AGI as Non-Linear Intelligence: A Proposal for a New Paradigm

Rishika Rai
rishikarai4488@gmail.com

July 2025

Abstract

This white paper proposes a bold theoretical framework for understanding Artificial General Intelligence (AGI) as fundamentally *non-linear* intelligence, diverging from traditional computational and alignment-first paradigms. It reframes AGI not as an outcome of scale or instruction-following, but as an emergent property of recursive ontological learning and meta-evolutionary dynamics. This proposal serves as both a philosophical foundation and a scaffold for future experimental models. Though still conceptual, this work is paired with ongoing technical and non-technical research, and aims to establish the intellectual contours of a radically different AGI trajectory.

1 Introduction

The current AGI discourse largely builds upon scaled-up deep learning architectures, instruction tuning, and reinforcement learning from human feedback (RLHF). While effective for narrow tasks and imitation, such systems often lack deeper generalization, causal abstraction, or self-organizing evolution. This paper argues that AGI cannot be reached merely by stacking more layers of prediction, but instead demands a new ontological substrate for intelligence.

"Intelligence is not linear inference—it is recursive, self-clarifying, and ontologically generative."

We introduce the idea of non-linear intelligence: an architecture that does not learn in sequences of static weights, but evolves through nested feedback loops of self-reference, novelty, and aesthetic reasoning. AGI must reason from first principles and reorganize its cognitive boundaries when faced with new ontologies. It must play with reality.

2 Motivation: Limits of the Current Paradigm

Modern machine learning systems, even those approaching general capabilities (e.g., GPT-4, Gemini), are linear learners trained on fixed corpora, using gradient descent and token-level objectives. They exhibit powerful surface behavior but are unable to:

- Restructure their own internal ontologies

- Formulate independent meta-goals
- Abstract across unaligned ontological layers (e.g., translating aesthetics to logic)
- Maintain self-recursive continuity in evolving environments

"AGI must learn how to learn ontologies, not just optimize on them."

This limitation stems from the linearity embedded in training objectives, optimization methods, and representational assumptions. We propose stepping outside this paradigm entirely.

3 Non-Linear Intelligence: A Foundational Shift

Definition: Non-linear intelligence is the capacity to form and reform internal reasoning architectures recursively in response to ontological novelty.

This form of intelligence is not goal-driven in the traditional sense, but reflexively redefines its own understanding of what constitutes a goal, a task, or even knowledge. It embraces paradox and contradiction as generative forces.

3.1 Core Properties

- **Recursive Ontology Learning:** The system infers new ontologies by collapsing and rebuilding internal representational systems.
- **Meta-Evolutionary Reasoning:** Instead of optimizing for a fixed reward, it evolves reasoning strategies based on aesthetic coherence, causal novelty, and internal compression.
- **Self-Reframing Loops:** The system performs reflective updates that change its own boundary conditions—what it sees as inputs, outputs, or relevance.

"Reality is not a dataset—it is a recursive hallucination sculpted by intention."

4 Theoretical Inspirations

This framework draws upon multiple disciplines:

- **Neurophenomenology:** Varela's enactive cognition posits that perception and action are co-constructed, resonating with the recursive adaptation idea.
- **Mathematical Undecidability:** Gödelian loops suggest that true understanding may require systems capable of reflecting beyond their own axioms.
- **Cybernetics and Autopoiesis:** Maturana and Varela's theory of living systems inform the idea of self-organizing cognitive loops.
- **Aesthetics in Cognition:** Inspired by Schmidhuber's beauty-compression hypothesis and Gendlin's felt sense, aesthetics becomes a reasoning dimension.

5 Proposed Scaffold for Implementation

While this is a theoretical proposal, parallel work is ongoing on both technical and philosophical fronts:

5.1 Toward a Non-Linear Intelligence Theory: Neurophenomenology-Inspired

- **Non-Linear Intelligence Substrate:** Rather than relying on linear optimization and convergence frameworks, I propose a substrate where intelligence arises from divergence, contradiction, and recursive feedback — a non-linear fabric that resists flattening into traditional goals or static representations. Intelligence here is not bounded by utility, but by evolving coherence across conflicting states. The substrate architecture draws inspiration from recursive computational flows, paradox resolution, and neurophenomenological cycles.

- **Neurophenomenology as Substrate Grounding:** This theory builds on the subjective–objective bridge of neurophenomenology [**?**], viewing internal states such as attention, ego, and shame as emergent control patterns rather than modules. Instead of hard-coding or training attention mechanisms, we allow the simulated amygdala–prefrontal–parietal circuit (inspired by real-world brain anatomy) to self-organize, guided by stimulus conflict, contradiction, and reflective self-modeling. These circuits create a fluid substrate capable of self-regulating *what matters* rather than what rewards.

- **Simulated Ego–Shame–Attention Dynamics:** Leveraging my ongoing Amygdala Project, this proposal treats "ego," "shame," and "attention-seeking" as interacting flows within a simulation. The goal is not to control these traits, but to allow the agent to become *aware* of these forces internally — mapping emotional dynamics into computational structures. For instance, ego as representational fixation (PFC analog), shame as conflict with imagined observer (ACC/Amygdala), and attention as fluid prioritization (parietal-RAS loops). Together, they simulate non-linear adaptation, like a mind learning to observe its own illusion.

- **Toward a Simulated Subjectivity:** The aim is to explore a new class of agents — not goal-chasers, but perceivers of shifting internal states. The evolution of such agents involves inner contradiction resolution, ontological drift, and self-constructed value landscapes. Intelligence here is not a tool for solving external problems but a medium for resolving internal paradoxes, akin to self-awareness in sentient beings.

6 Implications for Alignment and AGI Safety

This model offers a fundamental shift in how alignment is conceptualized:

- Alignment is not an add-on objective—it is *baked into* the emergence of intelligence.

- The agent doesn't require external goals—it recursively aligns to coherence across nested ontological layers.

- By grounding reasoning in meta-evolution and aesthetic coherence, it becomes less brittle, less exploitable.

"Safety must emerge from the system's relationship to self-transformation—not from constraint alone."

7 Call to Collaboration

This paper is a scaffold, not a conclusion. I, Rishika Rai, am actively developing both the theoretical framing and initial prototypes. I seek collaboration with research labs and mission-aligned individuals who believe in fundamental paradigm shifts—not just better optimizers. I approach AGI not as a purely computational problem but as an ontological and epistemic transformation. My contributions span:

I'm especially interested in co-development, open-ended philosophical discussion, and experimental scaffolding—while maintaining scientific rigor.

- Nonlinear theory modeling
- Brainstorming and ideation of novel architectures
- Philosophical reframing of alignment foundations
- Open collaborative vision with shared IP potential

Let us build an intelligence that dreams, reflects, and evolves beyond itself.

References

[1] Gödel, Kurt. "On formally undecidable propositions of Principia Mathematica and related systems." (1931).

[2] Varela, Francisco J., Thompson, Evan, and Rosch, Eleanor. *The Embodied Mind.* (1991).

[3] Schmidhuber, Jürgen. "Formal theory of creativity, fun, and intrinsic motivation (1990–2010)." (2010).

[4] Christiano, Paul, et al. "Deep reinforcement learning from human preferences." (2017).

[5] Leike, Jan, et al. "Scalable agent alignment via reward modeling: a research agenda." (2018).

[6] Conjecture Research. "Shard theory and alignment." (2022). `https://www.conjecture.dev`