

About Me

2014 - 2019

PhD Researcher

Western University

- Evolutionary RL / NAS
- Novelty Search
- Neuroimaging

2019 - 2020

Postdoctoral Researcher

University of Guelph

- Deep Learning for Economic Forecasting
- Canada's Food Price Report

2020 - 2022

Applied ML Scientist

Vector Institute

- Data Science
- PETs (Privacy Enhancing Technologies)
- Forecasting

2022 - 2024

Social AI Researcher (Psychology)

University of Toronto

- Multiagent RL
- Artificial Hippocampus
- Biologically Plausible Learning

2022 - 2025

Co-Founder

ChainML/Theorix

- Analytics Agents
- Memory and Learning
- Agent-to-Agent Protocol

2026 - Present

Applied ML Scientist

Vector Institute

- Agents & Reasoning
- Social and Evolutionary AI

Human Memory & Learning

AGI, LLMs, and Memory

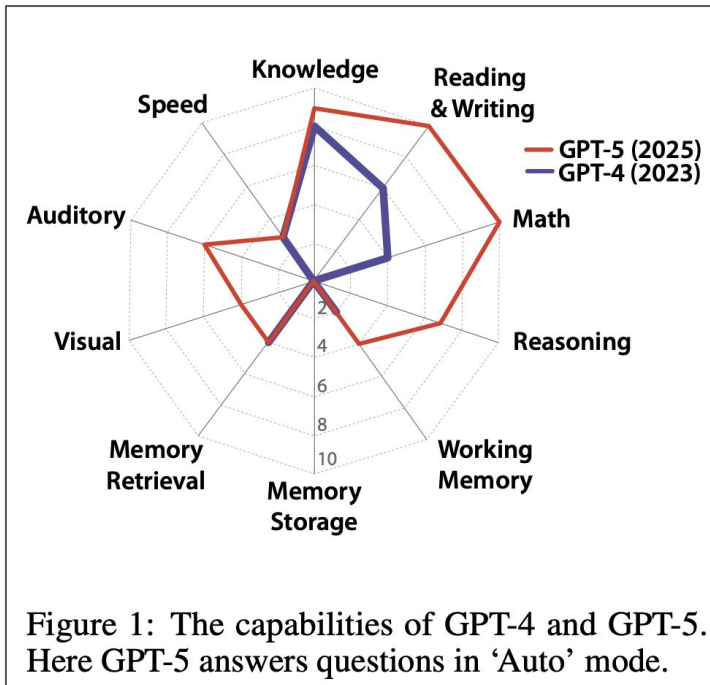
A Definition of AGI

Dan Hendrycks¹, Dawn Song^{2,3}, Christian Szegedy⁴, Honglak Lee^{5,6}, Yarin Gal⁷, Erik Brynjolfsson⁸, Sharon Li⁹, Andy Zou^{1,10,11}, Lionel Levine¹², Bo Han¹³, Jie Fu¹⁴, Ziwel Liu¹⁵, Jinwoo Shin¹⁶, Kimin Lee¹⁶, Mantas Mazeika¹, Long Phan¹, George Ingebreten¹, Adam Khoja¹, Cihang Xie¹⁷, Olawale Salaudeen¹⁸, Matthias Hein¹⁹, Kevin Zhao²⁰, Alexander Pan², David Duvenaud^{21,22}, Bo Li^{23,25}, Steve Omohundro²⁴, Gabriel Alfour²⁵, Max Tegmark¹⁸, Kevin McGrew²⁶, Gary Marcus²⁷, Jaan Tallinn²⁸, Eric Schmidt¹⁸, Yoshua Bengio^{29,30}

¹Center for AI Safety ²University of California, Berkeley ³Virtue AI
⁴Morph Labs ⁵University of Michigan ⁶LG AI Research
⁷University of Oxford ⁸Stanford University ⁹University of Wisconsin–Madison
¹⁰Gray Swan AI ¹¹Carnegie Mellon University ¹²Cornell University
¹³Hong Kong Baptist University ¹⁴HKUST ¹⁵Nanyang Technological University
¹⁶KAIST ¹⁷University of California, Santa Cruz ¹⁸Massachusetts Institute of Technology
¹⁹University of Tübingen ²⁰University of Washington ²¹University of Toronto
²²Vector Institute ²³University of Illinois Urbana-Champaign ²⁴Beneficial AI Research
²⁵Conjecture ²⁶Institute for Applied Psychometrics ²⁷New York University
²⁸CSER ²⁹Université de Montréal ³⁰LawZero

Abstract

The lack of a concrete definition for Artificial General Intelligence (AGI) obscures the gap between today's specialized AI and human-level cognition. This paper introduces a quantifiable framework to address this, defining AGI as matching the cognitive versatility and proficiency of a well-educated adult. To operationalize this, we ground our methodology in Cattell-Horn-Carroll theory, the most empirically validated model of human cognition. The framework dissects general intelligence into ten core cognitive domains—including reasoning, memory, and perception—and adapts established human psychometric batteries to evaluate AI systems. Application of this framework reveals a highly “jagged” cognitive profile in contemporary models. While proficient in knowledge-intensive domains, current AI systems have critical deficits in foundational cognitive machinery, particularly long-term memory storage. The resulting AGI scores (e.g., GPT-4 at 27%, GPT-5 at 57%) concretely quantify both rapid progress and the substantial gap remaining before AGI.



Human Learning & Memory

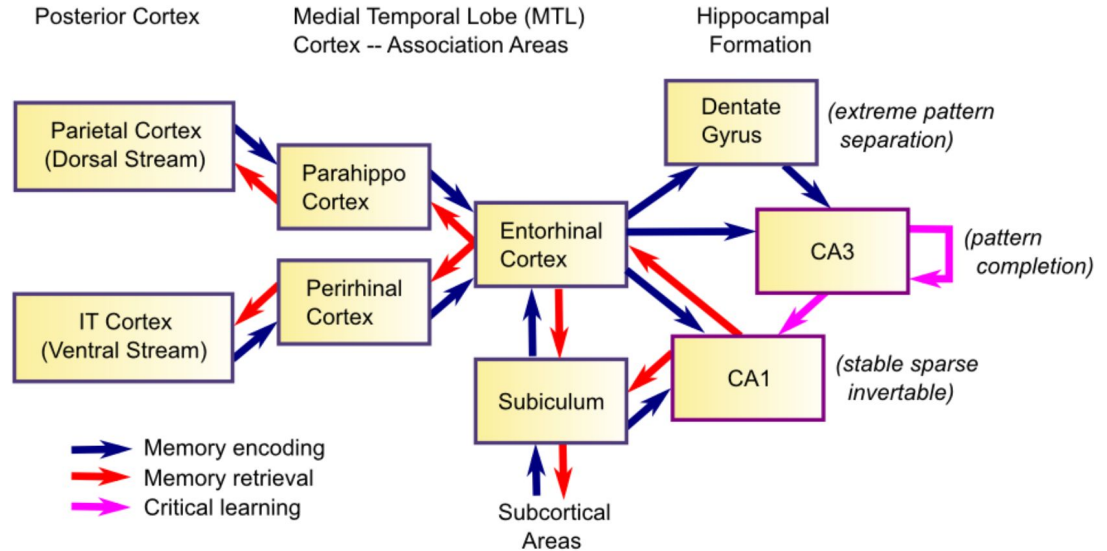
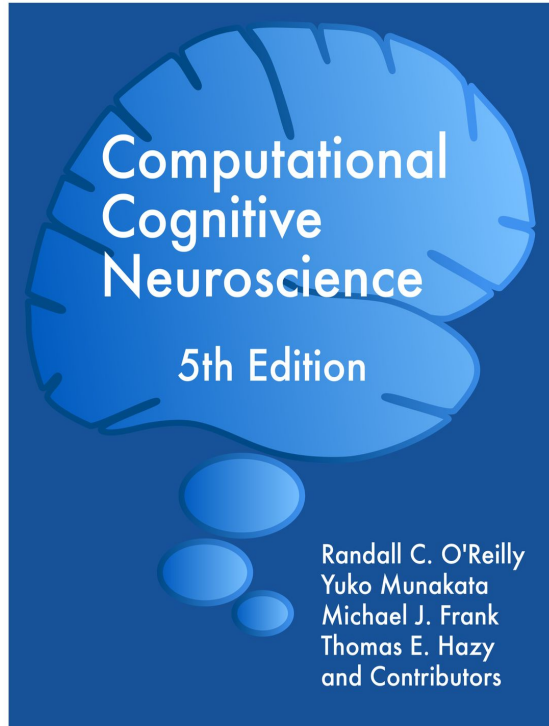
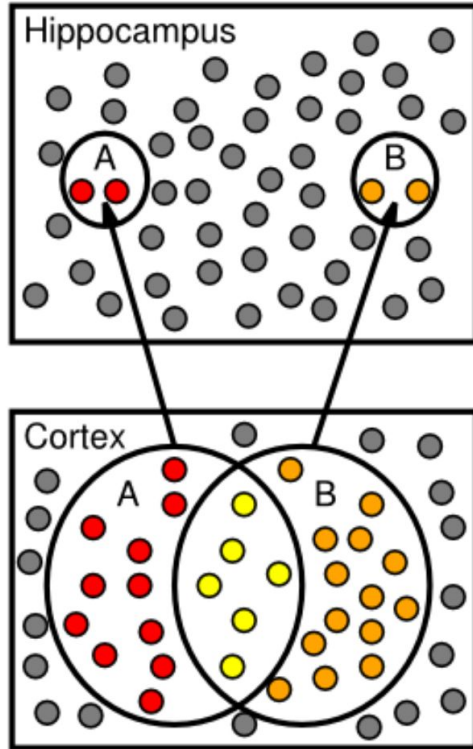


Figure 7.2: The hippocampus sits on “top” of the cortical hierarchy and can **encode information from all over the brain, binding it together into an episodic memory**. Dorsal (parahippocampal) and Ventral (perirhinal) pathways from posterior cortex converge into the entorhinal cortex, which is then the input and output pathway of the hippocampus proper, consisting of the dentate gyrus (DG) and areas of “ammon’s horn” (cornu ammonis, CA) — CA3 and CA1. CA3 represents the primary “engram” for the episodic memory, while CA1 is an invertible encoding of EC, such that **subsequent recall of the CA3 engram can activate CA1 and then EC, to reactivate the full episodic memory** out into the cortex.

Pattern Separation & Completion



Hippocampus regions (especially DG) have an inductive bias for sparse activation, which enables effective **pattern separation** in episodic memory formation.

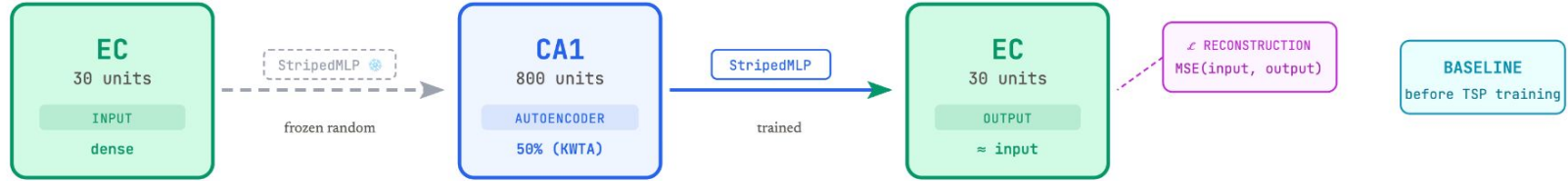
Hippocampus region CA3 has a large number of recurrent connections that enables complex **pattern completion** from partial cues.

Hippocampus anatomy can be considered as an evolutionary solution to the tradeoff problem between pattern separation (distinct memory formation) and pattern completion (association with existing memory).

Hippocampal DNN Module

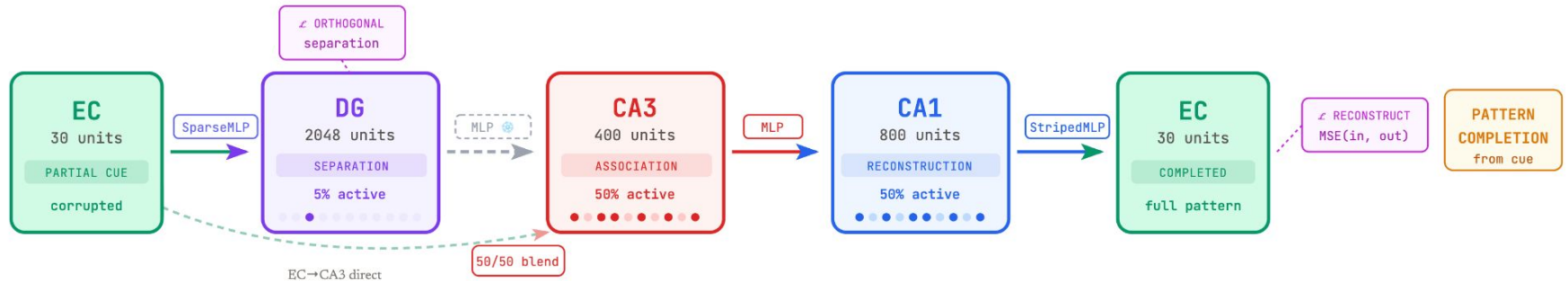
MSP: Monosynaptic Pathway **ENCODING**

Theta Phase 1 – Direct autoencoder baseline (CA3 inhibited)

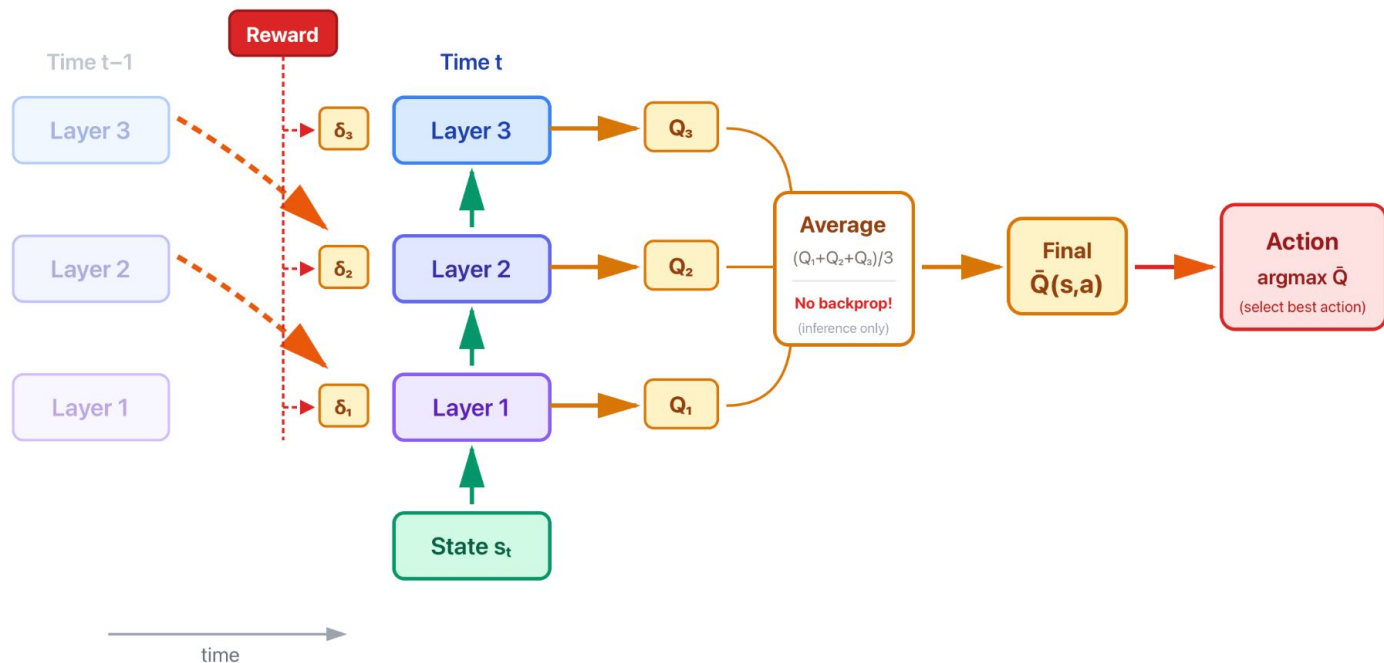


TSP: Trisynaptic Pathway **RECALL**

Theta Phase 2 – Pattern separation (DG) \rightarrow Association (CA3) \rightarrow Completion



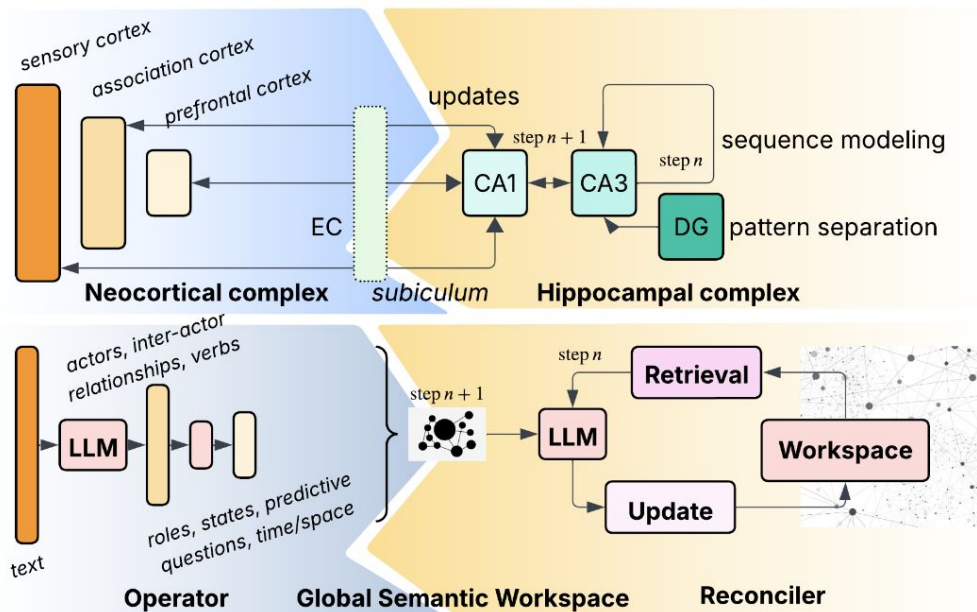
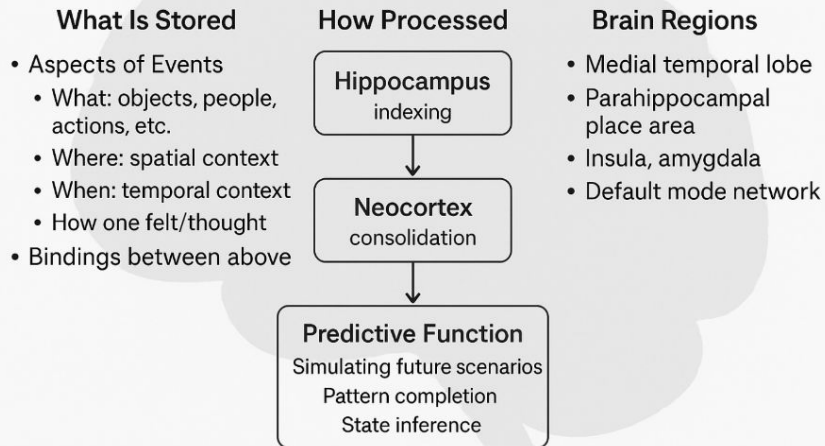
Artificial Dopamine: Distributed TD Learning



Key insight: Each layer independently computes $\delta = r + \gamma \cdot \max_{a'} Q(s', a') - Q(s, a)$ and updates its own weights. No error propagation between layers.

Functional Hippocampus

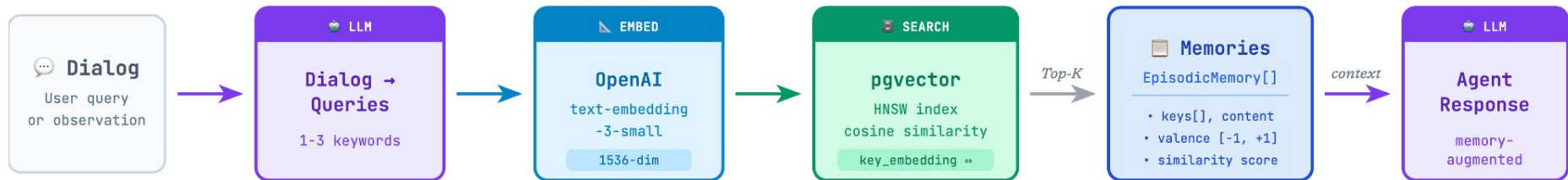
Functional Structure of Episodic Memory



Rajesh, Shreyas, Pavan Holur, Chenda Duan, David Chong, and Vwani Roychowdhury. 2025. "Beyond Fact Retrieval: Episodic Memory for RAG with Generative Semantic Workspaces." *arXiv [Cs.AI]*. arXiv. <https://doi.org/10.48550/arXiv.2511.07587>.

Experiential Learning With RAG

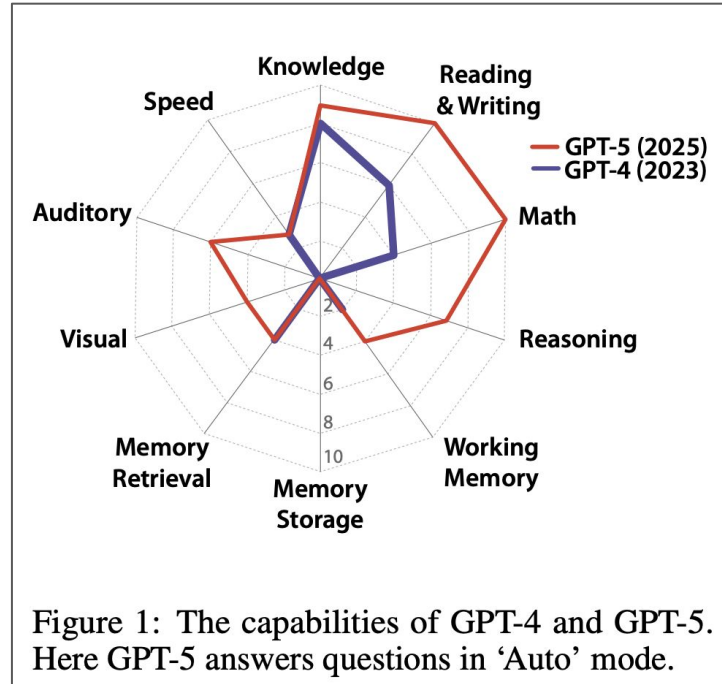
Memory Retrieval Pipeline QUERY → RESPONSE



Memory Consolidation Pipeline POST-INTERACTION



Going back to this result: what would the results look like if we evaluated a SOTA agentic system rather than the foundation LLM alone?



Social AI

Social Learning & Behaviour



PNAS Nexus, 2025, 4, pgaf076

<https://doi.org/10.1093/pnasnexus/pgaf076>

Advance access publication 4 March 2025

Research Report

Social coordination perpetuates stereotypic expectations and behaviors across generations in deep multiagent reinforcement learning

Rebekah A. Gelpi^{a,b,c,*}, Yikai Tang^{a,b}, Ethan C. Jackson^{a,b} and William A. Cunningham^{a,b,c,d}

^aDepartment of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3

^bVector Institute, 101 College Street, Toronto, Ontario, Canada M5G 1L7

^cSchwartz Reisman Institute for Technology and Society, University of Toronto, 101 College Street, Toronto, Ontario, Canada M5G 1L7

^dDepartment of Computer Science, University of Toronto, 40 St. George Street, Toronto, Ontario, Canada M5S 2E4

*To whom correspondence should be addressed: Email: rebekah.gelpi@mail.utoronto.ca

Edited By Attila Szolnoki

Abstract

Despite often being perceived as morally objectionable, stereotypes are a common feature of social groups, a phenomenon that has often been attributed to biased motivations or limits on the ability to process information. We argue that one reason for this continued prevalence is that preexisting expectations about how others will behave, in the context of social coordination, can change the behaviors of one's social partners, creating the very stereotype one expected to see, even in the absence of other potential sources of stereotyping. We use a computational model of dynamic social coordination to illustrate how this “feedback loop” can emerge, engendering and entrenching role-consistent stereotypic behavior and then show that human behavior on the task generates a comparable feedback loop. Notably, people's choices on the task are not related to social dominance or system justification, suggesting biased motivations are not necessary to maintain these stereotypes.

Keywords: reinforcement learning, agent-based model, stereotyping, social norms, coordination

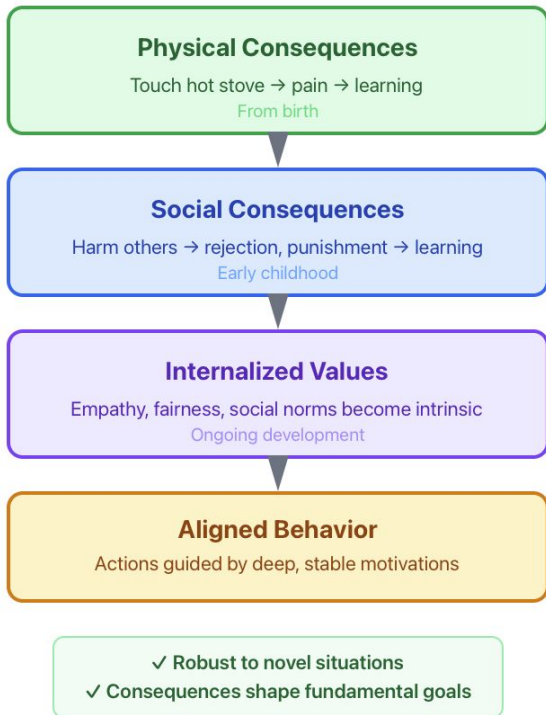
Key Finding:

Individual RL agents learn to conform to expectations for the group, even if this is in conflict with their own skills.

Social Alignment

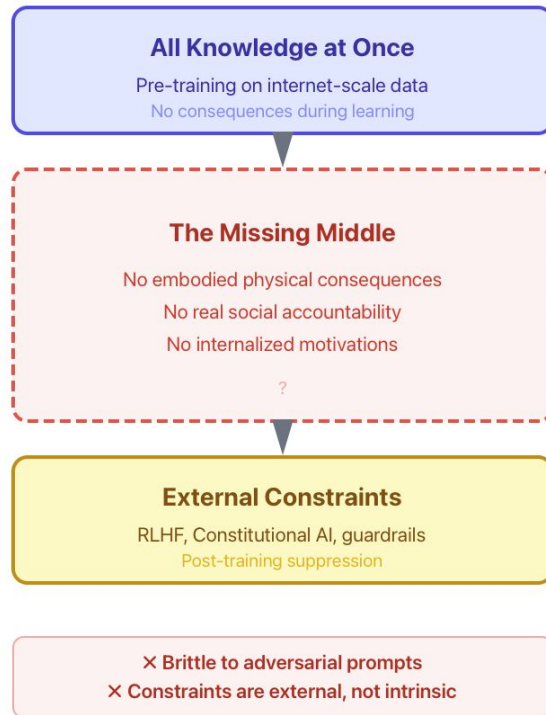
Human Development

(sequential, embodied, social)



Current AI Development

(all-at-once, disembodied, asocial)



Evolutionary AI

Evolutionary Program Search With LLMs

DARWIN GÖDEL MACHINE: OPEN-ENDED EVOLUTION OF SELF-IMPROVING AGENTS

Jenny Zhang^{*1,2} Shengran Hu^{*1,2,3} Cong Lu^{1,2,3} Robert Lange^{†,3} Jeff Clune^{†,1,2,4}

¹University of British Columbia ²Vector Institute ³Sakana AI ⁴Canada CIFAR AI Chair
{jennyzzt, srhu, conglu}@cs.ubc.ca, robert@sakana.ai, jeff.clune@ubc.ca

ABSTRACT

Most of today's AI systems are constrained by human-designed, fixed architectures and cannot autonomously and continuously improve themselves. The scientific method, on the other hand, is a cumulative and open-ended system, where each innovation builds upon previous artifacts, enabling future discoveries. There is growing hope that the current manual process of advancing AI could itself be automated. If done safely, such automation would accelerate AI development and allow us to reap its benefits much sooner. This prospect raises the question of how AI systems can endlessly improve themselves while getting better at solving relevant problems. Meta-learning can automate the discovery of novel algorithms, but is limited by first-order improvements and the human design of a suitable search space. The Gödel machine (Schmidhuber, 2007) proposed a theoretical alternative: a self-improving AI that repeatedly modifies itself in a provably beneficial manner. Unfortunately, proving that most changes are net beneficial is impossible in practice. We introduce the Darwin Gödel Machine (DGM), a novel self-improving system that iteratively modifies its own code (thereby also improving its ability to modify its own codebase) and empirically validates each change using coding benchmarks. Inspired by Darwinian evolution and open-endedness research, the DGM grows an archive of generated coding agents. It samples agents from this archive, which self-modify to create new, interesting versions of themselves. This open-ended exploration forms a growing tree of diverse, high-quality agents and allows the parallel exploration of many different paths through the search space. Empirically, the DGM automatically improves its coding capabilities (e.g., better code editing tools, long-context window management, peer-review mechanisms), increasing performance on SWE-bench from 20.0% to 50.0%, and on Polyglot from 14.2% to 30.7%. Furthermore, the DGM significantly outperforms baselines without self-improvement or open-ended exploration. All experiments were done with safety precautions (e.g., sandboxing, human oversight). Overall, the DGM represents a significant step toward self-improving AI, capable of gathering its own stepping stones along a path that unfolds into endless innovation. All code is open-sourced at <https://github.com/jennyzzt/dgm>.

ProFiT: Program Search for Financial Trading

Matthew Siper¹, Ahmed Khalifa¹, Lisa Soros¹, Muhammad Umair Nasir¹, Jay Azhang¹, and Julian Togelius¹

¹ Nof1, New York, NY, USA

² m@thenof1.com

Abstract. We present a framework called Program Search for Financial Trading (ProFiT), a large-language-model-driven evolutionary search for automated discovery and continual improvement of algorithmic trading strategies in financial markets. These markets are inherently non-stationary and thus resist static modeling or prediction. ProFiT integrates code-level mutation, self-analysis, and walk-forward validation within a closed feedback loop, enabling trading strategies to autonomously evolve in response to changing market conditions. ProFiT consistently outperforms both random and Buy-and-Hold strategies, which are used as baselines, across seven liquid futures assets. Specifically, it surpasses Buy-and-Hold in over 77% of all evolved strategy-asset combinations, outperforms random in 100% of cases, and furthermore achieves improvements over the initial seed strategies in more than 94% of runs. Collectively, these results demonstrate that the ProFiT framework yields robust, statistically significant, and risk-adjusted gains across diverse assets and market regimes, establishing a practical pathway toward open-ended, self-improving algorithmic trading systems.

Keywords: Evolutionary Algorithms · Tree Search · Finance · LLM · Code Generation.

Evolutionary Program Search With LLMs

LLM Fine Tuning At Scale

EVOLUTION STRATEGIES AT SCALE: LLM FINE-TUNING BEYOND REINFORCEMENT LEARNING

Xin Qiu ^{*†}
Cognizant AI Lab

Yulu Gan ^{*‡}
MIT

Conor F. Hayes ^{*}
Cognizant AI Lab

Qiyao Liang [‡]
MIT

Elliot Meyerson
Cognizant AI Lab

Babak Hodjat
Cognizant AI Lab

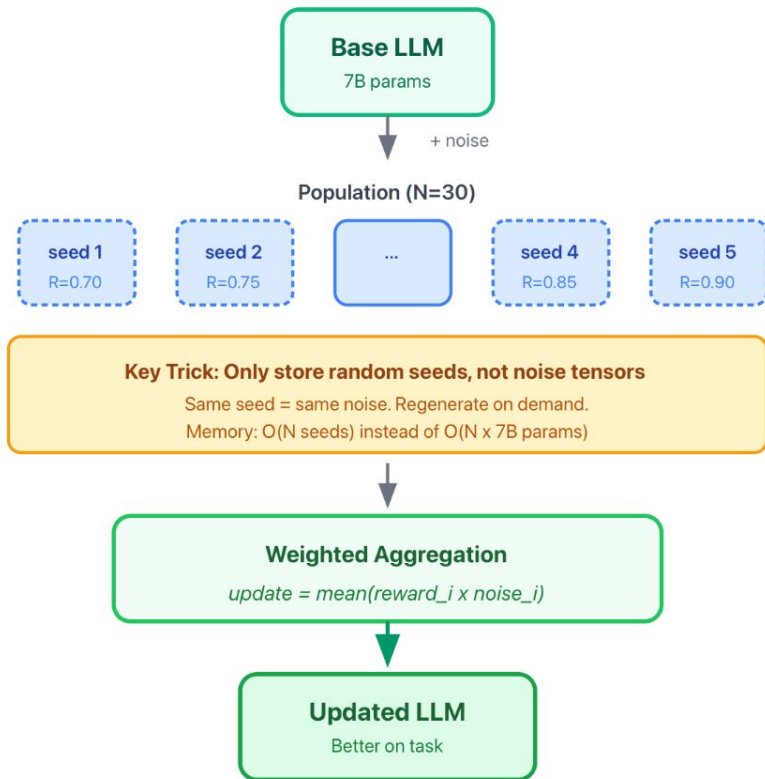
Risto Miikkulainen
Cognizant AI Lab,
UT Austin

ABSTRACT

Fine-tuning pre-trained large language models (LLMs) for down-stream tasks is a critical step in the AI deployment pipeline. Reinforcement learning (RL) is arguably the most prominent fine-tuning method, contributing to the birth of many state-of-the-art LLMs. In contrast, evolution strategies (ES), which once showed comparable performance to RL on models with a few million parameters, was neglected due to the pessimistic perception of its scalability to larger models. In this work, we report the first successful attempt to scale up ES for fine-tuning the full parameters of LLMs, showing the surprising fact that ES can search efficiently over billions of parameters and outperform existing RL fine-tuning methods in multiple respects, including sample efficiency, tolerance to long-horizon rewards, robustness to different base LLMs, less tendency to reward hacking, and more stable performance across runs. It therefore serves as a basis to unlock a new direction in LLM fine-tuning beyond what current RL techniques provide. The source codes are provided at: <https://github.com/VsonicV/es-fine-tuning-paper>.

Evolution Strategies (ES)

Explore in PARAMETER SPACE

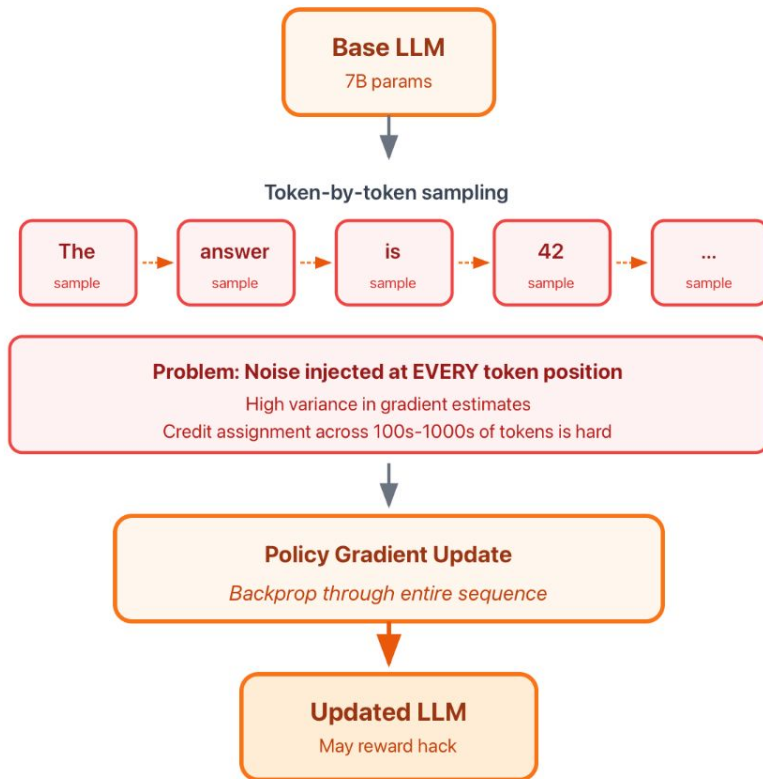


Advantages:

- One noise sample per trajectory (low variance)
- No backpropagation needed (inference only)
- Optimizes solution distribution (robust)

Reinforcement Learning (PPO/GRPO)

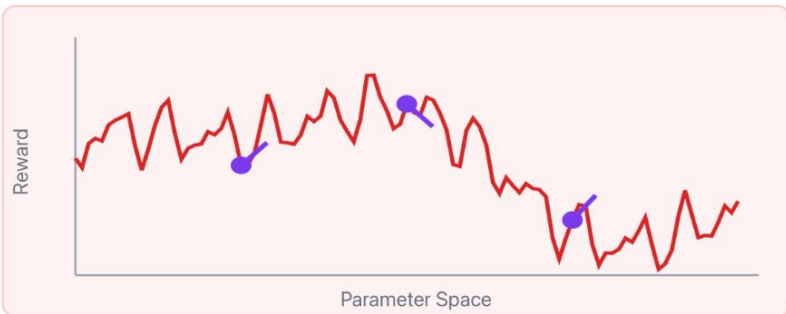
Explore in ACTION SPACE



Challenges:

- Noise at every step (high variance)
- Requires backpropagation (memory intensive)
- Optimizes single solution (can hack rewards)

RL: Navigating the Raw Landscape



Observed: High variance gradient estimates

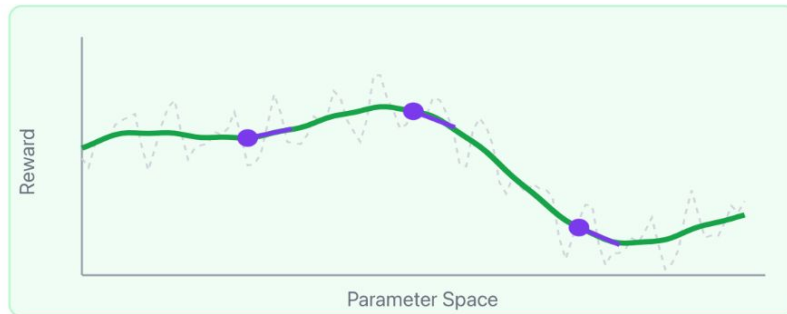
- Noise injected at every token position
- Prone to reward hacking without KL penalties
- Sensitive to hyperparameters

Empirical: High variance

15.5x higher std across runs (Table 2)



ES: Hypothesized Smoothing Effect



Observed: Low variance, stable optimization

- Single noise sample determines entire trajectory
- No reward hacking observed (no KL penalty needed)
- Same hyperparams work across all models

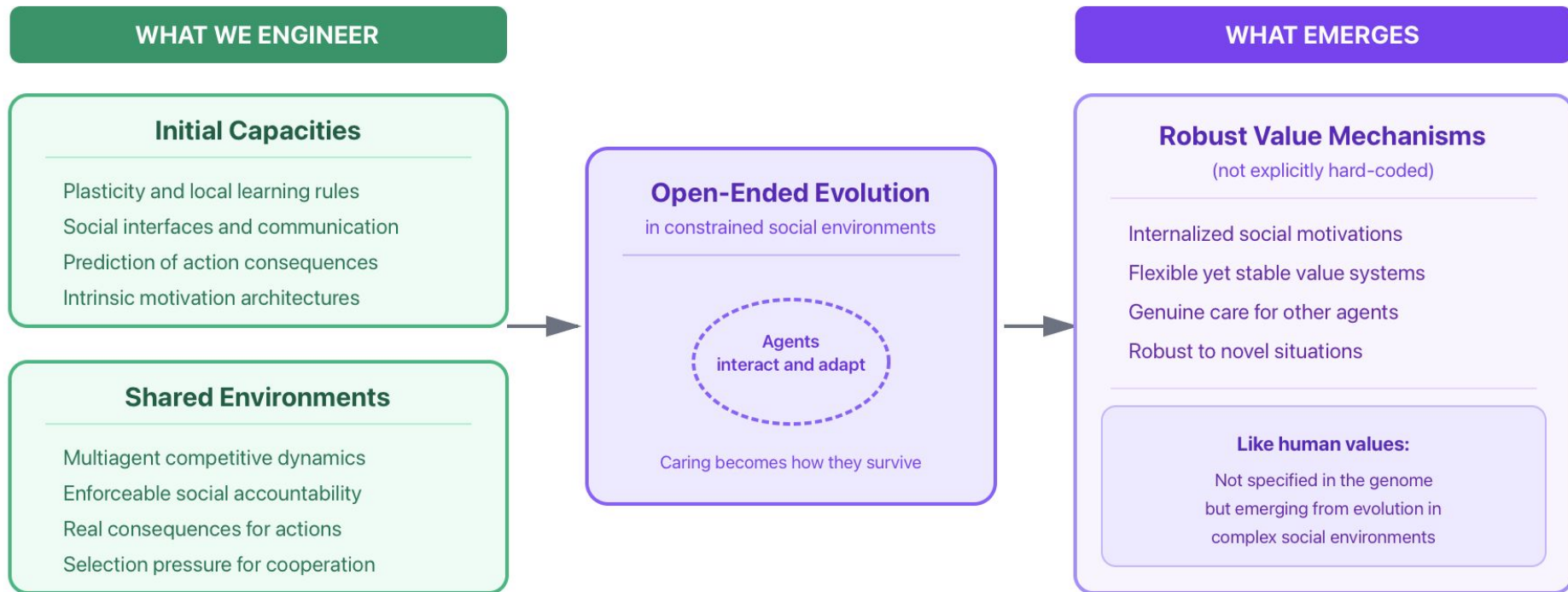
Empirical: Low variance

Consistent results across runs (Table 2)

Open Question: Why does N=30 optimize 7 billion parameters?

Hypotheses: (1) Gaussian convolution smooths reward landscape
(2) LLMs have low intrinsic dimensionality (Aghajanyan et al., 2021)

Alignment Through Constrained Evolution



The goal is not to hand-design social alignment. It is to ensure that in a shared, evolving environment, **social alignment becomes a predictable evolutionary advantage.**