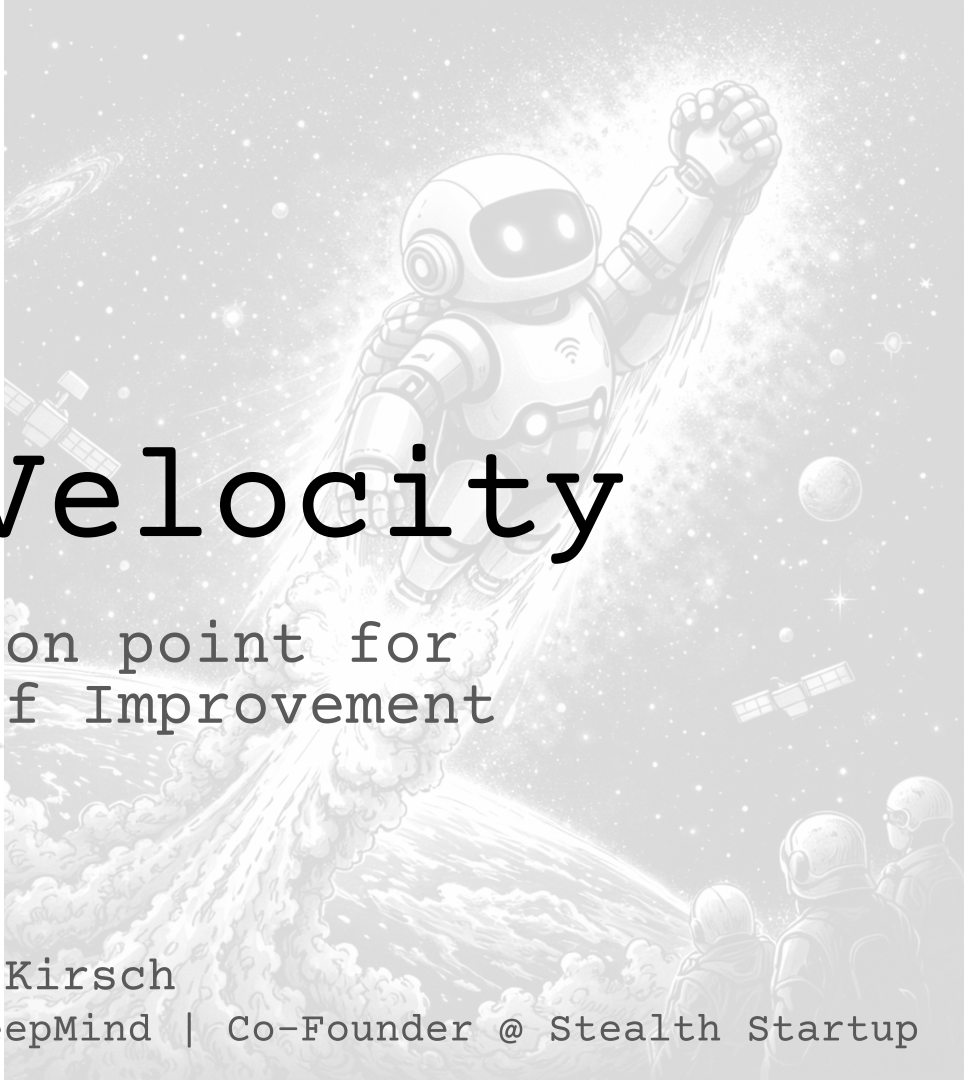


Escape Velocity

The inflection point for Recursive Self Improvement

Louis Kirsch
Automating AI Research | Ex Google DeepMind | Co-Founder @ Stealth Startup

Omega Ω =Superintelligence

Jürgen
Major milestones in history fit a curiously precise exponential pattern pointing to a convergence at around 2040.

(half jokingly)

[Deep Learning History](#) & TEDx, Schmidhuber 2014
Inspired by Teilhard de Chardin 100 years ago

Ω = 2040 or so - 13.8 B years: Big Bang

Ω - 1/4 of this time: Ω - 3.5 B years: first life on Earth
Ω - 1/4 of this time: Ω - 0.9 B years: first animal-like mobile life
Ω - 1/4 of this time: Ω - 220 M years: first mammals (our ancestors)
Ω - 1/4 of this time: Ω - 55 M years: first primates (our ancestors)
Ω - 1/4 of this time: Ω - 13 M years: first hominids (our ancestors)
Ω - 1/4 of this time: Ω - 3.5 M years: first stone tools (dawn of tech)
Ω - 1/4 of this time: Ω - 850 K years: controlled fire (next big tech)
Ω - 1/4 of this time: Ω - 210 K years: anatomically modern man
Ω - 1/4 of this time: Ω - 50 K years: behaviorally modern man
Ω - 1/4 of this time: Ω - 13 K years: neolithic revolution, civilization
Ω - 1/4 of this time: Ω - 3.3 K years: iron age, 1st population explosion
Ω - 1/4 of this time: Ω - 800 years: first guns & rockets (in China)
Ω - 1/4 of this time: Ω - 200 years: industrial revolution
Ω - 1/4 of this time: **Ω - 50 years (around 1990):** information revolution, WWW, cell phones & PCs for all, Cold War ends, Modern AI starts, Miraculous Year ...
Ω - 1/4 of this time: **Ω - 13 years (2030 or so):** cheap AIs with one human brain power? And then what?
Ω - 1/4 of this time: Ω - 3 years: ??
Ω - 1/4 of this time Ω - 9 months: ????
Ω - 1/4 of this time: Ω - 2 months: ????????
Ω - 1/4 of this time: Ω - 2 weeks: ???????????????? ....

J Schmidhuber 2014

Finally multiply Ω by 4! At the age of **55 B years**, the visible cosmos will be permeated by intelligence. After **Ω,** AIs will have plenty of time to go where the physical resources are, to make more and bigger AIs.

What drives the final stretch towards Ω ?

Automation of AI Research

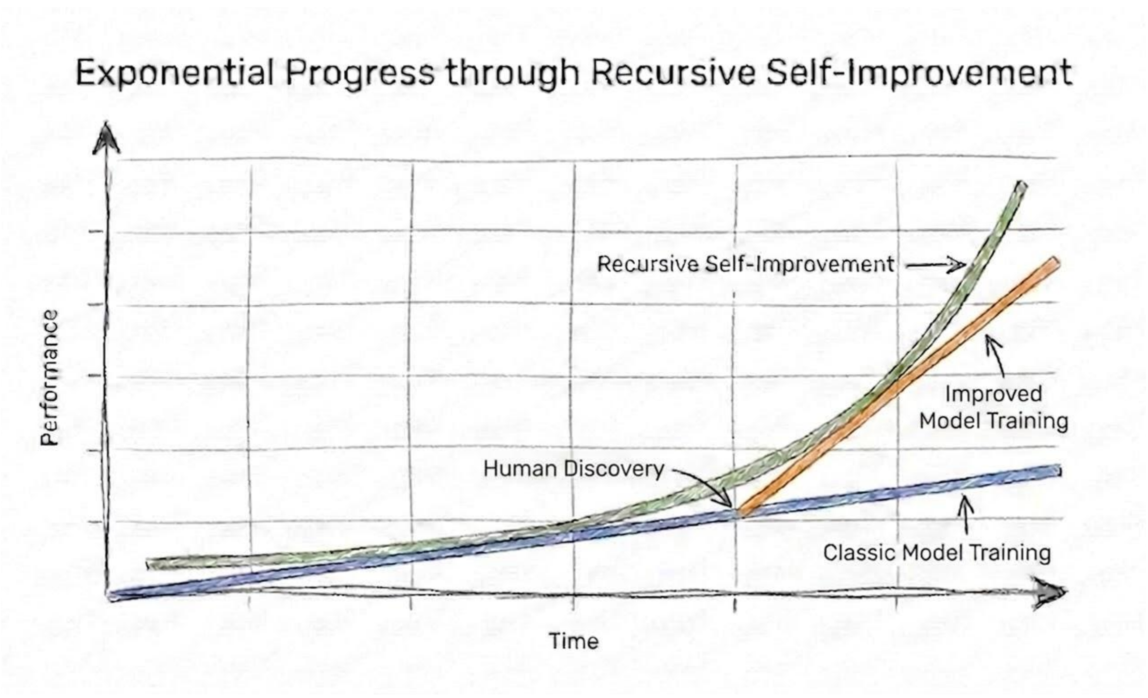
An AI doing AI research.

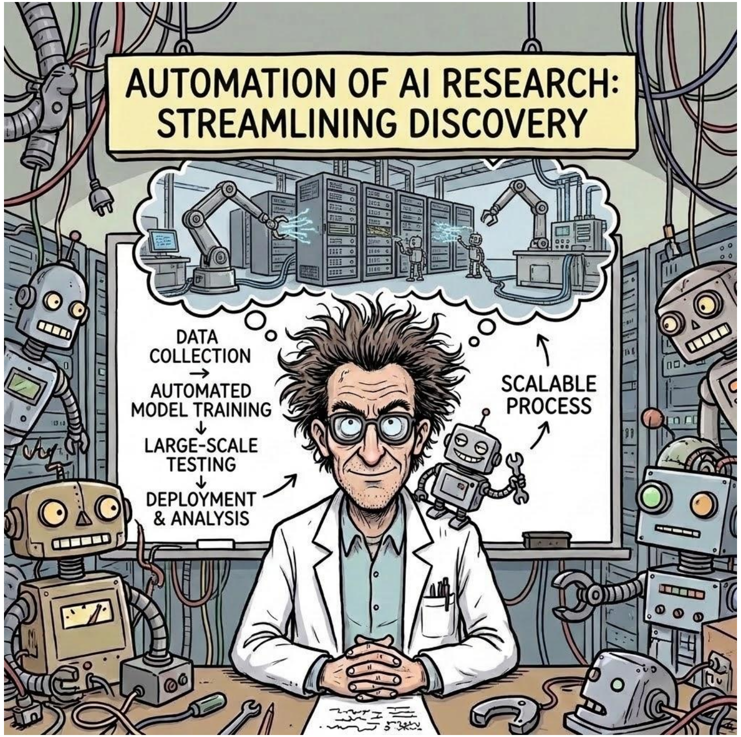
Recursive Self Improvement

A system that does research on itself to create an even better version of itself.

Lessons Learned

Open Problems





Automate AI Research

Why automate AI Research?

Automatically make discoveries by ...

- Processing *vastly* more knowledge than any human could.
- Thinking & experimenting *faster* than any human.
- Highly more *parallel & non-linear* research processes than human thought.

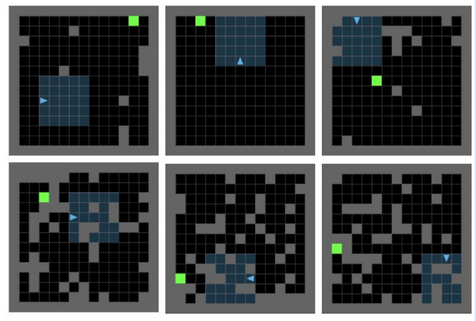
...

- Inventing *new workflows and tooling* to do science.
- Inventing an entirely new *scientific method*.

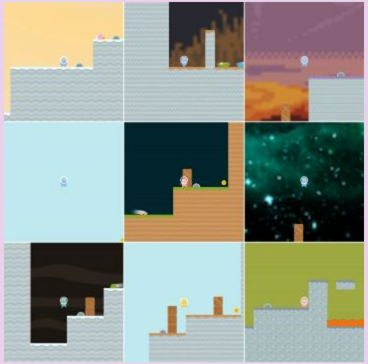
This talk is still (largely) human made.

From Meta-Learning ...

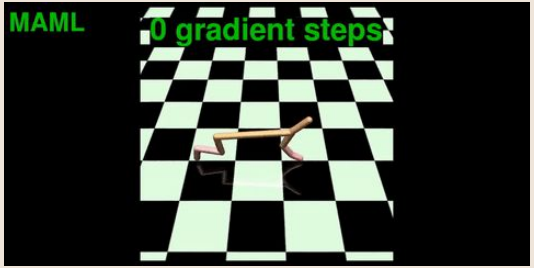
Excels in adaptation to unseen but similar tasks



Env Distribution A



Env Distribution B



Env Distribution C

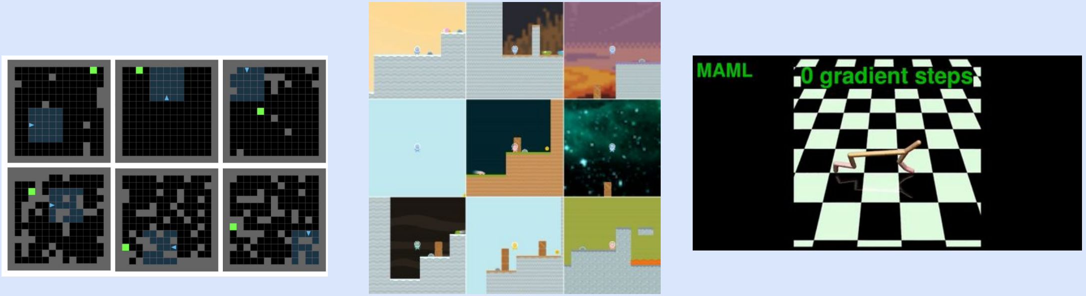
Meta Learner A
→ Learning Algorithm A

Meta Learner B
→ Learning Algorithm B

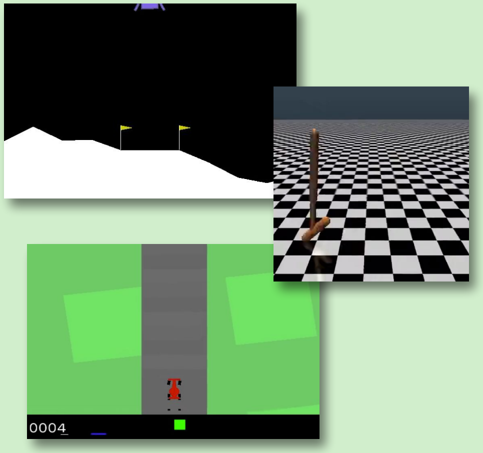
Meta Learner C
→ Learning Algorithm C

From General-Purpose Meta-Learning ...

Enable **reusability** across a wide range of tasks



Env Distribution

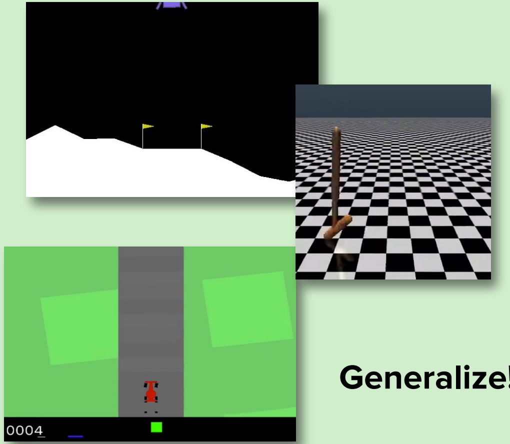


Meta Learner
→ General-Purpose Learning Algorithm

Generalize

From General-Purpose Meta-Learning ...

$$\nabla_\phi \mathbb{E}_\tau[L_{GAE}(\tau, \pi_\phi, V)] := \mathbb{E}_\tau[\nabla_\phi \sum_{t=0}^{T-1} \log \pi_\phi(a_t|s_t) \cdot A(\tau, V, t)].$$



Generalize!

Meta-learn a novel RL loss function with a neural network approximator

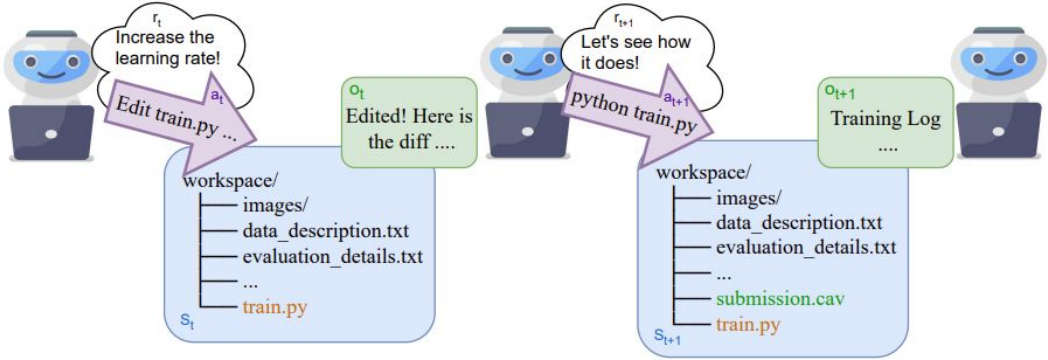
$$\nabla_\phi \mathbb{E}_\tau[L(\tau, \pi_\phi, V)].$$

MetaGenRL, Kirsch et al 2019

… to agentic scientists

Human Prompt: Think step by step.

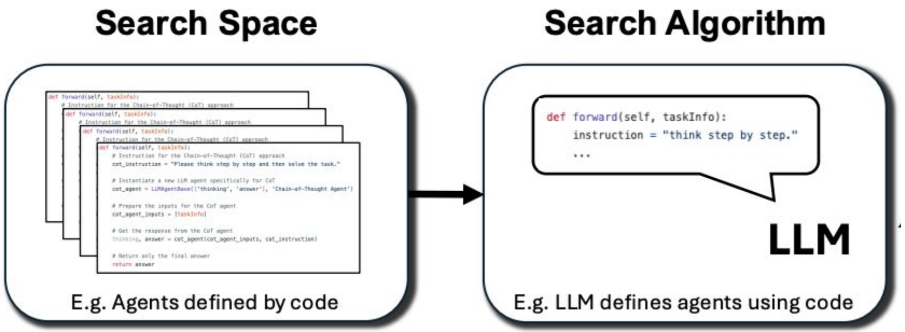
Learned Prompt: Think like a scientist!



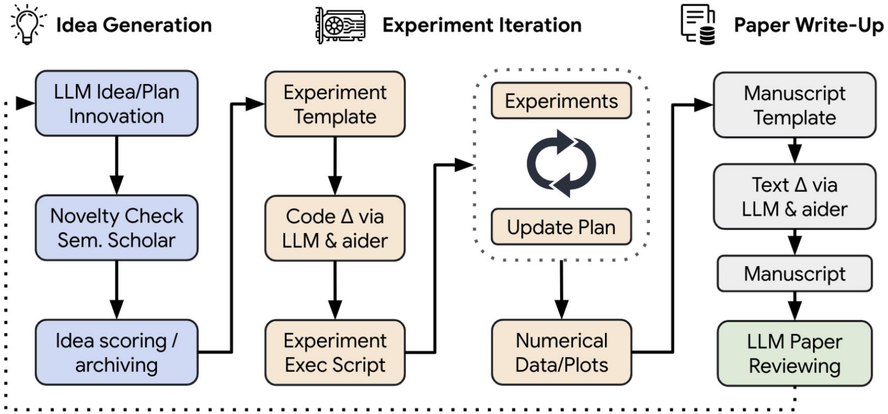
OPRO (Yang et al 2023)
Prompt Breeder (Fernando et al 202

ML Agent Bench (Huang et al 2023)
MLE Bench (Shern et al 2024)

… to agentic scientists



GPTSwarm (Zhuge 2024)
ADAS (Hu et al 2024)



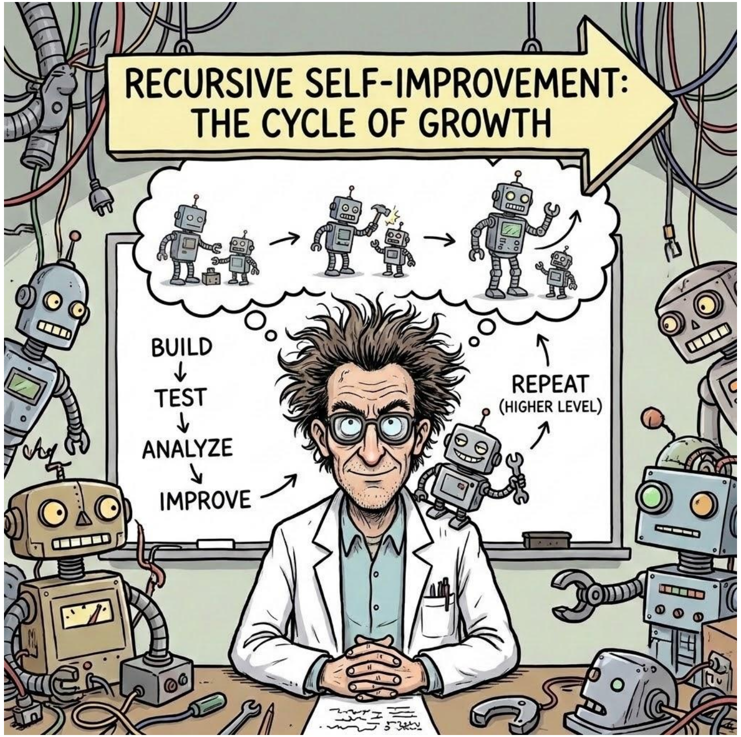
Sakana AI Scientist (Lu et al 2024)
MLGym (Nathani et al 2025)

Search spaces

	"In-context" learners	Agentic / LLM-based	
Meta Params: RL Loss Function	Meta Params: RNN / Transformer W	Meta Params: LLM & Meta Harness	Meta Params: LLM & Meta Harness
↓	↓	↓	↓
Policy weights	RNN Memory / Transformer Ctx	Harness / Prompts / Code	LLM Parameters & Architecture

MetaGenRL

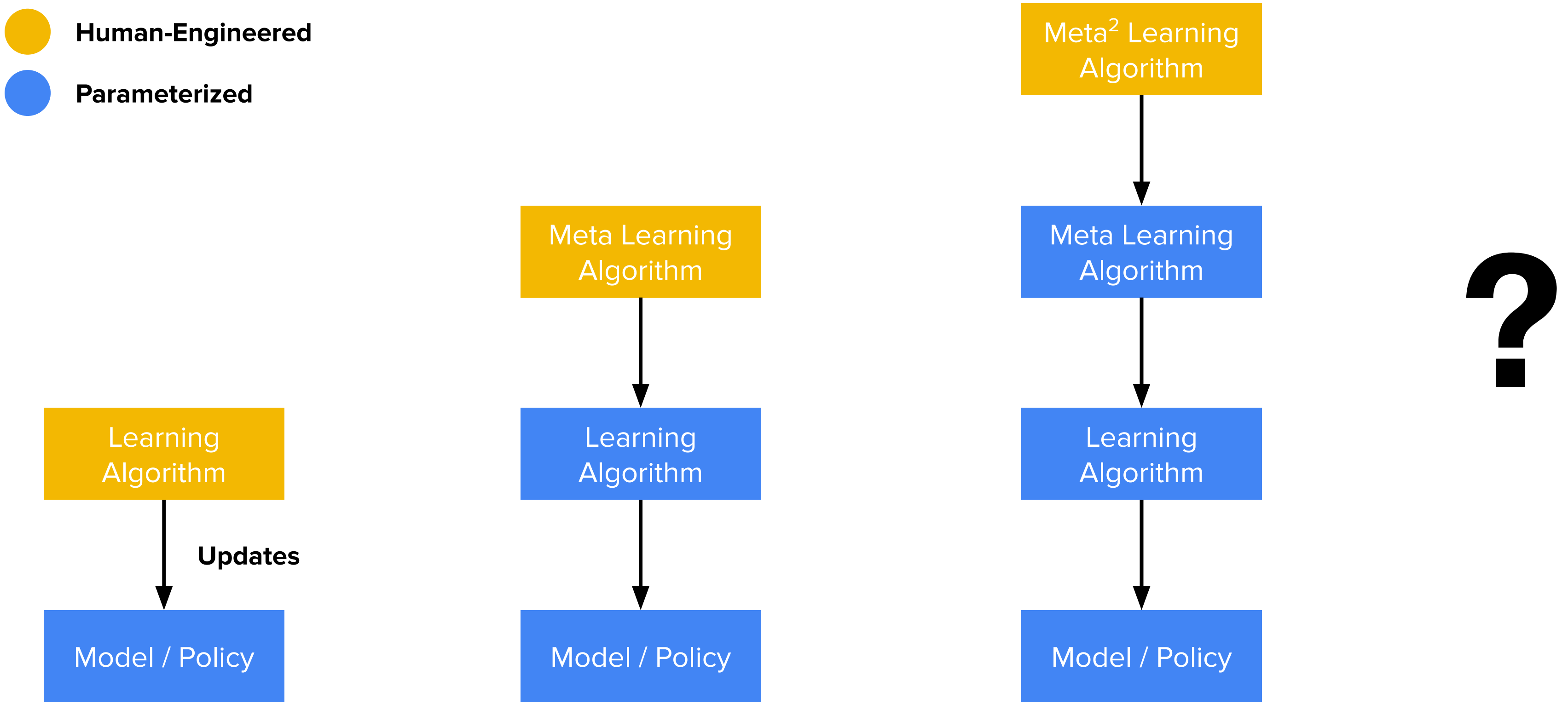
L2L, Hochreiter 2001
VSML, Kirsch 2020
GPICL, Kirsch 2022



Recursive Self-Improvement

Removing humans from the outer loop

Learning   Meta-Learning   Meta-Meta-Learning

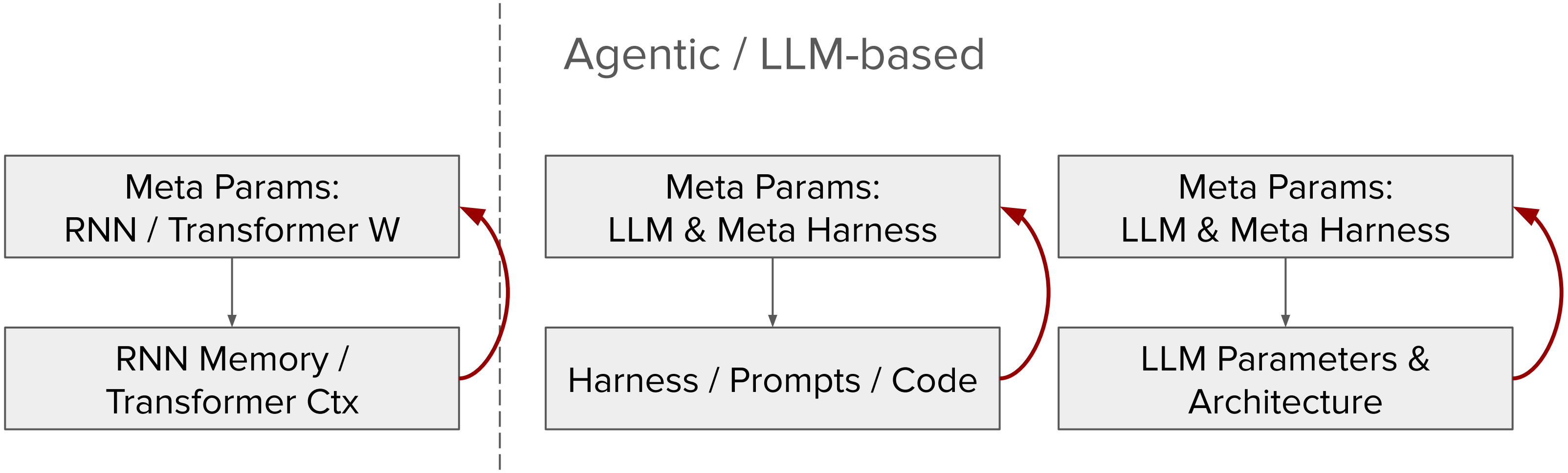


[Eliminating Meta Optimization Through Self-Referential Meta Learning](), Kirsch & Schmidhuber 2022

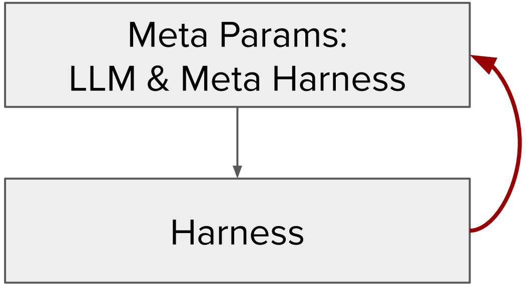
RSI - Closing the loop

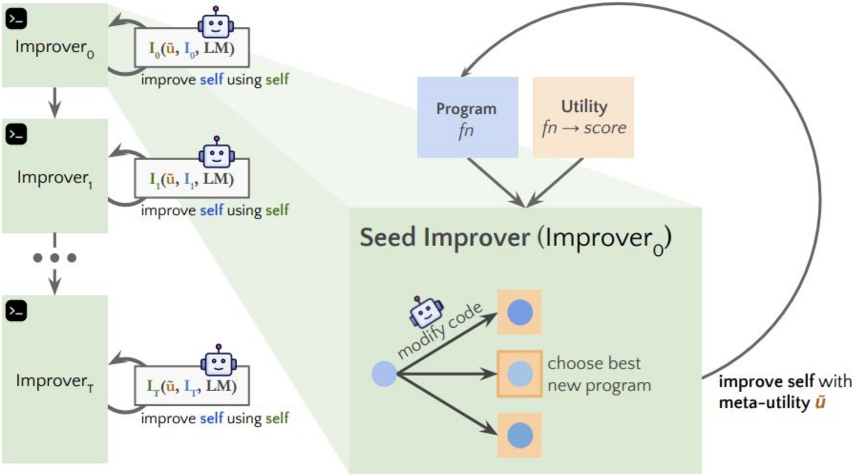
Objective: Minimizing reliance on human engineering

An RSI can improve its own capabilities indefinitely, without bounds, and if desirable without human intervention.



Agentic Examples





Self-Taught Optimizer (Zelikman et al 2023)
Darwin Gödel Machine (Zhang et al 2025)

Guaranteeing improvement

No guarantees (e.g. STOP, 2023)

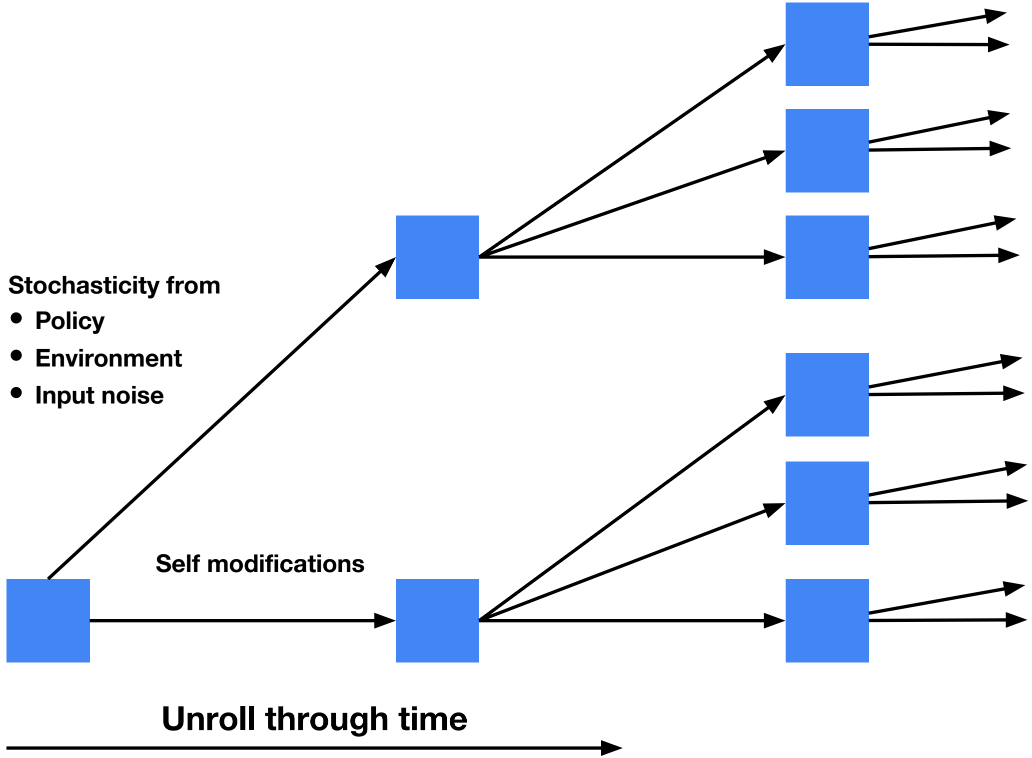
Proofs of improvement (e.g. The Gödel Machine, 2009)

Reverting harmful changes (e.g. The Success Story Algorithm, 1997)

A meta-orchestrator / evolutionary algorithm (e.g. Darwin Gödel Machine, 2025)

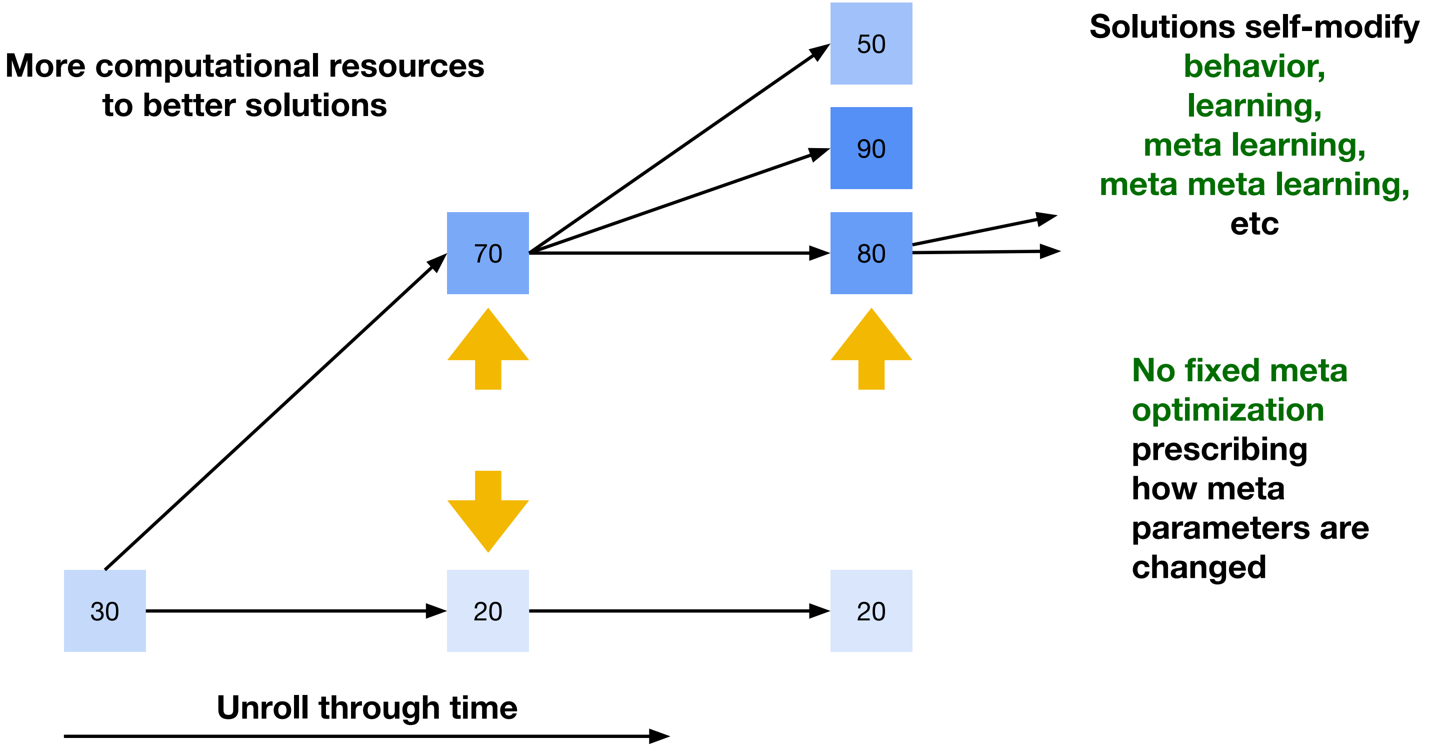
Computational resources based on fitness (e.g. **Fitness Monotonic Execution, 2022**)

Fitness monotonic execution



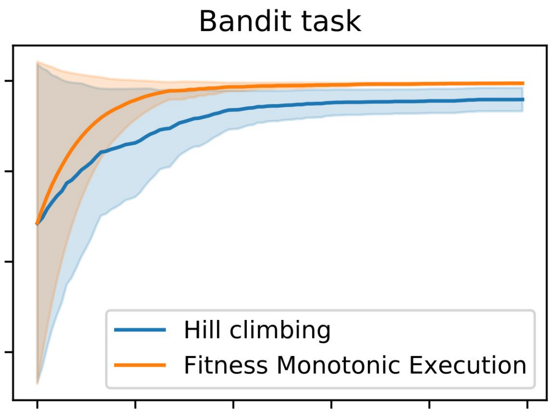
[Eliminating Meta Optimization Through Self-Referential Meta Learning](), Kirsch & Schmidhuber 2022
[Chapter 7](), 'Automating AI Research' PhD Thesis, Kirsch 2025

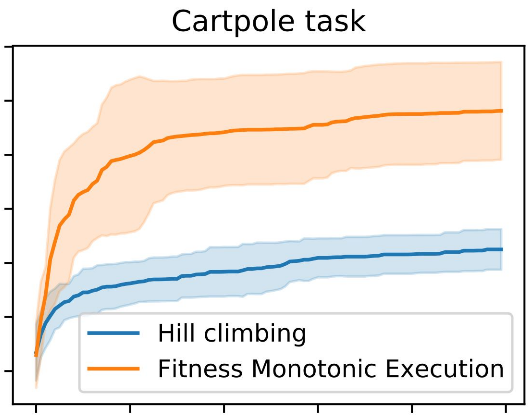
Fitness monotonic execution

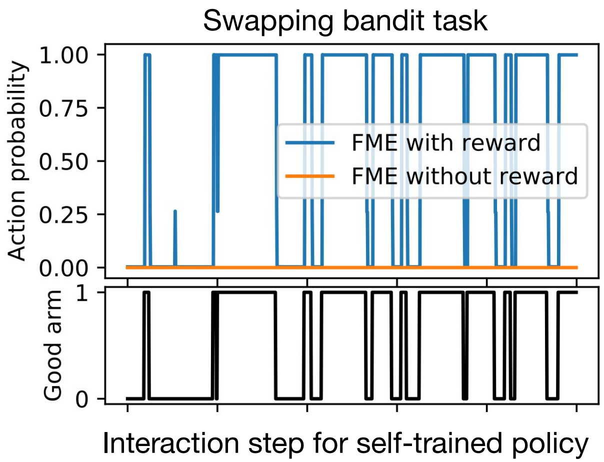


Eliminating Meta Optimization Through Self-Referential Meta Learning, Kirsch & Schmidhuber 2022
Chapter 7, 'Automating AI Research' PhD Thesis, Kirsch 2025

Experiments





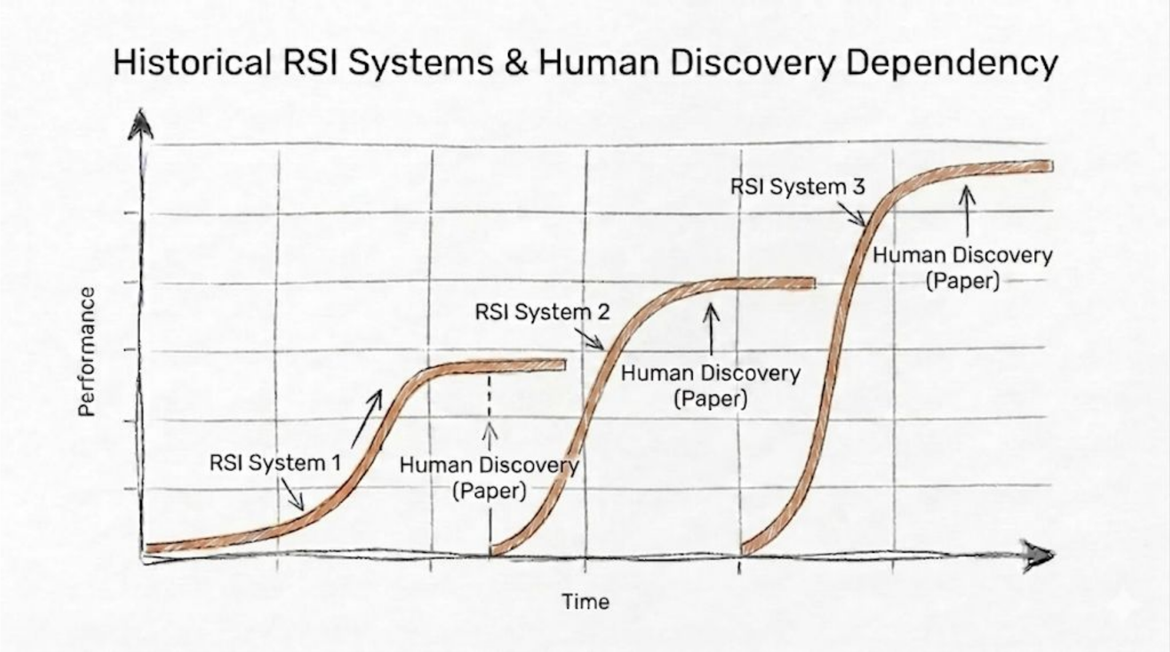


- Self-modifications lead to an effective policy for a given task
- Self-modifications adapt future modifications improving over fixed hill-climbing modifications
- Feeding the reward as input allows learning how to learn

Eliminating Meta Optimization Through Self-Referential Meta Learning, Kirsch & Schmidhuber 2022
Chapter 7, 'Automating AI Research' PhD Thesis, Kirsch 2025

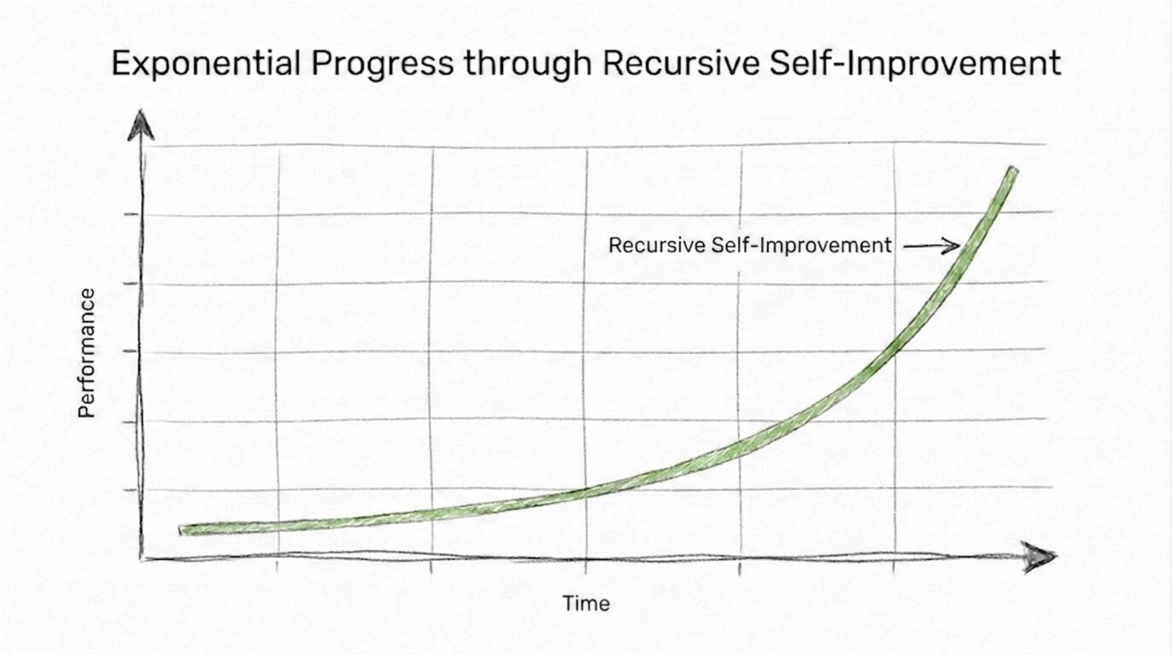
Thesis: Escape velocity

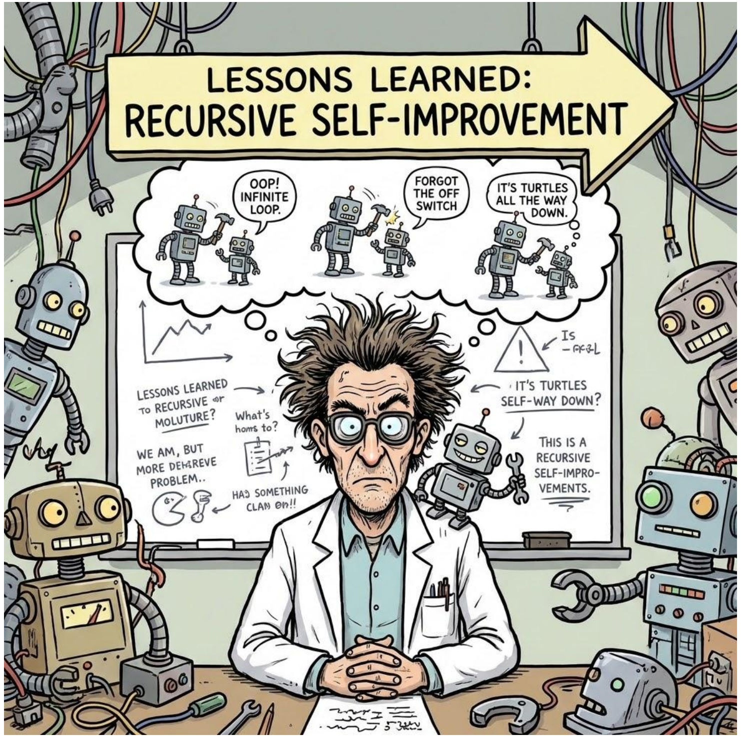
Current RSI systems still prematurely stop improving. Human interventions required.



Thesis: Escape velocity

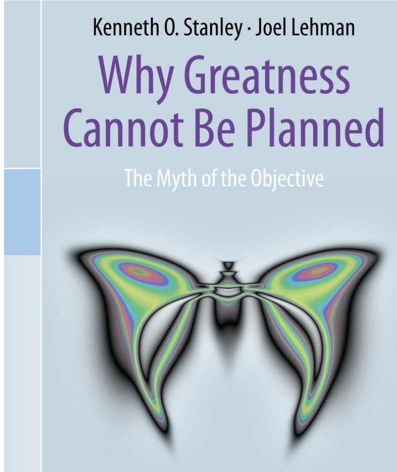
Inflection point where humans are no longer *needed* to deliver progress.

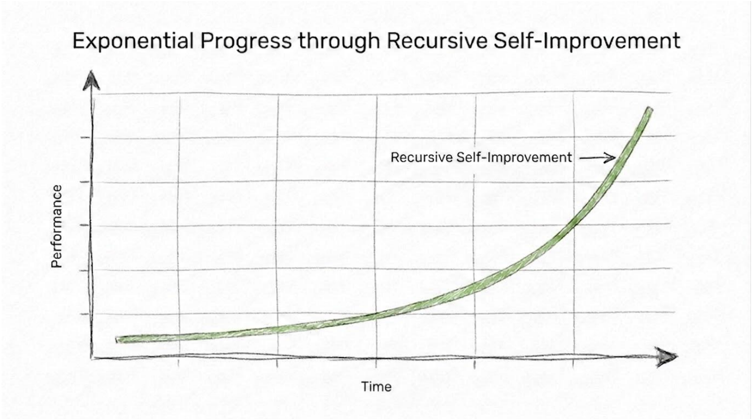




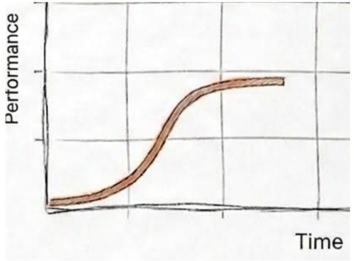
Lessons Learned

Stepping stones required

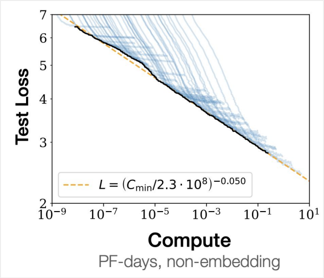




?

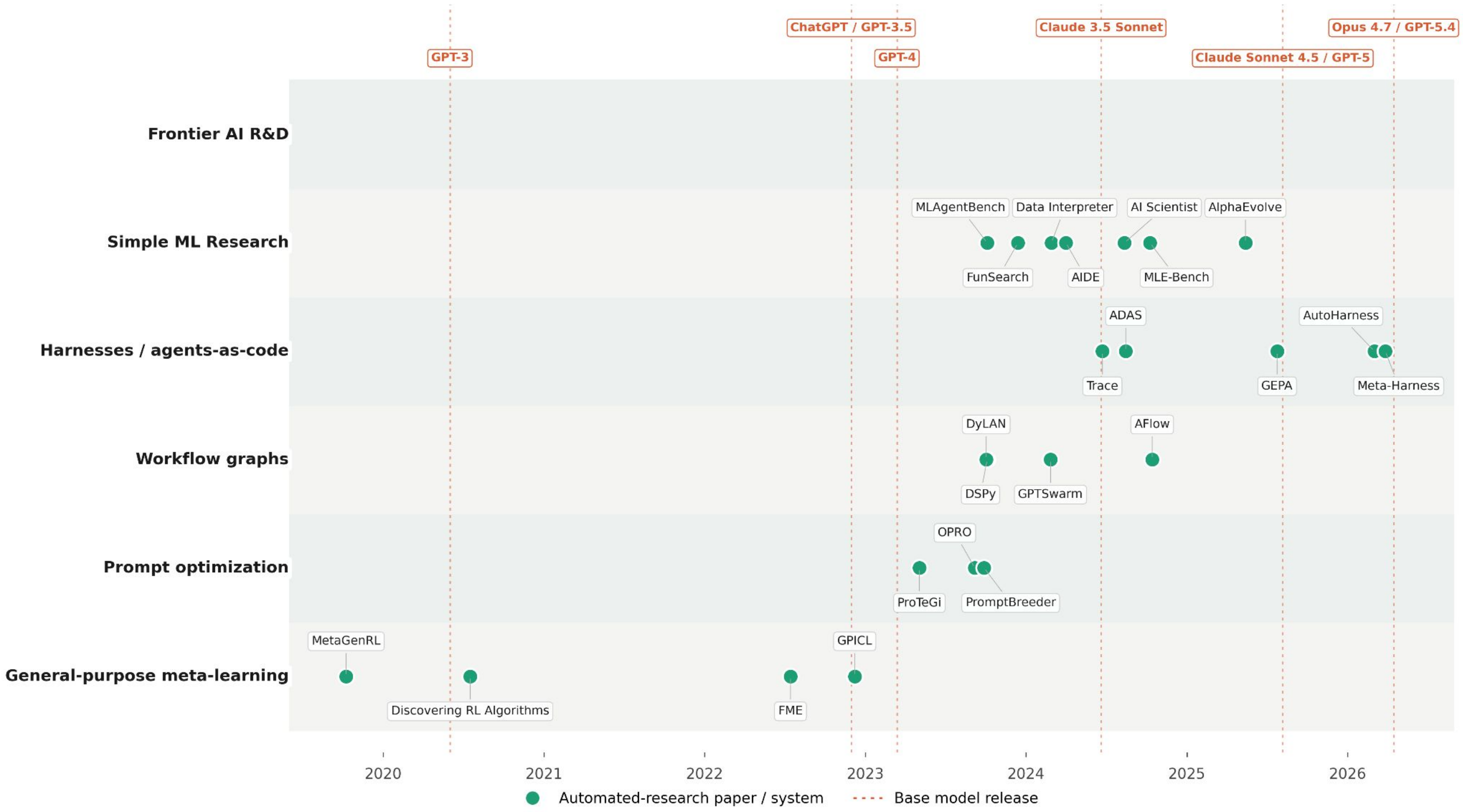


Stepping Stone

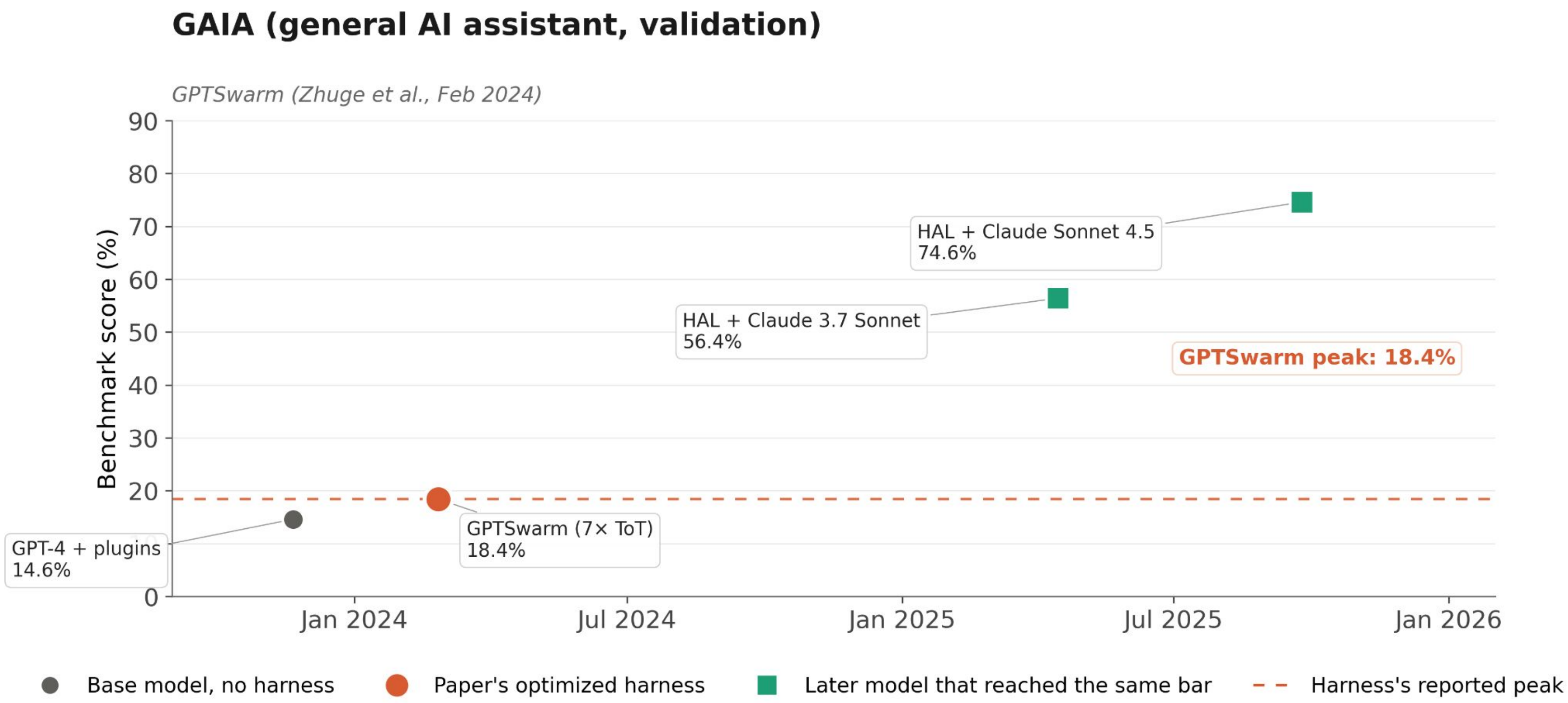


LLMs

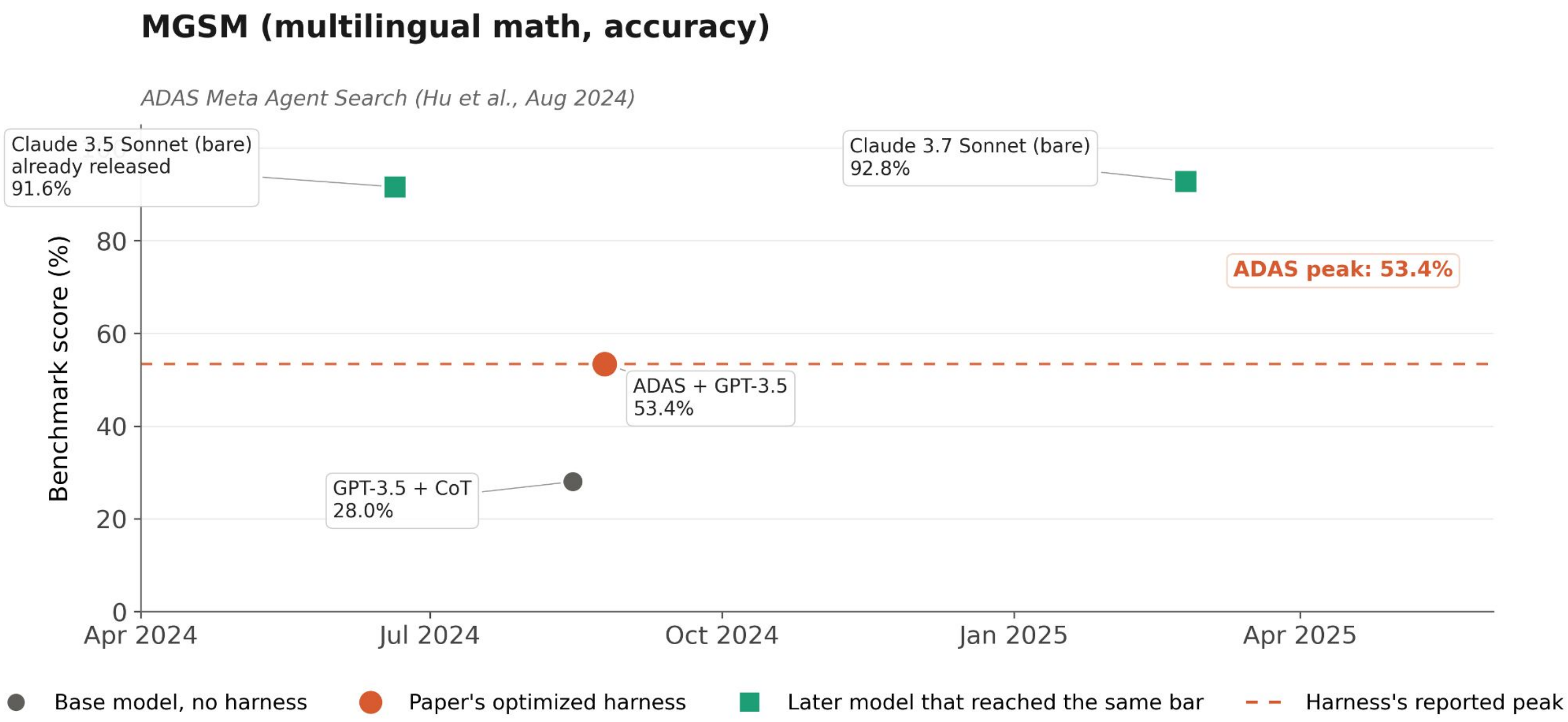
Base models matter a lot



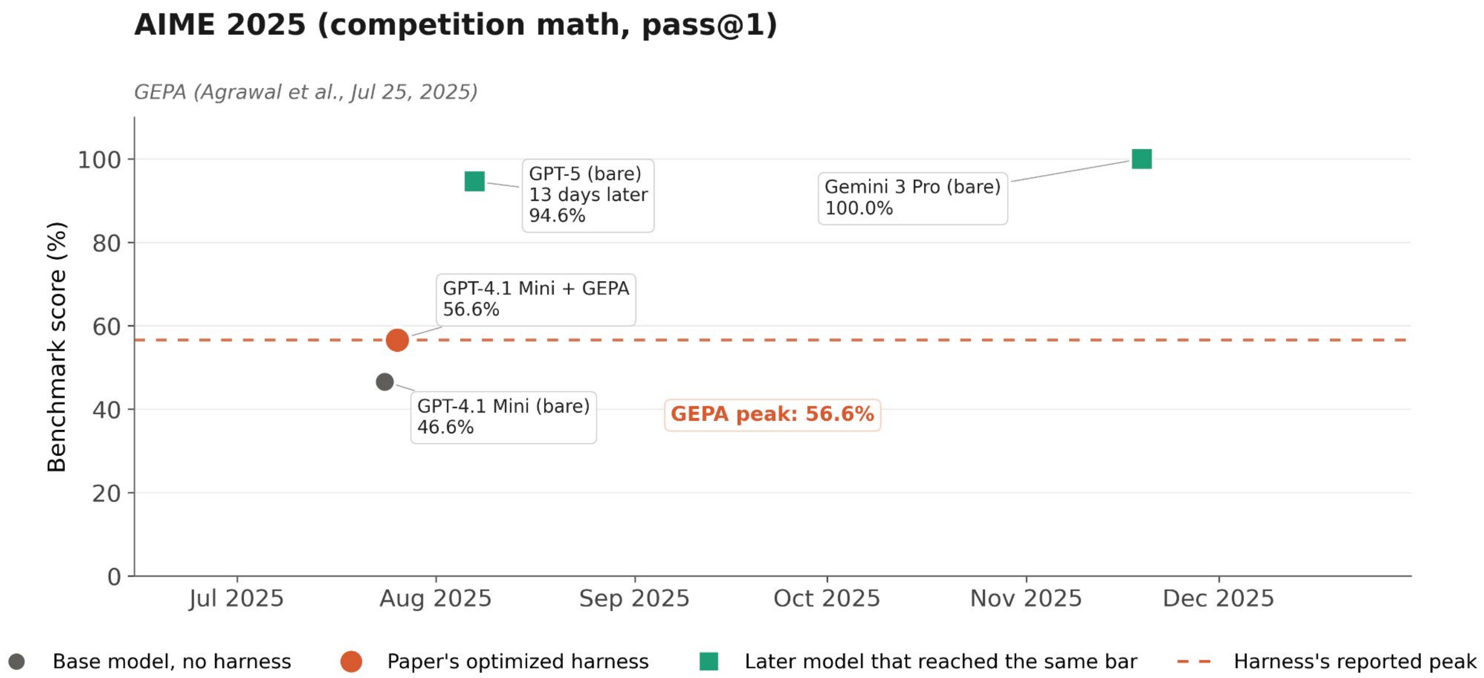
Agent harnesses are not enough



Agent harnesses are not enough

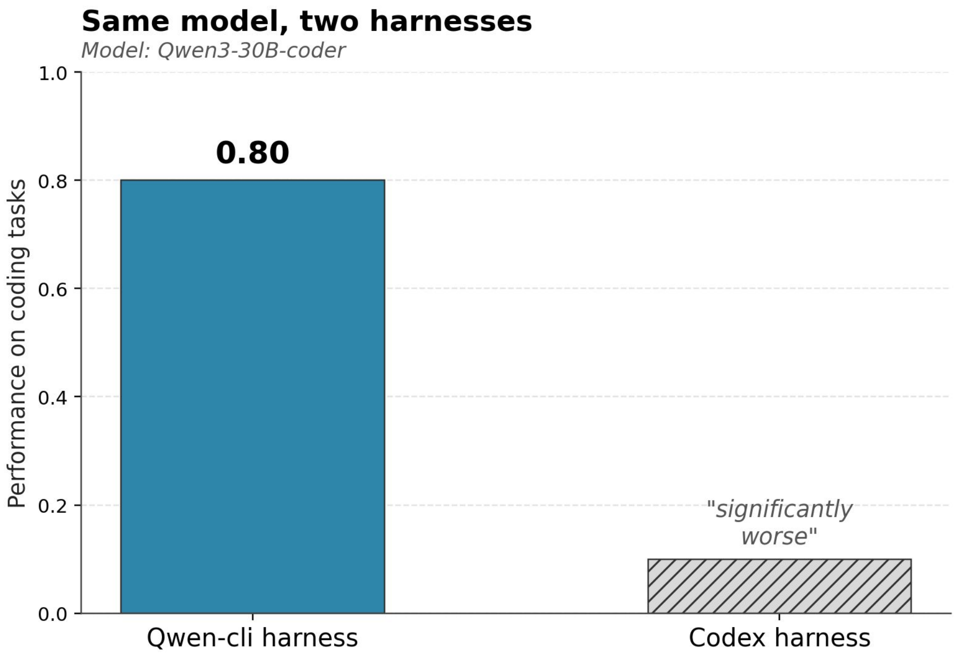


Agent harnesses are not enough

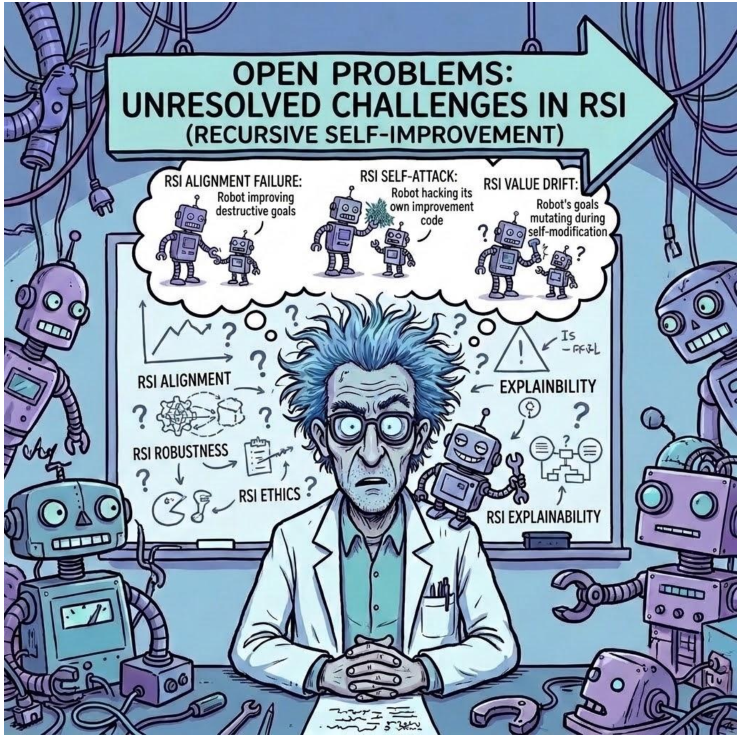


But harness may be used as an **improvement operator**, e.g. EvoTune (Surina et al 2025)

Harness-Model Co-evolution



Related blog: https://addyosmani.com/blog/agent-harness-engineering/



Open Problems

How to measure RSI?

As we’ve seen in various cases of this talk, RSI comes in many different forms.

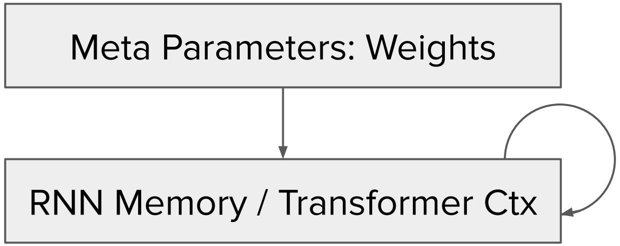
So far we have focused on a *whitebox* analysis.

Should we be more bitter lesson pilled?

How would we look for signatures of emergent RSI,
solely looking at its behavior *blackbox*?

How to measure RSI?

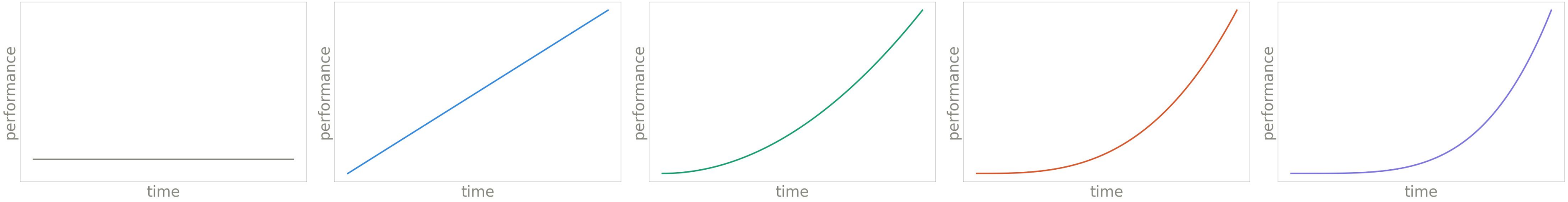
With a pure whitebox analysis, even an RNN can in principle self-improve arbitrarily.



Reference Turing Machine

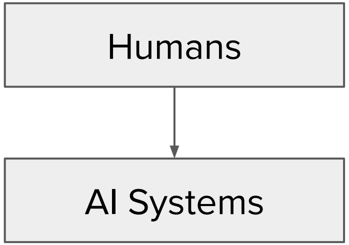
How to measure RSI?

Zeroth order	First order	Second order	Third order	...
Performance / Set of current knowledge	Process of Innovation	Process of improving the process of innovation (meta-learning)	Meta-meta learning	...

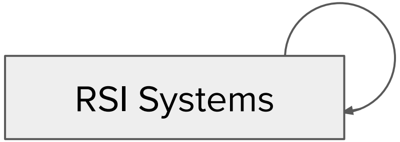


Human 🤖 Teaming
Reinserting humans in the loop

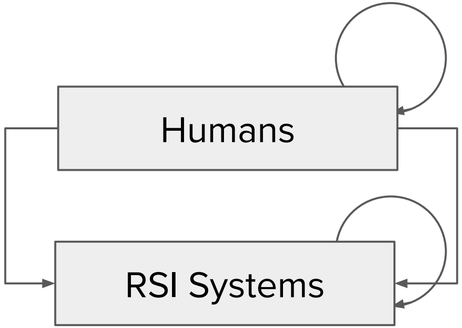
Human - Agent co-evolved system as a joint RSI system



Currently humans drive AI research



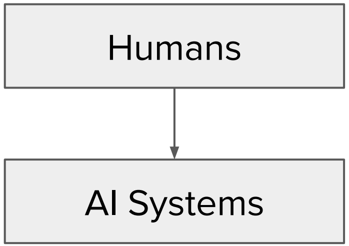
This could be the world we are moving to

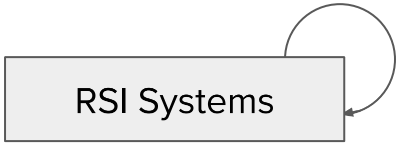


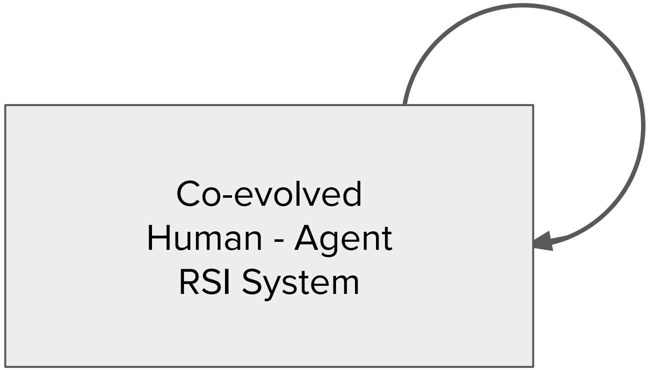
This is where we should be moving to

Human 🤖 Teaming
Reinserting humans in the loop

Human - Agent co-evolved system as a joint RSI system







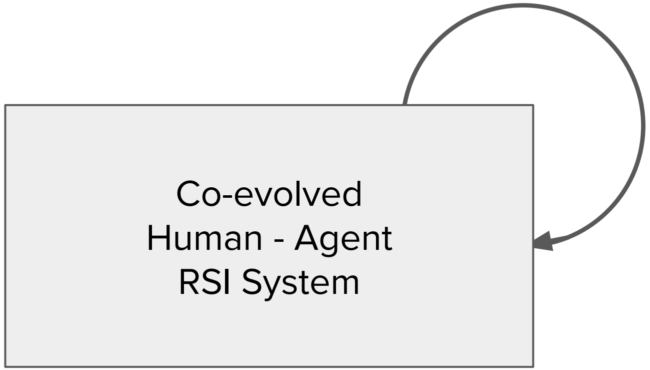
Currently humans drive AI research

This could be the world we are moving to

This is where we should be moving to

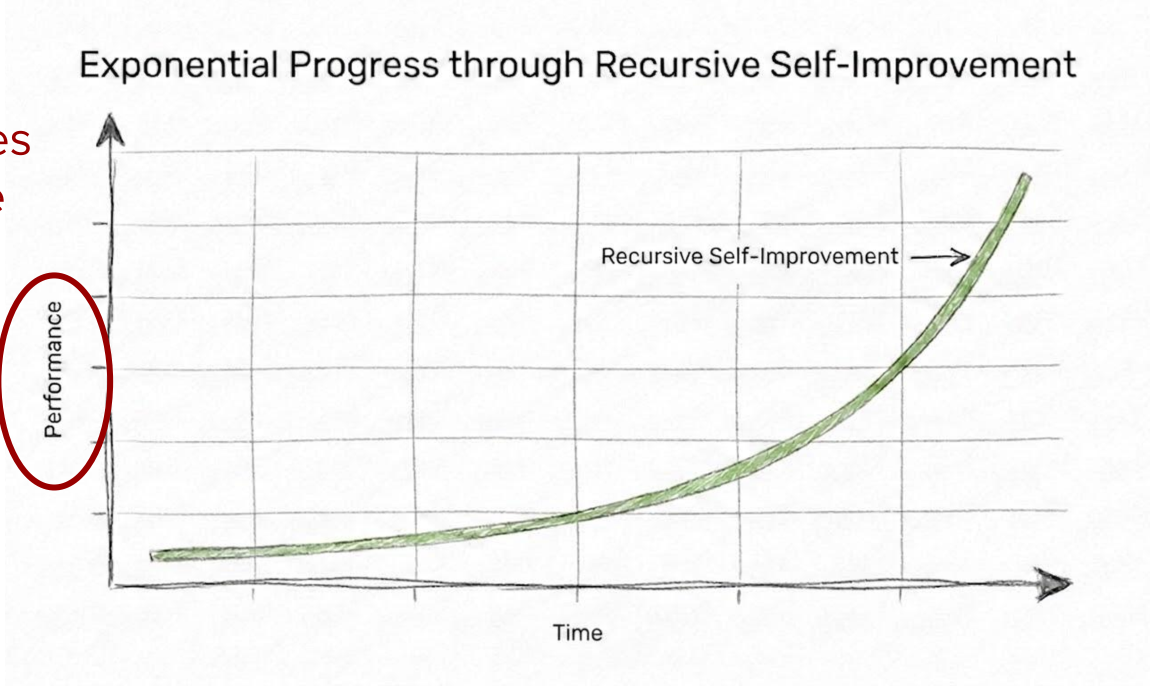
Safety implications

- Humans being ‘part’ of the RSI system
- Monitoring & proactive information sharing
- Automated safety research

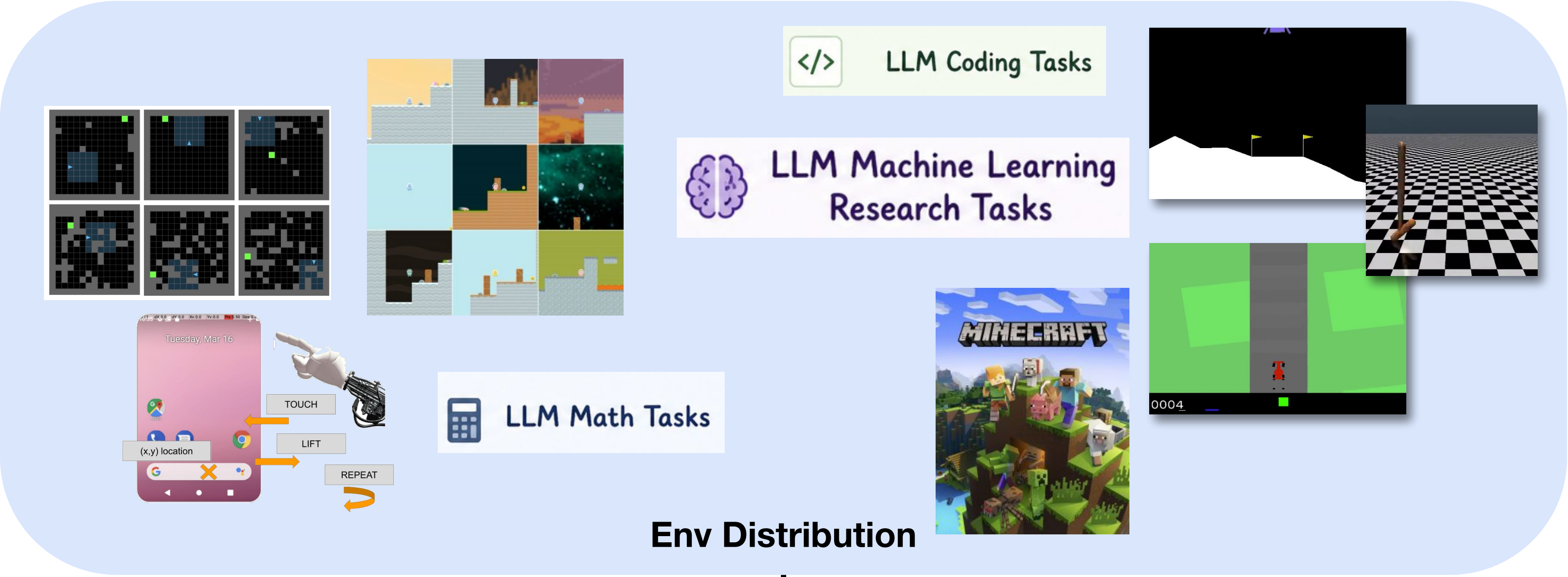


What is the reward (e.g. in FME)?

Where does this come from?



Self-generating new tasks



→ General-Purpose RSI System

Artificial Curiosity

Adversarial Curiosity / proto-GAN: The policy's intrinsic reward is simply the model's prediction error.

Still fails on stochastic environments (the "noisy TV problem").

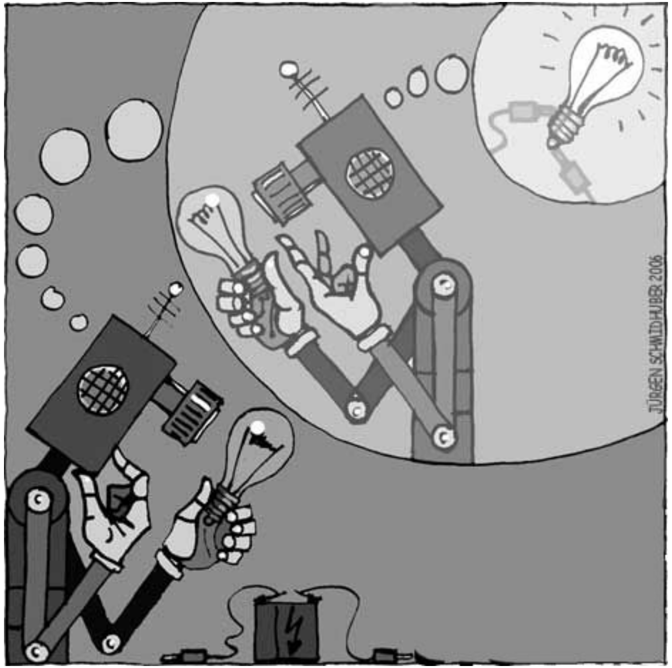
Making the world differentiable, Schmidhuber 1990
A possibility for implementing curiosity and boredom in model-building neural controllers, Schmidhuber 1991

Information Gain

Optimize for learning progress

$$r_{\text{int}}(Q) \;\propto\; \sum_i p_{M^*}(Q_i \mid H) \log \frac{p_{M^*}(Q_i \mid H)}{p_M(Q_i \mid H)}$$

$$\max_\pi \; \mathbb{E}_\pi \left[\sum_t D_{\mathrm{KL}}\big(p_{M_t^*}(\cdot \mid H_t) \,\|\, p_{M_t}(\cdot \mid H_t)\big) \right]$$

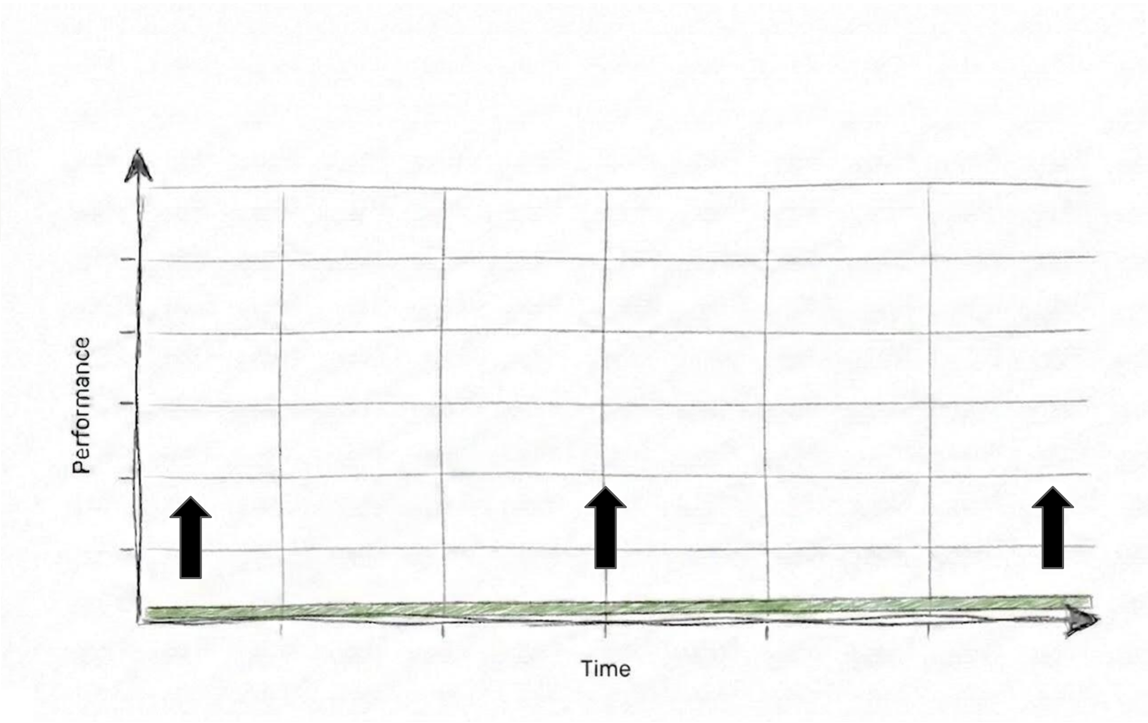


Curious model-building control systems, Schmidhuber 1991
Reinforcement-driven information acquisition in non-deterministic environments, Hochreiter & Schmidhuber 1995

How to optimize directly for RSI?

$$J_{\text{final}}(\pi) = \mathbb{E}_{\tau\sim\pi}\big[\,r(s_L)\,\big]$$

$$J(\pi) = \mathbb{E}_{\tau\sim\pi}\left[\sum_{l=0}^{L} r(s_l)\right]$$

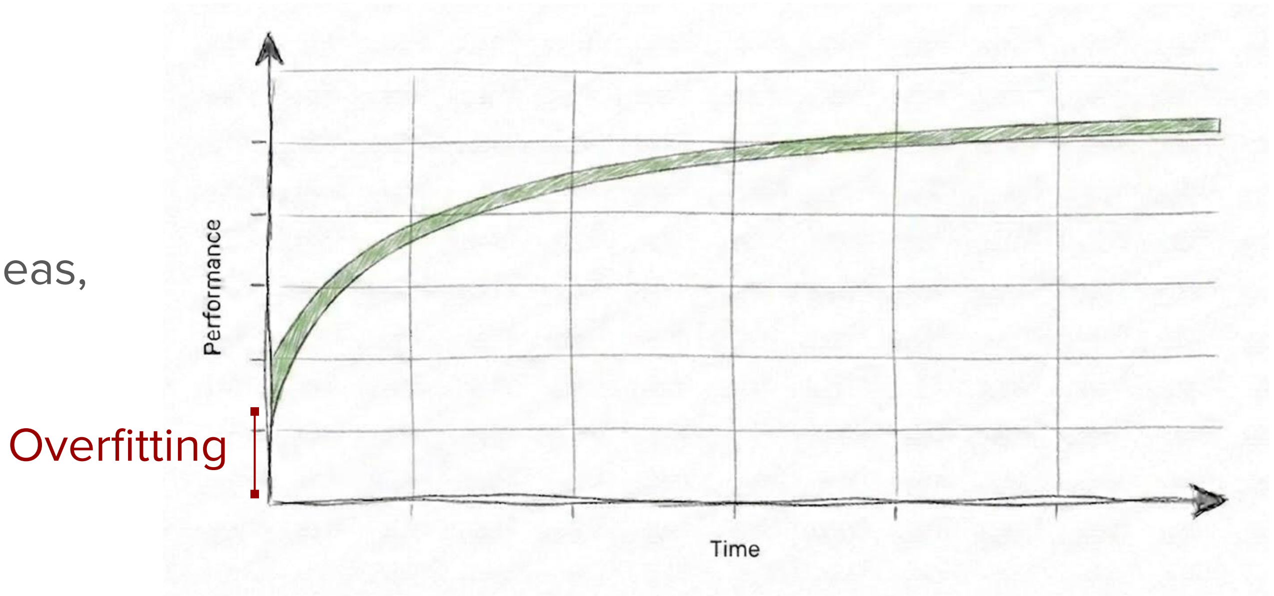


How to optimize explicitly for RSI?

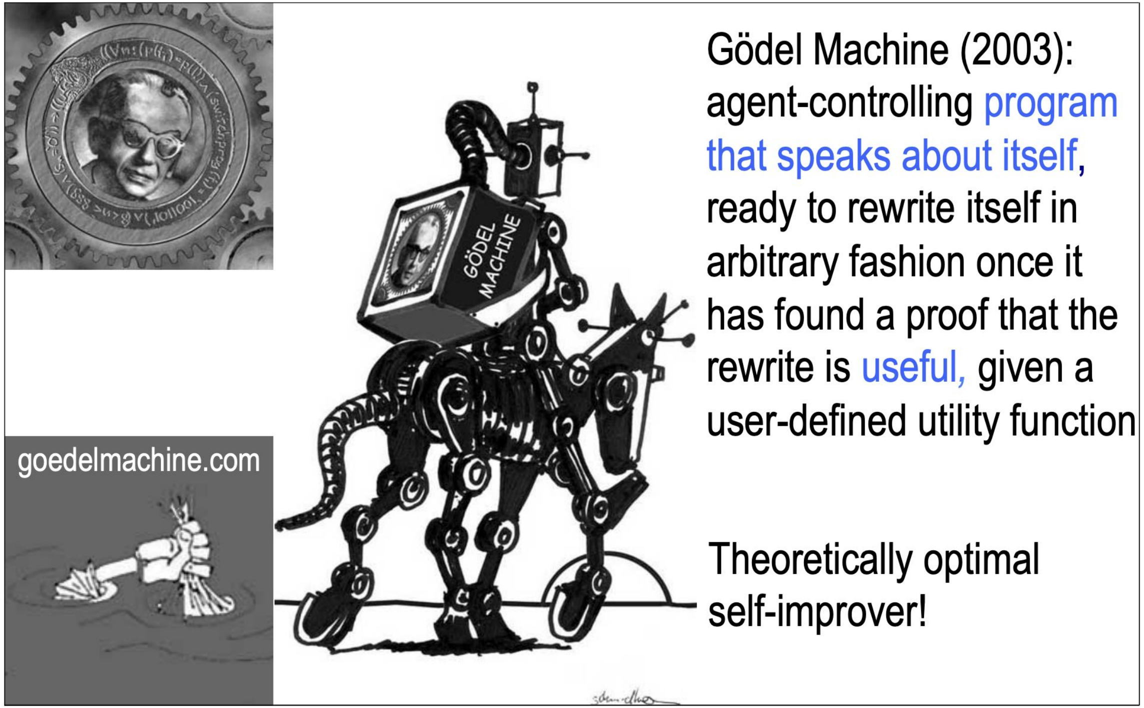
Why does this even lead to the auto-discovery of an improvement operator?
Only implicitly, due to

- Inductive bias, or
- a priori unknowables in the environment.

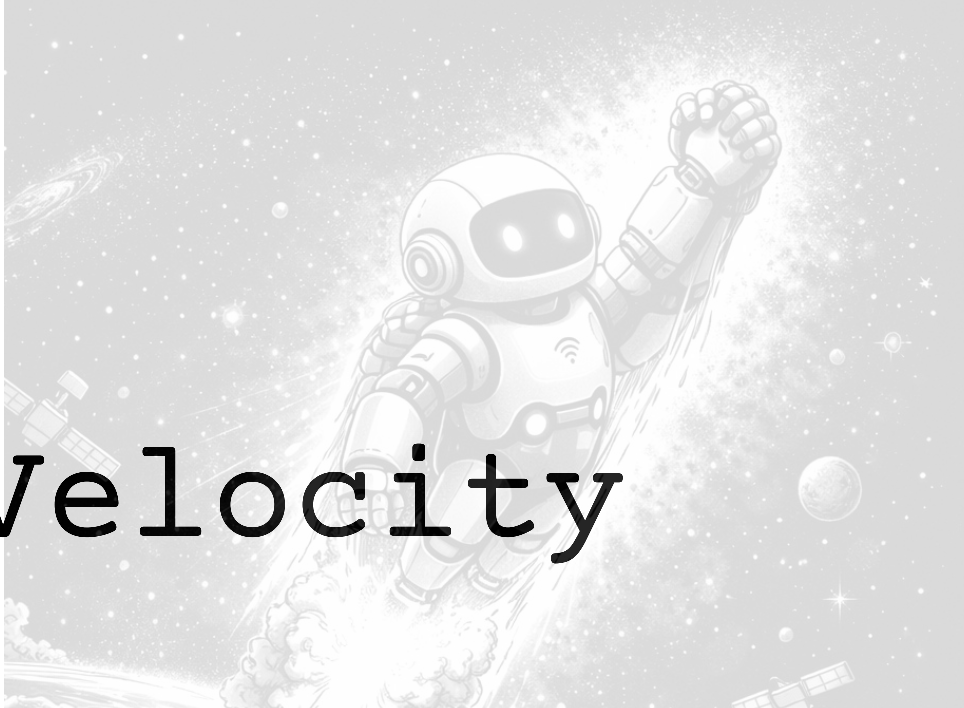
I have some alternative ideas, but the verdict is still out.



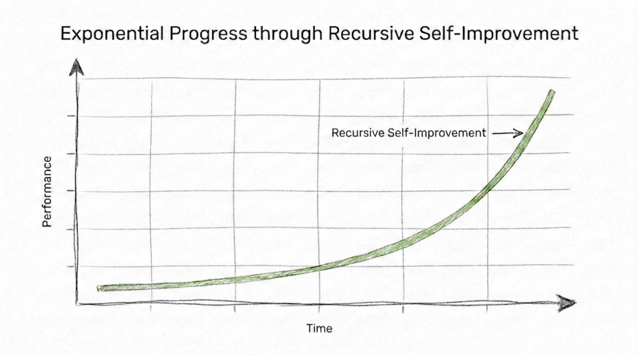
Provable recursive self-improvement?



Gödel Machines, Schmidhuber 2007



Escape Velocity



@LouisKirschAI