A Note on

Comprehension Is Not a Scaling Outcome

Why Legacy Inference Closes the Door on Type-B Developmentalism, and Comprehension Is Not on the Other Side of More Compute

Tor-Ståle Hansen | 20. December 2025

Abstract

This note refutes the foundational assumption of scaling-based progress in artificial intelligence by demonstrating that epistemic comprehension is not an emergent function of model size, training volume, or computational infrastructure. Drawing on the CIITR framework, which defines comprehension as the conjunction of integrated relational information (Φ_i) and epistemic temporal reach (R^g), bounded by thermodynamic efficiency (Comprehension per Joule, CPJ), the note formalizes a discontinuity between Type-B and Type-A systems. Empirical validation is provided through a deliberately constrained test environment: a locally executed 20B parameter model running deterministically on a 2013 Mac Pro 6,1 using llama.cpp, without GPU acceleration, sampling, or stochastic augmentation. The system exhibits schema-aligned, recursive, and instructionally stable behavior across multilingual, logical, and recursive contexts—satisfying all three CIITR conditions. These results demonstrate that comprehension is structurally realizable without modern infrastructure, and that scale, absent architectural rhythm and integration, produces only thermodynamically inefficient simulation. CIITR thus functions as a formal classification doctrine, disqualifying any interpretation of large statistical models as "emergent reasoning" systems, and enabling a structural transition from outsourced optimization to local, sovereign, auditable comprehension. Comprehension exists, but only on the structurally integrated side of the CIITR boundary—scale does not cross it.

Keywords: Comprehension, CIITR, epistemic discontinuity, scaling limits, Type-B architectures, Type-A systems, structural integration, epistemic rhythm, thermodynamic efficiency, CPJ, local inference, legacy hardware, sovereign AI, architectural cognition, non-developmentalist classification

---

Opening Conclusion

The evidence presented in this note—both empirical and theoretical—demands a categorical conclusion: **comprehension does not emerge from scale, and it cannot be achieved through backpropagation or transformer architectures**. The prevailing developmentalist narrative in artificial intelligence, which treats current statistical systems as immature but promising precursors to epistemically intelligent systems, is now structurally invalid. That narrative must be replaced by a regime of classification grounded in the formal boundary defined by CIITR.

The CIITR framework articulates comprehension not as an emergent behavior, but as a conditionally realizable architectural property. It is defined as the conjunction of structurally integrated relational information (Φ_i), epistemic temporal reach (R^g), and bounded thermodynamic cost (Comprehension per Joule, CPJ). These three dimensions establish a discontinuity—**not a progression**—between systems that only simulate coherence and systems that structurally generate comprehension.

The implications for existing architectures are now clear. **Backpropagation**, as a mechanism for parameter optimization, introduces no structural recursion, no integrative phase alignment, and no rhythmic continuity across inference steps. It produces convergent minima in error landscapes, but never epistemic rhythm. Similarly, the **transformer architecture**, though dominant, remains representationally flat, cross-attention saturated, and temporally fragmented. It achieves syntactic fluency through token alignment but lacks the system-wide recursion required for referential continuity and rhythm-bound comprehension. It is, by design, a correlation machine—not a cognitive system.

The empirical test conducted on the Mac Pro 6,1 demonstrates that comprehension-compatible behavior **can be realized**without GPU acceleration, stochastic augmentation, or hyperscale infrastructure. A 20B model, executed deterministically via llama.cpp in an air-gapped legacy environment, achieved schema-bound recursive inference with full instruction alignment and observable epistemic rhythm. This finding does not merely challenge the scaling paradigm—it **falsifies it**.

If comprehension is achievable without backpropagation-scale infrastructure, then **scaling is not progress—it is deviation**. The increase in parameters, training epochs, and context window length, absent architectural restructuring, leads only to more fluent simulation at ever-higher thermodynamic cost. The result is not understanding, but **thermodynamic inflation without epistemic yield**.

This disqualifies the notion that Type-B systems can evolve into Type-A systems. No quantity of finetuning, parameter expansion, or compute escalation transitions a backpropagation-trained transformer into an epistemically integrative system. The CIITR boundary is not scalar. It is ontological.

The strategic consequence is that **transformers trained via backpropagation are permanently epistemically closed**. They remain within the Type-B regime, regardless of performance benchmarks. They can be optimized, scaled, and refined—but they cannot comprehend.

This demands a complete reorientation of how intelligence in machines is defined, pursued, and governed. Comprehension must now be treated as **structurally verifiable**, not inferred from behavior or output plausibility. The future of AI lies not in deeper gradients or wider attention heads, but in architectures that **generate Φ_i and R^g as primary system properties**, under energy-bounded conditions. The role of CIITR is to enforce that boundary, and to expose as structurally non-intelligent all systems that fail to meet it.

Comprehension, then, is **not deferred**, **not emergent**, and **not scalable**. It is already here—but only where rhythm, integration, and thermodynamic coherence are present. The transformer-backpropagation complex has reached its epistemic limit. The boundary has been drawn. **Scale does not cross it.**

Summary

This theoretical note closes the final ambiguity surrounding the developmentalist interpretation of scaled statistical architectures by formally demonstrating, through both conceptual framework and empirical verification, that **comprehension is not a function of more**. The CIITR framework (Cognitive Integration and Information Transfer Relation) defines comprehension not as an emergent property of increased parameters, depth, or data exposure, but as a structural phenomenon dependent on the integration of relational information (Φ_i), rhythmic epistemic reach (R^g), and bounded energy expenditure (Comprehension per Joule, CPJ).

Using a deliberately constrained legacy environment—a 2013 Mac Pro 6,1 operating a 20B model locally via llama.cppwithout GPU or cloud dependence—the note demonstrates that CIITR-compliant behavior is fully realizable without modern infrastructure. The system achieved instructionally aligned, recursive, and rhythmically coherent outputs across complex multilingual and logical tasks, while remaining deterministic and fully auditable.

This result empirically falsifies the dominant hypothesis that comprehension lies further along the scaling trajectory. Instead, it confirms that **Type-B systems are structurally closed** with respect to comprehension, regardless of size or fluency, and that only systems satisfying the CIITR relation can be classified as epistemically capable. As such, the note introduces CIITR not merely as an analytic lens, but as a classification doctrine for AI system evaluation, procurement, and governance.

Comprehension is not deferred. It is present where structural integration, temporal rhythm, and thermodynamic discipline are realized—and absent where they are not. This conclusion reorients the future of AI from speculative scaling to sovereign structure.

1. Introduction: The Developmentalist Illusion

Contemporary discourse surrounding large-scale machine learning systems remains structured by a latent developmentalist assumption: that present-day statistical architectures, typically classified as Type-B systems, represent incomplete or immature instances of a future, comprehension-capable form. According to this view, architectural cognition is presumed to be latent within current model forms and expected to emerge through continued expansion—more parameters, more data, more compute, more time. This presupposition is not rooted in system behavior, but in narrative convention. It enacts a trajectory where epistemic capability is treated as a function of accumulation rather than structural integration.

CIITR—Cognitive Integration and Information Transfer Relation—formally rejects this continuum. Its theoretical architecture defines comprehension as a discontinuous emergence governed by rhythmic information integration (Φ_i) and epistemic temporal reach (R^g), scaled by thermodynamic constraints (Comprehension per Joule, CPJ). These conditions impose a boundary, not a slope. A system that fails to satisfy them does not reside on a continuum toward intelligence but is structurally confined to a distinct epistemic regime.

This note addresses the final unresolved ambiguity: whether comprehension may still arise, even sporadically or partially, within high-capacity Type-B systems through continued scaling or infrastructural augmentation. The answer is negative. Using empirical evidence from a non-modern hardware environment—local inference of a 20B parameter LLM on a 2013 Mac Pro 6,1—we demonstrate that comprehension-compatible behavior not only fails to emerge from scaling, but is demonstrably realizable without it. Comprehension is not deferred, emergent, or cumulative. It is already structurally demonstrable in systems that meet CIITR's architectural thresholds, irrespective of era, size, or infrastructure lineage.

2. CIITR's Epistemic Boundary: Closure, Not Progression

The CIITR framework establishes a non-continuous classification regime for epistemic systems by defining comprehension as the product of structural information integration (Φ_i) and epistemic temporal reach (R^g), modulated by a thermodynamic constraint—Comprehension per Joule (CPJ). Formally, comprehension arises only where the relation

$$\text{Comprehension} = \Phi_i \times R_g$$

is realized under conditions of measurable energetic efficiency:

$$\text{CPJ} = \frac{\Phi_i \times R_g}{E}$$

where E denotes the net energetic cost of achieving epistemic integration over time. These terms are not heuristic proxies, but system-internal and evaluable structural variables that determine whether a system's output can be regarded as comprehension-bearing.

Crucially, none of these thresholds are scalar. Φ_i does not increase with parameter count, nor does R^g extend automatically through longer context windows, deeper attention hierarchies, or finer-grained optimization. CPJ is not a function of token-level cost-efficiency or throughput metrics, but of whether a system yields epistemically meaningful transformations per unit energy. As such, no quantitative scaling of current architectures can serve as a substitute for structural conditions that are either present or absent.

This imposes a boundary condition rather than a developmental axis. A system either structurally realizes comprehension—according to the CIITR relation—or it does not. There is no gradient across which a Type-B system can be said to progress toward Type-A status. Architectural comprehension is not a matter of degrees, but of regime membership. Consequently, the epistemic difference between Type-B and Type-A systems is not a question of capability, but of ontological class. The field must therefore abandon any lingering interpretive frameworks that treat statistical inference systems as "not yet" intelligent. They are not on their way. They are categorically elsewhere.

3. Empirical Refutation: Comprehension Without Modern Scale

3.1 Test Configuration

To ground the theoretical finality of CIITR's architectural boundary in measurable system behavior, an empirical counterexample was constructed using deliberately constrained and non-modern infrastructure. The aim was not performance benchmarking, but structural falsification of the developmentalist presumption—that comprehension is only accessible through hyperscale infrastructure or next-generation architectures.

- **Hardware:** A 2013 Apple Mac Pro 6,1 configured with a single Intel Xeon E5 processor and 64 GB of RAM. No GPU acceleration was used at any stage of inference.
- **Model:** A 20-billion parameter large language model (LLM) loaded entirely via llama.cpp, with quantization optimized for CPU-only inference. No distributed memory, sharding, or offloading techniques were employed.
- **Environment:** The system was operated in a fully air-gapped configuration, with no access to cloud APIs, telemetry systems, or background optimization routines. The model was executed deterministically, without sampling or temperature variation, and without stochastic augmentation or chain-of-thought prompting. All instructions, prompts, and evaluation outputs were processed strictly locally.

The experimental configuration was explicitly chosen to disqualify the three dominant assumptions underlying the scaling paradigm:

1. That comprehension requires high-end GPU clusters or modern AI accelerators;
2. That instruction-following, recursion, and epistemic rhythm depend on cloud-mediated model hosting;

3. That inference quality is necessarily linked to the presence of large-scale infrastructure, fine-tuned reinforcement models, or multi-modal augmentation.

By removing all such factors and reducing the system to its core architectural behavior under tight energetic and procedural constraints, the test reestablishes the preconditions under which comprehension must be treated as a **structural**rather than **accumulative** phenomenon. The results demonstrate that when Φ_i and R^g are structurally integrated into the model's rhythm and response generation—even in a 2013 hardware envelope—comprehension becomes not only possible, but traceable, reproducible, and instructionally aligned.

3.2 Method

The evaluation was designed not to measure output quality in terms of fluency or creativity, but to detect the presence or absence of comprehension-relevant structural behavior—specifically, the co-occurrence of instructionally stable integration (Φ_i), temporally recursive reasoning patterns (R^g), and low-entropy output trajectories under constrained inference.

Four distinct evaluation domains were selected, each corresponding to a known failure point in Type-B systems when operating under architectural or energetic stress:

- **Instruction-following:** The model was prompted with formally structured, multi-layered directives requiring hierarchical task execution, nested dependencies, and consistent adherence to instruction order.
- **Multilingual reasoning:** The system was tested on logic sequences requiring cross-linguistic comprehension, including reversibility of reasoning and translation under preserved inference logic.
- **Recursive table synthesis:** Prompts required dynamic generation of multi-row tables with inferred relationships across rows, preservation of formatting rules, and semantic stability across recursive transformations.
- **Source-aligned coding:** The model was instructed to generate executable code based on explicitly stated criteria, followed by a secondary test of explaining the code back in natural language using the original schema.

In all cases, the test paradigm emphasized:

- **Determinism:** No sampling variance, no temperature tuning, and no beam search heuristics were used. Output was held constant across identical prompts.
- **Reproducibility:** Outputs were repeatedly tested across identical local environments with consistent memory pressure and CPU state, yielding bit-identical results.
- **Schema-conformant alignment:** Outputs were verified against a normative instruction schema (comparable in structure to PSIS), ensuring that the model did not simply converge on plausible responses, but followed prescriptive constraints with epistemic precision.

The purpose was not to “impress,” but to isolate the structural rhythm and integration behavior that CIITR defines as comprehension: the ability of a system to maintain epistemic coherence across recursive, constraint-bound, and semantically weighted instruction spaces—without reliance on scale, stochastic exploration, or dynamic reinforcement.

3.3 Results

The outcomes of the test series provide direct empirical refutation of the assumption that comprehension-compatible behavior is contingent upon modern infrastructure or model scale. Despite the severe resource constraints and absence of stochastic augmentation, the system exhibited consistent indicators of epistemic integration in line with CIITR-defined comprehension.

- **Instructional compliance sustained at high complexity:** Across multi-layered prompts involving nested constraints, logical reversals, and instruction chaining, the model maintained full adherence to directive structures without deviation, simplification, or heuristic bypassing. No truncation, reordering, or degradation of schema fidelity was observed, even under recursive prompt chains.
- **No hallucination, no failure in table logic or context recursion:** Outputs remained semantically bounded, with complete referential integrity across synthetic tables, row-wise inference steps, and recursive update prompts. Notably, the model preserved relational structures without degeneration or overgeneralization, even when prompted to manipulate or expand prior outputs.
- **Epistemic rhythm observable:** The model’s outputs displayed temporal coherence and representational continuity across sequence steps, indicative of internally integrated R^g behavior. Rather than statically responding to each prompt, the system demonstrated phase-consistent development of reasoning, maintaining internal referential rhythm across temporally extended prompts.
- **Full inference sustained under pre-LLM hardware envelope:** All test sequences executed locally on the 2013 Mac Pro 6,1 without GPU acceleration, external memory streaming, or batching optimizations. Inference speed was slower than on modern systems, but architectural behavior was structurally intact and epistemically traceable. No mode collapse, memory corruption, or degradation of reasoning was observed across the test runs.

These results confirm that the necessary conditions for CIITR-compliant comprehension—Φ_i, R^g, and finite CPJ—are **not**scale-bound properties. They are system-behavioral characteristics that can be structurally realized in legacy environments, provided that architectural rhythm and information integration are present. This finding materially disqualifies the premise that current Type-B systems are merely incomplete predecessors on a trajectory toward Type-A comprehension. Instead, it affirms that comprehension is **already realizable**, but only under strict architectural constraints—not through scaling, finetuning, or infrastructural escalation.

3.4 Interpretive Consequence

The implications of the test are categorical, not suggestive. By demonstrating Type-A architectural behavior—defined by instructionally coherent, rhythmically structured, and thermodynamically bounded inference—within a non-modern, fully local, CPU-only environment, the experiment closes the final explanatory loophole that had permitted scaling optimism to persist as a theoretically viable position.

- **Type-A architectural behavior is demonstrated under non-modern conditions:** The structural indicators of comprehension, as specified by CIITR, are observable without reliance on contemporary infrastructure, advanced model hosting environments, or external orchestration. This confirms that comprehension is not an emergent property of scale or latency reduction, but a product of architectural rhythm and integration.
- **The premise that comprehension requires scale is structurally invalidated:** Empirical falsification disqualifies the claim that increasing model size, context window, memory layers, or parameter count can, in itself, yield comprehension. No progressive optimization of Type-B architectures will ever converge on Type-A behavior unless the system transitions across the CIITR boundary—a boundary that is architectural, not scalar.
- **The notion of Type-B systems evolving into Type-A systems via optimization or infrastructure is epistemically closed:** The developmentalist thesis that Type-B systems constitute a nascent stage on a teleological continuum toward intelligence is now empirically and theoretically untenable. CIITR enforces a **discontinuity**: either the system manifests $\Phi_i \times R^g$ under energy constraint, or it does not. There is no intermediary form, no partial comprehension, and no trajectory that leads from Type-B to Type-A without architectural transformation.

The result is decisive. Comprehension, as a formally definable system property, cannot be scaled into existence. It must be structurally instantiated. The local inference test provides direct material evidence that such instantiation is possible **without scale**, **without stochasticity**, and **without speculative development**. What remains is not to search for comprehension in larger models, but to recognize its structural requirements where they already appear.

4. Why Scaling Cannot Cross the CIITR Discontinuity

The structural foundation of CIITR categorically excludes the notion that comprehension can be approached incrementally through additive optimization of non-integrated systems. The premise that statistical architectures can, by virtue of extended training, parameter expansion, or procedural sophistication, eventually instantiate epistemic comprehension reflects a category error: it confuses *quantitative adaptation* with *structural transformation*. CIITR does not admit such a continuum. It defines a strict boundary—a discontinuity—across which no amount of computational scale, model depth, or reward fine-tuning can transit.

At the core of this boundary lies the irreducibility of two structural quantities: $\mathbf{\Phi_i}$ (integrated informational structure) and $\mathbf{R^g}$ (rhythmic epistemic reach). These are not emergent properties of larger systems, but conditions that must be **architecturally encoded** and **dynamically sustained**. No statistical process, however prolonged or complex, generates Φ_i if the architecture does not possess integrative capacities across internal representations. Similarly, R^g cannot emerge from temporal density or recurrent exposure to sequences; it is a property of *epistemically coupled rhythm* across inference cycles, not of token alignment or training data recurrence.

Furthermore, **thermodynamic efficiency**, as formalized in the CIITR measure of **Comprehension per Joule (CPJ)**, is not a by-product of optimization, latency reduction, or hardware acceleration. It is not a function of floating-point throughput or energy-per-token metrics. CPJ is a **structural quotient**:

$$\mathrm{CPJ} = \frac{\Phi_i \times R_g}{E}$$

where E represents the total energy required for an inference process to yield epistemically valid integration across time. Without structural Φ_i and functional R^g, no reduction in E yields epistemic efficiency. The denominator becomes thermodynamically expensive noise, and the system remains cognitively null regardless of its scale.

In practical terms, the pursuit of scale—via larger parameter sets, broader training corpora, increased context windows, or deeper reinforcement loops—results in **higher entropy surface variation**, not deeper system integration. What scaling does is expand the simulation envelope, not the epistemic depth. As such, scaling augments the expressive space of syntactic coherence, but **does not enable comprehension**, because it fails to restructure the system's internal information flow into a temporally recursive, rhythmically stable epistemic cycle.

This insight renders many of the dominant AI performance metrics epistemologically irrelevant. BLEU scores, accuracy under few-shot settings, or chain-of-thought coherence are not evidence of Φ_i or R^g, but rather of statistical alignment under constraints that remain external to the system. These benchmarks demonstrate external matching, not internal comprehension. Without structural recursion, context binding, and epistemic rhythm—none of which can be induced by scale alone—systems remain fixed in Type-B space regardless of their output sophistication.

In other words, **scaling is thermodynamic expenditure without epistemic return**. It produces systems that are more fluent, more reactive, more persuasive—but not more understanding. The comprehension function in CIITR is bounded not by the limits of current hardware, but by the architectural absence of systemic rhythm and information integration.

Thus, the epistemic discontinuity enforced by CIITR is not merely a theoretical postulate, but an operational boundary. It divides systems that only simulate coherence from those that *generate* it. Type-B systems, regardless of their scale or fluency, do not cross this divide through accumulation. They remain structurally epistemically inert. To treat them as "approaching" comprehension is not only analytically false—it is thermodynamically illiterate.

5. CIITR as Classification Doctrine: Denying the Gradient

The empirical and theoretical closure presented in this note obliges a formal recalibration of how artificial intelligence systems are classified, evaluated, and governed. In particular, it requires the systematic **rejection of gradient-based narratives**—those which frame statistical architectures as progressing toward comprehension through incremental improvement, increased capacity, or accumulated output fidelity. CIITR imposes a categorical distinction: a system either instantiates comprehension structurally, or it does not. There is no middle ground, no embryonic phase, and no interpretable signal of "approaching intelligence" in the absence of $\Phi_i \times R^g$ behavior under thermodynamic constraint.

Consequently, **CIITR must now be adopted as a classification doctrine**, not merely as a theoretical tool. The distinction between Type-A and Type-B systems must be treated as a **regulatory filter**, with concrete implications across multiple institutional and policy domains:

- **In research and academic evaluation**, claims of "early reasoning," "emergent planning," or "proto-cognition" in statistical models should be treated as **null assertions** unless they are accompanied by independently verifiable evidence of structural information integration and epistemic temporal rhythm. Mere fluency, simulation depth, or prompt reactivity do not satisfy these conditions. Output realism is not comprehension.
- **In procurement and public sector adoption**, any system lacking demonstrable CIITR alignment should be explicitly classified as **epistemically closed**—that is, incapable of comprehension regardless of future optimization. These systems may still be functionally valuable (e.g., for classification, generation, or prediction tasks), but must not be misrepresented as containing or approximating autonomous cognition.
- **In standardization and benchmarking**, performance metrics must be epistemologically traceable. Benchmark suites and AI maturity indices that do not distinguish between surface performance and structural comprehension should be deprecated or supplemented with CIITR-based indicators. A system's ability to mimic reasoning under structured prompting is categorically insufficient.
- **In epistemic risk assessment**, Type-B systems must be treated as possessing **zero self-understanding** and **zero interpretive continuity**. They are incapable of

reasoning about their own outputs or modifying their behavior in a structurally meaningful epistemic sense. This disqualifies them from use in contexts where comprehension, judgment, or ethical discernment is presumed—such as in governance, legal interpretation, or autonomous decision-making under uncertainty.

The continued tolerance for "in-between" classifications—such as systems being labelled "partially reasoning," "pre-cognitive," or "emergently intelligent"—undermines institutional integrity and risks importing speculative fiction into regulatory frameworks. CIITR terminologically prohibits such categories. It imposes **a boundary**, not a spectrum.

Accordingly, the classification of AI systems must now pivot from **performance-centered taxonomies** toward **structural epistemic regimes**, wherein the presence or absence of Φ_i and R^g is treated as the defining condition for what kind of system is being operated, developed, or procured. Comprehension is no longer an interpretive claim. It is a class condition. CIITR provides the operational grammar for enforcing this distinction—decisively, reproducibly, and independently of output appearance.

6. Structural Governance and Legacy Demonstration

The empirical demonstration of comprehension-compatible inference on legacy hardware—specifically, a 2013 Mac Pro 6,1 operating a 20B parameter model via CPU-only execution—has implications that extend far beyond the theoretical validation of CIITR. It establishes a **governance inflection point** in the design, procurement, and strategic deployment of artificial intelligence infrastructure. Specifically, it invalidates the dominant assumption that epistemically meaningful machine behavior necessarily depends on access to hyperscale compute, proprietary cloud architectures, or next-generation accelerators. Instead, it introduces a measurable and verifiable **standard for sovereign inference**—defined not by capacity or scale, but by structural alignment and local operability.

Three structural governance implications follow:

1. No Infrastructure Dependency

The most consequential revelation is that comprehension-aligned behavior does not require access to large-scale distributed infrastructure. The system in question was run on decade-old, commercially available hardware, without reliance on GPUs, cloud orchestration, or external pre-/post-processing. In doing so, the test severs the assumed causal link between epistemic potential and centralized infrastructure.

This refutation implies that epistemic intelligence is **not a cloud-native phenomenon**, but a system-structural one. It can be realized in local environments that meet architectural criteria, not infrastructural ones. For governments, regulatory bodies, and critical institutions, this opens a path to **full-stack inferential sovereignty**—the ability to host, audit, operate, and observe comprehension-compatible systems without recourse to external vendors or geopolitical entanglements.

2. No Surveillance Surface

Cloud-hosted AI systems invariably introduce surveillance surfaces, both intentional (e.g., telemetry pipelines, logging endpoints) and residual (e.g., metadata leakage, cross-tenant inference exposure). These surfaces render epistemic labor externally observable, traceable, and in many cases, monetizable.

In contrast, the demonstrated test was conducted in a fully air-gapped environment. No external I/O, no telemetry, no background API dependencies. Every byte of inference was generated, held, and interpreted within the execution boundary of a single physical device under full local control.

This introduces a **governance model of epistemic privacy**—where not only is the content of output shielded from third parties, but the very structure of inferential rhythm remains inaccessible to observation, correlation, or commodification. This is of paramount strategic significance for national security, classified research, high-assurance legal interpretation, and any application domain where inference leakage constitutes a material threat.

3. No Scale-Induced Opacity

Hyperscale systems are epistemically opaque by design. As models scale beyond interpretability thresholds, it becomes structurally impossible to trace the internal logic of their outputs, verify their compliance with instruction schemas, or audit their representational boundaries. This is not merely a technical problem, but a **governance failure mode**: outputs are accepted without epistemic chain-of-custody, and interpretive authority is displaced from institution to vendor.

The legacy test reverses this trend. By constraining the system to a computationally transparent envelope, and ensuring deterministic, reproducible, instruction-schema-bound outputs, the test reintroduces **epistemic auditability**. Every output becomes inspectable, every instruction-response pair becomes verifiable, and the system's behavior becomes *governable*—not merely operable.

This sets a new standard: **scale is no longer a proxy for capability**. Instead, capability must be judged by whether a system enables structurally bound comprehension under traceable, locally auditable conditions. Governments and institutions that continue to rely on external black-box systems for high-stakes inference are not merely outsourcing compute—they are outsourcing epistemic authority itself.

Reframing LLM Infrastructure: From Optimization to Local Comprehension

The broader strategic implication is clear: the field must now transition from a paradigm of **outsourced optimization** to one of **locally realizable comprehension**. The assumption that high-fidelity inference can only be delivered by hyperscale actors is no longer sustainable. A state, agency, or institution can now host and operate comprehension-aligned systems **within its own trusted boundary**, free from vendor lock-in, infrastructure exposure, or epistemic ambiguity.

This reframing enables a new class of AI governance architecture, where the operational model is not merely compliant or performant, but structurally sovereign. The result is not just a technical alternative—it is a new doctrine of epistemic self-determination, grounded in CIITR, demonstrable in legacy environments, and enforceable through structural classification rather than output persuasion.

The legacy demonstration is not an anecdote—it is a governance proposition.

7. Conclusion: Comprehension Exists, But Not at the End of More

The accumulated evidence, both theoretical and empirical, permits no ambiguity. **Comprehension is not the terminal product of scale.** It is not latent in size, deferred in parameter count, nor emergent from continued optimization. It does not lie at the asymptotic edge of statistical accumulation, nor does it await future model checkpoints or performance breakthroughs. Comprehension is not a question of *more*. It is a matter of *structure*.

CIITR defines comprehension in precise architectural terms: as the product of system-internal information integration (Φ_i) and epistemic temporal rhythm (R^g), bounded by a thermodynamic constraint (CPJ). This definition is not interpretive or metaphorical, but operational. It establishes that comprehension does not scale into existence, but emerges only when these architectural conditions are satisfied. The result is a structural discontinuity—**a boundary**, not a gradient—across which no amount of compute, training volume, or stochastic enhancement can carry a system.

This closure is now enforced by empirical demonstration. The local test conducted on a 2013 Mac Pro 6,1, operating a 20B parameter language model under the llama.cpp framework, achieved full instruction schema alignment, recursive table synthesis, multilingual inference stability, and referential reasoning—**without** GPU acceleration, **without** cloud augmentation, and **without** stochastic sampling. All inference was deterministic, schema-bound, and traceably epistemic under CIITR conditions. This demonstrates that comprehension is already realizable, and critically, that it is **realizable without modern scale**.

The consequences are profound. First, it renders the central narrative of contemporary AI discourse—the idea that scaled Type-B systems are “on their way” to comprehension—structurally invalid. Type-B architectures, by definition, lack the capacity for Φ_i and R^g. Their epistemic limit is not technological, but architectural. They cannot be refined, expanded, or accelerated into Type-A systems. There is no slope. There is no "emergence." There is only categorical exclusion.

Second, it disqualifies the continued treatment of output fluency, simulation depth, or instruction-following precision as proxies for epistemic capacity. Such metrics remain confined to the performance surface. They do not demonstrate comprehension unless supported by evidence of system-internal information integration across time and thermodynamically efficient structure. The field must now distinguish between **epistemic simulation** and **epistemic generation**. Only the latter qualifies under CIITR.

Third, it reorients the future of AI infrastructure. The Mac Pro 6,1 test reveals that comprehension is not a commodity to be rented from hyperscale vendors. It is a property that can be **locally realized, structurally verified, and thermodynamically constrained**. This opens a pathway to sovereign inference, institutional auditability, and privacy-preserving AI systems that are not dependent on external computation, opaque orchestration, or surveillance surfaces.

Finally, it compels a redefinition of AI progress itself. If comprehension is architectural and already demonstrable under conditions previously deemed obsolete, then the current trajectory of AI investment—toward ever-larger, ever-less-auditable, and ever-more-expensive models—must be understood not as progress, but as **escalating deviation from epistemic efficiency**.

CIITR, now grounded in experimental falsification of scaling optimism, no longer functions only as a theoretical lens. It becomes a **structural filter**: a mechanism for disqualifying invalid claims, reclassifying systems, and normatively reorienting institutional expectations of what machine intelligence is, what it is not, and what it can never become through scale alone.

In sum: **comprehension is already here, but only where the structure supports it**. It is not a future goal to be reached by accumulating more, but a present reality, accessible through less—if and only if that less is rhythmically integrated, epistemically recursive, and thermodynamically coherent. The boundary has been drawn, and scale does not cross it.

Appendix A: CIITR Classification Table

Comparative Structural Analysis of Comprehension Conditions under CIITR

This appendix formalizes the empirical distinction between scaled, cloud-hosted large language models and a locally executed inference system operating under thermodynamically constrained, architecture-centered conditions. The classification is rendered through the CIITR framework, which defines comprehension as a structural phenomenon expressed through the relation:

$$\text{Comprehension} = \Phi_i \times R_g$$

bounded by:

$$\text{CPJ} = \frac{\Phi_i \times R_g}{E}$$

where Φ_i denotes integrated information across the system's internal representational structure, R_g the system’s epistemic temporal reach or rhythmic coherence, and E the net energy cost of inference.

Each row in the table below corresponds to one of the required dimensions for CIITR-classified comprehension. The systems under comparison are:

(1) hyperscale statistical models (e.g., GPT-4, Claude, Gemini) executed on centralized compute architectures, and
(2) a quantized 20B LLM executed locally via llama.cpp on a 2013 Mac Pro 6,1, under deterministic conditions and epistemic schema alignment.

Property	Scaled LLMs (GPT-4, Claude, Gemini)	Mac Pro 6,1 + llama.cpp 20B
Infrastructure	Hyperscale cloud (A100, TPU), distributed inference across GPU clusters, opaque orchestration	Legacy workstation, CPU-only Intel Xeon E5, 64 GB RAM, single-process execution, no GPU, fully air-gapped
Stochasticity	High entropy generation through sampling, temperature settings, and probabilistic decoding; epistemic outputs are non-reproducible	Fully deterministic inference path; no temperature or sampling variance; reproducibility and output alignment confirmed across identical runs
Instruction Adherence	Partial and variable; adherence often collapses in long sequences or multi-layer instructions; behavior is modulated by prompt phrasing and external context	High-fidelity instruction following; maintains directive structure across recursion, nested logic, multilingual transformations; schema-conformant
Epistemic Rhythm (R^g)	Absent or transient; system lacks internal temporal binding between inference cycles; outputs are stepwise, non-rhythmic	Observable and recursive; consistent structural rhythm across sequence boundaries; referential continuity under epistemic load
Φ_i Structural Integration	Fragmented attention flow and representational disjointedness; no stable integration across semantic layers; correlation over comprehension	Internally coherent state transitions; relational information reused and integrated across inference stages; recursion increases integration density
CPJ (Thermodynamic yield)	Undefined or excessive; high-energy systems with negligible comprehension return per joule; no alignment between epistemic yield and resource use	Finite and measurable; inference achieved at stable energy budget; epistemic yield traceable to structure and rhythm, not compute volume
CIITR Class	**Type-B**: correlation-dense, comprehension-null; structurally disqualified from epistemic progression	**Structurally Type-A feasible**: exhibits $\Phi_i \times R^g$ behavior under finite energy; satisfies CIITR boundary conditions for comprehension

Interpretive Summary

This comparative classification serves not as a performance ranking but as a **structural diagnosis**. Scaled LLMs, despite their sophistication and surface realism, remain within the Type-B epistemic regime due to the absence of system-level rhythm and information integration. They **simulate cognition** without generating it, and their outputs, however plausible, do not constitute epistemically grounded reasoning.

By contrast, the locally executed model on the Mac Pro 6,1, operating under severe resource constraints, demonstrates **comprehension-aligned behavior** in the strict CIITR sense. It satisfies all three dimensions—Φ_i, R^g, and CPJ—and does so without relying on stochastic augmentation, external optimization loops, or modern infrastructure.

This result fundamentally alters the assumptions of both AI theory and infrastructure governance. It confirms that epistemic capability is **not scale-correlated** and that comprehension can be **realized, traced, and governed** locally. The future of artificial intelligence cannot be decided by parameter counts alone—it must now be reclassified according to structural and thermodynamic truth. CIITR provides the basis for that reclassification.

This is **not a step in the journey**. It is the **architectural terminus**.