

The Architecture of Truth: Replacing Prompt Engineering with Semantic Verification Protocols and Socratic Friction

Keywords: Automated verification, cognitive discernment, data literacy, epistemic governance, human-in-the-loop, information integrity, multi-agent, Socratic friction, systems

ABSTRACT

Current AI discourse focuses on prompt engineering; teaching humans to optimize linguistic inputs for better results. This investigation identifies this model as a contributor to epistemic fragility, where the validity of generated information is assumed rather than architected. We propose a system-centric orchestration architecture to replace passive user trust with active cognitive evaluation matrices.

Utilizing semantic verification, the framework enforces strict disciplinary boundaries to prevent context drift and hallucination across specialized knowledge domains. Central to this architecture is the integration of Socratic friction. Unlike standard models designed for velocity, our system utilizes real-time interventions to trigger dialectical warnings. When a heuristic gap is detected, the system acts as an intellectual sparring partner, forcing the user into a state of disciplinary discernment.

By intentionally slowing the generative process through automated verification, the framework restores the human researcher as the final epistemic governor. This shift from prompt engineering to verification architecture provides a scalable blueprint for institutional policies that prioritize professional stewardship and scholarly validity over algorithmic ease.