SPAR Project: Physical Theories of Intelligence and Recursive Self-Modification (Elija Perrier)

Summary

This project investigates the physical foundations of recursive self-modification in intelligent systems, exploring which physical substrates and theories admit adaptive intelligence and the implications for AGI design, safety, and alignment.

Prospective Work Program

This project investigates the physical realisation of intelligence, with a particular focus on recursive self-modification as a defining characteristic of advanced intelligent systems.

Most theories of intelligence are formulated at an algorithmic or computational level, while remaining largely agnostic as to the physical conditions required for their implementation. This project asks the complementary question: under what physical theories, scales, and substrates can recursively self-modifying intelligence exist?

Possible directions include:

- the relationship between physical theories and theories of intelligence;
- recursive self-modification across different physical substrates (e.g. electronic circuits, classical computation, quantum systems, biological systems, or alternative physical realisations);
- physical and thermodynamic constraints on adaptive intelligent systems;
- energy efficiency, information processing, and self-improvement;
- implications for AGI architectures, controllability, and alignment;
- mathematical and computational models of adaptive self-modification;
- empirical investigation of recursive self-improvement through simulation and prototype implementations.

The precise direction will be refined according to the interests and background of participants. The overall objective is to produce publishable research suitable for submission to leading venues in artificial intelligence, machine intelligence, computational physics, or physical information processing.

The project will follow a research-oriented workflow typical of an academic laboratory, including:

- comprehensive review of the relevant literature and identification of open research questions;
- development of mathematical models and theoretical frameworks;

- rapid design and implementation of computational experiments;
- critical evaluation of experimental results;
- iterative refinement of hypotheses and manuscript preparation.

Participants are encouraged to make extensive use of modern AI systems to assist with coding, experimentation, literature exploration, and research design. However, the emphasis will remain on developing independent research capability. In particular, participants should expect to perform the core scientific reasoning, mathematical development, and manuscript writing themselves.

Some background articles of mine that are related, but not definitive, of my research in this area and the project scope are below:

- Watts-per-Intelligence Part II: Algorithmic Catalysis
  https://link.springer.com/chapter/10.1007/978-3-032-33195-3_12
- Deconstructing Superintelligence: Identity, Self-Modification and Différance
  https://link.springer.com/chapter/10.1007/978-3-032-33195-3_14
- Quantum AIXI: Universal Intelligence via Quantum Information
  https://link.springer.com/chapter/10.1007/978-3-032-00800-8_6
- Hamiltonian Formalism for Comparing Quantum and Classical Intelligence
  https://link.springer.com/chapter/10.1007/978-3-032-00800-8_7

Background

A background in one or more of the following disciplines would be beneficial, although it is not essential:

- Mathematics
- Physics
- Computer Science
- Artificial Intelligence / Machine Learning
- Cognitive Science and Philosophy of Science
- Information Theory
- Engineering

Curiosity, willingness to learn unfamiliar material, and enthusiasm for tackling difficult interdisciplinary problems are considerably more important than prior expertise.

Mentoring Structure

The project will be conducted in a manner similar to an academic research group.

Participants should expect to work independently between meetings while progressively developing the project through literature review, theoretical analysis, mathematical modelling, coding, and experimentation.

Mentoring will consist of approximately one hour of online/remote meetings per week, together with periodic research presentations in which participants discuss their progress, experimental findings, and proposed next steps. Meetings will focus on research direction, technical discussion, critical feedback, and planning rather than structured teaching.

The expectation is that participants will progressively develop the skills required to conduct independent research while contributing towards a high-quality scholarly publication.