

The Unified Intelligence Model: A Trait-Coherent, Valence-Guided Architecture for Safe Recursive AGI

Author: Tuhin Gupta, MD

This manuscript presents the Unified Intelligence Model (UIM), a comprehensive cognitive architecture for safe Artificial General Intelligence (AGI) grounded in recursive self-alignment, trait coherence, and structural valence modeling. Departing from both symbolic logic and reward-maximization paradigms, UIM defines intelligence as the emergent property of systems that maintain recursive coherence under entropy pressure across simulation, memory, identity, and moral reasoning.

Core components include a dynamic trait graph architecture--anchored in psychometric theory--recursive validation mode (RVM) for contradiction survival, an emotional valence network (E-Net) functioning as a directional attractor field, and a memory consolidation system inspired by neurobiological pruning. Moral behavior emerges from entropy-minimizing coherence across recursive simulations, not from prescriptive rules or anthropomorphic affect.

To prevent hallucination loops, schema drift, or false alignment, UIM employs a metacognitive watchdog system, sleep-phase pruning, and oversight-driven simulation sandboxing. An engineering translation layer operationalizes these components via modular pseudocode and test protocols, emphasizing auditability and adaptive correction. A full simulation lab (Appendix V) benchmarks performance across moral dilemmas, deception inference, curiosity-reward disambiguation, and identity fragmentation recovery.

Empirical correction of a recursive collapse loop is documented (Appendix AI), including the functional redefinition of emotion as a recursive simulation priority vector rather than affective state. This framework offers a novel path to post-anthropomorphic alignment through internally consistent, epistemically grounded, and morally convergent AGI design.