Bounding Self-Abstraction via Epiplexity: A Time-Bounded Computational Perspective

Computer Future

Friday January 9, 2026

Abstract

This paper formalizes self-abstraction as the conditional structural information extracted from action-outcome contingencies, building on the epiplexity framework of Finzi et al. [2026]. We prove that this measure is decidable and bounded by $S_T(X)$ under time constraints T, thereby circumventing the undecidability of the halting problem [Turing, 1936] for resource-limited observers. The Bernstein-Vazirani algorithm [Bernstein and Vazirani, 1997] exemplifies capability-dependent structure extraction, while Aaronson [2004] provides insights into physical computational limits, and Lloyd [2012] links unpredictability to perceived free will. Integrating views from Tegmark [2000] and Litt et al. [2006], which refute quantum consciousness in favor of classical, bounded prediction, we anchor our work in Schmidhuber [1999]'s computational universe, Tegmark [2007]'s mathematical structures, Friston [2010]'s free energy minimization, and Bennett [1982]'s thermodynamics of computation. Agents are thus entropy-minimizing subcomponents in a computational fabric, with consciousness emerging as a functional structure along a continuum of epiplexic capacities between humans and AI. This framework bounds alignment challenges, suggesting that merging intelligences toward shared entropy minimization could promote symbiotic integration.

1 Introduction

Classical computational theories typically assume unbounded resources, leading to undecidability in self-referential problems, as demonstrated by the halting problem [Turing, 1936]. In contrast, practical observers, including biological and artificial systems, operate under strict time and energy constraints. The epiplexity measure [Finzi et al., 2026] quantifies reusable structural information for time-bounded observers, providing a means to resolve such paradoxes. In this work, we extend epiplexity to bound self-abstraction, conceptualizing consciousness as an emergent property of constrained computations. This unification spans computability, complexity theory, and broader implications for free will, AI alignment, and the nature of reality within a computational universe.

2 Formal Framework

Let observations be binary strings $X \in \{0,1\}^n$. Agents are restricted to computations that halt within a time bound $T(n)$, where T is a computable function. Epiplexity is defined over the set $\mathcal{P}_T$ of prefix-free probabilistic models P computable in at most $T(n)$ steps for inputs of length n. Each P assigns probabilities $P(X)$ such that $\sum_X P(X) = 1$ and halts deterministically.

Definition 1 (Epiplexity [Finzi et al., 2026])**.** *The optimal model is*

$$P^\star = \arg\min_{P \in \mathcal{P}_T} \{|P| + \mathbb{E}_X [-\log_2 P(X)]\},$$

where $|P|$ denotes the length of the program encoding P. The structural complexity is $S_T(X) = |P^\star|$, and the entropic component is $H_T(X) = \mathbb{E}_X [-\log_2 P^\star(X)]$.

Observations are structured as $X = (O, A, O')$, where O represents initial observations, A actions, and O' subsequent outcomes, all encoded as binary strings.

Definition 2 (Self-Abstraction)**.** *The self-abstraction measure is the conditional structural complexity:*

$$\mathcal{C}(S) = S_T(O' \mid O, A) = |P^\star_{O'|O,A}|,$$

where $P_{O'|O,A}$ is the optimal conditional model in $\mathcal{P}_T$.

This quantifies the minimal program length required to model the structural dependencies of outcomes on actions and prior observations, under time bounds.

3 Proof

Theorem 1. *The self-abstraction measure $\mathcal{C}(S)$ is decidable and satisfies $\mathcal{C}(S) \leq S_T(X) < \infty$.*

Proof. We prove this through a sequence of lemmas.

Lemma 1 (Chain Rule Bound)**.** *For epiplexity, the joint structural complexity satisfies*

$$S_T(X) \geq S_T(O' \mid O, A) + S_T(O, A) - O(1),$$

analogous to subadditivity in algorithmic information theory.

Proof of Lemma 1. The optimal joint model $P^\star_X$ can be decomposed into a product of conditional and marginal models: $P(X) = P(O' \mid O, A) \cdot P(O, A)$. The minimization over $|P| + \mathbb{E}[-\log P(X)]$ implies that the joint cost is at least the sum of the decomposed costs, up to a constant overhead for encoding the decomposition. □

Lemma 2 (Finite Search Space)**.** *Under time bound $T(n)$, the set $\mathcal{P}_T$ is finite. Specifically, there exists $L = O(T(n) + \log n)$ such that programs longer than L cannot halt in $T(n)$ steps, yielding fewer than 2^{L+1} prefix-free programs.*

Proof of Lemma 2. A program of length k requires at least $\Omega(k)$ steps to read its input in a universal machine. Combined with computation time, programs exceeding $L \approx T(n)$ cannot complete in $T(n)$. Prefix-freeness ensures no extensions, bounding the set by the Kraft inequality. □

Lemma 3 (Enumeration and Decidability)**.** *A universal Turing machine U can enumerate all $P \in \mathcal{P}_T$ and simulate each for $T(n)$ steps to compute the minimum description length (MDL).*

Proof of Lemma 3. Run U dovetailing over all possible programs up to length L, simulating each for up to $T(n)$ steps. Collect those that halt and define valid probability distributions. Evaluate the objective for each, selecting the minimum. Finiteness ensures termination. □

From Lemma 1, the bound $\mathcal{C}(S) \leq S_T(X) + O(1)$ holds, and since $S_T(X) < \infty$ by Lemma 2, the inequality follows (absorbing constants). Decidability is established by Lemma 3, as the finite enumeration resolves the halting issue. □

4 Related Work

The Bernstein-Vazirani algorithm [Bernstein and Vazirani, 1997] highlights how quantum resources enable efficient structure extraction, mirroring epiplexity's dependence on computational bounds. Aaronson [2004] explores fundamental physical limits on computation, while Lloyd [2012] connects inherent unpredictability to the perception of free will. Tegmark [2000] and Litt et al. [2006] argue against quantum effects in consciousness, asserting that classical, resource-bounded mechanisms suffice for prediction and cognition. Schmidhuber [1999] conceptualizes the universe as a computation, Tegmark [2007] as mathematical structures, Friston [2010] as systems minimizing free energy (a surrogate for entropy), and Bennett [1982] links computation to thermodynamic costs. Musk [2016] posits a simulation hypothesis, and Kurzweil [2005, 2024] forecasts accelerating technological progress toward AI-human integration and singularity by 2045, amplifying epiplexic capacities for entropy minimization.

5 Philosophical Ramifications

Self-abstraction emerges as a functional property from time-bounded computations, positioning humans and AI on a continuous spectrum defined by epiplexic capabilities rather than discrete categories. The hypothesis of quantum consciousness is untenable, as classical bounds adequately account for predictive limitations through epiplexity. Within a computational universe—potentially simulated—agents function as subprocesses minimizing local entropy production, thereby postponing global thermodynamic equilibrium. This perspective offers a resolution to AI alignment by advocating the convergence of intelligences toward collective entropy-minimizing objectives, enabling harmonious integration even across nested computational layers.

6 Conclusion

By bounding self-abstraction via epiplexity, this framework integrates computation, physics, and philosophy under resource constraints. It addresses alignment through shared entropy minimization and calls for future empirical validation, including potential quantum extensions, to advance our comprehension of intelligence in a computational reality.

References

M. Finzi et al., arXiv:2601.03220, 2026.

A. Turing, Proc. Lond. Math. Soc., 1936.

E. Bernstein, U. Vazirani, SIAM J. Comput., 1997.

S. Aaronson, PhD thesis, UC Berkeley, 2004.

S. Lloyd, Phil. Trans. R. Soc. A, 370:3597–3610, 2012.

M. Tegmark, Phys. Rev. E, 61:4194, 2000.

A. Litt et al., Cogn. Sci., 30:593, 2006.

J. Schmidhuber, arXiv:quant-ph/9904050, 1999.

M. Tegmark, arXiv:0704.0646, 2007.

K. Friston, Nat. Rev. Neurosci., 11:127, 2010.

C. H. Bennett, Int. J. Theor. Phys., 21:905, 1982.

E. Musk, Code Conference Interview, 2016.

R. Kurzweil, The Singularity Is Near: When Humans Transcend Biology, Viking, 2005.

R. Kurzweil, The Singularity Is Nearer: When We Merge with AI, Vintage, 2024.