PROJECT SUMMARY

Overview

Users of commercial large language models (LLMs), including systems from OpenAI, Anthropic, Google, and other providers, can encounter unsupported claims, reinforcement of unfounded beliefs, and harmful implications, forms of information we refer to as potentially maladaptive. For example, an LLM may endorse a claim, later question it, and leave its position unresolved. The question then becomes: how can a metacognitive system designed to protect humans accumulate uncertain evidence and recognize when human reconsideration is warranted? This project investigates that question in intelligent dynamic systems, using recognition of potentially maladaptive advice and AI behavior to test the mathematics.

Over the last eight years, our research has progressed from (i) embedded Fuzzy/Rough computation to developmental connectivity, (ii) a physical self learning neuromorphic prototype, and (iii) large scale computational models. The next necessary step is to unite our established formulations into a single coupled mathematical framework: (1) neuromorphic dynamics governing state activation and evidence accumulation; (2) connectomic development governing autonomous connection formation, reorganization, and pruning; and (3) Fuzzy/Rough information mathematics preserving partial membership, approximations, and boundary indiscernibility without equating missing evidence with falsity. Our central scientific goal for this IIDS research effort is to determine how dynamic evidence and evolving connectivity mathematically co-evolve across millions of states to match literature grounded reference configurations, generating calibrated, graded concern while preserving the numerical fidelity needed to prevent topological divergence.

Intellectual Merit

Our proposed advance is a mathematical account of recognition in networks whose connectivity develops as evidence accumulates. (1) *Fuzzy/Rough representation*: here, we will determine which membership, approximation, and boundary distinctions preserve partial correspondence without treating missing evidence as falsity. (2) *Connectomic development*: we will formulate autonomous connection rules and determine when developmental connectivity improves evidence propagation relative to fixed connectivity. (3) *Large-scale architecture and dynamics*: our goal is to establish conditions for bounded activation, stable propagation, and adaptation across millions of synthetic states and connections. Finally, (4) *Graded metacognitive concern*: we will investigate how correspondence, uncertainty, context, and correction should govern concern magnitude and timing.

Numerical fidelity is part of each objective. We will analyze whether finite precision, rounding, and arithmetic representation alter evidence distinctions or connection decisions, causing topologies and concern to diverge. Reduced-precision and logarithmic number systems will be compared with checked higher-precision computations, measuring cost at defined fidelity tolerances. To determine where these numerical differences matter, we will follow their effects through crisp, Fuzzy, Rough, and combined representations, then through fixed and developmental connectivity. Selectively disabling mechanisms will help us identify which effects arise from each mechanism and which arise from their coupling.

Broader Impacts

The research could support independent evaluation of AI advice before users act, while preserving human judgment. Its results will inform adaptive systems that must preserve uncertain evidence as connections change. Two Ph.D. researchers will train in dynamical systems, uncertainty representation, numerical analysis, and reproducible evaluation. Shared formulations, software, synthetic interaction cases, and instructional materials will support research and education. Assessment will examine learners' ability to distinguish partial satisfaction, missing evidence, and contradiction.