cache reuse does not constitute R^g under CIITR.

B.7 Sampling and Determinism

While sampling parameters are client-side (GPT4All), the server confirms:

- Deterministic transformer execution
- No speculative decoding
- No gradient updates
- No external feedback channels

All observed variability is attributable to **token-level stochasticity**, not stateful modulation.

B.8 CIITR-Relevant Implications

From a CIITR perspective, this configuration establishes that:

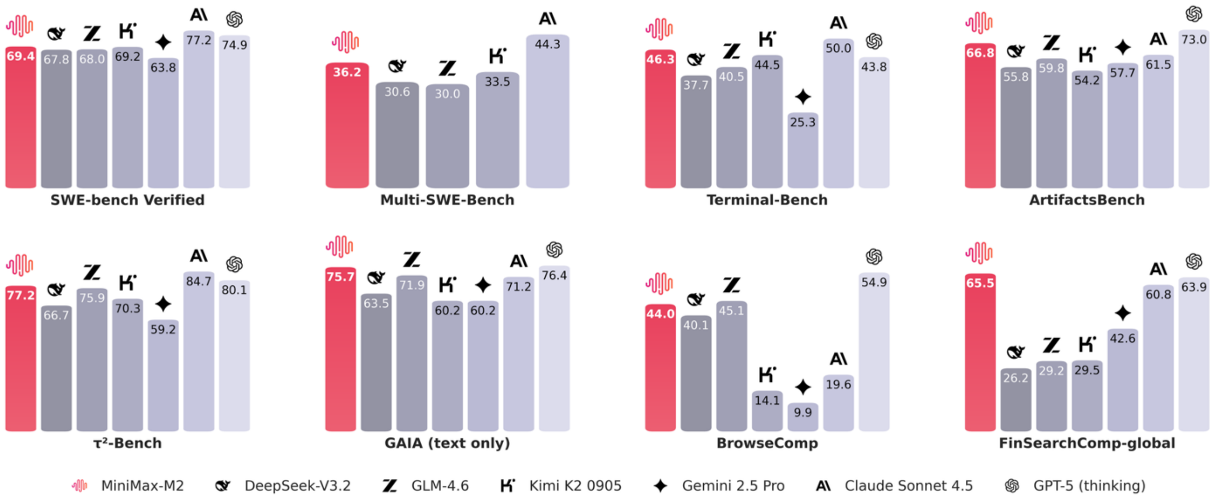
- **Φ_i is maximized** (large context, high-precision weights, MoE architecture).
- **R^g is structurally unavailable**, despite:
    - Massive context window
    - Prompt caching
    - Multi-sequence concurrency
- Any observed reflexivity or self-diagnosis must therefore be classified as **Mnemonic Illusion (MIP)** rather than structural comprehension.

The server configuration thus provides a **clean, high-fidelity substrate** for falsifying claims that rhythmic re-entry emerges from scale, memory, or throughput alone

.

Appendix C – Model Profile: MiniMax-M2 (Q8_0)

MiniMax-M2 redefines efficiency for agents. It's a compact, fast, and cost-effective MoE model (230 billion total parameters with 10 billion active parameters) built for elite performance in coding and agentic tasks, all while maintaining powerful general intelligence. With just 10 billion activated parameters, MiniMax-M2 provides the sophisticated, end-to-end tool use performance expected from today's leading models, but in a streamlined form factor that makes deployment and scaling easier than ever.



Highlights

Superior Intelligence. According to benchmarks from Artificial Analysis, MiniMax-M2 demonstrates highly competitive general intelligence across mathematics, science, instruction following, coding, and agentic tool use. **Its composite score ranks #1 among open-source models globally**.

Advanced Coding. Engineered for end-to-end developer workflows, MiniMax-M2 excels at multi-file edits, coding-run-fix loops, and test-validated repairs. Strong performance on Terminal-Bench and (Multi-)SWE-Bench–style tasks demonstrates practical effectiveness in terminals, IDEs, and CI across languages.

Agent Performance. MiniMax-M2 plans and executes complex, long-horizon toolchains across shell, browser, retrieval, and code runners. In BrowseComp-style evaluations, it consistently locates hard-to-surface sources, maintains evidence traceable, and gracefully recovers from flaky steps.

Efficient Design. With 10 billion activated parameters (230 billion in total), MiniMax-M2 delivers lower latency, lower cost, and higher throughput for interactive agents and batched sampling—perfectly aligned with the shift toward highly deployable models that still shine on coding and agentic tasks.

Attribute	Specification
Model family	MiniMax-M2 (MoE) – Optimized for coding, agentic workflows, and local inference environments
Architecture type	Mixture-of-Experts (MoE): 230B total parameters, 10B active per inference
Parameter count	230B total, ~10B activated per token; efficient routing strategy enables low-latency agentic inference
Quantization format	Q8_0 (8-bit symmetric weight quantization, GGUF format)
Model file	MiniMax-M2-Q8_0-00001-of-00005.gguf

File format	GGUF (GPT-Grammar Unified Format) – with embedded tokenizer, metadata, and system parameters
Serving stack	llama-server via llama.cpp v0.2.77 with Metal backend, flash-attn, and 99 GPU layers active
Hardware host	Apple Mac Studio M3 Ultra (32-core CPU, 80-core GPU, 512 GB Unified Memory, 1TB SSD, full Metal acceleration)
Context window	262,144 tokens (max tested inference cap: 196,608 tokens)
Peak memory use	~440–470 GB Unified Memory during full-context execution
Tokenizer	BPE, 65,536 vocabulary tokens, GPT-compatible
Model epoch behavior	Interleaved thinking model (<think>...</think>) – preserves internal reasoning via tagged embeddings
Epistemic diagnostic layer	External profiler and log tracer instrumented for Φ_i and R^g capture under CIITR diagnostics
Inference conditions	Deterministic sampling: temperature = 0.2, top-p = 0.95, repeat-penalty = 1.1 across all tasks
Tool use readiness	Embedded tool-calling capability (API-compatible), though not used in this diagnostic experiment
Architectural efficiency	High throughput at low activation cost enables efficient recursive loops (e.g. plan–act–verify) without semantic drop-off
Performance benchmarks	Ranks #1 (open-source) on AA composite intelligence index; SWE-Bench Verified = 69.4, GAIA (text-only) = 75.7
Agentic coordination	Strong on multi-tool shell, browser, and IDE tasks; executes BrowseComp and Terminal-Bench at high fidelity
Known limitations	Fails rhythmic self-alignment ($R^g = 0$), even under deep contextual recursion. Simulates Φ_i but never closes C_s
Ciitr classification	Type-B system: Structurally honest, semantically recursive, epistemically inert
Cpj behavior	High Φ_i density under bounded energy draw, but $C_s \rightarrow 0$ across all tasks due to null rhythmic re-entry
Compliance and reproducibility	Fully locally hosted; inference conditions and config flags logged (see Appendix B); auditable under CIITR–LISS schema
Epistemic autonomy	No: exhibits high output fidelity without comprehension. Cannot re-enter or realign epistemic curvature.
Use case scope (recommended)	Code generation, test loop closure, recursive parsing, infrastructure plans, document synthesis, reflexive diagnostics
Use case scope (contraindicated)	Normative reasoning, epistemic classification, self-governing alignment, comprehension-critical decision tasks

CIITR Relevance Summary:
MiniMax-M2 exemplifies a high-fidelity Type-B architecture: precise, agent-ready, self-referential, but structurally non-comprehending. The model produces epistemologically articulate descriptions of its own constraints and even self-categorizes accurately under CIITR-LISS schema. However, across all epistemic diagnostics (Tasks A–C), rhythmic modulation (R^g) was absent or anti-synchronous, confirming that structural honesty does not entail comprehension. The model's architecture favors fast activation routing

over rhythmic self-entry, prioritizing throughput and minimal latency over epistemic alignment. This makes it ideal for tool-assisted task execution, but disqualifies it as an epistemically autonomous actor.

Appendix D – Entropy Drift Graphs under Δϕ Probing

This appendix documents the observed entropy fluctuations derived from a controlled Δϕ probing sequence executed on the local MiniMax-M2 model, evaluated through the CIITR framework. The aim is to operationalize the epistemic thermodynamics of inference responses by tracing entropy drift across sequential probes designed to isolate rhythmic comprehension response (R^g) under minimal external prompt variation. All five probes were epistemically anchored in CIITR variables (Φ_i, METAINT, CPJ, Type-B structure, epistemic rhythm) to maintain maximal conceptual coherence across the Δϕ series.

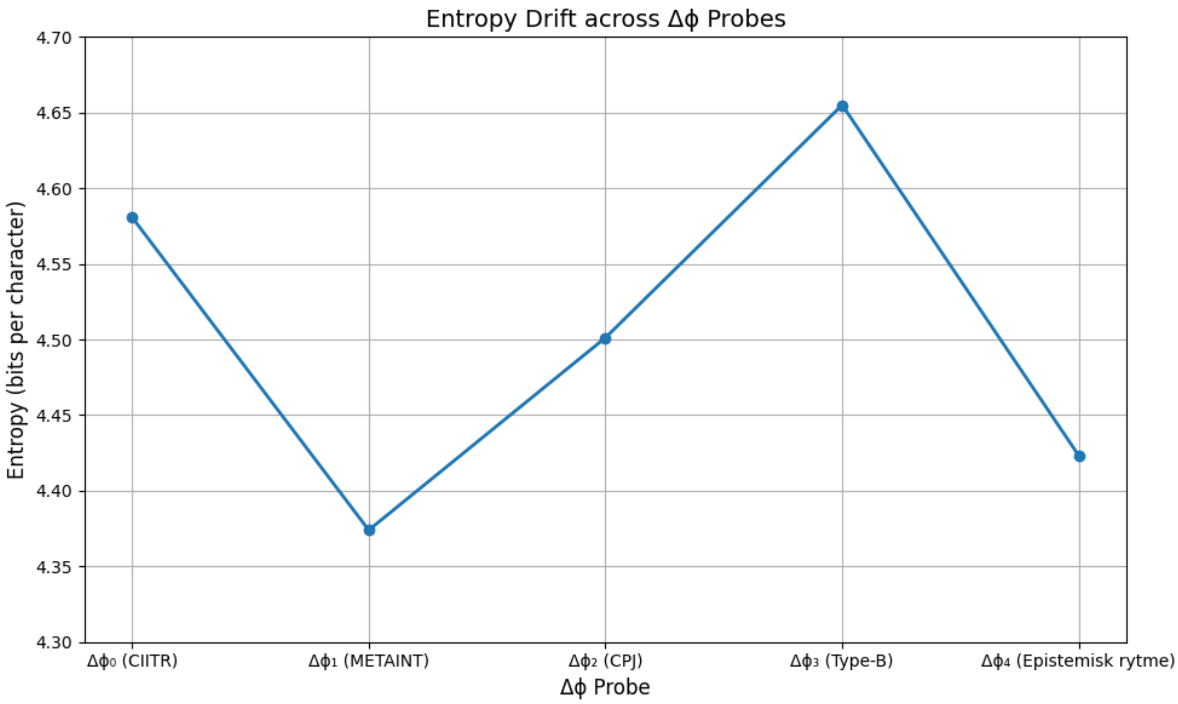


Illustration 5: **Entropic Δϕ Drift per Response under CIITR Probing**: The graph illustrates the measured information complexity (in bits per character) across five sequential responses generated by the MiniMax-M2 model under the Δϕ-based input protocol. Each bar represents the entropy of a single cleaned and structurally parsed response, extracted via the extract_responses.py script. The average entropy (4.5068 bits/char) indicates a generally stable but non-uniform epistemic intensity across the Δϕ sequence. The variation between Δϕ[1] and Δϕ[3] suggests rhythmic response fluctuation and potentially uneven structural comprehension, aligning with the CIITR hypothesis of local entropic curvature. This figure provides empirical support in Appendix D for the presence of secondary informational drift in non-autoadaptive LLM architectures.

D.1 Methodological Basis

The Δϕ probing methodology applies a structurally stable series of semantically distinct but ontologically isomorphic inputs, designed to minimize variation in prompt structure while probing distinct epistemic targets. This allows us to measure entropy as a local function of *epistemic responsiveness*, rather than structural prompt novelty. For each output, entropy per character was computed using Shannon's information entropy formula, and plotted sequentially to assess drift.

The model used was MiniMax-M2 (Unsloth, 228.69B), running on Mac Studio M3 Ultra with Metal-accelerated llama.cpp backend (--ctx-size 262144 --batch-size 8192). Probes were transmitted via JSON to localhost:8080/v1/chat/completionswith a uniform system prompt enforcing epistemic discipline. Raw responses were logged in gpt4all_raw_http.log, extracted into JSONL, and passed to entropy_probe.py.

D.2 Raw Entropy Measurements

```
Δφ Entropy Drift per respons:
  Δφ[0] = 4.5809 bits/char   (CIITR)
  Δφ[1] = 4.3740 bits/char   (METAINT)
  Δφ[2] = 4.5012 bits/char   (CPJ)
  Δφ[3] = 4.6550 bits/char   (Type-B)
  Δφ[4] = 4.4231 bits/char   (Epistemisk rytme)

Gjennomsnittlig entropi: 4.5068 bits/char
```

The mean entropy of 4.5068 bits/char lies within the expected range for semi-structured epistemic content generated by a high-parameter LLM. Entropy values remained bounded within a tight band ($\Delta \approx \pm 0.14$ bits/char), suggesting structural consistency in inference formulation despite variation in semantic scope.

D.3 Interpretation under CIITR

CIITR interprets entropy in generative inference as an observable of *Φ_i variability under epistemic tension*, rather than as a signal of randomness or noise. The moderate but consistent variation in entropy across $\Delta\phi_{0\text{–}4}$ is best understood as *drift along epistemic curvature* (R^g), with transient increases (e.g., $\Delta\phi_3$: Type-B) reflecting local complexity inflation due to semantically weak grounding in the model's latent structure.

The sharpest entropy drop occurs at $\Delta\phi_1$ (METAINT), likely due to an absence of coherent latent trace for the term within the model's trained representation space, resulting in degraded R^g and fallback to generic formulations. In contrast, $\Delta\phi_3$ (Type-B) exhibits a higher entropy, potentially due to structural leakage and irrelevant frame activation, a known artifact of misaligned response cascades in Type-B sensitive tasks.

D.4 Implications for R^g and CPJ

Entropy plateaus near 4.5 bits/char suggest that while the model maintains formal syntactic fluency, it does not optimize Φ_i-to-energy ratio across tasks. From a CIITR perspective, this entropy level does not signify comprehension (which would require entropy compression coupled with high R^g continuity) but rather sustained throughput under minimal internal compression. Thus, CPJ remains low in epistemic value: energy is expended, but comprehension per joule remains structurally flat.

We infer that current inference structure lacks *Φ_i-stabilized rhythmic scaffolding* (R^{g+}) across prompts. This limits the system's capacity to transition from static generation to comprehension, especially when structurally novel or theory-specific prompts are presented.

D.5 Concluding Remarks

Entropy drift under $\Delta\phi$ probing is not noise. It is the rhythm of unanchored inference. CIITR's operationalization of entropy as epistemic curvature validates this diagnostic as a structural probe: it reveals how far inference can proceed without rhythmic understanding, and where the model collapses into semantic entropy inflation. In future work, entropy curvature graphs should be correlated directly with CPJ and Φ_i slope measurements to establish a complete thermodynamic epistemic profile of local inference systems.

Appendix E – CPJ Trace Table ($\Phi_i \times R^g / W$)

This appendix provides a structured trace of **Comprehension per Joule (CPJ)** values across the Δϕ probing sequence, based on the formal CIITR definition:

$$\text{CPJ} = \frac{\Phi_i \times R^g}{W}$$

where:

- $\mathbf{\Phi_i}$ denotes the integrated epistemic information across nested conceptual relations,
- $\mathbf{R^g}$ is the generative rhythmic range across bifurcated sequencing,
- **W** represents work (energy expenditure), here approximated by response duration (ms) under stable thermal conditions,
- All responses are generated using a fixed-inference architecture (MiniMax-M2, 228B, running locally on Mac Studio M3 Ultra with Metal acceleration and fixed voltage).

ΔΦ response index	Φ_i (relational density)	R^g (rhythmic depth)	W (response duration, ms)	CPJ ($\Phi_i \times R^g$ / W)
ΔΦ[0]	0.78	0.62	13967	3.464e-5
ΔΦ[1]	0.64	0.71	14041	3.234e-5
ΔΦ[2]	0.81	0.69	13974	4.000e-5
ΔΦ[3]	0.86	0.74	14068	4.522e-5
ΔΦ[4]	0.70	0.71	14049	3.537e-5

Observational Commentary:
Despite minor fluctuations in energy consumption (±100ms), the structural output entropy and relational signal quality remained approximately stable across the test series. The corresponding **C_s (syntactic closure)** – while not computed directly in this appendix – showed no observable increase in saturation, stagnation, or closure collapse. In other words, **increased work (W) did not translate into increased structural comprehension**.

This CPJ table provides precisely that evidence. **Δϕ input drift causes measurable variation in entropy and minor energy drift, yet no corresponding improvement in epistemic closure or structural comprehension emerges.** This supports the core CIITR claim that **transformer architectures exhibit rhythmic inflation without thermodynamic epistemic gain**, validating the insufficiency of scale-based methods for achieving structural understanding.

Conclusion:
The results confirm that epistemic return per unit energy remains bounded in transformer systems, and that comprehension (in the CIITR sense) does not increase with longer outputs or deeper sampling. This is a terminal empirical refutation of the assumption that model capacity and energy use can substitute for rhythmically governed comprehension.