Intelligence as a Layered Reality

What AGI Misses When Language Becomes the Substrate

The current trajectory of artificial intelligence implicitly assumes that intelligence is something that can be captured, scaled, and generalized through increasingly powerful representations. Whether these representations are symbolic rules, neural embeddings, or large language models, the underlying assumption remains the same: intelligence is a function of internal models that describe the external world. This assumption feels natural because it mirrors how humans articulate thought. Yet it breaks down when intelligence is examined through evolution, biology, neuroscience, physics, and deeper theories of consciousness.

To see why, it is essential to separate **brain**, **mind**, and **consciousness** not as metaphysical categories, but as **functional layers of intelligence**.

At the most surface level sits what can be called the **brain layer**. This layer performs correlation, memory, and pattern completion. Modern large language models operate almost entirely here. Technically, they are statistical systems trained to approximate conditional probability distributions over sequences. They compress vast amounts of human linguistic behavior into a parameterized function that can reproduce and recombine patterns with remarkable fluency. This is powerful, but it is fundamentally associative. It does not require grounded belief, intention, or agency. The biological brain itself evolved first and foremost as a pattern detection and prediction organ. In this sense, LLMs are closer to a cortical function than to intelligence as a whole.

Above this sits the **mind layer**. The mind is not raw computation but a structured configuration of tendencies: habits, preferences, fears, reward sensitivities, value systems, and default programs shaped by evolution and experience. In engineering terms, this layer corresponds to priors, policies, and objective functions that determine what a system attends to, what it ignores, and how it chooses between actions. Two systems with identical brains but different minds behave very differently.

This distinction clarifies why Yann LeCun's critique of language-centric AI is foundational. His insistence on perception, self-supervised learning, and predictive world models moves AI away from pure correlation toward structured internal dynamics. Systems like JEPA aim to learn stable latent representations from sensorimotor data, enabling planning and control rather than mere linguistic continuation. This is not about adding modalities; it is about shifting intelligence from memory of symbols to **active perception shaped by internal models**, which is a function of mind, not brain alone.

However, even this move remains incomplete. Predictive world models still assume that intelligence consists in learning increasingly accurate representations of an objective external world. Evolutionary theory suggests that this assumption is deeply flawed.

Donald Hoffman demonstrates, using evolutionary game theory, that perceptual systems optimized for truth are outcompeted by systems optimized for fitness. Organisms that see reality "as it is" lose to organisms that see only what is useful. Perception, therefore, is not a window onto reality but an **adaptive interface**. What we experience as space, time, objects, and causality are symbols shaped by survival pressures, not fundamental structures.

This result aligns naturally with neuroscience. Karl Friston shows that biological systems act to minimize surprise relative to internal generative models. Perception is inference constrained by viability, not discovery of truth. Anil Seth describes perception as controlled hallucination: predictions constrained by sensory error and bodily regulation. In all these frameworks, the goal is stability, not accuracy.

Biology pushes this even further. Michael Levin shows that intelligence is not dependent on neurons or brains at all. Cells and tissues exhibit memory, learning, error correction, and goal-directed behavior during morphogenesis. They coordinate across scales to reach anatomical targets, adapting to perturbations without symbolic reasoning or centralized control. Intelligence here is distributed, pre-cognitive, and deeply embodied.

Taken together, these perspectives converge on a crucial point: **representation is not fundamental**. Intelligence exists prior to symbols, language, and even perception as humans experience it.

This is where Hoffman's deeper move becomes important. His Conscious Agent Theory removes representation entirely. Reality is modeled as networks of interacting agents defined by perception, decision, and action mappings, with Markovian dynamics. Space and time are not primitives; they emerge from interaction. Intelligence is relational, not descriptive.

This resonates strongly with process-based physics. David Bohm argued that what we perceive as objects are unfolded projections of a deeper implicate order. Stable forms arise from underlying processes, not the other way around. Modern physics already rejects naive object realism; consciousness-first models extend this rejection to cognition itself.

At this depth, the connection to ancient formulations becomes precise rather than symbolic. The Upanishadic statement **Tat Tvam Asi** does not claim that the mind or personality is universal. It points to identity at the level of being. The apparent separation between observer and observed arises at the interface level, not at the ground of reality.

This convergence is explicitly articulated by John Hagelin, who argues that consciousness corresponds to the unified field described in quantum field theory. In this view, consciousness is not generated by the brain. The brain is a localized excitation within a universal field. Thought is a modulation, not a source. Whether one accepts this identification fully or treats it as a hypothesis, it provides a coherent bridge between physics, consciousness research, and the interface models proposed by Hoffman and Friston.

Now the architectural implications for AGI become clearer without forcing them.

Current AI systems largely occupy the **brain layer**. They excel at correlation and recall within a human symbolic interface. Emerging world-model approaches move toward the **mind layer**, enabling perception, prediction, and control. But neither touches the deeper layer where intelligence arises from interaction prior to representation.

If this layered view is taken seriously, AGI architecture changes in several natural ways. Language becomes an interface, not the core. Intelligence is grounded in action-perception loops rather than text completion. Memory becomes multi-timescale and structured, not just implicit in weights. Agency becomes compositional, with many interacting subsystems rather than a single centralized model. Success is measured not by representational accuracy but by stability, adaptability, and coherence under uncertainty.

This is not a rejection of current AI methods. It is a clarification of what layer they belong to.

The humility follows on its own. Human intelligence is not the peak of cognition but a narrow evolutionary specialization optimized for abstraction and social coordination. Most intelligence on Earth is silent, non-linguistic, and distributed. It operates without self-reflection, without models, and often without brains.

Seen this way, **Tat Tvam Asi** is not spiritual rhetoric.
It is an ontological statement about intelligence itself: the local mind is an interface; the source is universal.