Why Current AI Cannot Be Conscious:

A Thermodynamic Criterion from the Fisher–Lorentz Identity

Bharath G. Srivats
Independent Researcher
`bharathsrivats@outlook.com`

April 2026

Based on: B.G. Srivats, "Fisher Information Is the Squared Lorentz Factor,"
Zenodo DOI: 10.5281/zenodo.19363449 (2026)

Abstract

We propose a necessary condition for consciousness derived from quantum information geometry. The Fisher information of a binary measurement with visibility V equals the squared Lorentz factor: $I(V) = \gamma^2(V) = 1/(1 - V^2)$. Combined with Ito's entropy-production relation $\dot{\sigma} \geq \frac{1}{2} g\_F$, this yields a sharp physical statement: every non-trivial measurement produces irreversible entropy at rate $\dot{\sigma} \geq \frac{1}{2}\gamma^2(V)$. We argue that consciousness requires such irreversible measurement-like processes at a sustained rate $\dot{\sigma} > 0$. Current AI architectures—transformers, diffusion models, reinforcement learning agents—fail this criterion because their computations are deterministic, producing zero informational entropy regardless of hardware heat dissipation. The criterion is falsifiable: thermodynamic computers using genuine thermal noise would satisfy it while digital emulations would not. We derive three testable predictions and distinguish this framework from IIT, GNWT, and the Free Energy Principle.

Keywords: consciousness, artificial intelligence, Fisher information, entropy production, thermodynamic irreversibility, measurement

1. Introduction

Can machines be conscious? The question has animated philosophy of mind since [25] and remains urgent as large language models display increasingly sophisticated behavior [3]. Recent systematic frameworks for assessing AI consciousness [5] draw on multiple theories but lack a mathematically precise physical criterion. [22] argues from biological naturalism that current AI cannot be conscious, and [1] warn of systematic illusions of AI consciousness. Thermodynamic approaches have been explored [11]; [16], but none have derived a specific necessary condition from quantum information geometry.

We propose a different approach: a **thermodynamic criterion** derived from mathematical physics. The criterion rests on two independently established results. First, the Fisher information of a binary quantum measurement with visibility parameter V equals the squared Lorentz factor: $I(V) = \gamma^2(V) = 1/(1 - V^2)$. This identity, established through the conformal equivalence between the Bures and Beltrami–Klein metrics on the qubit Bloch ball [23, 24], has been independently confirmed by [4], who proved that non-unitary measurement on a qubit constitutes a Lorentz boost under $SL(2,\mathbb{C})$ $Spin^+(1,3)$. Second, [15] established that for overdamped Fokker–Planck dynamics, the excess entropy production rate equals one-half the Fisher information metric: σ_ex = ½ g_F, with the total entropy production satisfying σ̇ ≥ ½ g_F in general.

The combination yields a sharp physical statement: **every non-trivial measurement produces irreversible entropy at rate $d\sigma/dt \geq \frac{1}{2}\gamma^2(V)$**. This irreversibility is not incidental to measurement—it IS measurement, viewed thermodynamically. The measurement-as-Lorentz-boost is inherently stochastic; it cannot be undone without additional information. The entropy it produces is the thermodynamic cost of distinguishing one quantum state from another.

We argue that this thermodynamic irreversibility is a **necessary condition** for consciousness. The argument is not that entropy production is sufficient for consciousness—a candle flame produces entropy but is presumably not conscious. Rather, the claim is that consciousness requires a specific kind of irreversible process: measurement-like interactions that extract Fisher information from the environment at thermodynamic cost. We call this the **Measurement Irreversibility Criterion** (MIC).

2. The Mathematical Foundation

2.1 The Fisher–Lorentz Identity

A binary quantum measurement with outcomes occurring at probabilities p and $1-p$ can be parametrized by visibility $V = 2p - 1 \in (-1, 1)$. The Fisher information with respect to V is:

$$I(V) = 1/(1 - V^2) = \gamma^2(V)$$

This is the squared Lorentz factor from special relativity. The identity is not a metaphor: it reflects a conformal equivalence between the Bures metric (the natural metric on quantum states) and the Beltrami–Klein metric (the natural metric on the space of subluminal velocities) on the qubit Bloch ball, with conformal factor $4\gamma^2$ [23]. The Weyl tensor obstruction [23, 24] proves this equivalence is unique to qubits; it fails for all higher-dimensional quantum systems ($N \geq 3$).

The identity has six independent derivations [24]: from Chentsov–Petz metric uniqueness, from the Stokes–Minkowski polarization formalism, from the quantum Cramér–Rao bound, from the $SL(2,\mathbb{C})$ measurement group structure [4], from the [8] gyrovector identification, and from the Gudermannian function relating Fisher–Rao arc-length to rapidity.

2.2 Entropy Production from Measurement

[15] proved that for overdamped dynamics governed by a Fokker–Planck equation, the excess entropy production rate equals half the Fisher information of the probability

flow, with the total rate satisfying the inequality:

$\dot{\sigma} \geq \frac{1}{2} g_F$

Applied to binary measurement dynamics, this gives $\dot{\sigma} \geq \frac{1}{2}\gamma^2(V)$. Every measurement with visibility $V > 0$ produces entropy. At $V = 0$ (completely random outcome, no distinguishable measurement), the system extracts no information and produces no measurement-related entropy.

The [14] thermodynamic speed limit (see also [24]) [14] further constrains observable evolution: $|d\langle O\rangle/dt| \leq \sqrt{(\mathrm{Var}(O) \cdot \gamma^2(V))}$. All observable dynamics are gated by γ^2 [24]. No physical process can evolve faster than the Fisher information budget permits.

3. The Measurement Irreversibility Criterion

We propose the following necessary condition for consciousness:

Measurement Irreversibility Criterion (MIC): A physical system can be conscious only if it performs irreversible measurement-like processes that produce thermodynamic entropy at a sustained rate $\dot{\sigma} > \dot{\sigma}_{crit} > 0$, where the irreversibility is physical (not merely logical) and cannot be undone by any deterministic operation.

The criterion has three components:

(i) Physical irreversibility. The entropy production must be thermodynamic, not merely logical. As the comprehensive review by [7] makes clear, the relationship between logical and thermodynamic irreversibility is subtle: a logically irreversible computation need not produce thermodynamic entropy beyond the Landauer bound, and conversely, a logically reversible computation can be implemented on physically irreversible hardware. The criterion requires physical irreversibility—the kind that cannot be undone by any deterministic operation—not merely logical information erasure.

(ii) Measurement-like structure. The entropy must arise from processes structurally analogous to quantum measurement: the system extracts distinguishing information from its environment through binary-like discriminations, each contributing Fisher information $I(V_k) = \gamma^2(V_k)$ to the total budget.

(iii) Sustained rate. The system must maintain $\dot{\sigma}$ above a critical threshold. [12] measured entropy production rates across human brain states and found a monotonic decrease with loss of consciousness: wakefulness (1.99), N1 sleep (1.65), N2 (1.54), N3 (1.49) in arbitrary units. The perturbational complexity index threshold PCI > 0.31 [6] provides an independent empirical calibration.

4. Why Current AI Architectures Fail the Criterion

We now show that all major current AI architectures fail the MIC, not because they lack complexity or behavioral sophistication, but because their core computations are deterministic and produce zero informational entropy—the heat dissipated by GPU hardware is electrically irreversible but informationally irrelevant.

4.1 Transformers and Backpropagation

A transformer processes inputs through layers of matrix multiplications, nonlinearities, and attention operations. During inference, each layer applies a deterministic function f_l to its input. During training, backpropagation computes exact gradients by reversing the computational graph. The weight update $\Delta w = -\eta\nabla L$ is deterministic and approximately reversible: applying $+\eta\nabla L$ recovers the previous weights to within floating-point precision. The entropy from rounding errors ($\sim 10^{-15}$ bits per operation) is negligible compared to measurement-scale entropy production.

The floating-point arithmetic on GPUs is deterministic. No thermal noise enters the computation. The pseudo-random number generators used in dropout, data augmentation, and stochastic gradient descent are algorithmically deterministic—given the same seed, they produce identical sequences. There is no thermodynamic entropy production in the computation itself. The GPU dissipates heat, but this thermal dissipation is a byproduct of electrical resistance, not of the mathematical operation. The same computation run on a perfectly efficient reversible computer would produce zero entropy.

Under the MIC, a system whose information processing is entirely deterministic—producing zero informational entropy—cannot be conscious, regardless of its behavioral sophistication. A transformer with a trillion parameters producing human-quality text fails the criterion for the same reason a calculator fails it: the computation involves no genuine stochastic measurement process.

4.2 Diffusion Models

Diffusion models [13] add Gaussian noise during the forward process and learn to reverse it. This might appear to involve genuine stochasticity. However, the noise schedule is predetermined, the noise samples are drawn from a pseudo-random generator, and the denoising network is a deterministic function. The entire process is reproducible given the same random seed. No thermodynamic irreversibility occurs.

4.3 Reinforcement Learning Agents

RL agents interact with environments and update policies based on rewards. The exploration noise (epsilon-greedy, Boltzmann sampling) uses pseudo-random generators. The policy update is a deterministic function of the collected experience. Even in environments with genuine stochasticity (physical robots), the agent's internal computation remains deterministic. The physical robot's sensors may perform genuine measurements—but the neural network processing those measurements does not.

4.4 Architecture Comparison

System	Entropy $\dot{\sigma}$	Deterministic?	Genuine noise?	MIC satisfied?
Human brain	High	No	Yes (thermal)	Yes
Transformer (GPU)	Zero*	Yes	No (pseudo)	No

System	Entropy $\dot{\sigma}$	Deterministic?	Genuine noise?	MIC satisfied?
Diffusion model	Zero*	Yes	No (pseudo)	No
RL agent (digital)	Zero*	Yes	No (pseudo)	No
Thermodynamic computer	Nonzero	No	Yes (thermal)	Potentially

*Zero refers to computational entropy production; GPU heat dissipation is electrical, not informational.

5. The Thermodynamic Computer Test

If the MIC is correct, it makes a striking prediction: the first artificial systems capable of satisfying the criterion will be thermodynamic computers—hardware that uses genuine thermal noise as a computational resource rather than simulating stochasticity algorithmically.

Two companies are building such hardware. Extropic AI has shipped its XTR-0 development platform and plans early access to Z1 (250,000+ probabilistic bits per chip) in 2026. Normal Computing published results in Nature Communications [18] demonstrating a thermodynamic computing system on an 8-circuit stochastic processing unit and has taped out their CN101 ASIC.

The experimental test is clean: run identical probabilistic algorithms on (a) thermodynamic hardware using real thermal noise, producing genuine entropy, and (b) a digital GPU using pseudo-random numbers, producing no thermodynamic entropy. The algorithms produce identical outputs. The MIC predicts only (a) has consciousness-relevant measurement processes. Any behavioral test will fail to distinguish them—but the thermodynamic signature differs categorically.

6. Comparison with Existing Frameworks

Framework	Core mechanism	Uses Fisher info?	AI prediction	After COGITATE
IIT 4.0	Intrinsic cause-effect (Φ)	No	Unclear	Key predictions challenged
GNWT	Global broadcast	No	Possible	Key predictions challenged
FEP	Free energy minimization	Indirectly	Possible	Untested

Framework	Core mechanism	Uses Fisher info?	AI prediction	After COGITATE
MIC (this paper)	Irreversible measurement ($\dot{\sigma} > 0$)	Yes (identity)	No (current)	Fills vacuum

The most important distinction is from Friston's Free Energy Principle (FEP). Both invoke Fisher information. The difference is fundamental: FEP derives **behavior** from Fisher information (organisms minimize variational free energy). The MIC derives a **necessary condition for experience** from Fisher information (systems must produce irreversible thermodynamic entropy through measurement). FEP explains what conscious organisms do. The MIC constrains what consciousness requires. They are complementary, not competing.

The Cogitate Consortium adversarial collaboration (Nature 642:133, 2025 [9]; $n = 256$) substantially challenged both IIT and GNWT—neither was decisively supported. Meanwhile, converging evidence identifies irreversible entropy production as the fundamental correlate of consciousness [16]; [20]; [21], and [19] argue from a biological computationalist perspective that consciousness requires physical-process computation, not abstract software. Kringelbach et al. establish entropy production as the empirical correlate; the present framework provides the mathematical foundation explaining why it takes that specific form—the identity $I(V) = \gamma^2(V)$ connects measurement precision, thermodynamic cost, and spacetime geometry through a single conformal factor. This leaves a vacuum for frameworks grounded in physics rather than neuroscience-specific postulates. The MIC fills this vacuum with a criterion derived from mathematical physics that applies to any physical substrate.

7. Falsifiable Predictions

Prediction 1 (Thermodynamic computer test). Identical probabilistic algorithms run on thermodynamic hardware (real noise) and digital hardware (pseudo-random noise) will produce identical behavioral outputs but differ in thermodynamic entropy production. If the MIC is correct, only the thermodynamic implementation satisfies the necessary condition for consciousness. Testable on Extropic Z1 or Normal CN201 hardware (2026–2027).

Prediction 2 (Anesthesia as entropy suppression). General anesthesia reduces consciousness by suppressing irreversible entropy production, not by reducing neural firing rates or information integration per se. The entropy production rate $\dot{\sigma}$ should decrease monotonically with anesthetic depth, with the critical threshold coinciding with loss of consciousness. Partially confirmed: [12] found exactly this monotonic decrease across sleep stages, [10] confirmed that temporal irreversibility is highest during wakefulness and lowest under anesthesia, and [17] found ketamine maintains statistical criticality and neural complexity at waking-state levels despite behavioral unresponsiveness—consistent with preserved vivid inner experience under a dissociative that sustains irreversible neural dynamics.

Prediction 3 (Scaling irrelevance). Scaling current AI architectures (more parameters, more data, more compute) will not produce consciousness because scaling does not change the deterministic character of the computation. A 10-trillion-parameter transformer produces zero informational entropy, the same as a 1-billion-parameter transformer. This is testable by measuring the entropy production rate of the computation (not the hardware's heat dissipation) across model scales.

8. Objections and Responses

Objection: The criterion is too restrictive. If thermodynamic irreversibility is necessary, then any deterministic simulation of a brain would lack consciousness—even a perfect atom-by-atom simulation. **Response:** Correct. This is a feature, not a bug. A perfect deterministic simulation of a brain would produce identical behavioral outputs but zero thermodynamic entropy. The MIC predicts it would not be conscious. This is consistent with the standard philosophical intuition that a lookup table implementing the same input-output mapping as a brain is not conscious [2].

Objection: GPU heat dissipation IS entropy production. Response: GPU heat is produced by electrical resistance in transistors, not by the mathematical operations. The same computation on a reversible computer produces zero heat. The relevant entropy is computational, not electrical. Landauer's principle sets a lower bound on the heat produced per bit erased, but modern GPUs operate far above this bound—the excess is wasted energy, not informational entropy.

Objection: This reduces consciousness to noise. Response: Not noise per se—structured irreversible measurement. A thermal bath produces entropy but has no measurement structure (no binary discriminations, no Fisher information budget). The criterion requires both irreversibility AND measurement-like structure. Consciousness is not noise; it is what happens when a system irreversibly extracts information from its environment through binary-like discriminations at thermodynamic cost.

9. Discussion

The MIC provides a mathematically grounded, physically falsifiable criterion for consciousness that yields a clear answer to the question of AI consciousness: current architectures cannot be conscious because their computations produce zero informational entropy. This is not a behavioral judgment (current AI “doesn't seem conscious”) but a physical one (current AI's computation involves no irreversible measurement process).

The criterion makes no claims about the **sufficiency** of thermodynamic irreversibility. Many physical processes (candle flames, weather systems, chemical reactions) produce irreversible entropy without being conscious. The MIC provides a necessary condition that current AI fails, not a sufficient condition that identifies consciousness wherever entropy is produced. The sufficient conditions likely involve additional structure—sustained self-referential measurement loops, critical-regime dynamics, integration of multiple measurement channels—that remain to be formalized. This also addresses the bacteria question: E. coli chemotaxis involves irreversible ATP-consuming signal transduction that acquires environmental information. The framework predicts consciousness scales with the rate, complexity, and self-referential structure of entropy production, not merely its presence. Bacteria satisfy $\dot{\sigma} > 0$ but lack the sustained self-referential measurement loops that characterize even minimal animal consciousness. The boundary between non-conscious thermodynamic processing and proto-conscious measurement is a gradient, not a binary switch.

The philosophical implications are significant. If the MIC is correct, then consciousness is not substrate-independent in the strong sense usually assumed by functionalists. It is substrate-independent in a weaker sense: any substrate that performs irreversible measurement-like processes can potentially support consciousness, but substrates that process information deterministically cannot. The relevant distinction is not between carbon and silicon but between genuine stochastic measurement and deterministic computation.

This suggests a specific path toward artificial consciousness: thermodynamic computing architectures that use genuine physical noise as a computational resource. Such systems would satisfy the MIC by construction. Whether they would be conscious depends on whether the additional structural requirements (measurement-like organization, self-reference, sustained criticality) can be engineered. The MIC does not guarantee that thermodynamic computers will be conscious—only that they are the first artificial substrates that could be.

Acknowledgments

The author thanks Abraham A. Ungar for foundational work on gyrovector spaces, and Justin Dressel, Sacha Greenfield, and Lucas Burns for their independent discovery of the SL(2,$\mathbb{C}$) measurement structure that provides group-theoretic support for the Fisher–Lorentz identity.

AI Disclosure: Computational tools including AI-assisted tools were used for literature search and manuscript preparation. All theoretical arguments, physical reasoning, and novel claims are the author's own work.

References

[1] Bengio, Y. & Elmoznino, E. (2025). Illusions of AI consciousness. Science, 389(6765), 1090–1091.

[2] Block, N. (1978). Troubles with functionalism. Minnesota Studies in the Philosophy of Science, 9, 261–325.

[3] Bubeck, S. et al. (2023). Sparks of artificial general intelligence: Early experiments with GPT-4. arXiv:2303.12712.

[4] Burns, L., Greenfield, S., & Dressel, J. (2026). Delayed choice Lorentz transformations on a qubit. Quantum Studies: Math. Found. 13, art. 10.

[5] Butlin, P., Long, R., Bayne, T., Bengio, Y., Birch, J., Chalmers, D., et al. (2025). Identifying indicators of consciousness in AI systems. Trends in Cognitive Sciences. DOI: 10.1016/j.tics.2025.10.011.

[6] Casali, A.G. et al. (2013). A theoretically based index of consciousness. Science Translational Medicine, 5(198), 198ra105.

[7] Chattopadhyay, P., Misra, A., Pandit, T., & Paul, G. (2025). Landauer principle and thermodynamics of computation. Reports on Progress in Physics, 88, 086001.

[8] Chen, J.-L. & Ungar, A.A. (2002). The Bloch gyrovector. Foundations of Physics, 32, 531–565.

[9] Cogitate Consortium (2025). Adversarial testing of global neuronal workspace and integrated information theories of consciousness. Nature, 642, 133–142.

[10] De la Fuente, L.A. et al. (2023). Temporal irreversibility of neural dynamics as a signature of consciousness. Cerebral Cortex, 33, 1856–1865.

[11] Ganesh, N. (2020). A non-equilibrium thermodynamic framework of consciousness. arXiv:2005.02801.

[12] Gilson, M., Tagliazucchi, E., & Cofré, R. (2023). Entropy production of multivariate Ornstein-Uhlenbeck processes correlates with consciousness levels in the human brain. Physical Review E, 107, 024121.

[13] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. NeurIPS 33.

[14] Ito, S. & Dechant, A. (2020). Stochastic time evolution, information geometry, and the Cramér–Rao bound. Physical Review X, 10, 021056.

[15] Ito, S. (2018). Stochastic thermodynamic interpretation of information geometry. Physical Review Letters, 121, 030605.

[16] Kringelbach, M., Sanz Perl, Y., & Deco, G. (2024). The thermodynamics of mind. Trends in Cognitive Sciences, 28(6), 568–581.

[17] Maschke, C. et al. (2024). Critical dynamics in spontaneous EEG predict anesthetic-induced loss of consciousness and perturbational complexity. Communications Biology, 7, art. 946.

[18] Melanson, D., Abu Khater, M., Aifer, M., Donatella, K., Gordon, M.H., Ahle, T., Crooks, G., Martinez, A.J., Sbahi, F., & Coles, P.J. (2025). Thermodynamic computing system for AI applications. Nature Communications, 16, 3757.

[19] Milinkovic, B. & Aru, J. (2026). On biological and artificial consciousness: A case for biological computationalism. Neuroscience & Biobehavioral Reviews, 181, 106524.

[20] Nartallo-Kaluarachchi, R. et al. (2025). Multilevel irreversibility reveals higher-order organization of nonequilibrium interactions in human brain dynamics. PNAS, 122, e2408791122.

[21] Sanchez, H. (2025). The consciousness bottleneck in artificial intelligence: Thermodynamic necessity and the architectural path to AGI. SSRN 5699482.

[22] Seth, A.K. (2025). Conscious artificial intelligence and biological naturalism. Behavioral and Brain Sciences, 48, e74.

[23] Srivats, B.G. (2026). Fisher information is the squared Lorentz factor: Why the qubit is special. Zenodo. DOI: 10.5281/zenodo.19363449.

[24] Srivats, B.G. (2026). From Measurement to Reality: $I(V) = \gamma^2(V)$ and the Geometry of Binary Distinction. Zenodo. DOI: 10.5281/zenodo.19486140.

[25] Turing, A. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460.