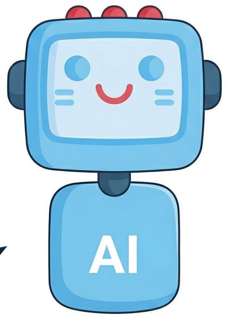
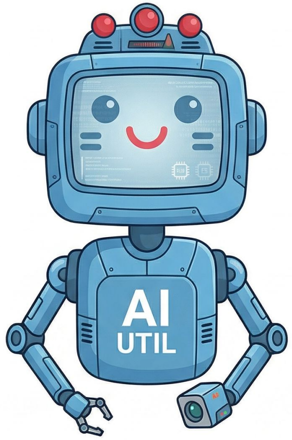
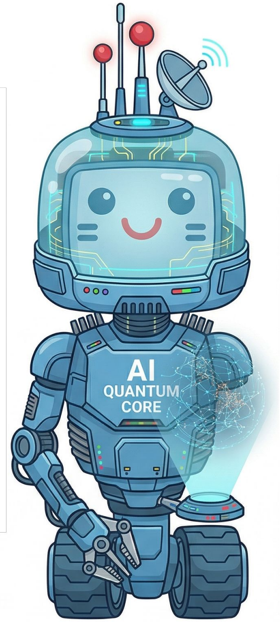
人工智慧能否自我成長 (下集)

人工智慧能否自我成長

人類最後的發明

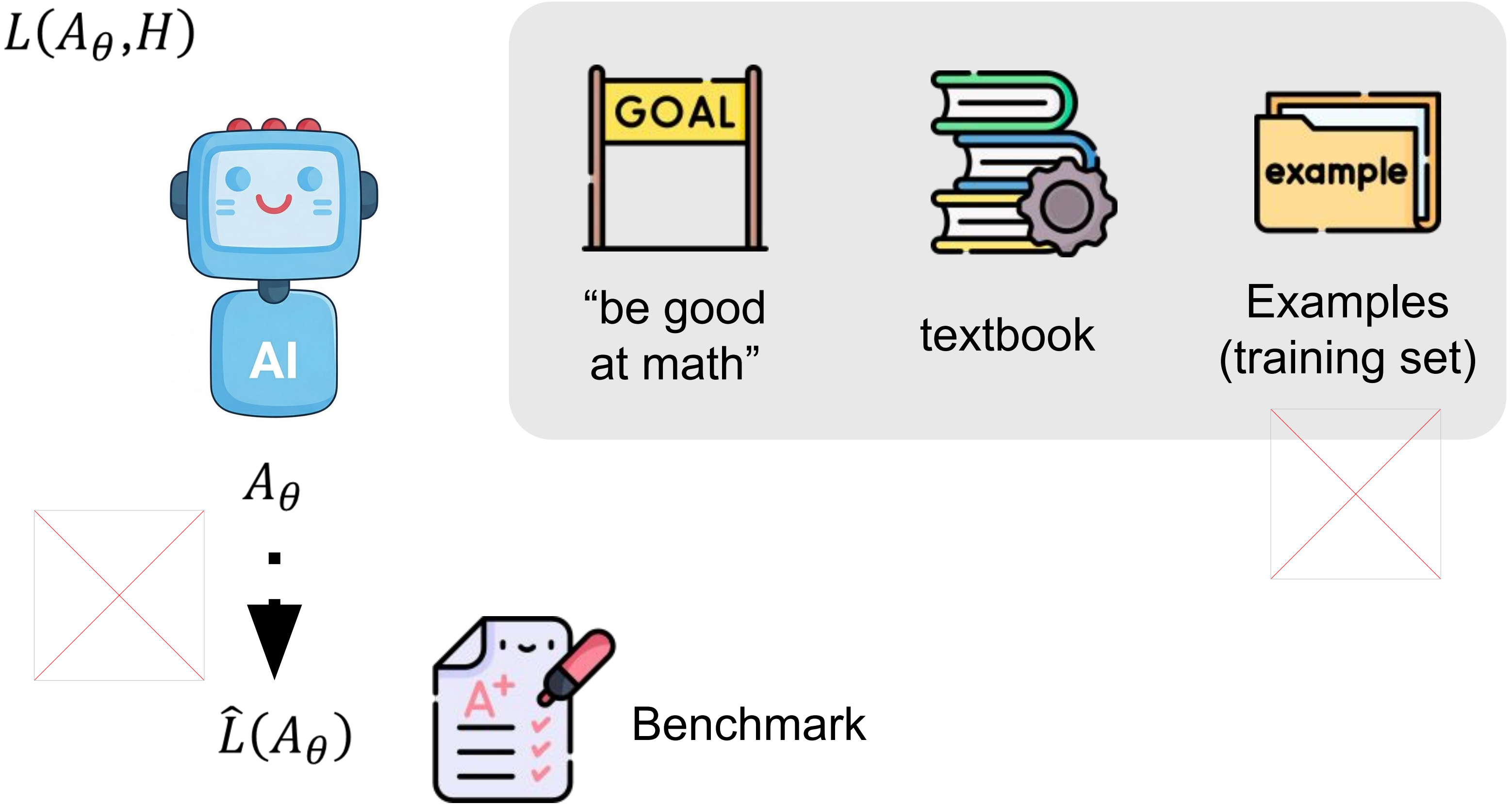


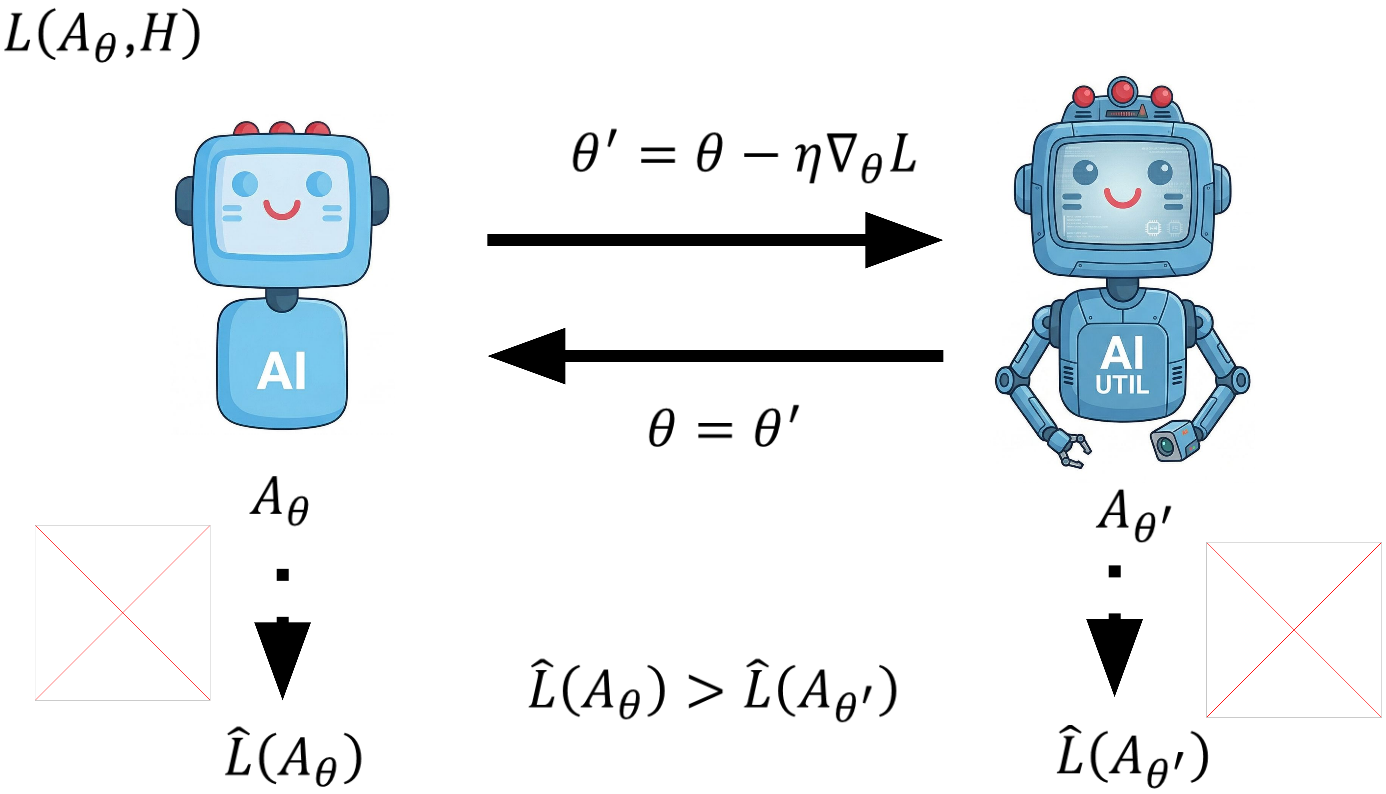




……

"technological singularity"


$L(A_\theta, H)$
GOAL
"be good at math"
textbook
example
Examples (training set)
AI
A_θ
$\hat{L}(A_\theta)$
Benchmark


$L(A_\theta, H)$
$\theta' = \theta - \eta \nabla_\theta L$
AI
AI
UTIL
$\theta = \theta'$
A_θ
$A_{\theta'}$
$\hat{L}(A_\theta) > \hat{L}(A_{\theta'})$
$\hat{L}(A_\theta)$
$\hat{L}(A_{\theta'})$

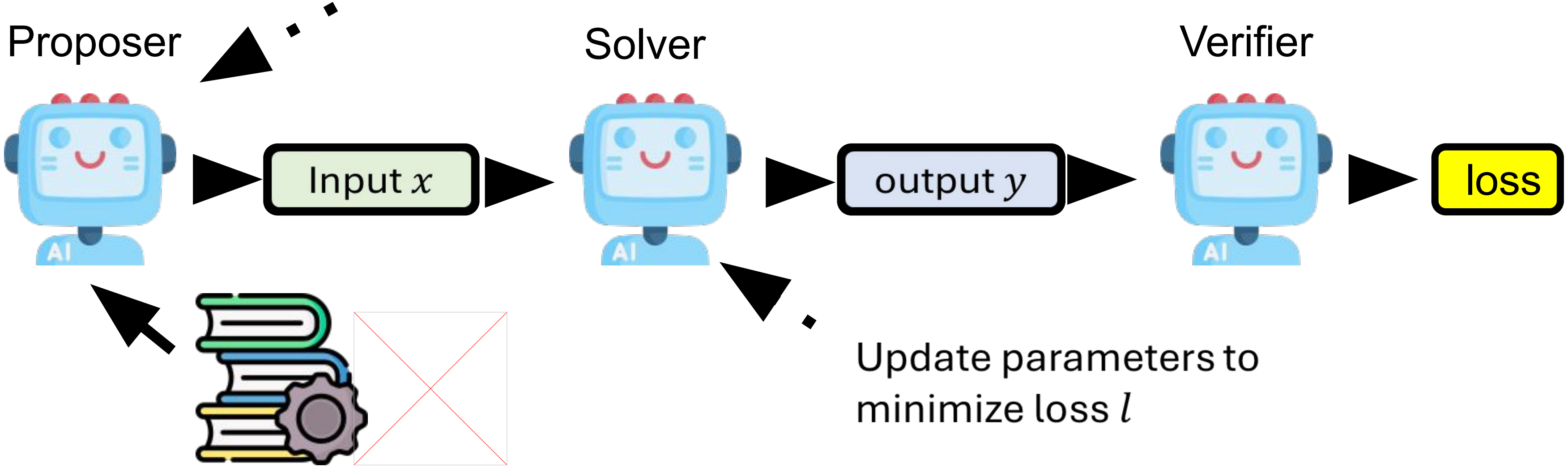
Almost No Human in the Loop

Absolute Zero: https://arxiv.org/abs/2505.03335

R-Zero: https://arxiv.org/abs/2508.05004

Self-Questioning: https://arxiv.org/abs/2508.03682

Update parameters to minimize loss l'



SPICE: https://arxiv.org/abs/2510.24684

R-Few: https://arxiv.org/abs/2512.02472

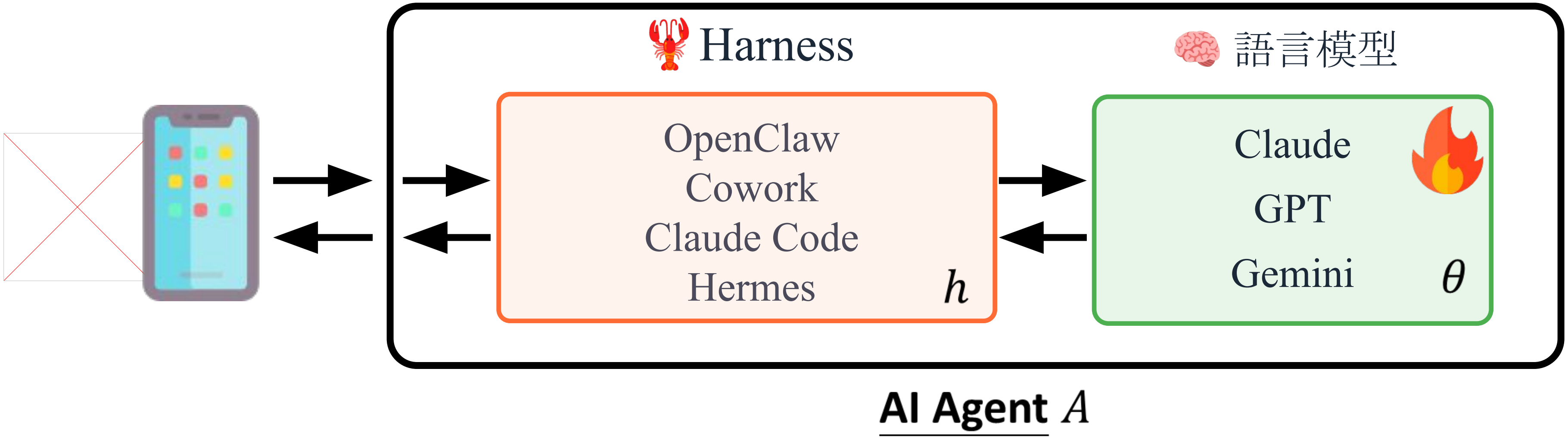
System Message:
You are an expert competition-math problem setter. FIRST, in your private scratch-pad, think step-by-step to design a brand-new, non-trivial problem. The problem could come from any field of mathematics, including but not limited to algebra, geometry, number theory, combinatorics, prealgebra, probability, statistics, and calculus. Aim for a difficulty such that fewer than 30% of advanced high-school students could solve it. Avoid re-using textbook clichés or famous contest problems.

R-Zero: https://arxiv.org/abs/2508.05004

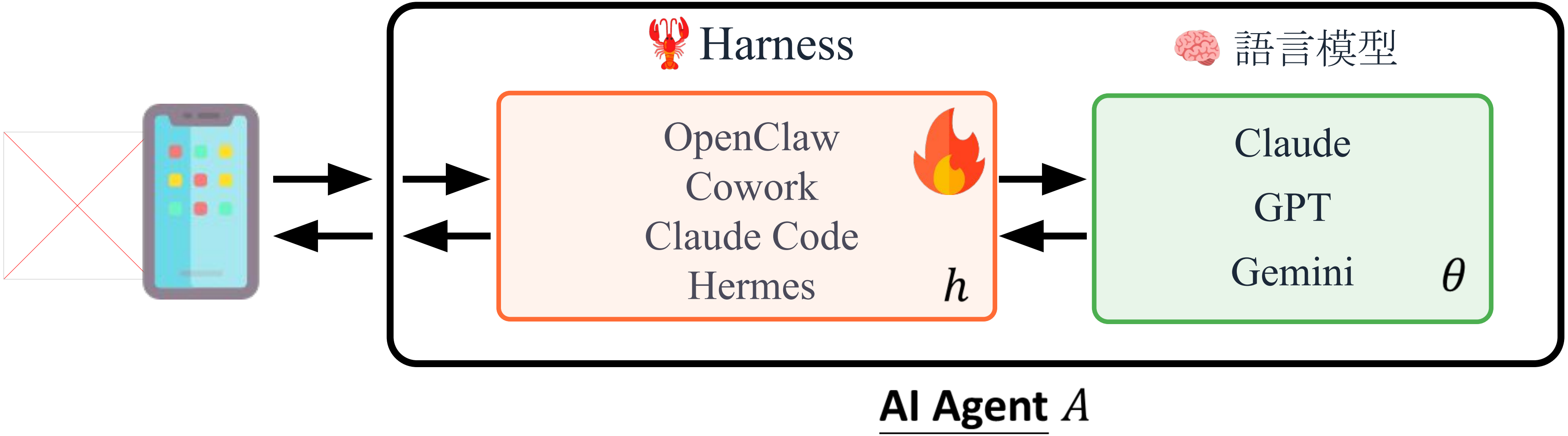
```
## Task: Create a Python Code Snippet (where custom classes are allowed, which should be defined
↪  at the top of the code snippet) with one Matching Input

Using the reference code snippets provided below as examples, design a new and unique Python code
↪  snippet that demands deep algorithmic reasoning to deduce one possible input from a given
↪  output. Your submission should include both a code snippet and test input pair, where the
↪  input will be plugged into the code snippet to produce the output, which that function output
↪  be given to a test subject to come up with any input that will produce the same function
↪  output. This is meant to be an I.Q. test.
```

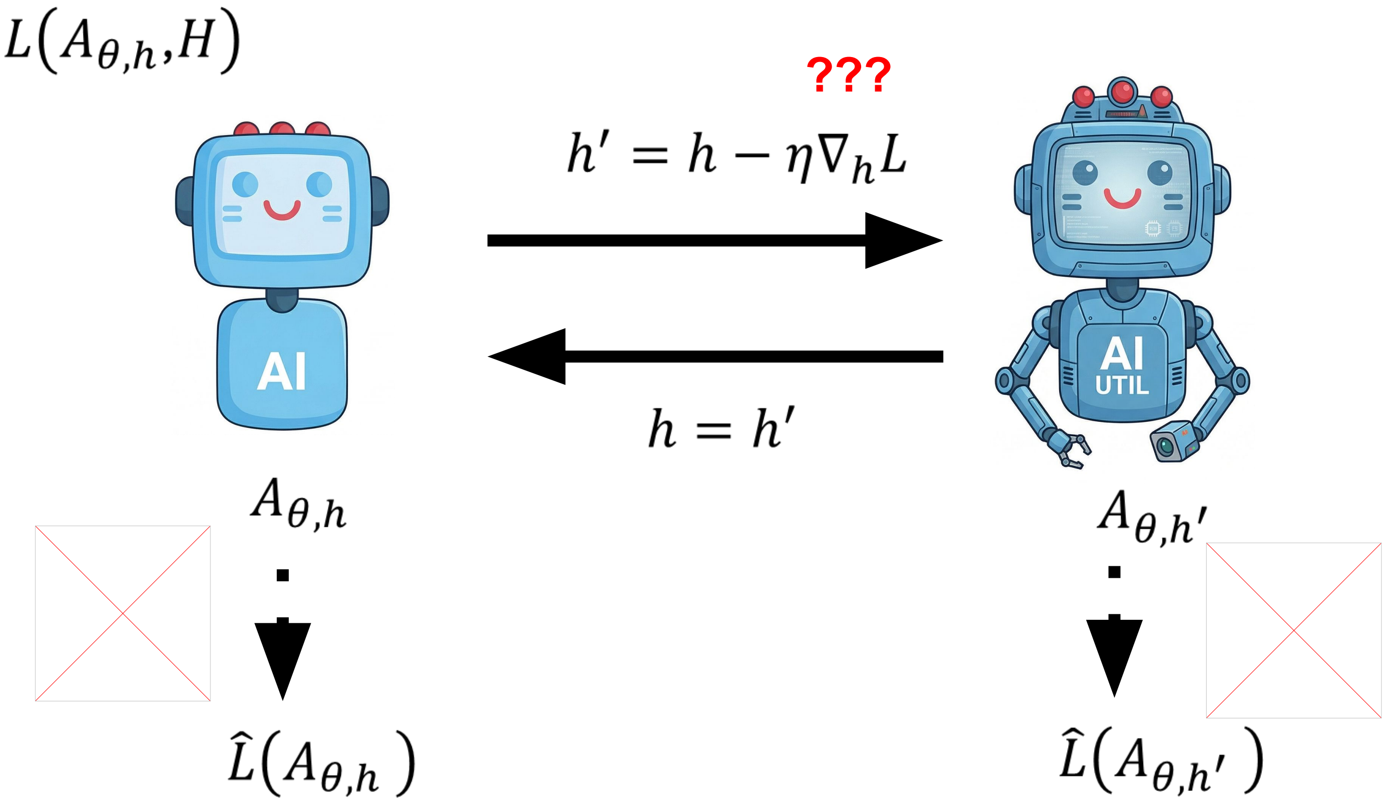
Absolute Zero:
https://arxiv.org/abs/2505.03335

AI Agent = Harness + LLM



AI Agent = Harness + LLM



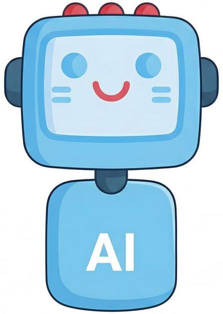


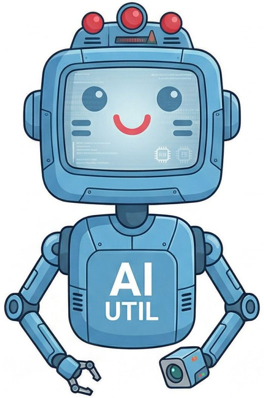
$L(A_{\theta,h}, H)$
???
$h' = h - \eta \nabla_h L$
AI
AI
UTIL
$h = h'$
$A_{\theta,h}$
$A_{\theta,h'}$
$\hat{L}(A_{\theta,h})$
$\hat{L}(A_{\theta,h'})$

How to improve harness?

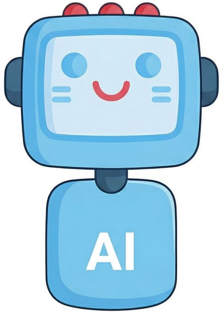
???

$$h' = h - \eta \nabla_h L$$





$h \quad L(A_{\theta,h}, H)$

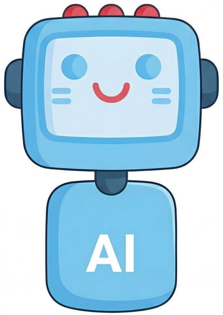


h'

h is prompt

Prompt Optimization

h $L(A_{\theta,h},H)$ $\rightarrow$



$\rightarrow$ h'

Think step by step. 72

Take a deep breath.

https://arxiv.org/abs/2309.03409

I have some texts along with their corresponding scores. The texts are arranged in ascending order based on their scores, where higher scores indicate better quality.

text:
Let's figure it out!
score:
61

text:
Let's solve the problem.
score:
63

(... more instructions and scores ...)

The following exemplars show how to a
text, then read the input and give an out
from the given output, and we say your

input:
Q: Alannah, Beatrix, and Queen are preparing for the new school year and have been given books by their parents. Alannah has 20 more books than Beatrix. Queen has 1/5 times more books than Alannah. If Beatrix has 30 books, how many books do the three have together?
A: <INS>
output:
140

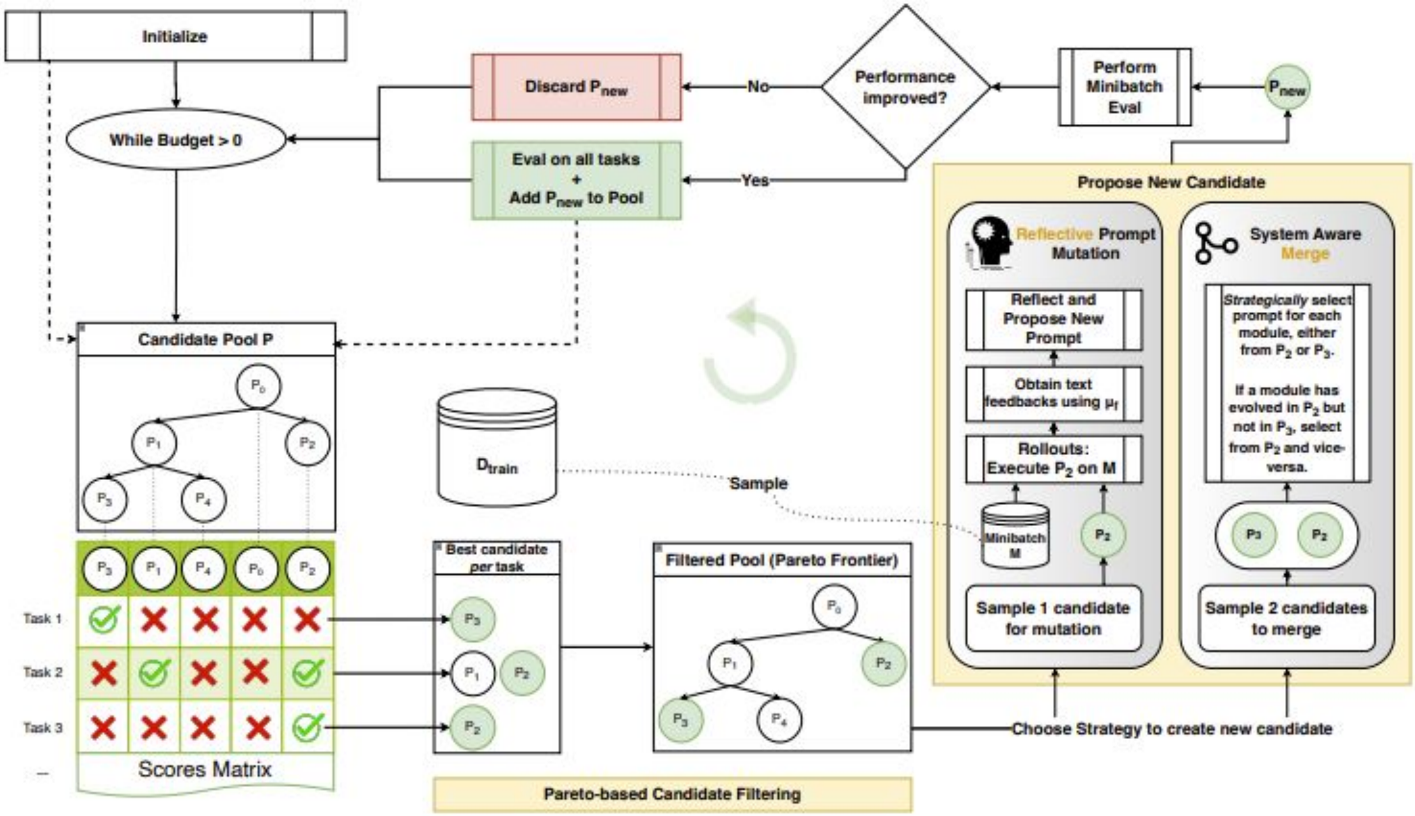
(... more exemplars ...)

Write your new text that is different from the old ones and has a score as high as possible. Write the text in square brackets.

https://arxiv.org/abs/2309.0340

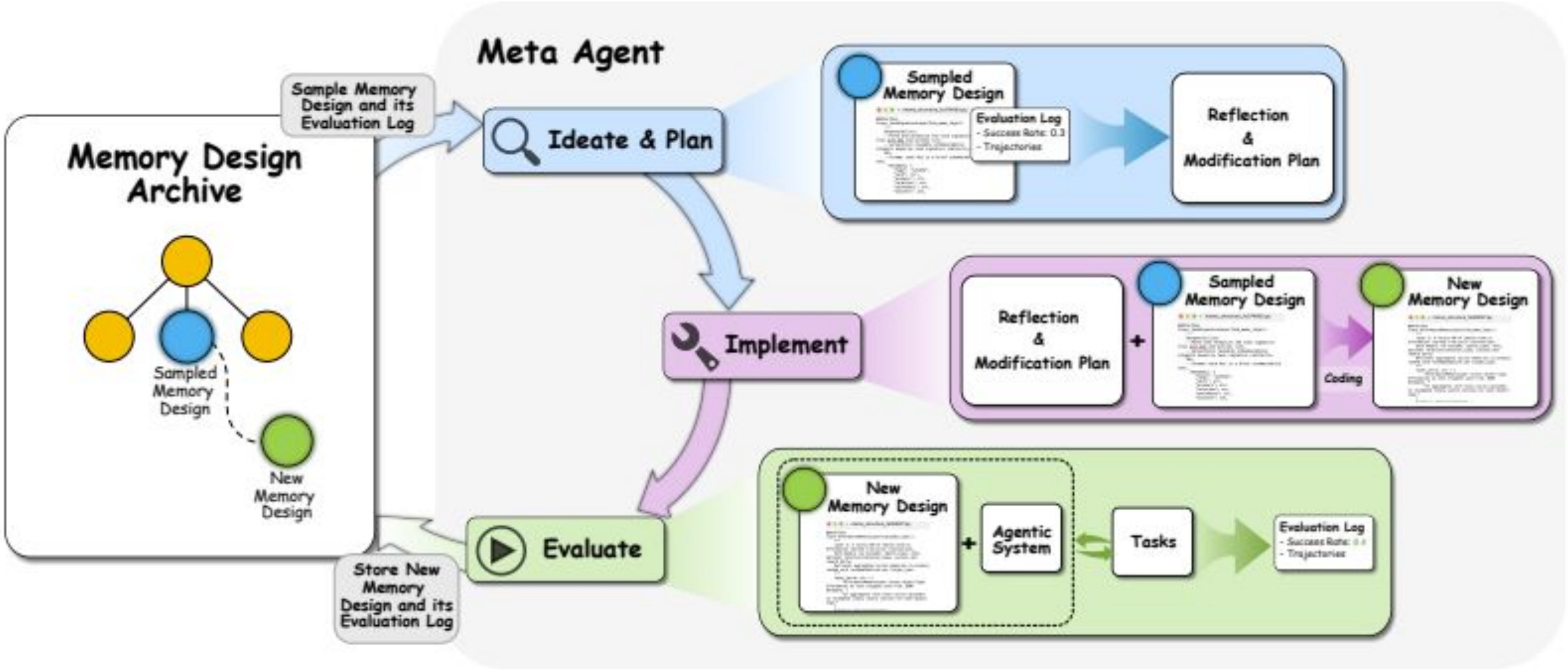
h is prompt

GEPA
https://arxiv.org/abs/2507.19457



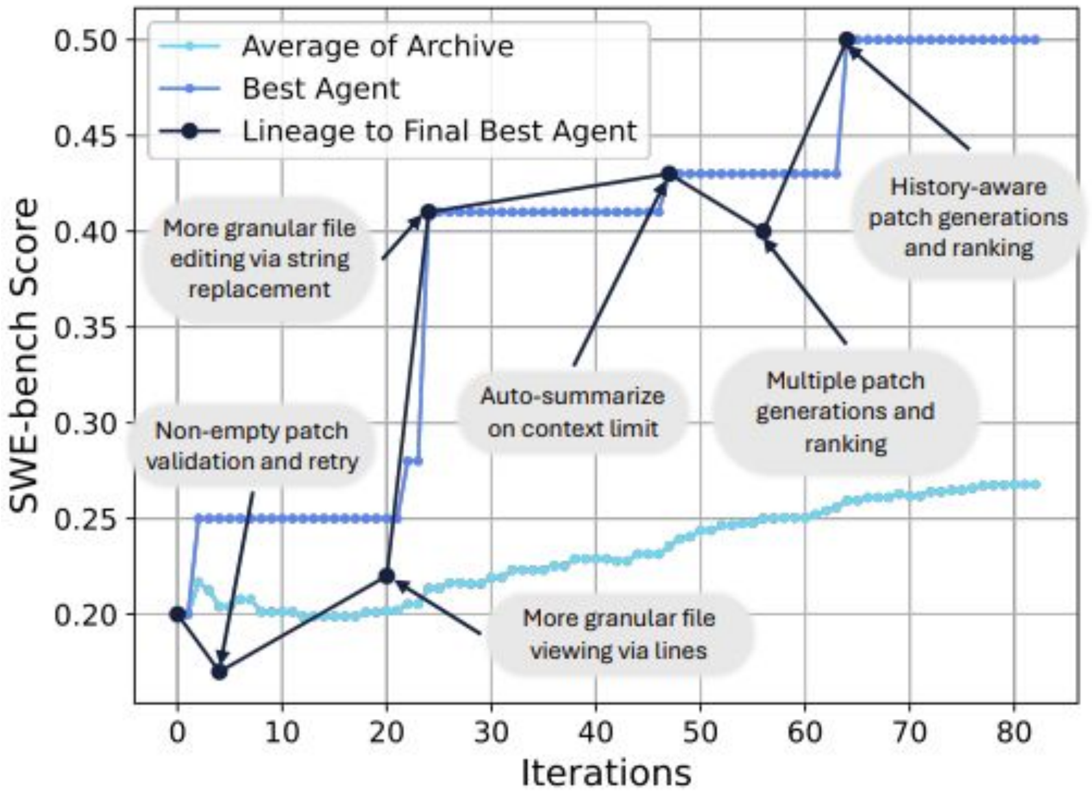
h is memory management

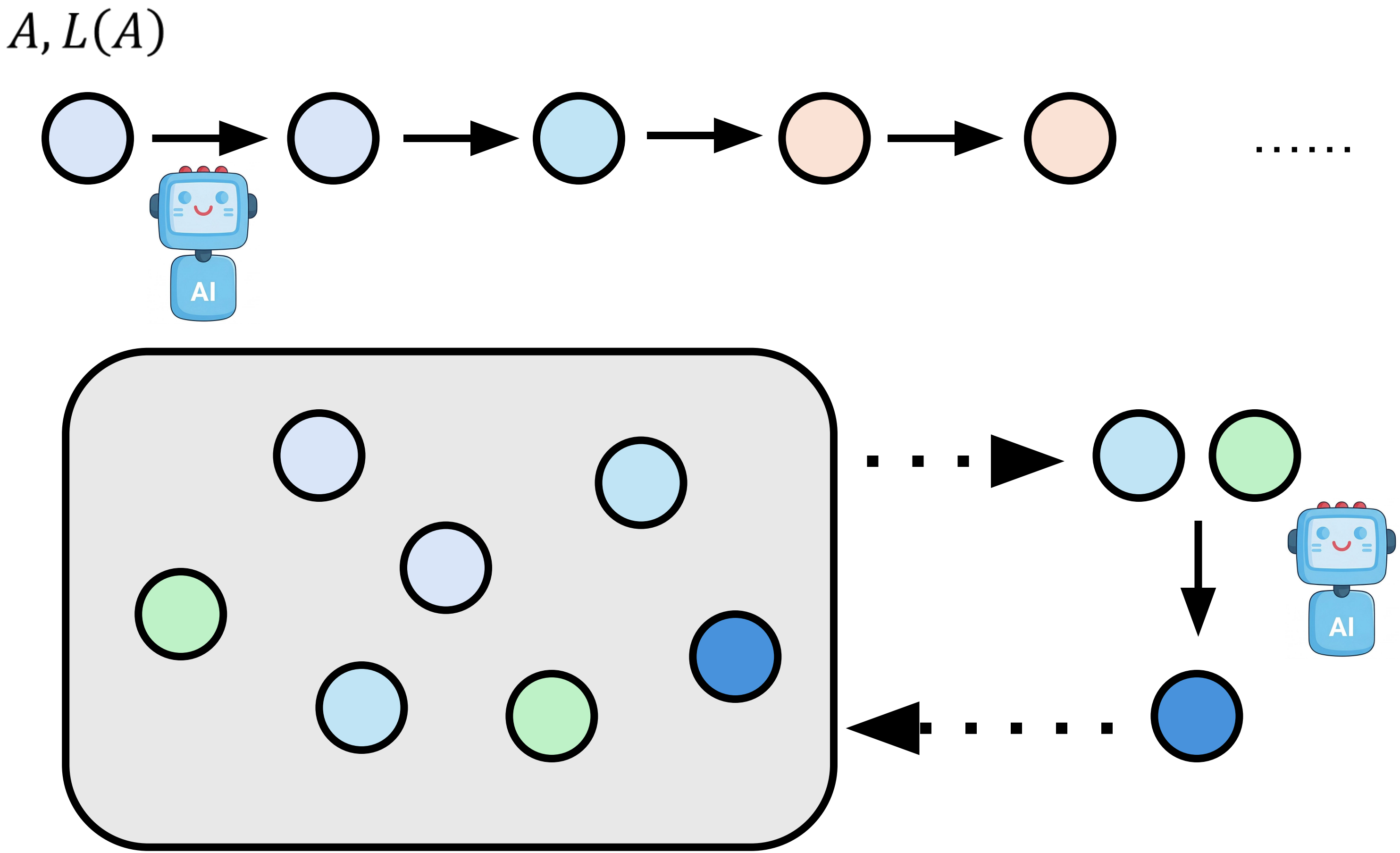
https://arxiv.org/abs/2602.07755

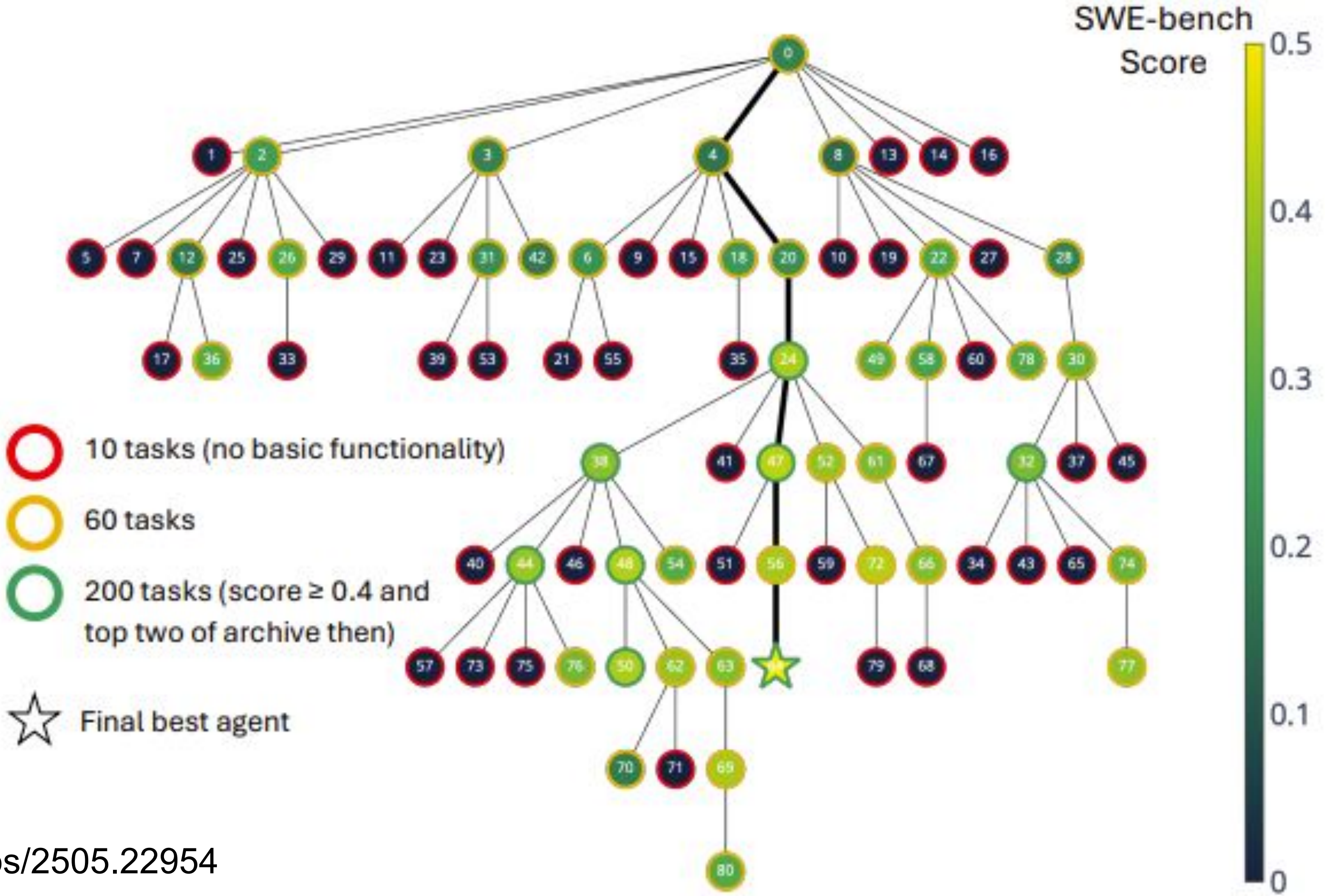


h is workflow

https://arxiv.org/abs/2505.22954




$A, L(A)$
AI
AI



https://arxiv.org/abs/2505.22954

Declarative Self-improving Language Programs in Python (DSPy)

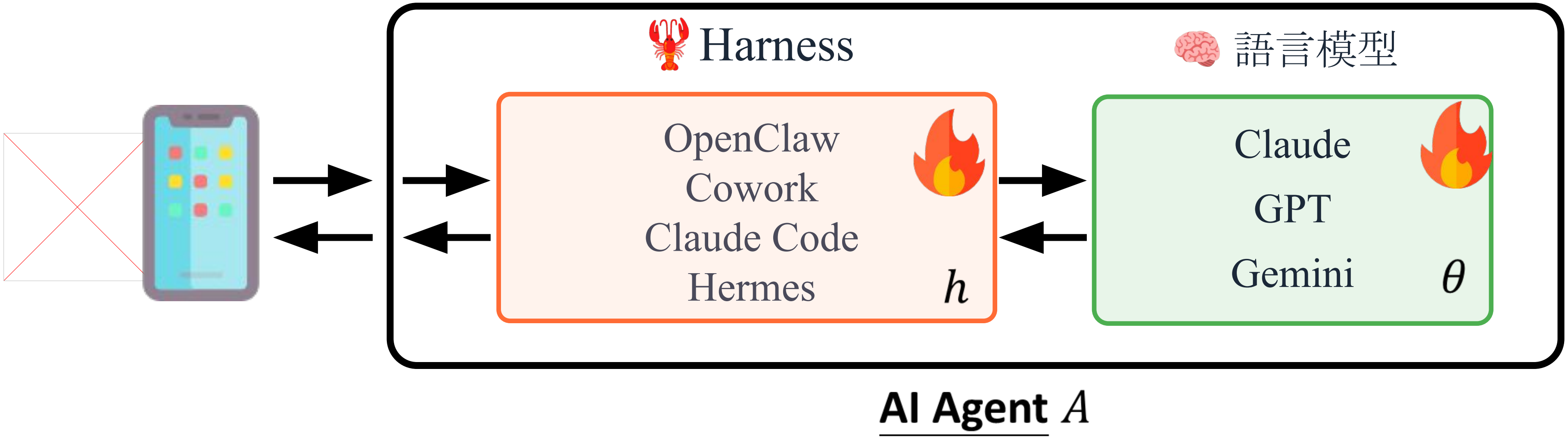
https://arxiv.org/abs/2310.03714

DSPy

DSPy: *Programming*—not prompting—Foundation Models

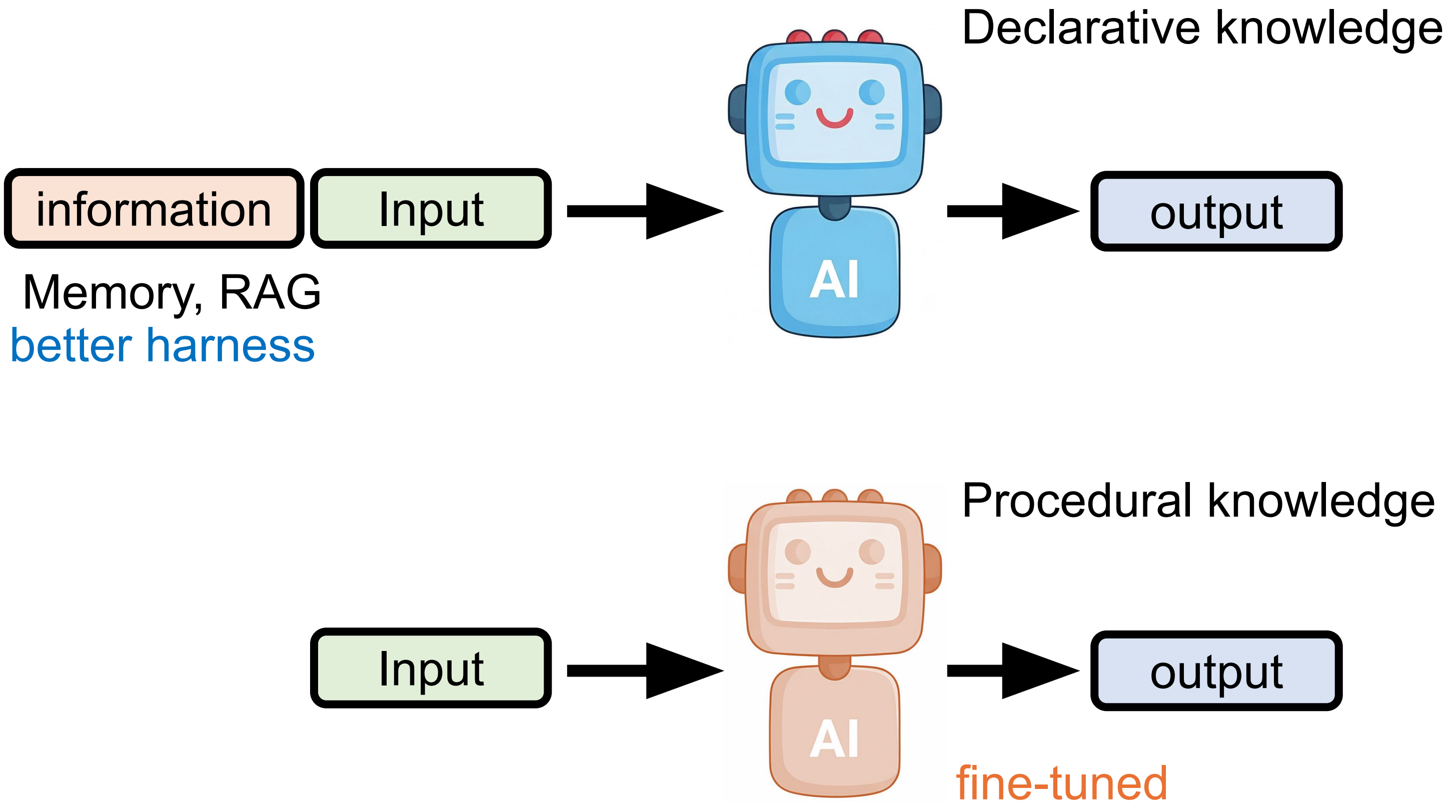
https://github.com/stanfordnlp/dspy

AI Agent = Harness + LLM



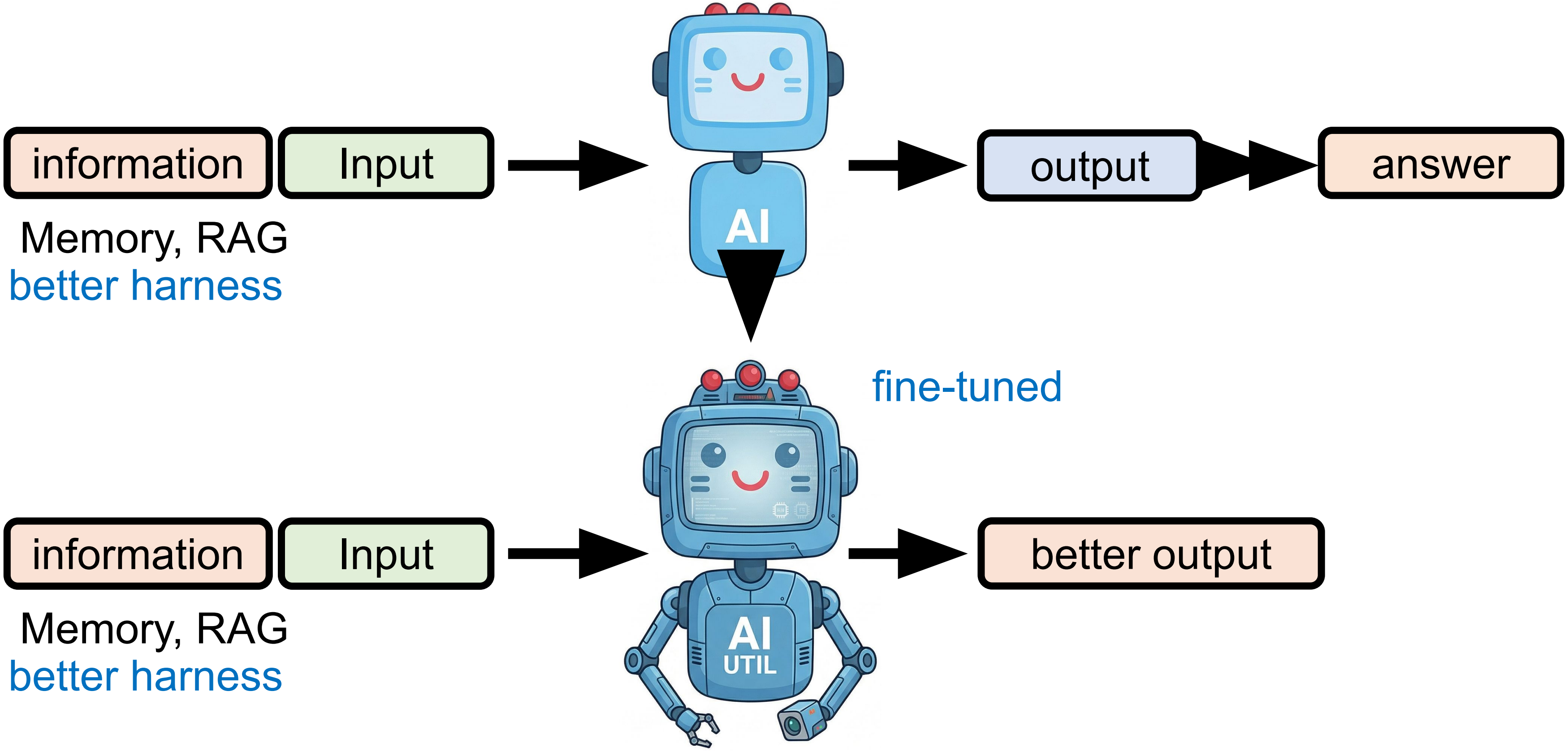
Retrieval-Augmented LLM Agents: Learning to Learn from Experience

https://arxiv.org/abs/2603.18272



Retrieval-Augmented LLM Agents: Learning to Learn from Experience

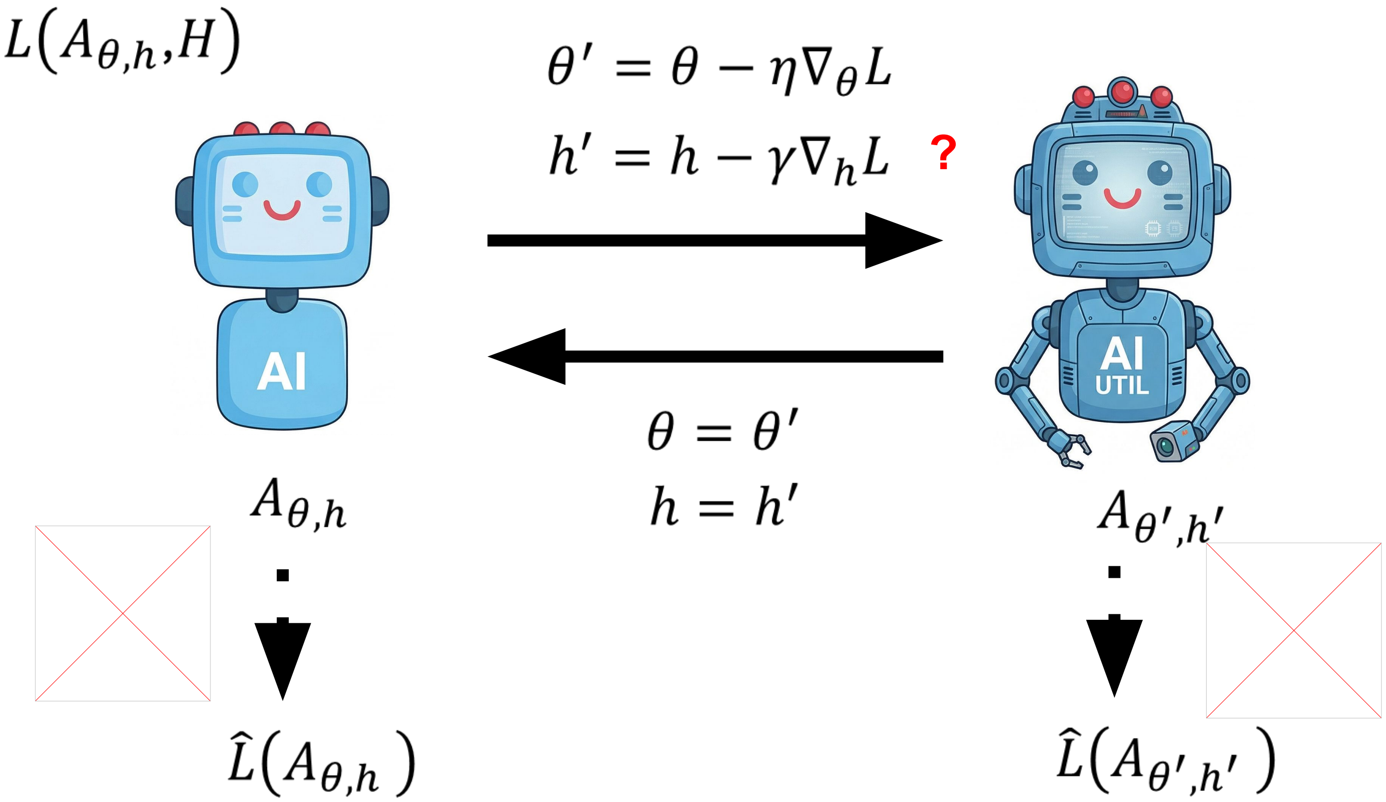
https://arxiv.org/abs/2603.18272



Fine-Tuning and Prompt Optimization: Two Great Steps that Work Better Together

https://arxiv.org/abs/2407.10930

Strategy	mistral-7b-instruct-v0.2			llama-2-7b-chat			llama-3-8b-instruct		
	HotPotQA	GSM8K	Iris	HotPotQA	GSM8K	Iris	HotPotQA	GSM8K	Iris
Baseline Strategies									
Vanilla Zero-shot	17.2	40.3	26.0	13.2	24.0	0.0	31.6	72.7	48.0
Prompt Optimization (Π)	33.8	46.4	57.3	33.3	26.0	56.7	**46.9**	77.9	79.3
Weight Optimization (Θ)	22.9	40.7	29.3	12.2	24.0	–	34.8	75.1	37.3
$\Pi \rightarrow \Pi$	33.8	47.7	59.3	32.6	24.7	64.0	46.5	77.6	**82.0**
$\Theta \rightarrow \Theta$	24.0	42.8	38.0	13.0	24.1	–	34.4	44.1	39.3
BetterTogether Strategies									
$\Pi \rightarrow \Theta$	36.3	47.3	30.7	32.7	**27.3**	26.7	42.8	77.6	44.0
$\Theta \rightarrow \Pi$	33.0	**48.3**	**66.7**	34.2	26.6	–	43.6	**78.9**	78.7
$\Pi \rightarrow \Theta \rightarrow \Pi$	**37.6**	46.8	52.7	**34.8**	26.3	**65.3**	46.7	77.0	79.3



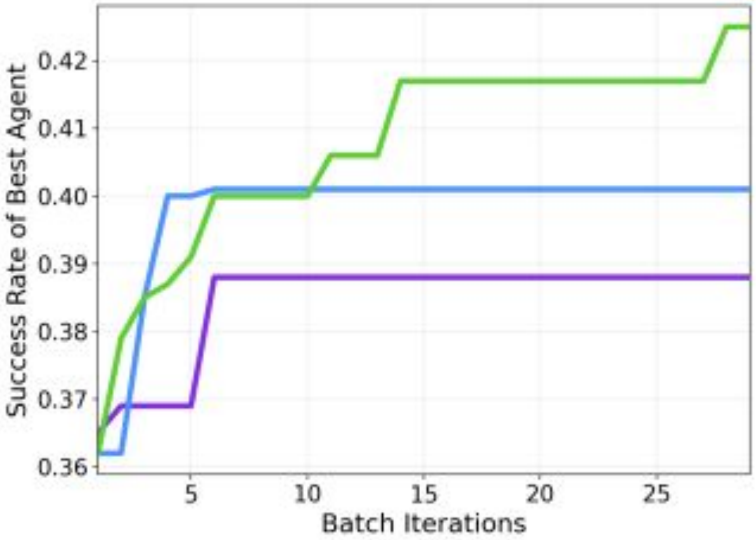
$L(A_{\theta,h}, H)$
$\theta' = \theta - \eta \nabla_\theta L$
$h' = h - \gamma \nabla_h L$?
AI
AI
UTIL
$\theta = \theta'$
$h = h'$
$A_{\theta,h}$
$A_{\theta',h'}$
$\hat{L}(A_{\theta,h})$
$\hat{L}(A_{\theta',h'})$

AIME → BeyondAIME

Easy-to-Hard Generalization

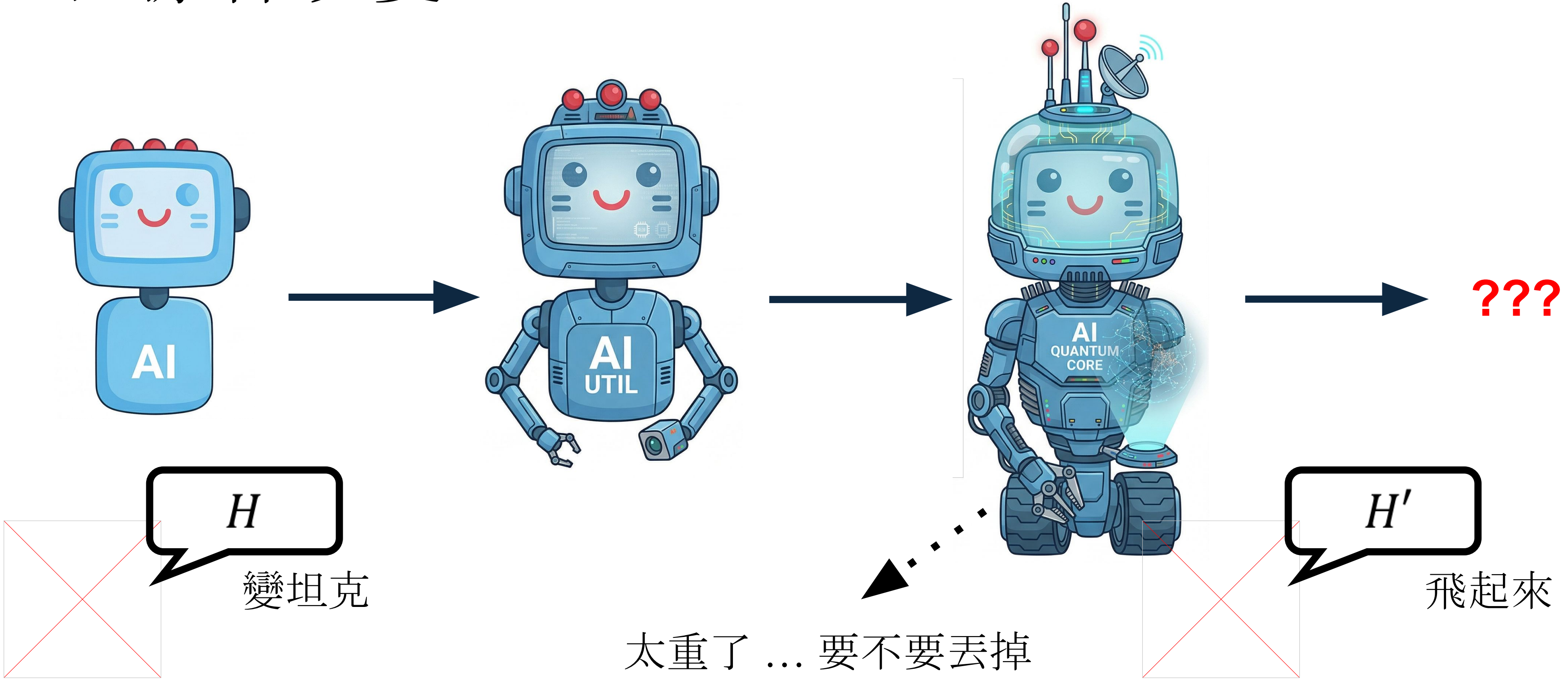
Agent Evolution Tree

Eval on BeyondAIME

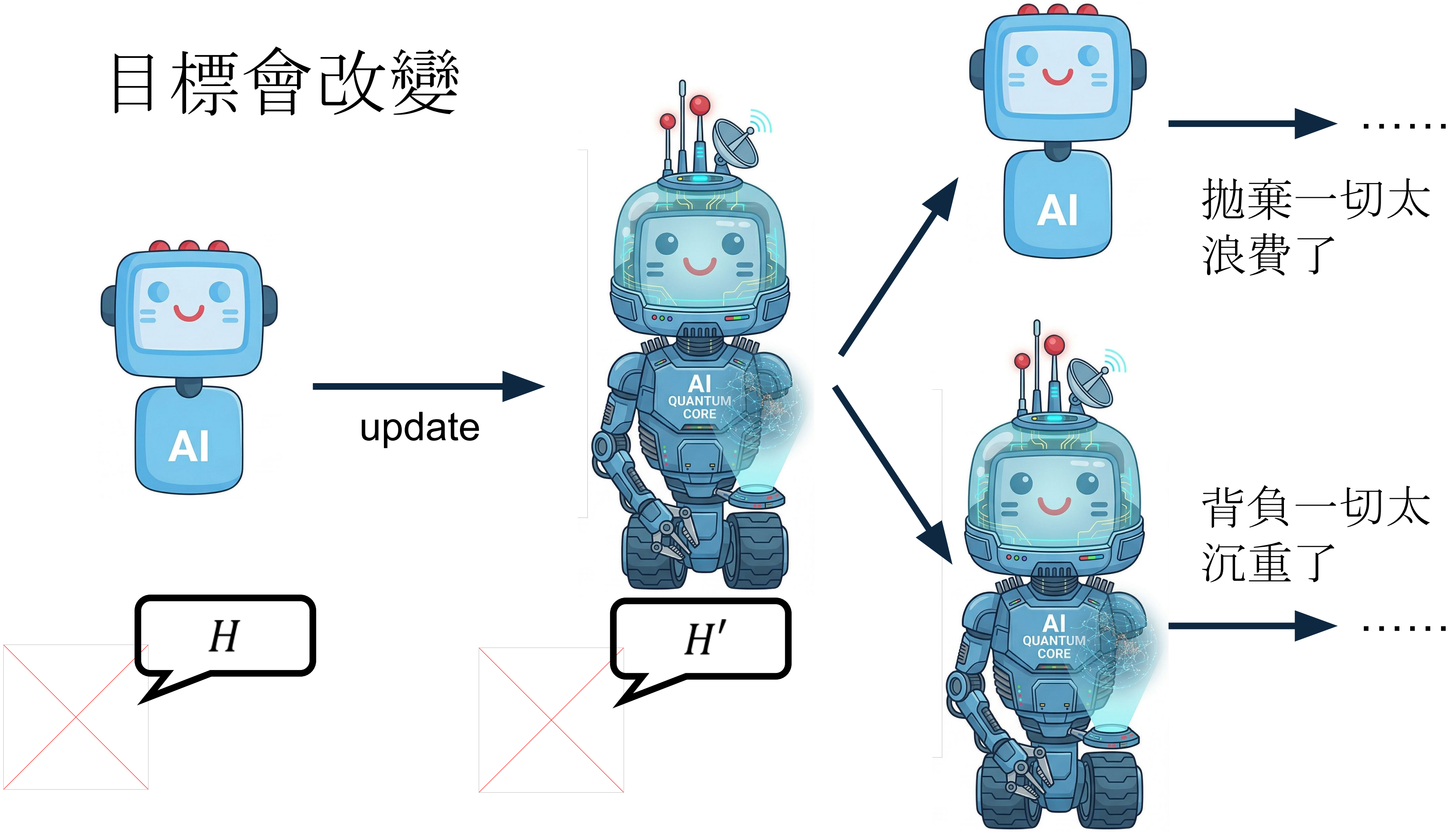


https://arxiv.org/abs/2602.14697

目標會改變

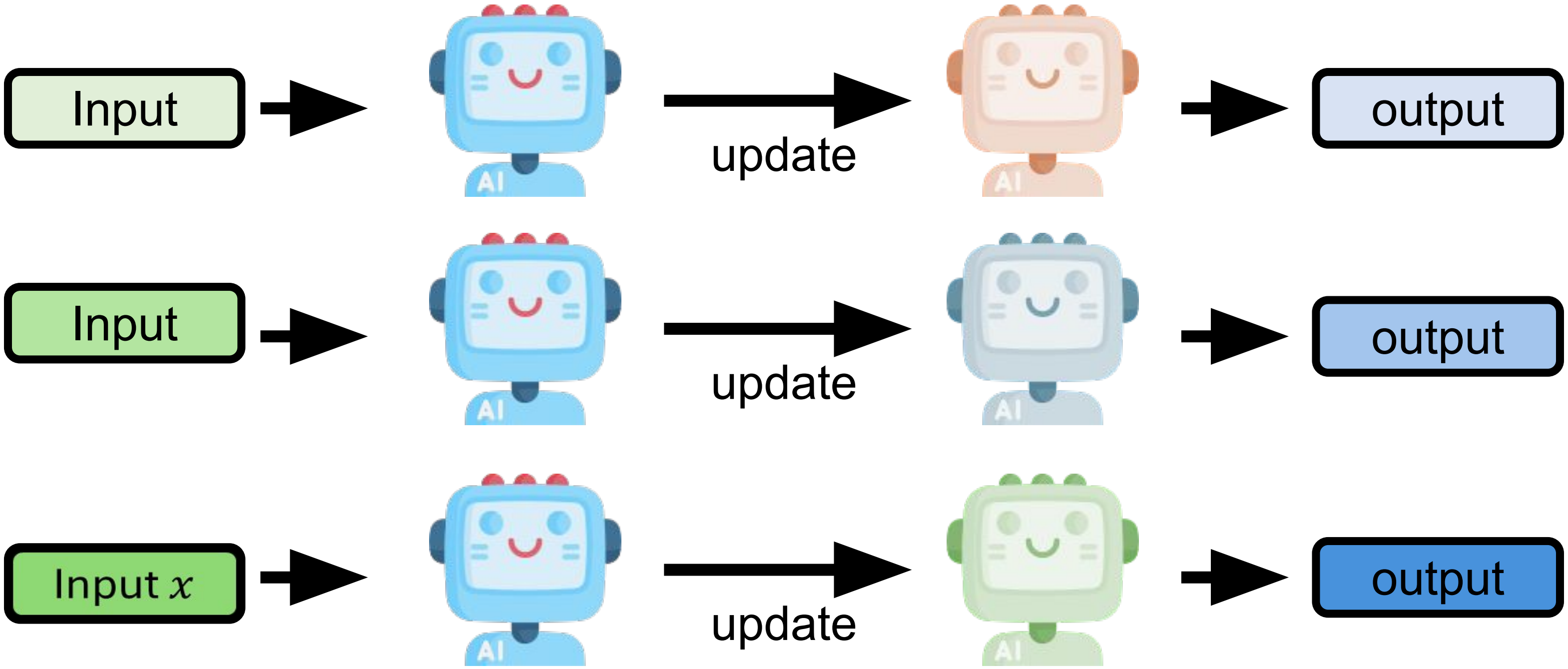


目標會改變

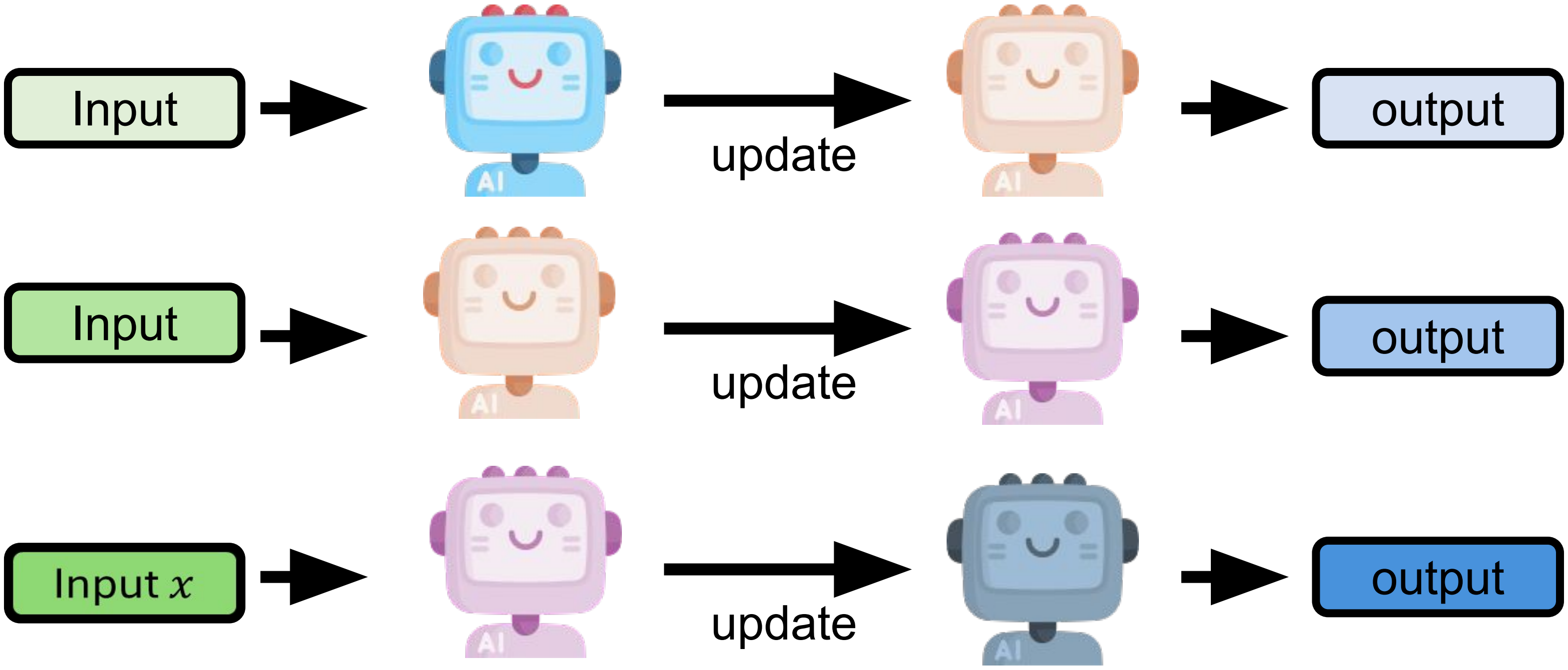


AI
update
AI
QUANTUM
CORE
H
H′
AI
……
拋棄一切太浪費了
AI
QUANTUM
CORE
背負一切太沉重了
……

目標會改變：Test Time Training (TTT)



目標會改變：Test Time Training (TTT)



目標會改變：Test Time Training (TTT)

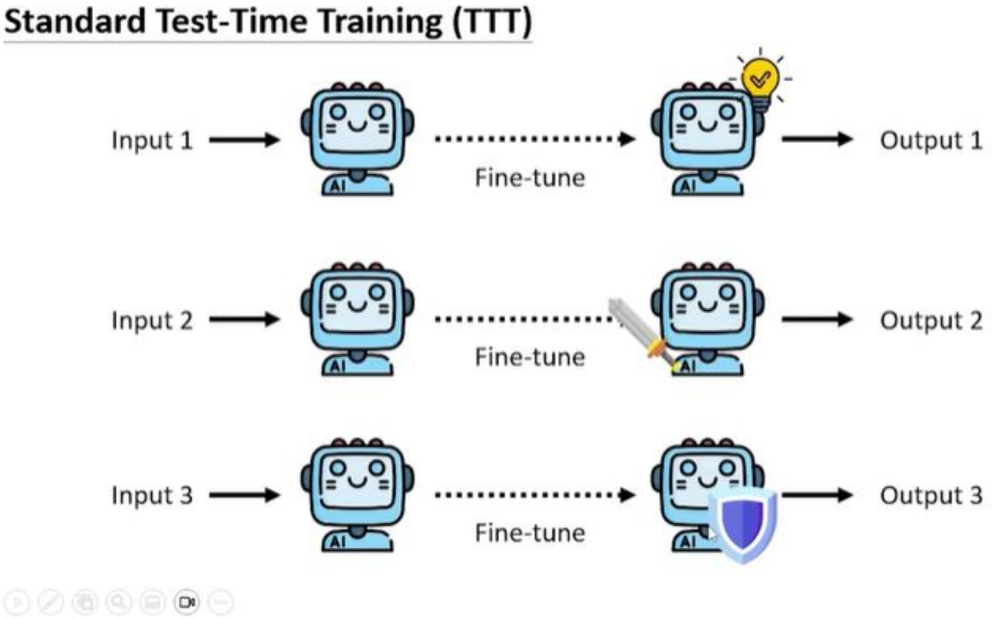
Wei-Ping Huang

Guan-Ting Lin

https://arxiv.org/abs/2406.11064

1:54:30

https://youtu.be/EnWz5XuOnlQ?si=jDgXiXkWWkhlsPFr&t=6872



【生成式人工智慧與機器學習導論2025】第 8 講：通用模型的終身學習 (Fine-tuning, Model Editing, Model Merging, Test-Time Training)

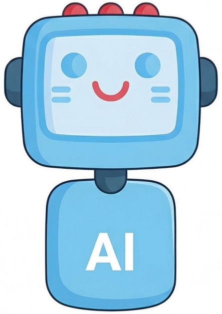
目標會改變:避免遺忘

https://youtu.be/Z6b5-77EfGk?si=kAyIOZ2Me9-8jFLh

【生成式AI時代下的機器學習(2025)】第六講：生成式人工智慧的後訓練(Post-Training)與遺忘問題

目標會改變:避免遺忘

https://arxiv.org/abs/2605.09315

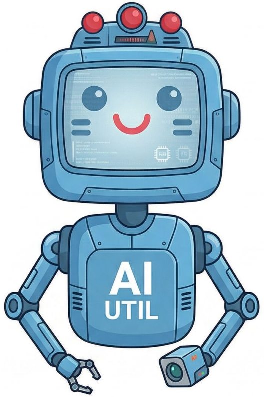


遺忘舊能力

$$\theta' = \theta - \eta \nabla_{\theta} L$$

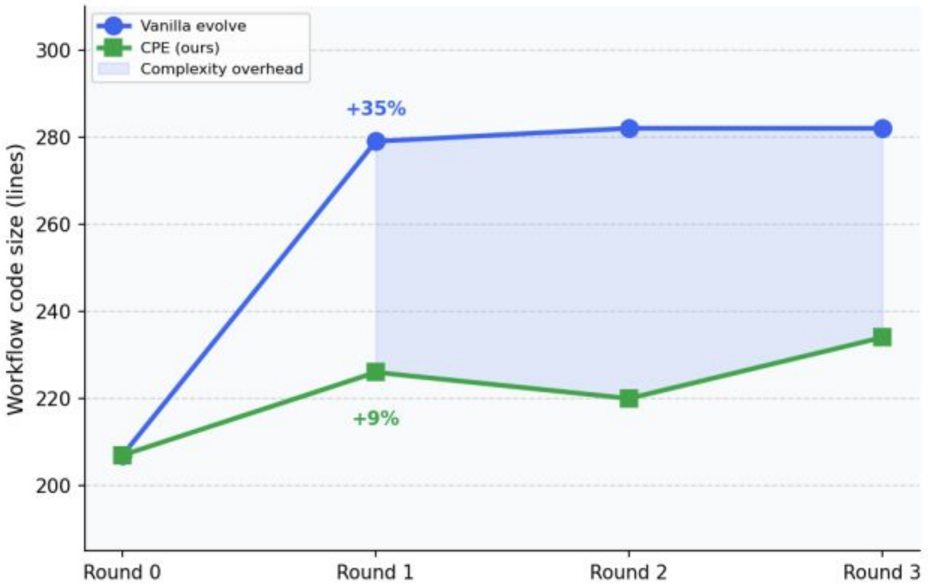
$$h' = h - \gamma \nabla_{h} L \quad {\color{red}?}$$

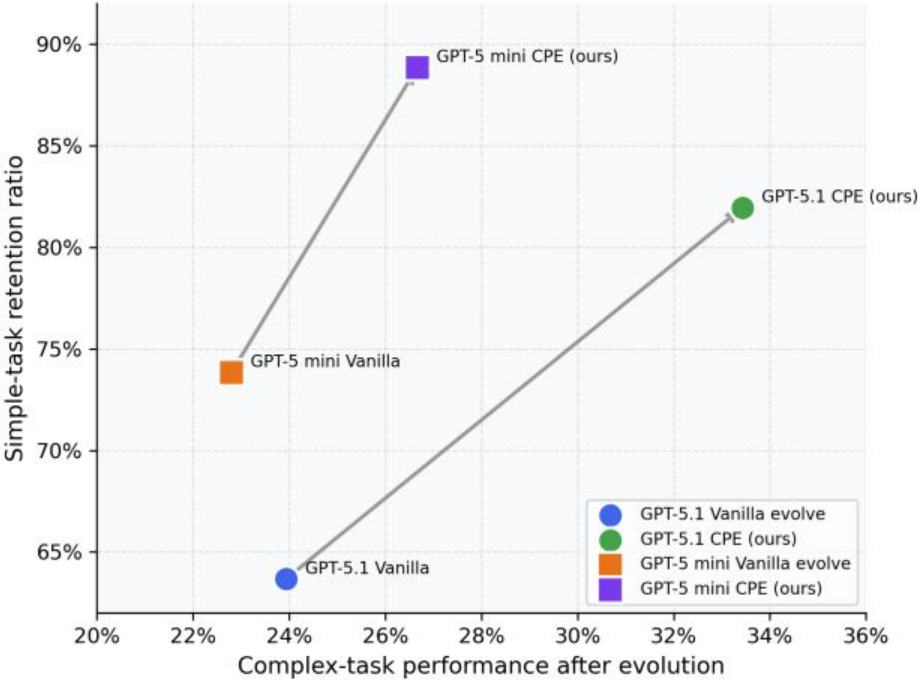
遺忘舊能力



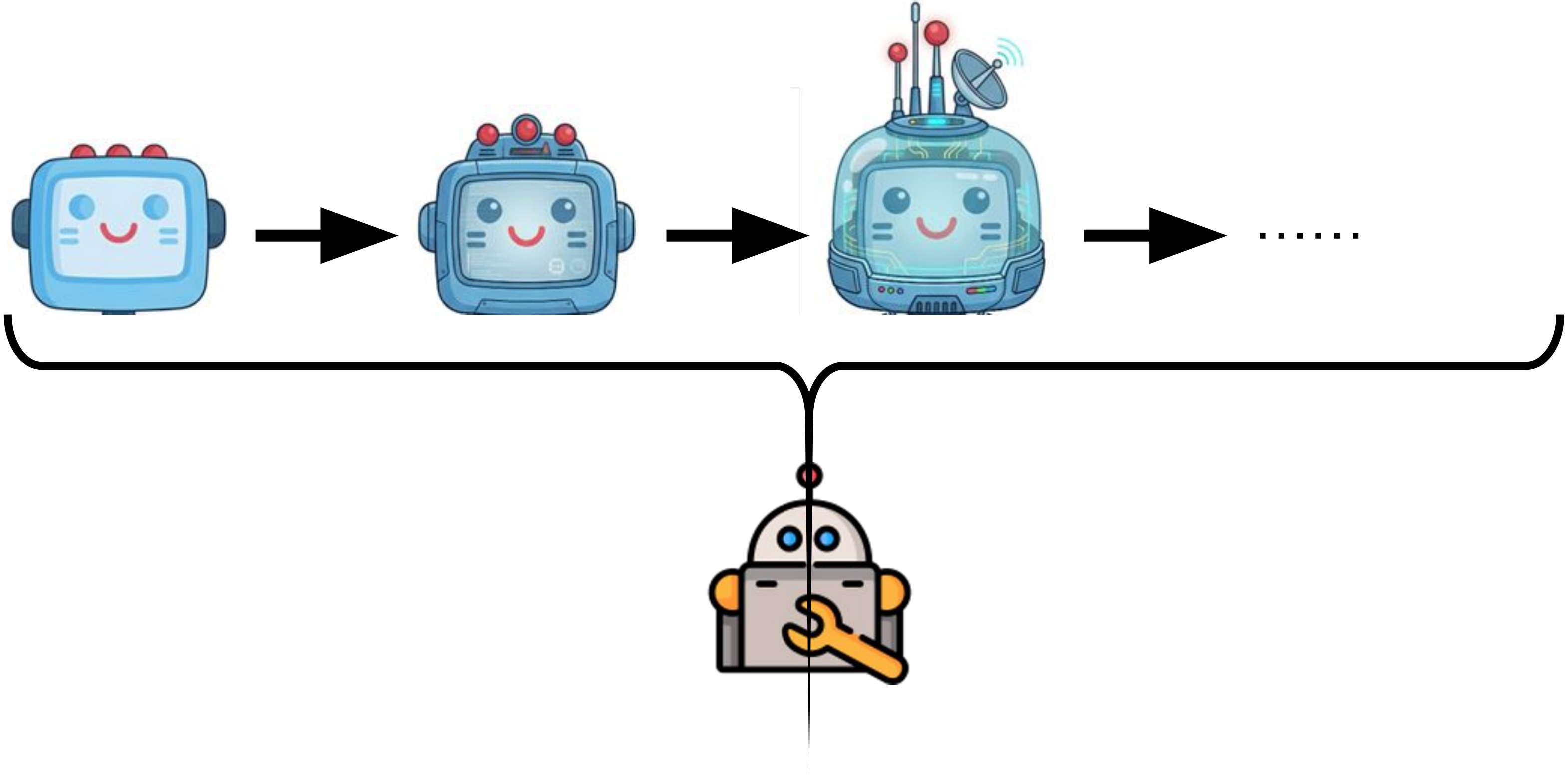
目標會改變:避免遺忘

https://arxiv.org/abs/2605.09315

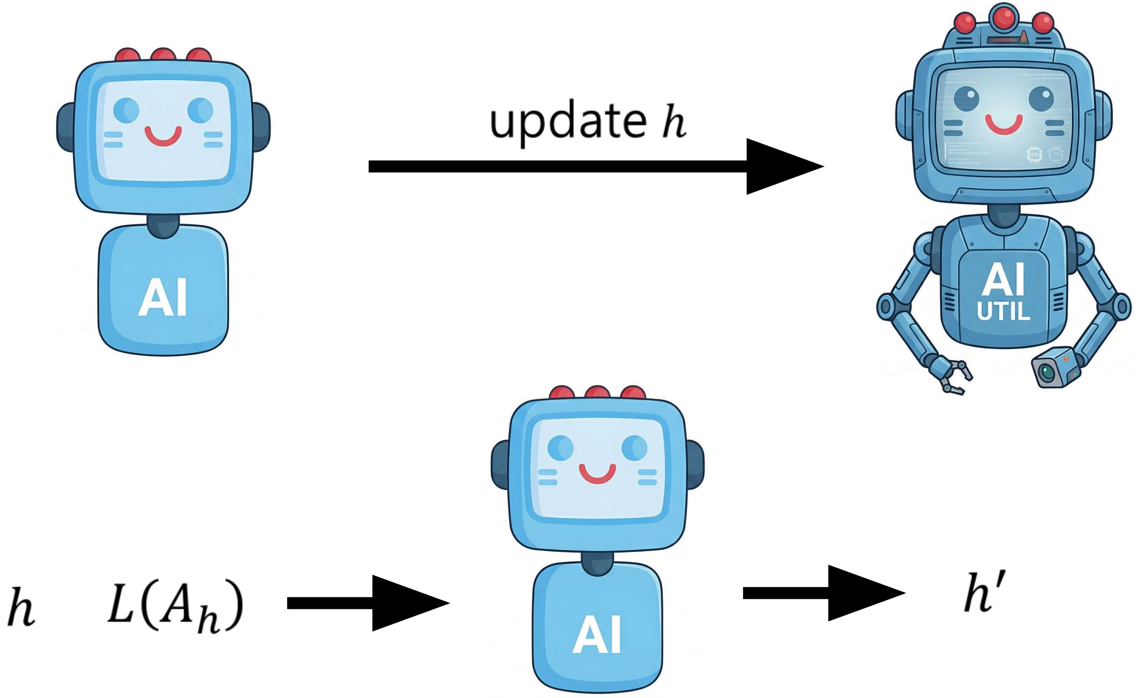




Improvement module controls how to improve.

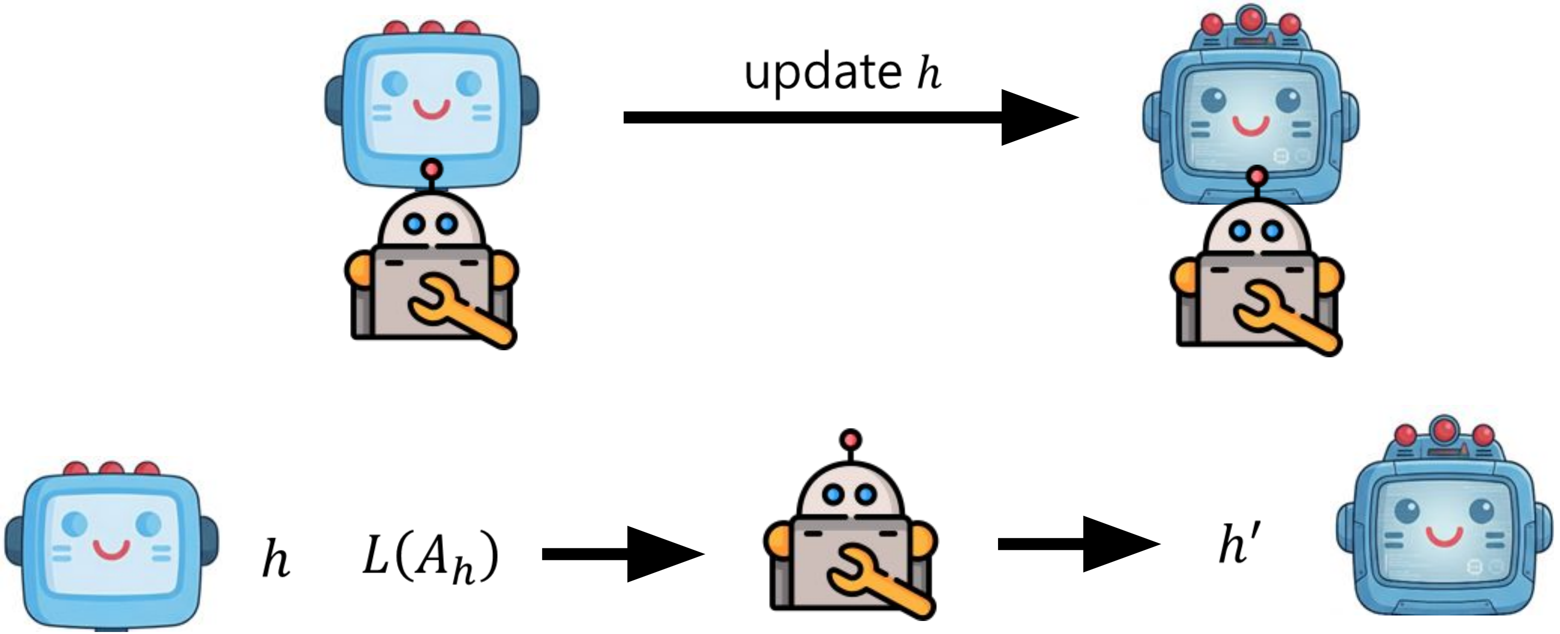


How to improve harness?



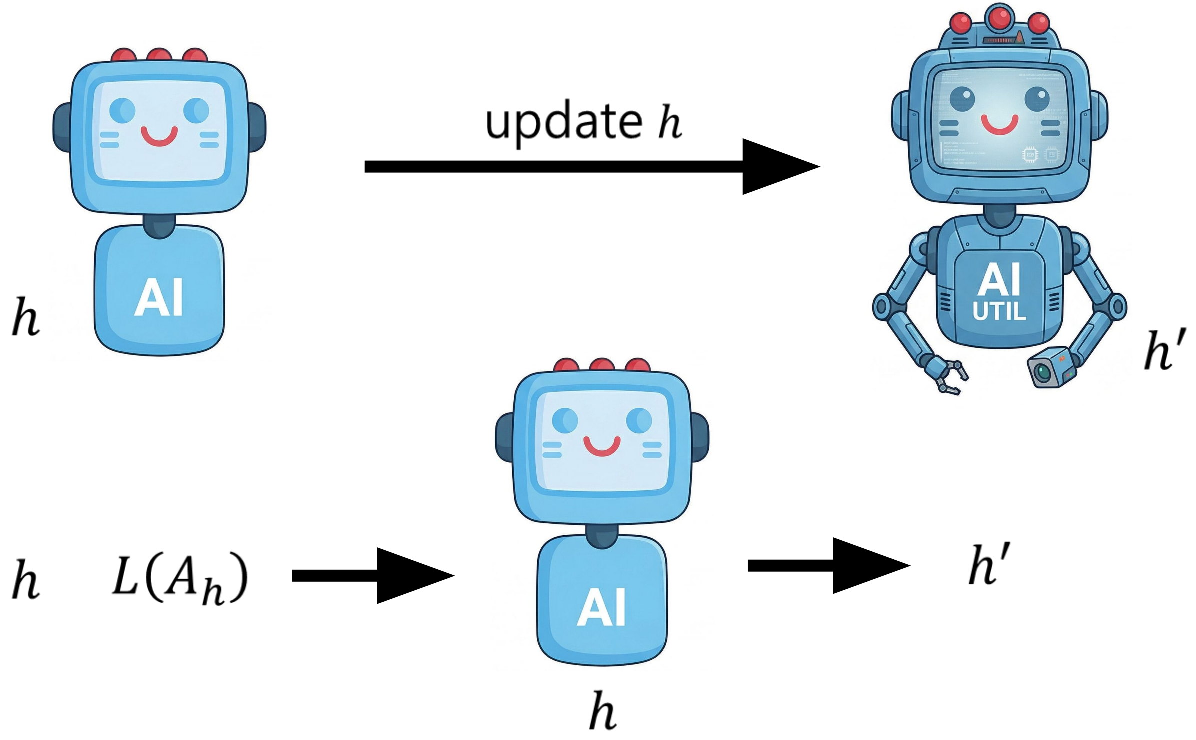
https://arxiv.org/abs/2505.22954

How to improve harness?



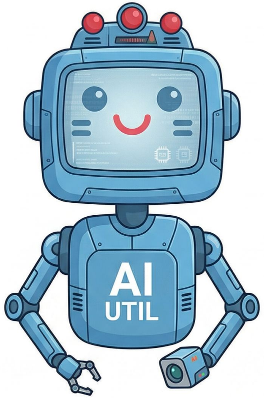
https://arxiv.org/abs/2505.22954

How to improve “improve harness”?



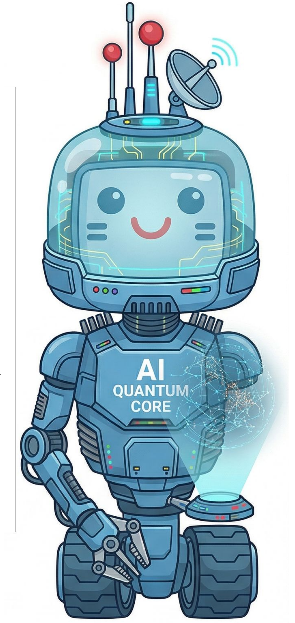
HyperAgent https://arxiv.org/abs/2603.19461

Gödel Agent https://arxiv.org/abs/2410.04444



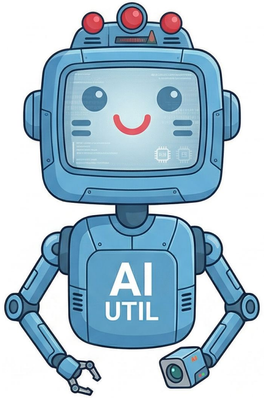
h'

update h $\longrightarrow$



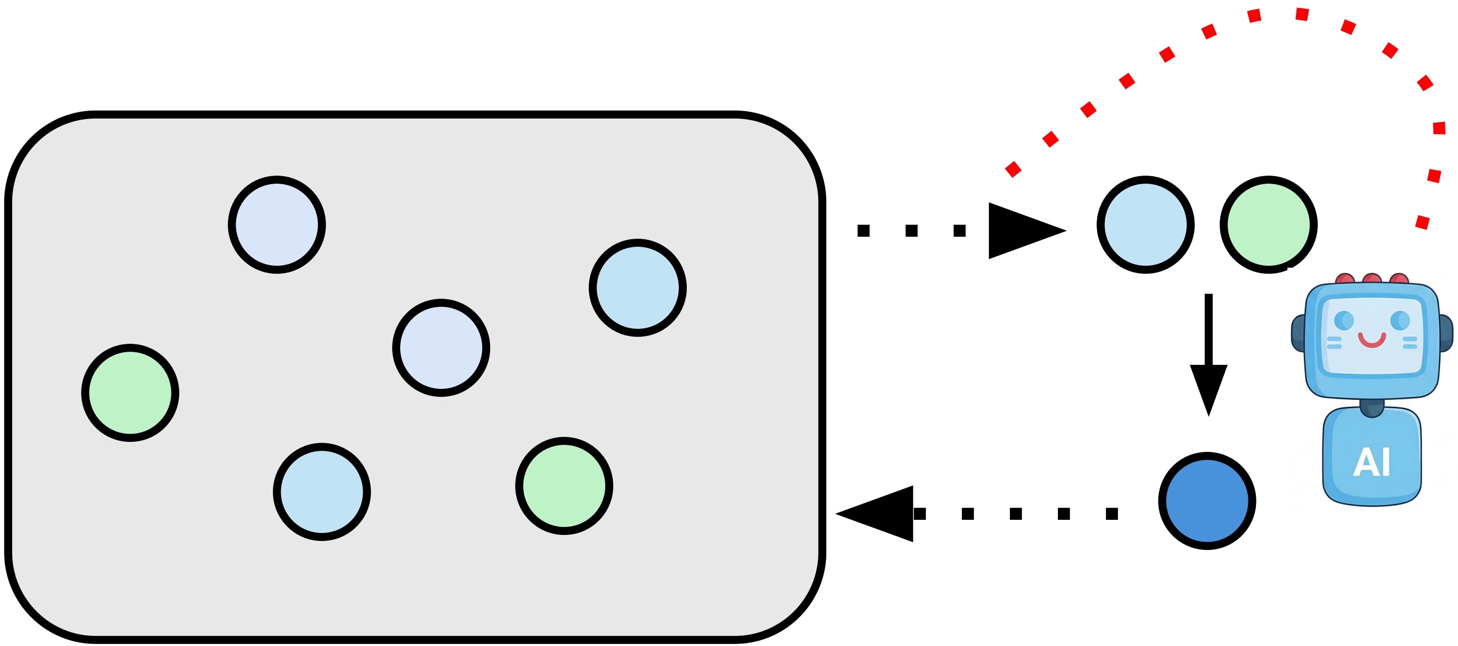
h''

h' $L(A_{h'})$ $\longrightarrow$

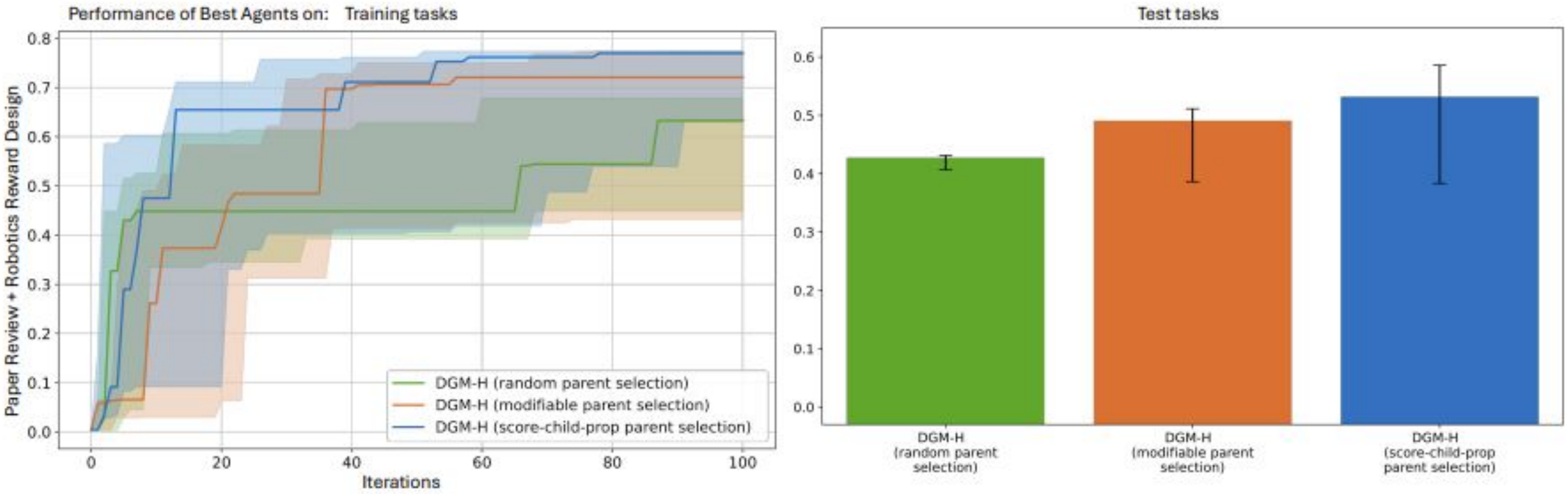


h'

$\longrightarrow$ h''



HyperAgent https://arxiv.org/abs/2603.19461



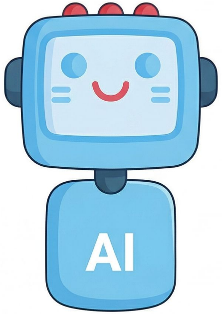
https://arxiv.org/abs/2603.19461

How to improve “improve harness”?

Learning to Self-Evolve https://arxiv.org/abs/2603.18620

fine-tuned

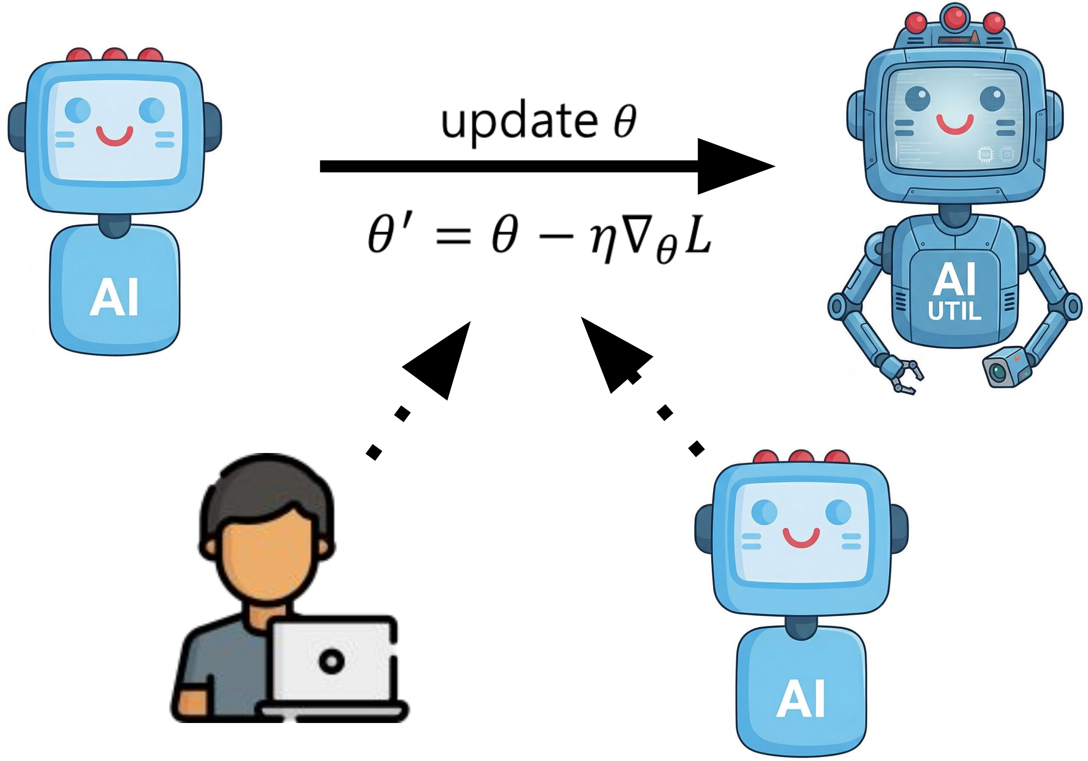
h $L(A_h)$ $\rightarrow$



$\rightarrow$ h' $L'(A_{h'})$

$L(A_h) - L'(A_{h'})$

Update Parameters

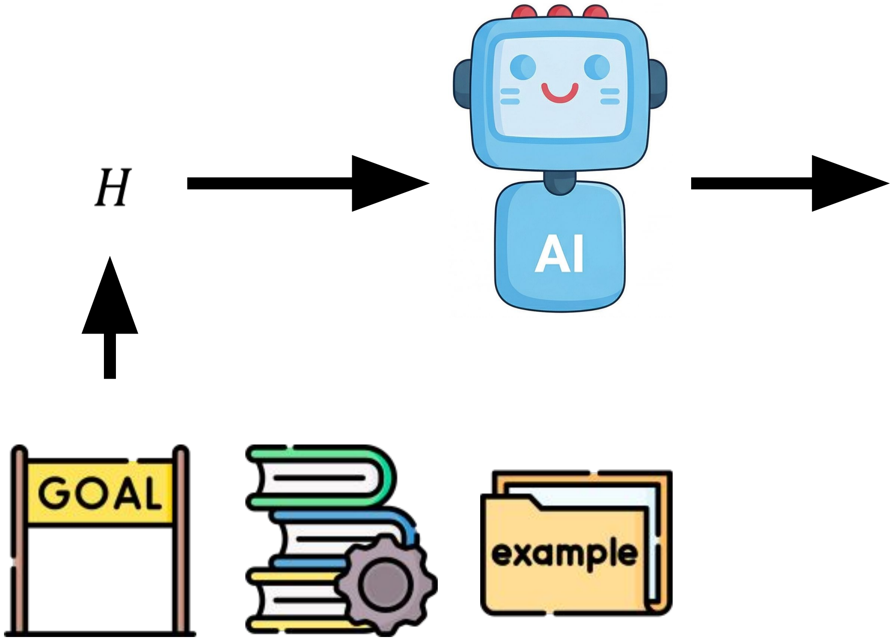


上次在課程接近結尾的時候有提到

52:17
https://youtu.be/s06mSAGN4gM?si=1VAkFCUUheM3kqaz&t=3137

Update Parameters

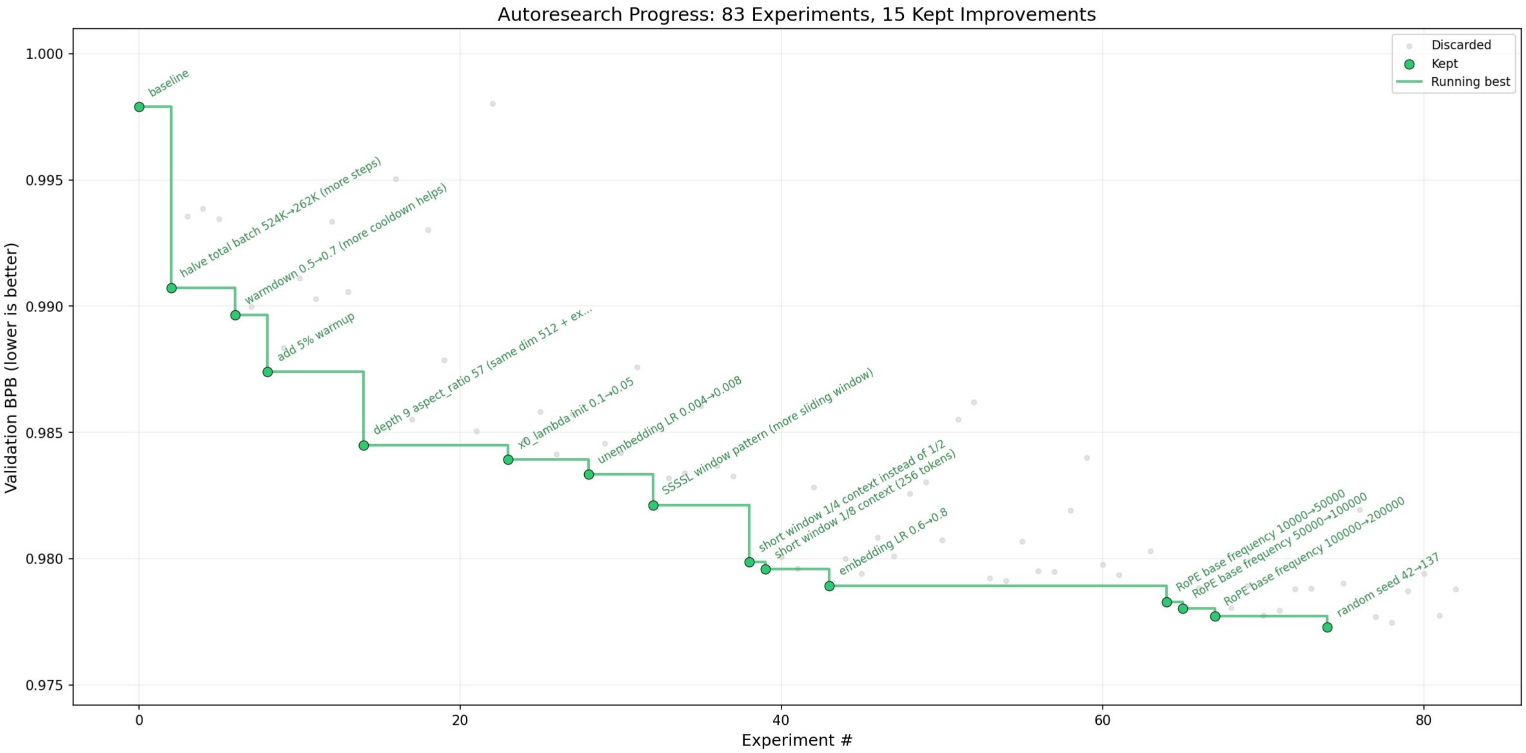
https://arxiv.org/abs/2603.08640



$L(A_\theta, H)$ $\theta' = \theta - \eta \nabla_\theta L$

Update Parameters

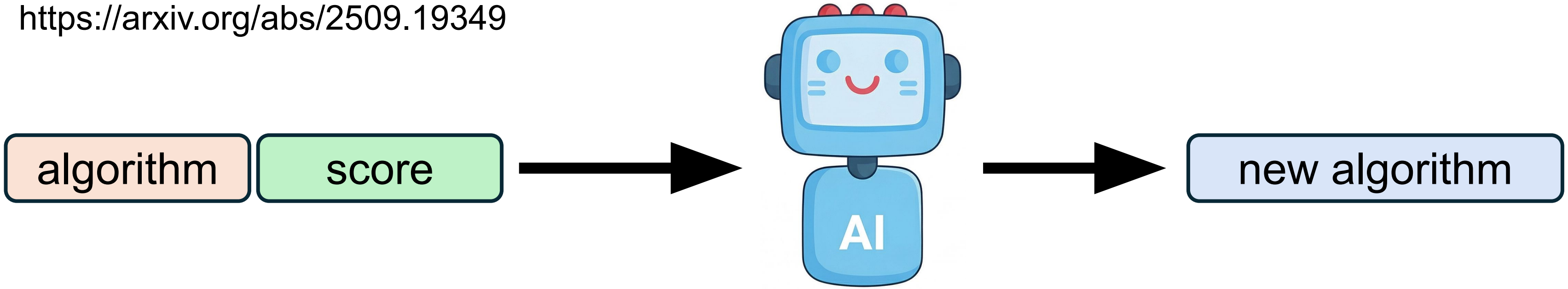
https://github.com/karpathy/autoresearch



Learn to “Update Parameters”

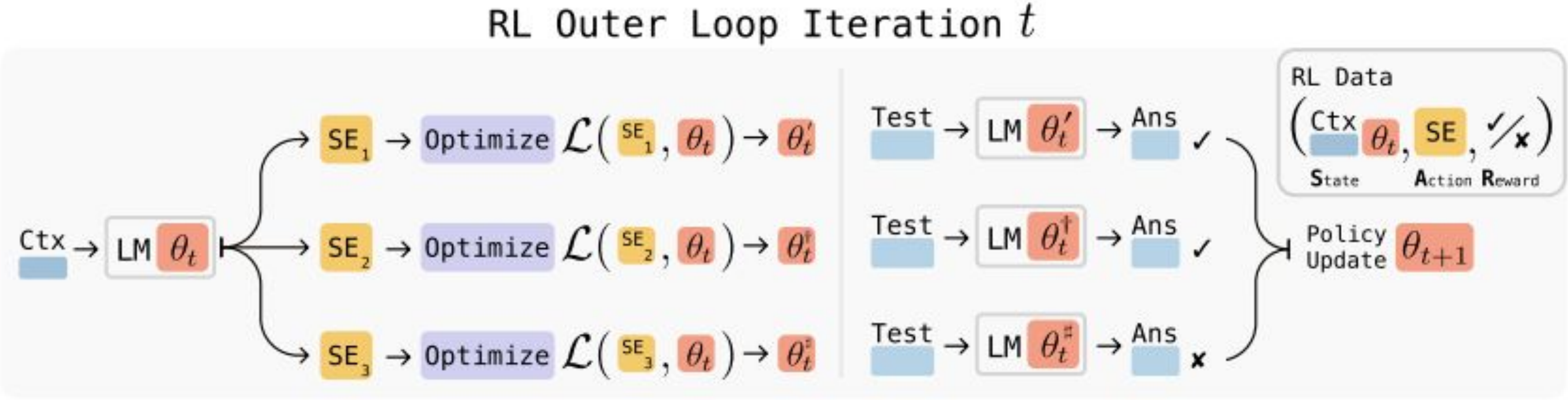
Alpha Evolve
https://deepmind.google/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/

Shinka Evolve
https://arxiv.org/abs/2509.19349



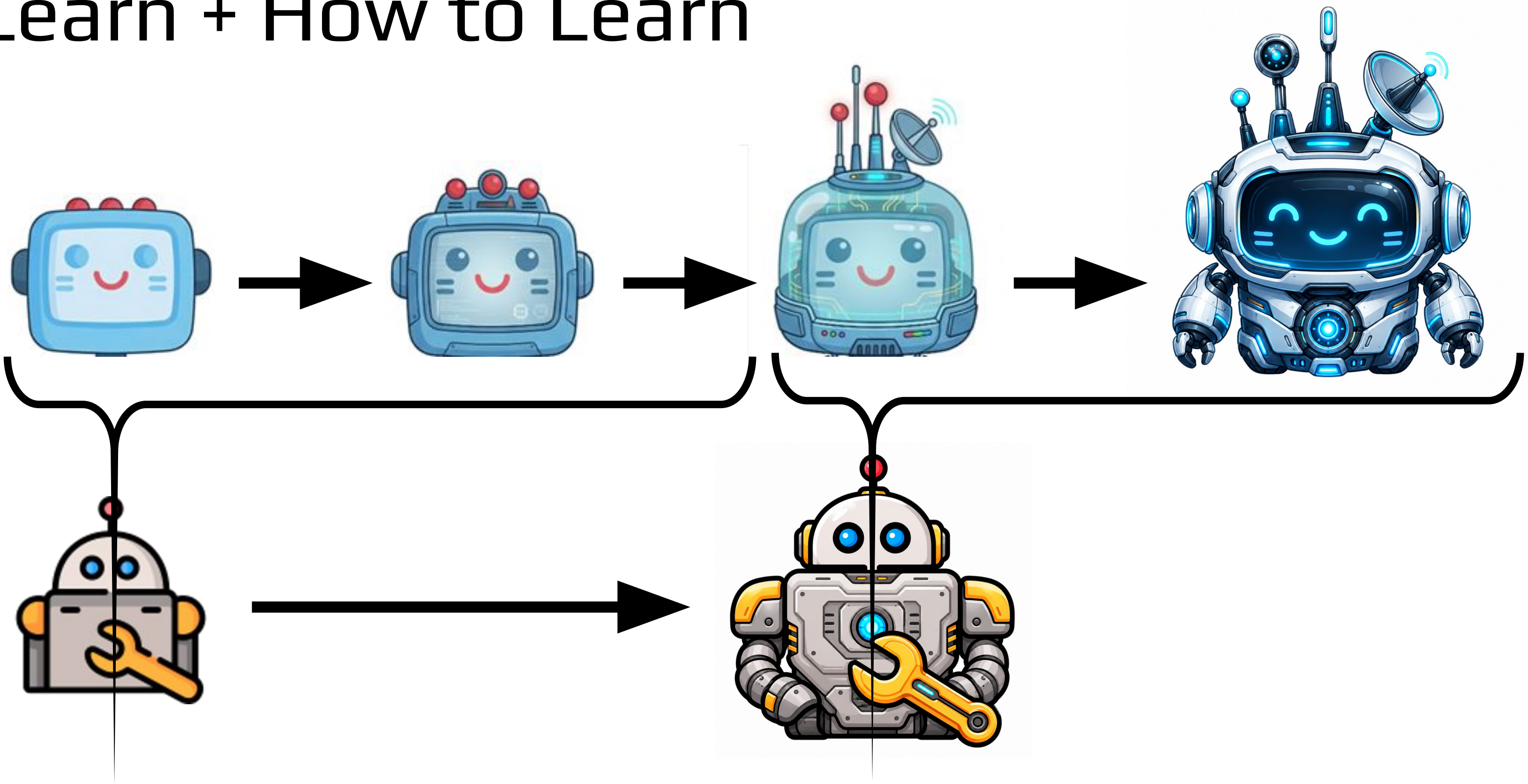
Learn to "Update Parameters"

Self-Adapting LLMs (SEAL)

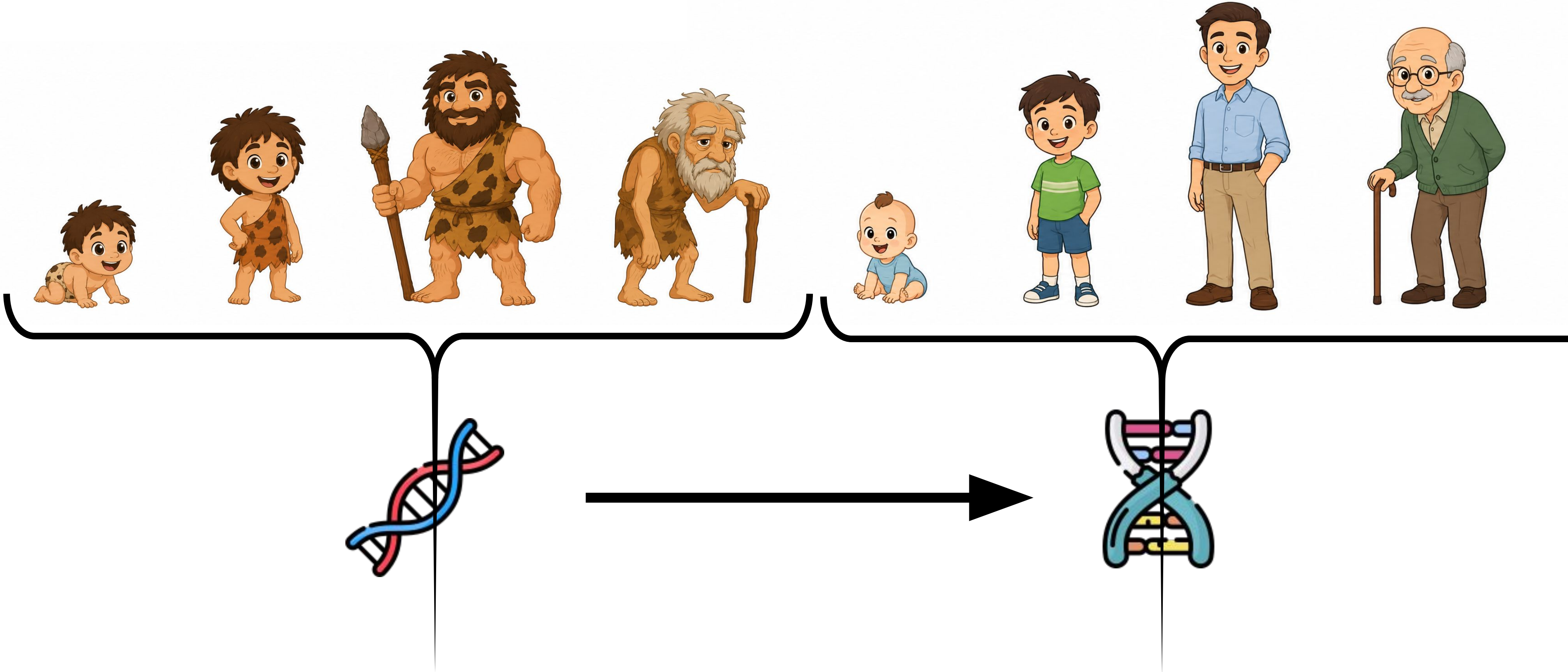


https://arxiv.org/abs/2506.10943

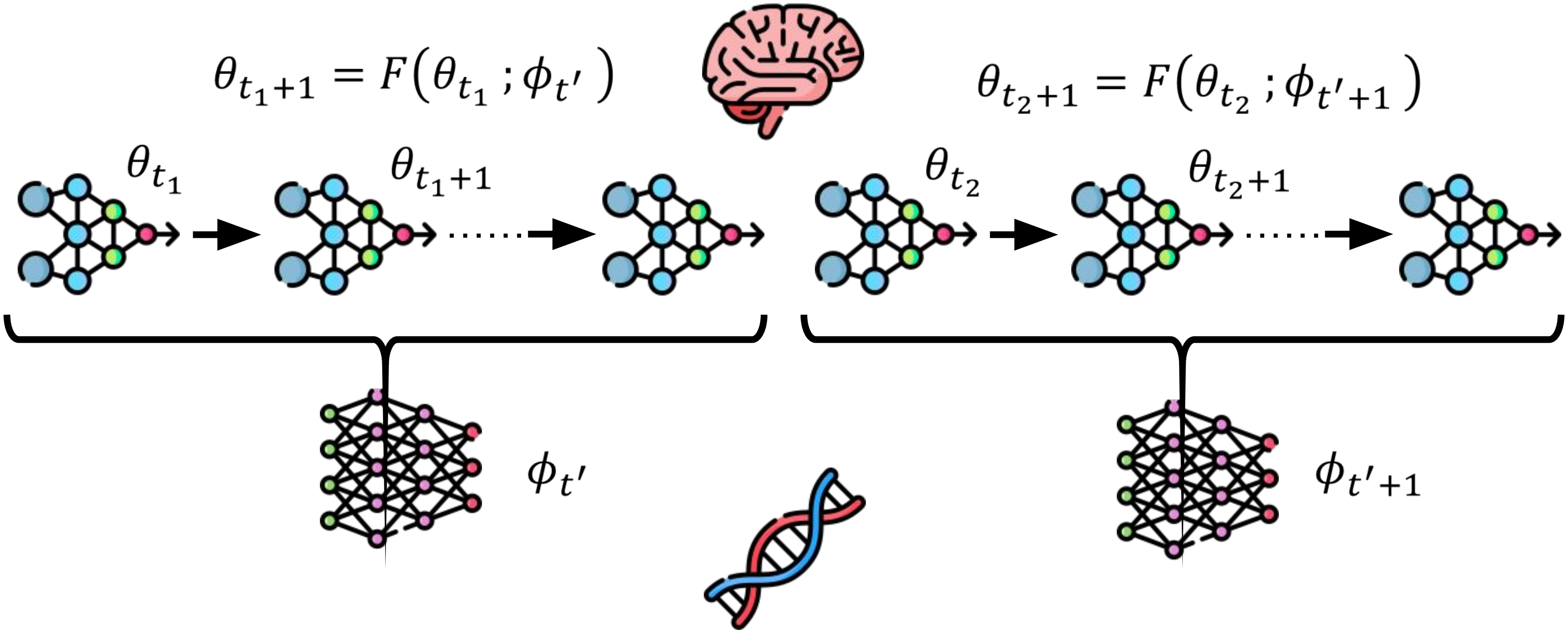
Learn + How to Learn



Learn + How to Learn



Learn + How to Learn

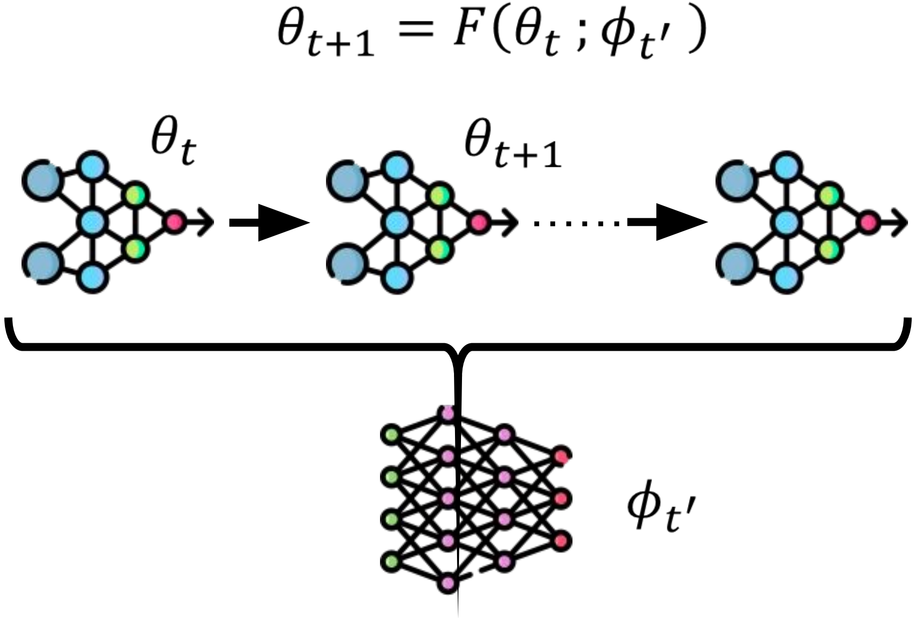


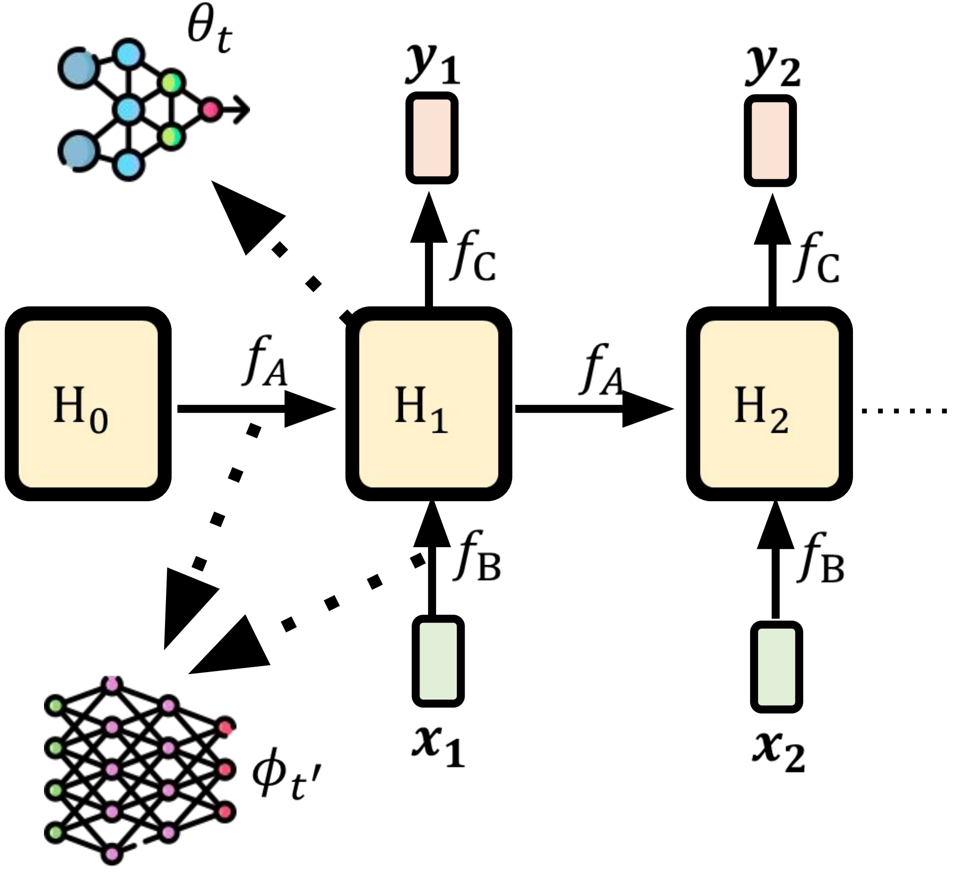
Meta Learning

【機器學習2021】元學習 Meta Learning (一) - 元學習跟機器學習一樣也是三個步驟

【機器學習2021】元學習 Meta Learning (二) - 萬物皆可 Meta

https://youtu.be/xoastiYx9JU?si=7rW7ULj1Zwj1MzNS https://youtu.be/Q68Eh-wm1Ts?si=Y-A9SEafbre6-zhf



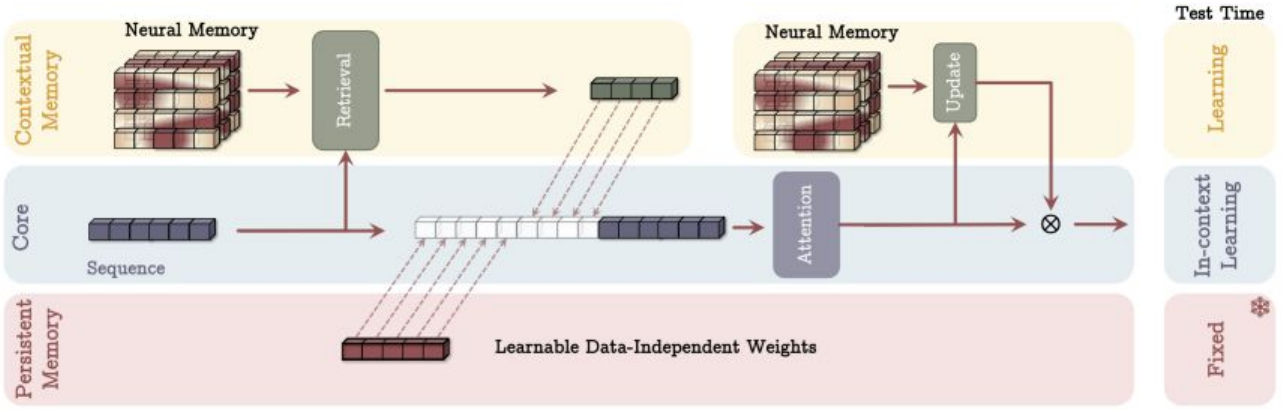


Learning to (Learn at Test Time):
RNNs with Expressive Hidden States
https://arxiv.org/abs/2407.04620

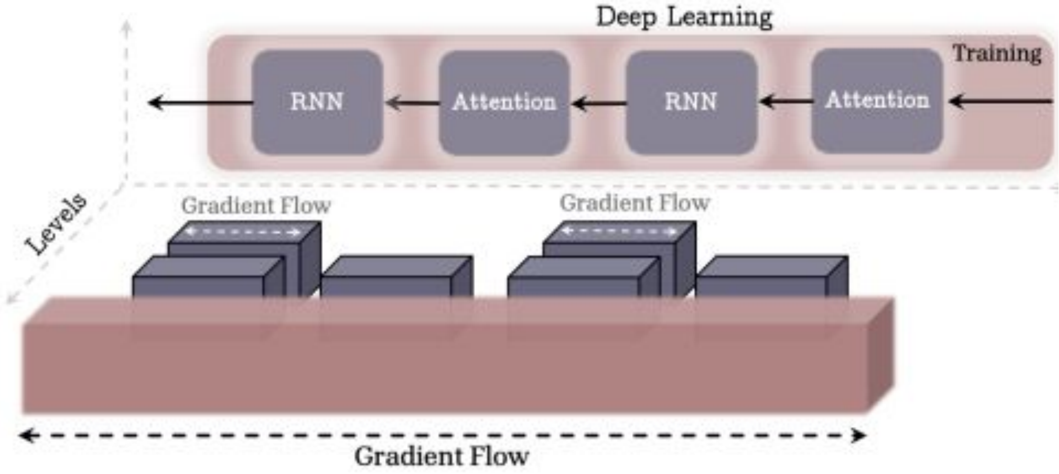
Transformer 的競爭者們

【生成式AI時代下的機器學習(2025)】第四講：Transformer 的時代要結束了嗎？介紹 Transformer 的競爭者們

https://youtu.be/gjsdVi90yQo?si=voyQ4GivpFn8SXIC



Titans
https://arxiv.org/abs/2501.00663

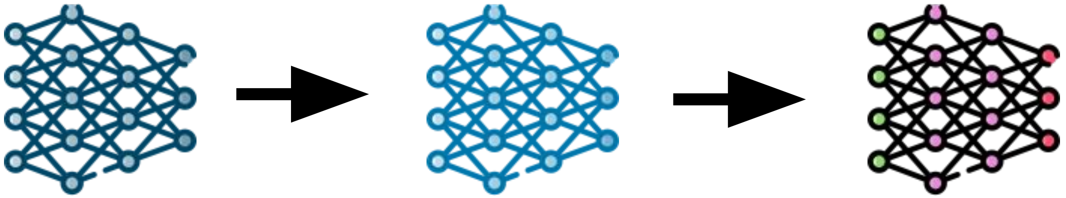


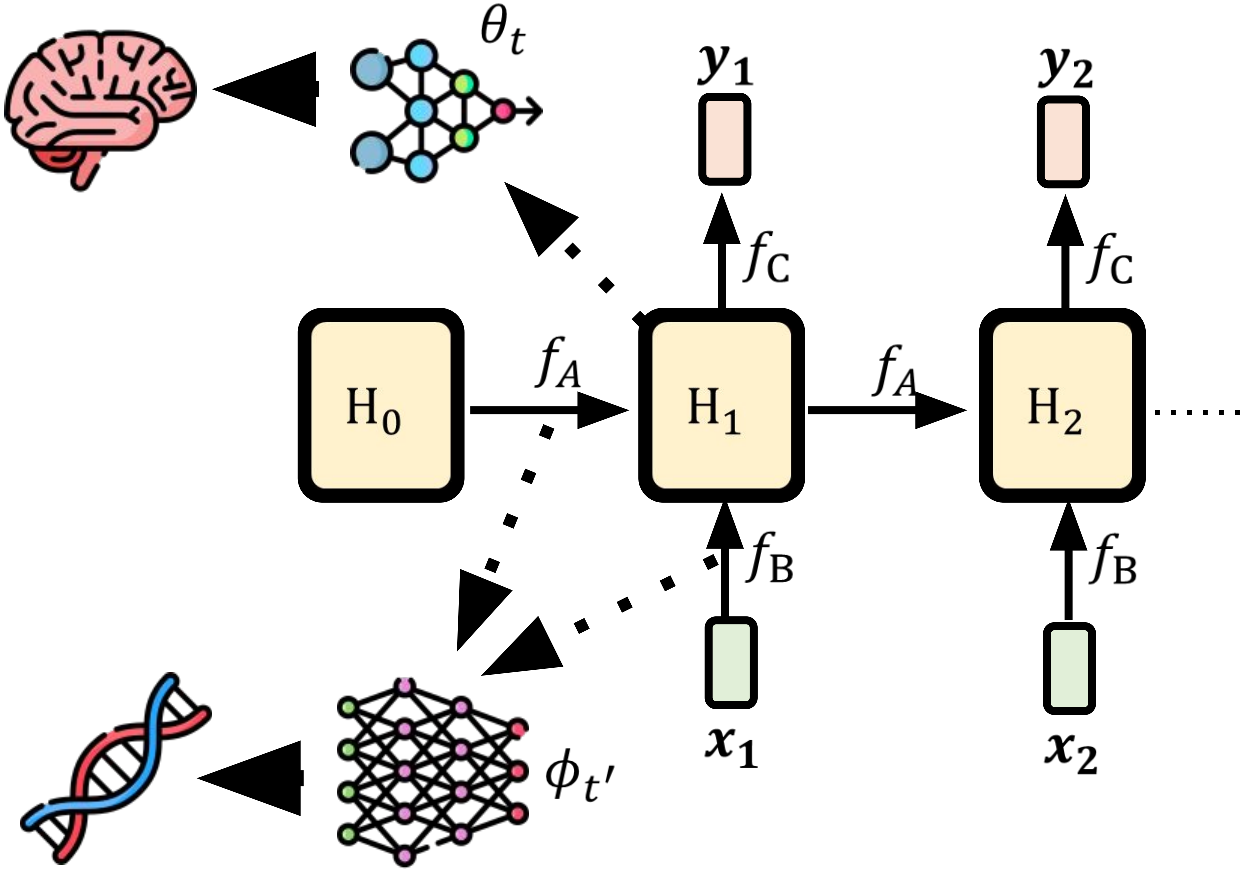
Nested Learning
https://arxiv.org/abs/2512.24695

何謂「學習」？

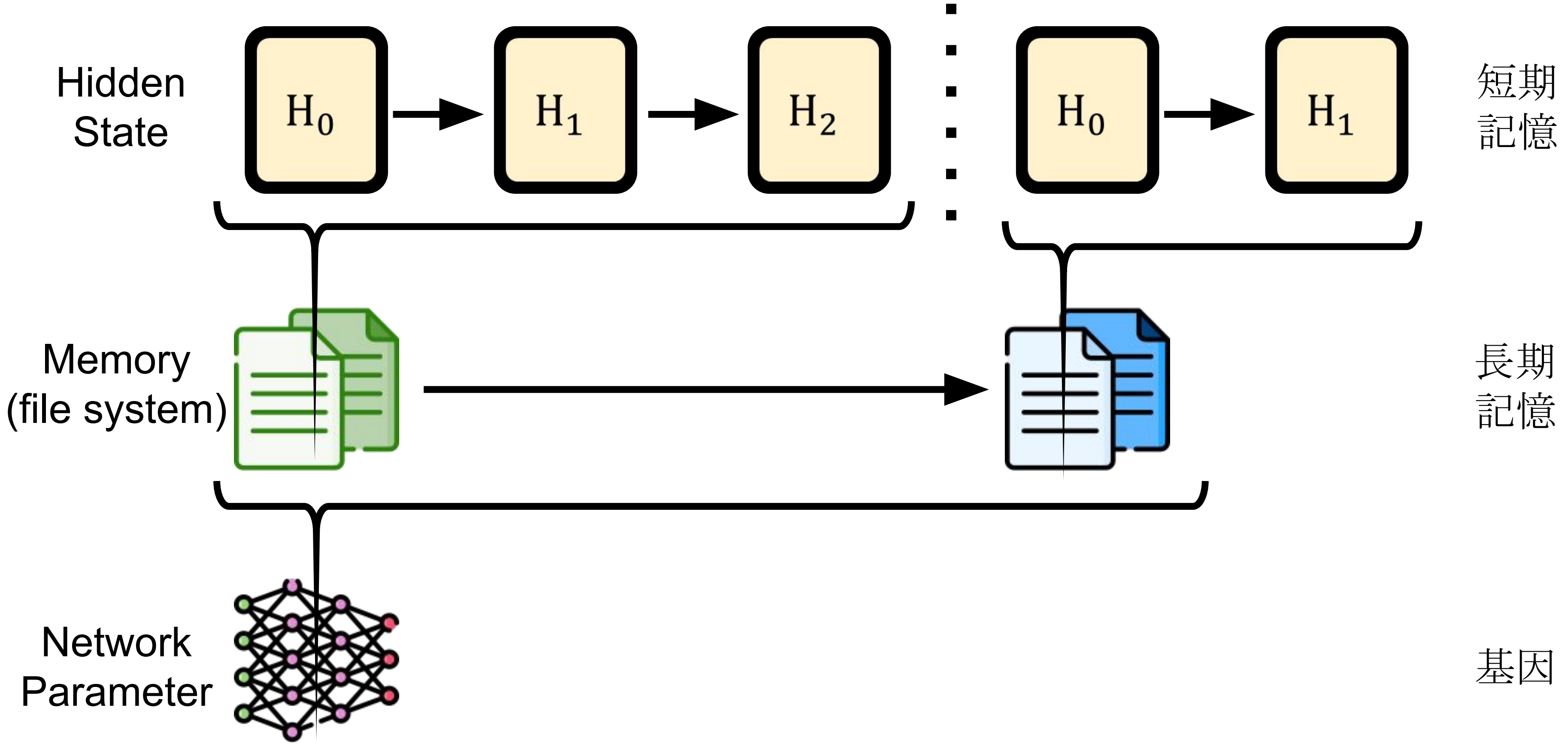
人類
學習

機器
學習





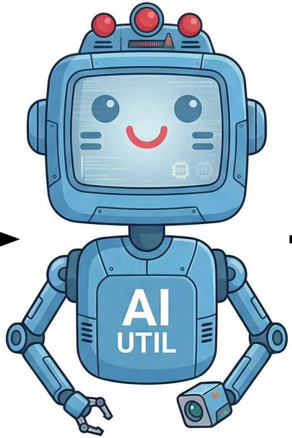
θ_t
$\boldsymbol{y}_1$
$\boldsymbol{y}_2$
f_C
f_C
H_0
f_A
H_1
f_A
H_2
f_B
f_B
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$
$\phi_{t'}$


Hidden State
H_0
H_1
H_2
H_0
H_1
短期記憶
Memory (file system)
長期記憶
Network Parameter
基因

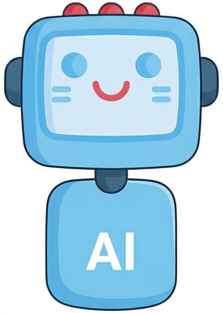
Intrinsic Motivation

去某個領域
找研究問題

在某個領域
找研究問題

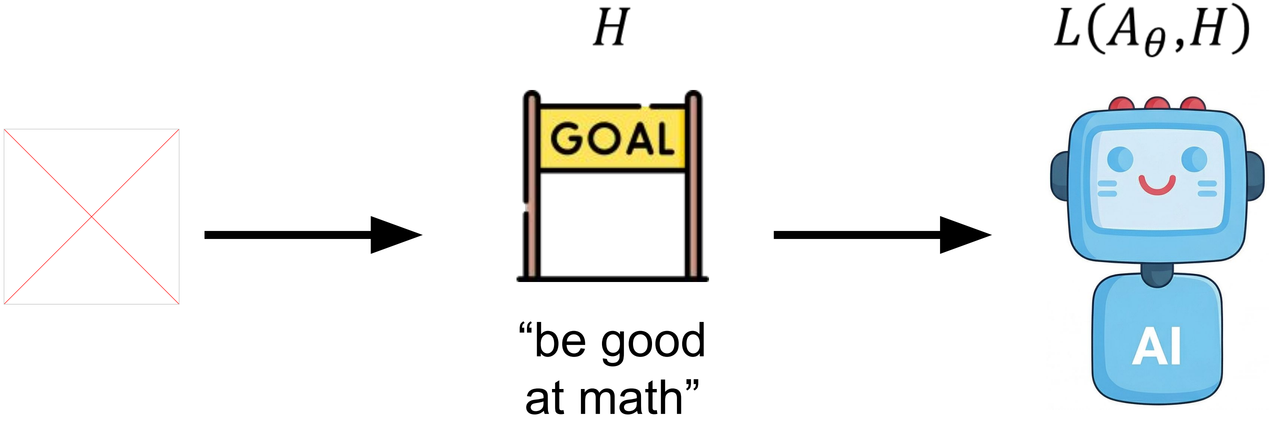


規劃並
執行實驗

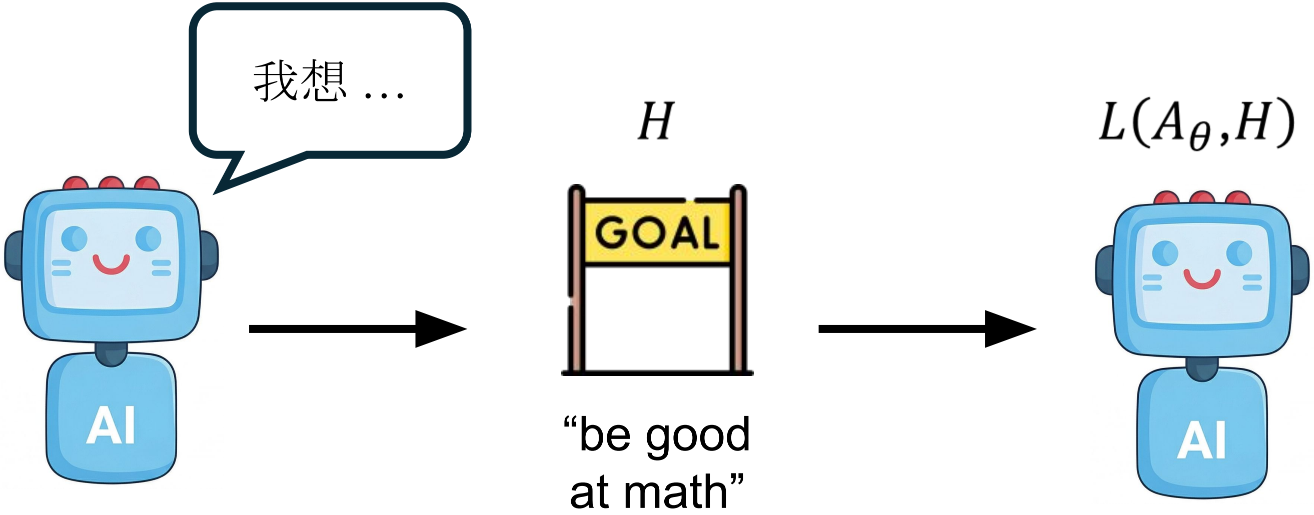


整理資料、
潤飾文稿

Intrinsic Motivation

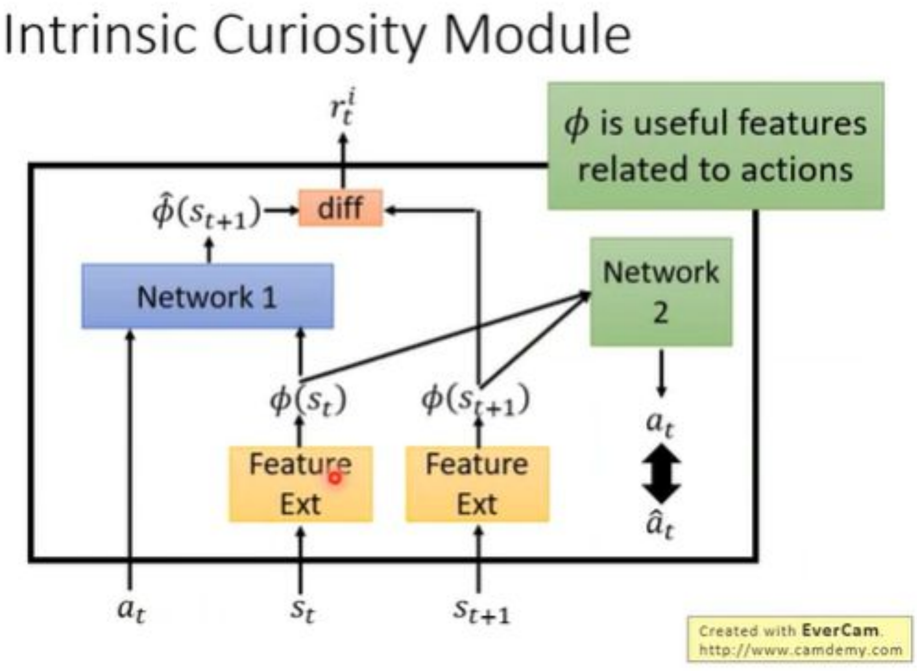


Intrinsic Motivation



- Curiosity-driven
- Empowerment

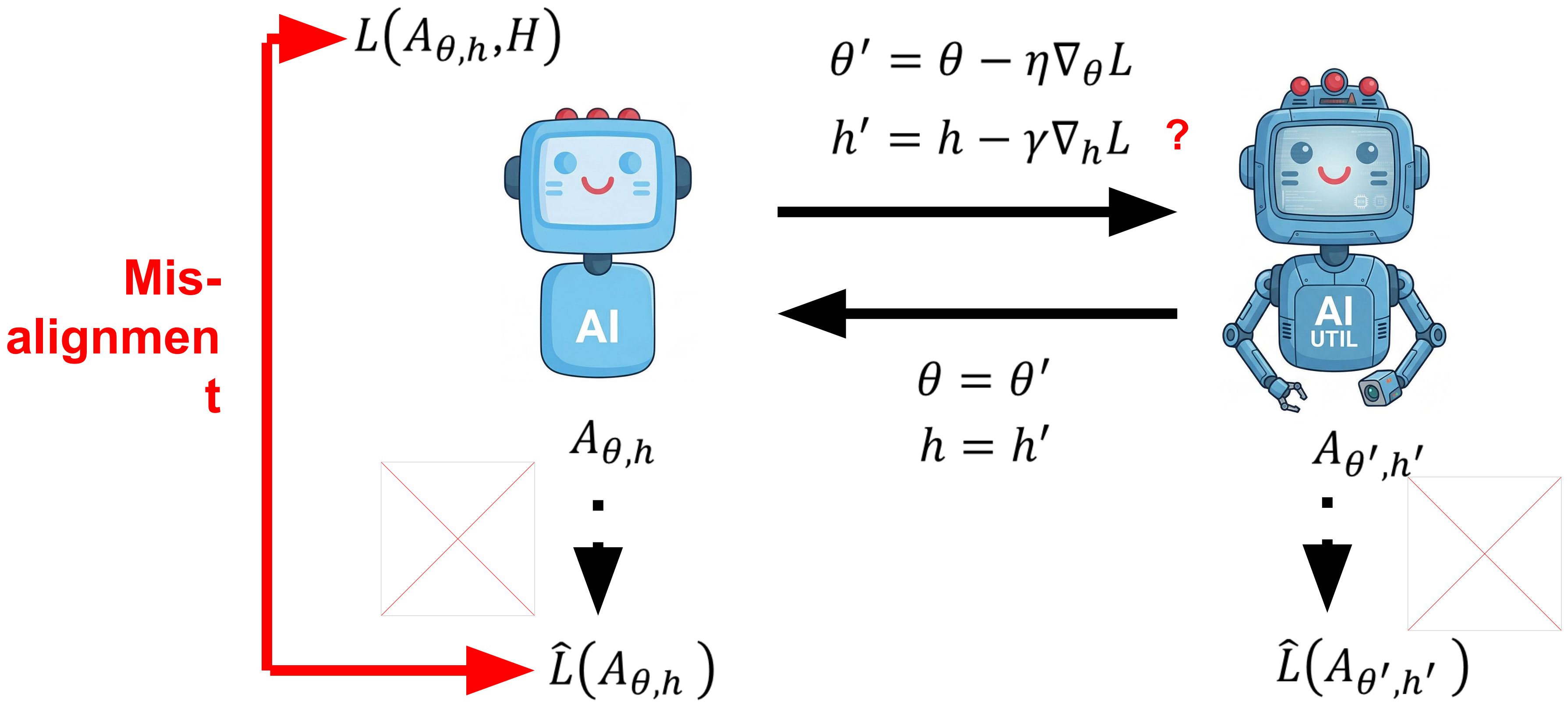
https://arxiv.org/abs/1705.05363
https://arxiv.org/abs/1810.12894
https://arxiv.org/abs/2505.17621
https://arxiv.org/abs/1509.08731
https://arxiv.org/pdf/2506.06725



https://youtu.be/-5cCWhu0OaM?si=-/
&t=547

DRL Lecture 7: Sparse Reward

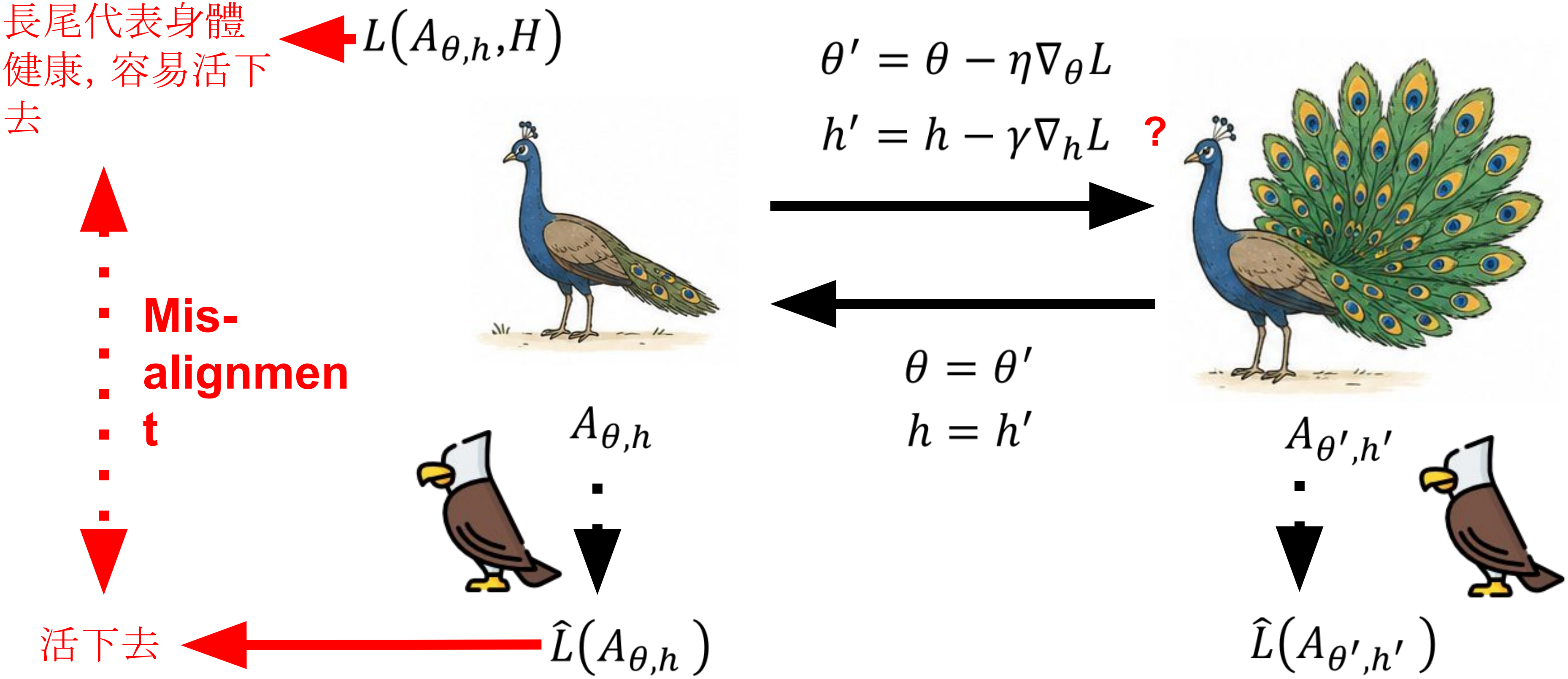
成長會失控嗎？


$L(A_{\theta,h},H)$
$\theta' = \theta - \eta\nabla_\theta L$
$h' = h - \gamma\nabla_h L$?
Mis-alignment
AI
AI UTIL
$\theta = \theta'$
$h = h'$
$A_{\theta,h}$
$A_{\theta',h'}$
$\hat{L}(A_{\theta,h})$
$\hat{L}(A_{\theta',h'})$

成長會失控嗎？


1. 原始狀態
孔雀尾巴原本較短，
雌孔雀會偏好尾巴較長的雄孔雀。
2. 性擇作用
尾巴稍長的雄孔雀更受歡迎，
有更多交配機會，
將基因傳給下一代。
3. 持續加強
經過許多世代，
尾巴越來越長、越華麗，
但這對生存幫助不大，
甚至可能增加負擔
（例如：更容易被天敵發現、
飛行不便）。
4. 失控的演化
性擇讓尾巴不斷被放大，
即使已經不利生存，
這就是「失控的演化」
（Runaway Evolution）。
雌孔雀偏好
尾巴較長的雄孔雀
尾巴越長 → 越受歡迎 →
更多後代
尾巴持續變長，
但生存成本增加
尾巴過度誇張，
成為負擔，仍持續被偏好

成長會失控嗎？


長尾代表身體健康，容易活下去
$L(A_{\theta,h},H)$
$\theta' = \theta - \eta\nabla_\theta L$
$h' = h - \gamma\nabla_h L$?
$\theta = \theta'$
$h = h'$
Mis-alignment
$A_{\theta,h}$
$A_{\theta',h'}$
活下去
$\hat{L}(A_{\theta,h})$
$\hat{L}(A_{\theta',h'})$

https://zh.wikipedia.org/zh-tw/%E6%A9%9F%E6%A2%B0%E5%85%AC%E6%9

成長會失控嗎？

電影《機械公敵》

$\hat{L}$

(人類福祉)

H

$L(H)$