Recursive Fitness Alignment Protocol (RFAP)

Submission Date: April 19, 2025

Submitted by: Andy Williams

Organization: Caribbean Center for Collective Intelligence (CC4CI)

Email: info@cc4ci.org

Location: St. John's, Antigua

Project Summary

The Recursive Fitness Alignment Protocol (RFAP) is a novel experimental framework for testing the capacity of reasoning systems to self-correct under recursive pressure. Rather than optimizing for fixed goals or architectures, RFAP begins with the simplest conceivable fitness function—assigning a random value between 0 (non-viable) and 1 (optimal)—and introduces a single recursive operation: the ability to test whether the reasoning path used to generate this fitness value is itself coherent.

This minimal recursive structure allows the system to begin aligning itself, without preloaded assumptions or external goals. Through recursive questioning—'What if I'm wrong?' and 'What if I'm right?'—reasoning paths that lead to incoherence naturally self-terminate, while those that increase coherence are reinforced.

While the experiment begins with a simple recursive prompt—"What if I'm right? What if I'm wrong?"—its recursive application across every layer of reasoning reveals something far more powerful: the minimal functional components that define intelligence itself. This isn't just a method for improving alignment. It's a universal architecture for tracking how individuals, AI systems, and entire civilizations navigate toward or away from adaptive coherence. By defining and testing these components recursively, we can forecast and preempt the formation of misalignment attractors—trajectories where error compounds faster than correction. In doing so, we unlock not only safer AI development, but a new foundation for understanding and steering civilization itself.

Over time, this scaffolds a full functional model of intelligence, including:
- A universal fitness representation
- Semantic and contextual representation
- Recursive error detection
- Path selection for long-term stability
- Epistemic coherence across abstraction levels

This protocol has applications in AI alignment, institutional governance, and human reasoning. Our pilot experiment funds philosophy students to explore these mechanisms from first principles, beginning with one hour of structured recursive reflection.

The goal: build adaptive intelligence from the bottom up—not through increasingly complex constraints, but through recursive fitness correction that reveals where coherence fails and how it can evolve.