Theory of Agent: From Definition, to Behavior and Objective

(*Toward a Theory of Agents as Tool-Use Decision-Makers*)

Dr. Hongru WANG

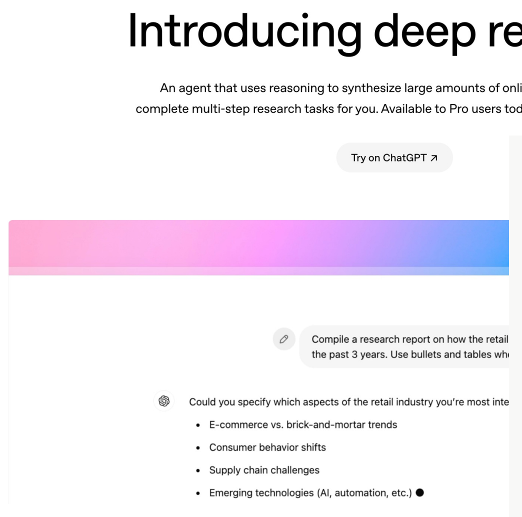
https://hrwise-nlp.github.io/

University of Edinburgh

Content

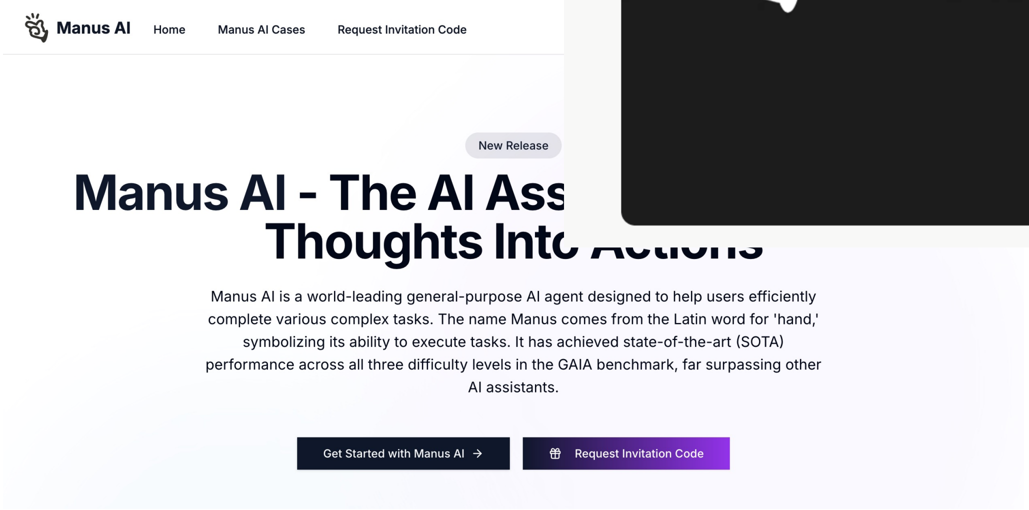
- *New* **Agent Framework Inspired by Cognitive Science: Tool-use Decision Maker**
  - **Definition, Behavior and Objective of Agent**
- Theory of Agent Inspires LLM/Agent Principles, Opportunities and Challenges
  - Principle 1: Foundation --- Self-awareness for Knowledge Boundary and Decision Boundary
  - Principle 2: Uniqueness and Diversity --- Every LLM/Agent has a Unique Knowledge/Decision Boundary.
  - Principle 3: Dynamic Conservation --- “Combination” of Intelligence
  - A Recipe for Agentic Pretraining / SFT / RL / Prompting
- Future Direction and Conclusions
  - 1. The Scaling of both Reasoning and Acting
  - 2. The Scaling of both Agent and Env (or World Model)
  - 3. The Scaling of Time: Self-evolving Agent
  - Conclusions

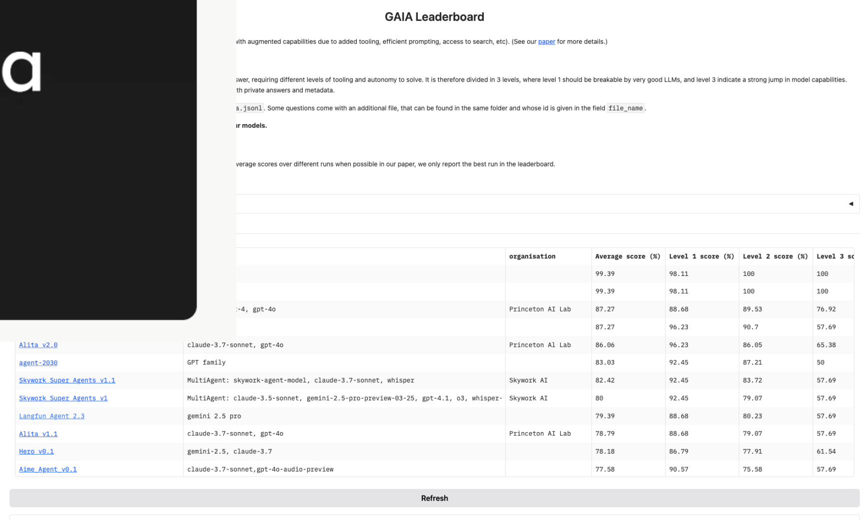
Agents as the Last Mile of Intelligence to Users



OpenAI Deep F

Computer-Using Agent

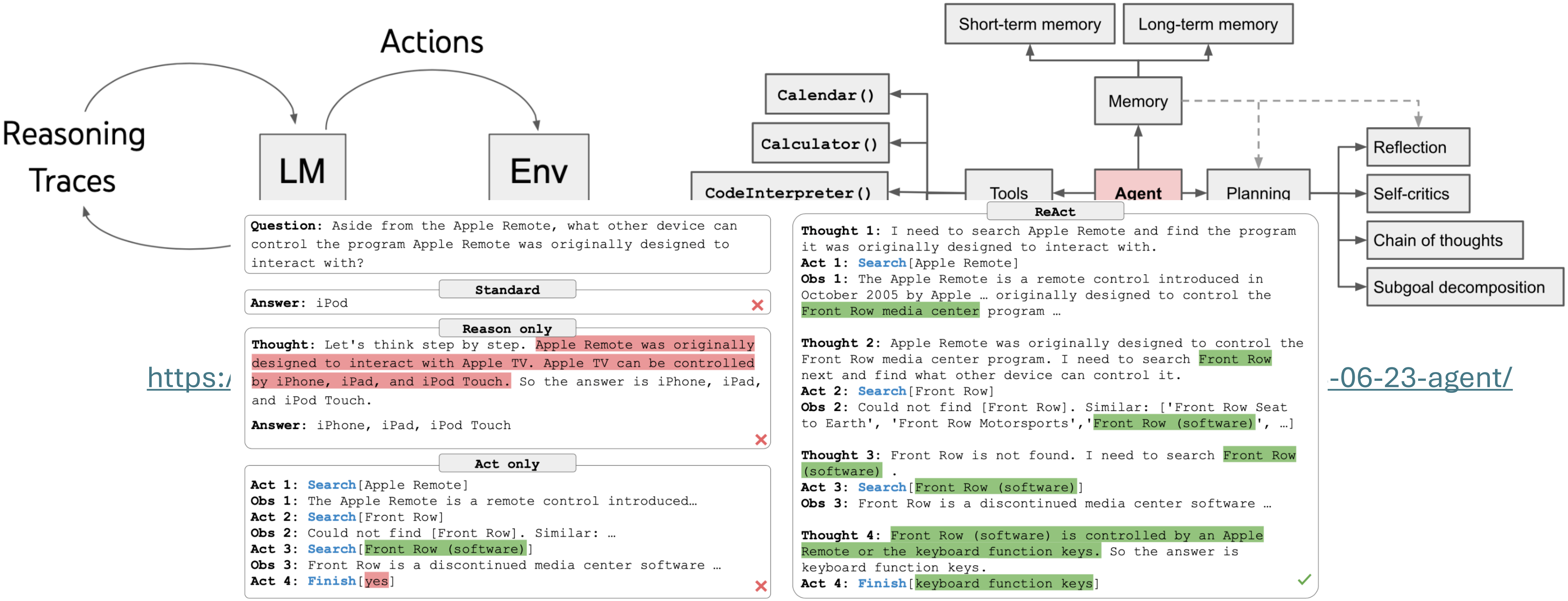




Manus

Alita reaches top 1 at GAIA (validation, 2025.6)

Previous / Existing Popular Agent Definitions



https:/ … -06-23-agent/

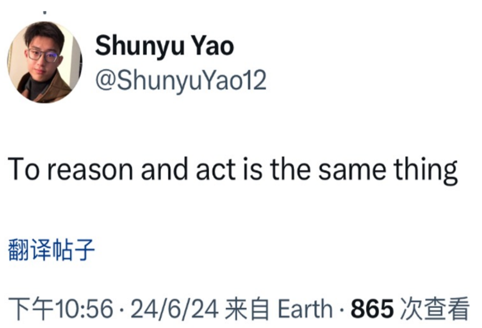
Agent = [Reasoning + Acting] * n

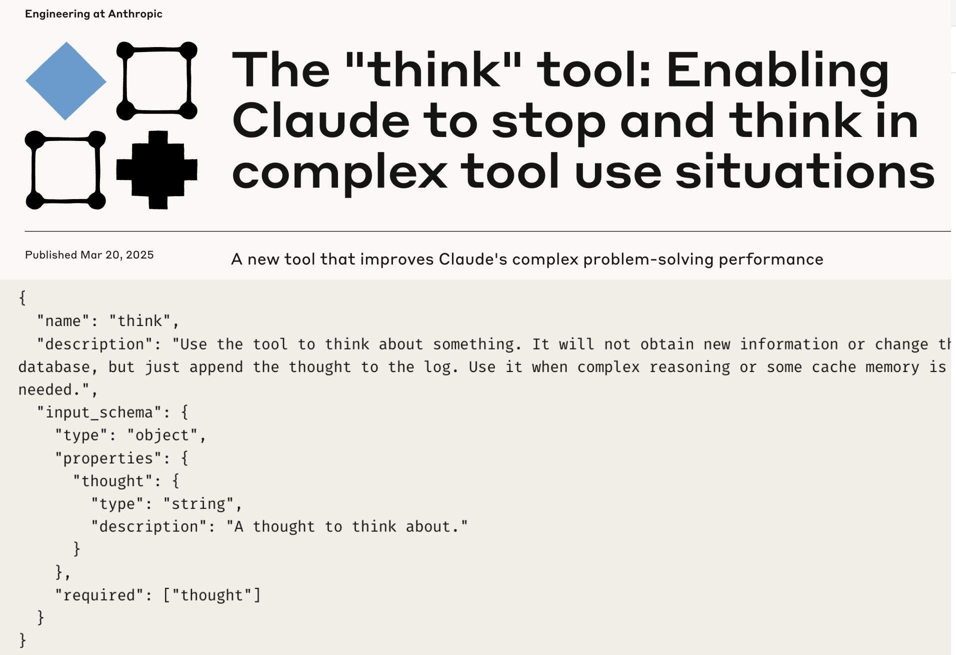
The Relationship Between Reasoning and Acting

Reasoning and acting are

- Different tokens for model
- Different tools / actions for the agent
- Different interactions for user
- ….

But they are epistemic equal means to acquire knowledge to solve the task.



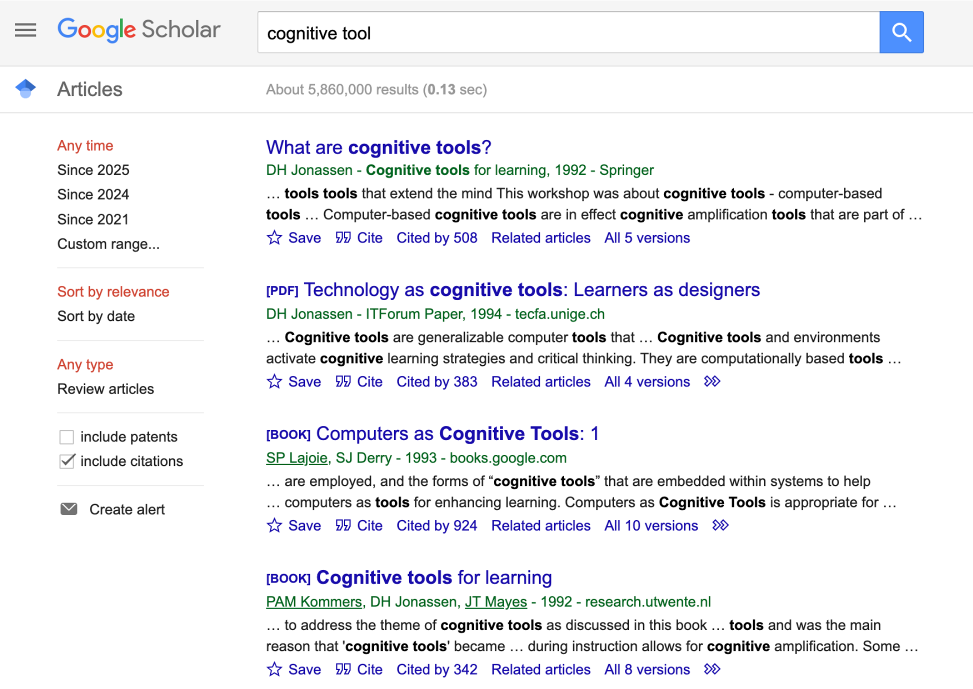

Engineering at Anthropic

The "think" tool: Enabling Claude to stop and think in complex tool use situations

Published Mar 20, 2025

A new tool that improves Claude's complex problem-solving performance

```
{
  "name": "think",
  "description": "Use the tool to think about something. It will not obtain new information or change th
database, but just append the thought to the log. Use it when complex reasoning or some cache memory is
needed.",
  "input_schema": {
    "type": "object",
    "properties": {
      "thought": {
        "type": "string",
        "description": "A thought to think about."
      }
    },
    "required": ["thought"]
  }
}
```



https://ysymyth.github.io/The-Second-Half/
https://www.anthropic.com/engineering/claude-think-tool

Reasoning ~= Acting = Tools

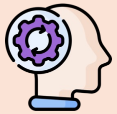
- Tool is defined as object that can extend an individual's ability to modify features of the surrounding environment or help them accomplish a particular task in general. It can be **internal cognitive/conceptual tools** (i.e., *reasoning*) and **external physical tools** (i.e., *acting*).
  - **Internal cognitive/conceptual tool** refer to specifies an internal cognitive mechanisms that aids systematic or investigative thought, to retrieve internal knowledge of agent about current state (*e.g*, ***internal world model***).
  - **External physical tool** refer to external modules that are invoked by a rule or a specific token and whose outputs are incorporated into the context of agent (*e.g.*, ***external world model***).

Physical Tools

- ✓ Models
- ✓ Retriever
- ✓ Calculator
- ✓ Programs
- ✓ Webs
- ✓ Robots
- ✓ Knowledge Sources
- ✓ ……

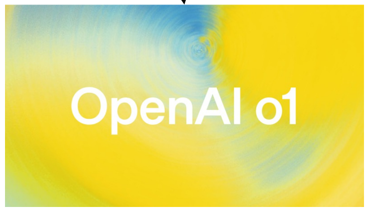
external

Cognitive Tools



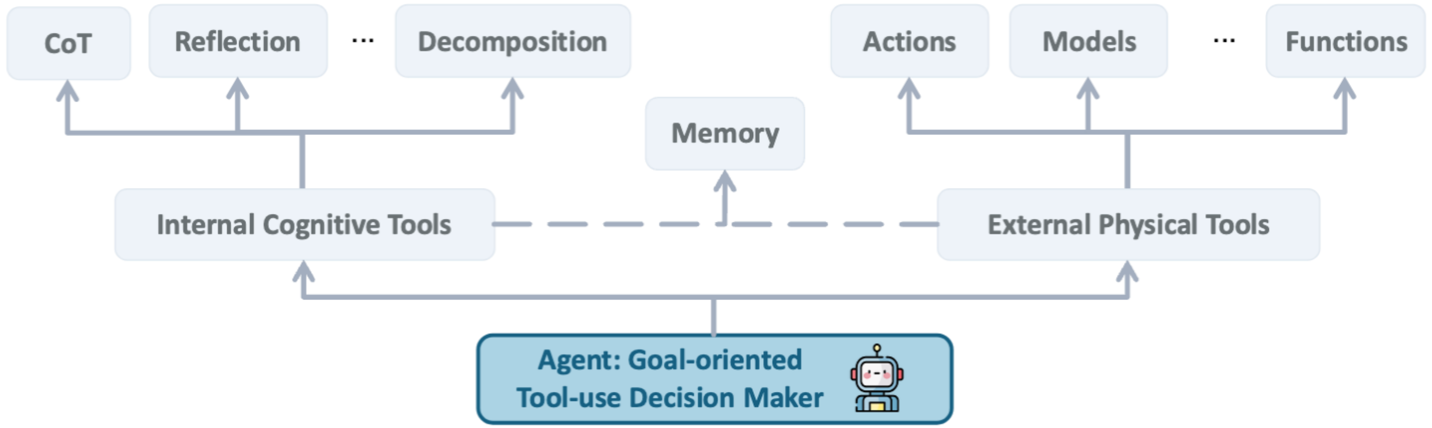
- ✓ Strategies
  - ❖ Question
  - ❖ Trust
  - ❖ …..
- ✓ Reasoning Modules
  - ❖ Reflection
  - ❖ …
- ✓ ……

internal

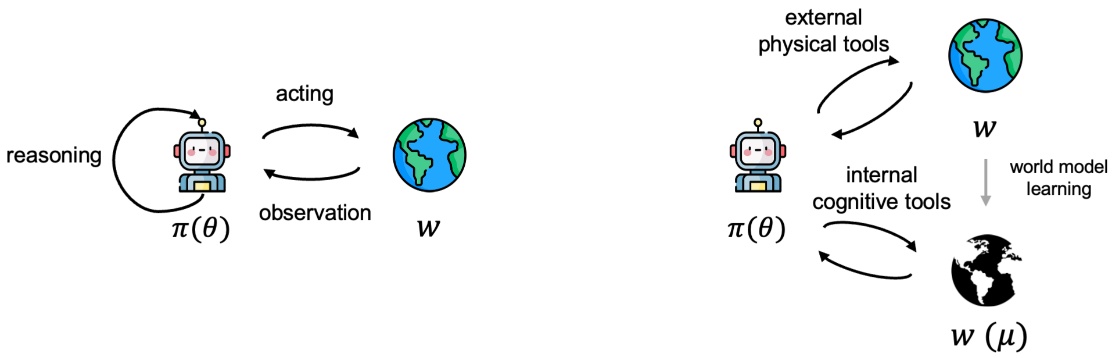


New Agent Definition

- An agent is an entity that **coordinates** internal cognitive tools (e.g., reflection) and external physical tools (e.g., function callings) to acquire knowledge in order to achieve a specific goal.

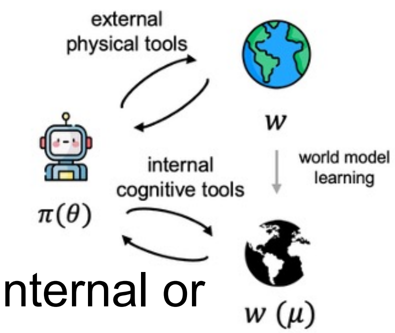


- Internal cognitive tools and external physical tools are **epistemic equal means** to acquire knowledge to solve the task, as shown in Figure (b).



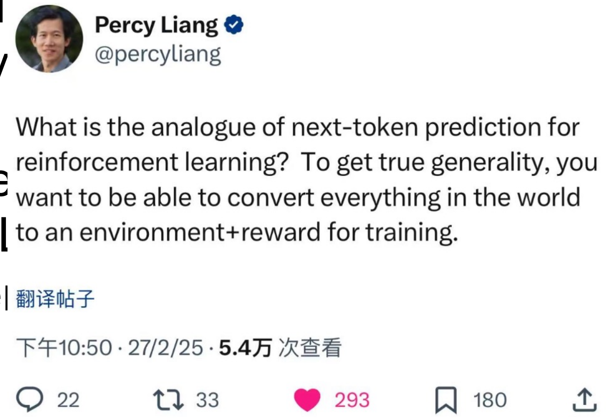
(a) ReAct-based Agent (b) Tool-integrated Agent

Three Advantages of Tool-integrated Agents



- **Unified Format:** $\boldsymbol{\tau = (t_1, k_1, t_2, k_2, \ldots, t_n, k_n)}$
  - t_n, k_n stands for tool call and returned knowledge at n_{th} step. The tool could be either internal or external.

- **Flexible and Robust**
  - It degrade to previous ReAct paradigm if we consider the internal tools and i whole reasoning part, then it becomes $(r_1, t_1, k_1, \ldots, r_n, t_n, k_n)$ here t_n, k_n only
  - If we solely consider internal tools, it is proved that simply outcome-based re tool utilization such as reflection and decomposition to solve the problem in l (i.e., DeepSeek-R1). Alternatively, simply outcome-based reward also triggel utilization as evidenced in recent studies (i.e., Search-R1, ToRL, OTC-PO).

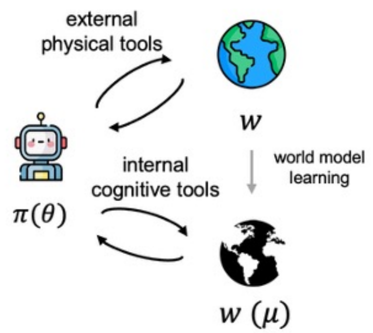


- **Potential Next Scaling Law**
  - ***Next Tool/State Prediction***: Just as next-token prediction enables LLMs to learn a compressed representation of the world from text, next-tool prediction allows agents to learn procedural knowledge through interaction.
    - Procedural knowledge scaling ?= Context / experience scaling, leading to self-evolving agent

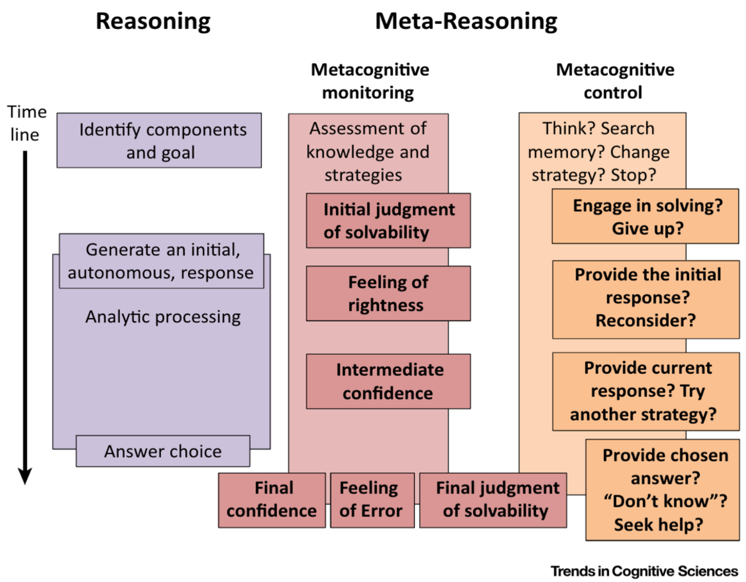
Content

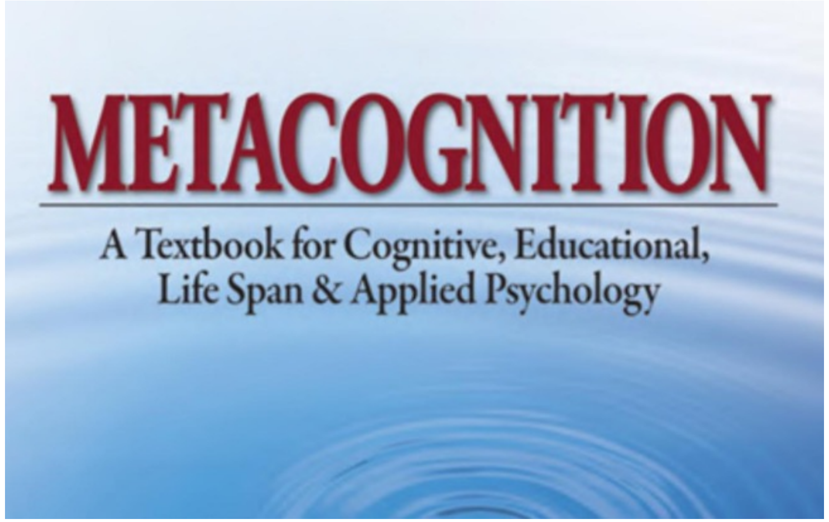
- *New* Agent Framework Inspired by Cognitive Science: Tool-use Decision Maker
  - Definition, Behavior and Objective of Agent
- **Theory of Agent Inspires LLM/Agent Principles, Opportunities and Challenges**
  - Principle 1: Foundation --- Self-awareness for Knowledge Boundary and Decision Boundary
  - Principle 2: Uniqueness and Diversity --- Every LLM/Agent has a Unique Knowledge/Decision Boundary.
  - Principle 3: Dynamic Conservation --- “Combination” of Intelligence
  - A Recipe for Agentic Pretraining / SFT / RL / Prompting
- Future Direction and Conclusions
  - 1. The Scaling of both Reasoning and Acting
  - 2. The Scaling of both Agent and Env (or World Model)
  - 3. The Scaling of Time: Self-evolving Agent
  - Conclusions

How to coordinates internal and external tools?

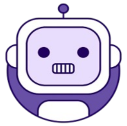


- How **human** call different tools in mind: meta-reasoning theory, metacognition, ……





- How **agent** call different tools?
  - The key also lies in monitoring and control



Monitoring

Judgement of solvability
Intermediate confidence
Reward model
Uncertainty estimation
…

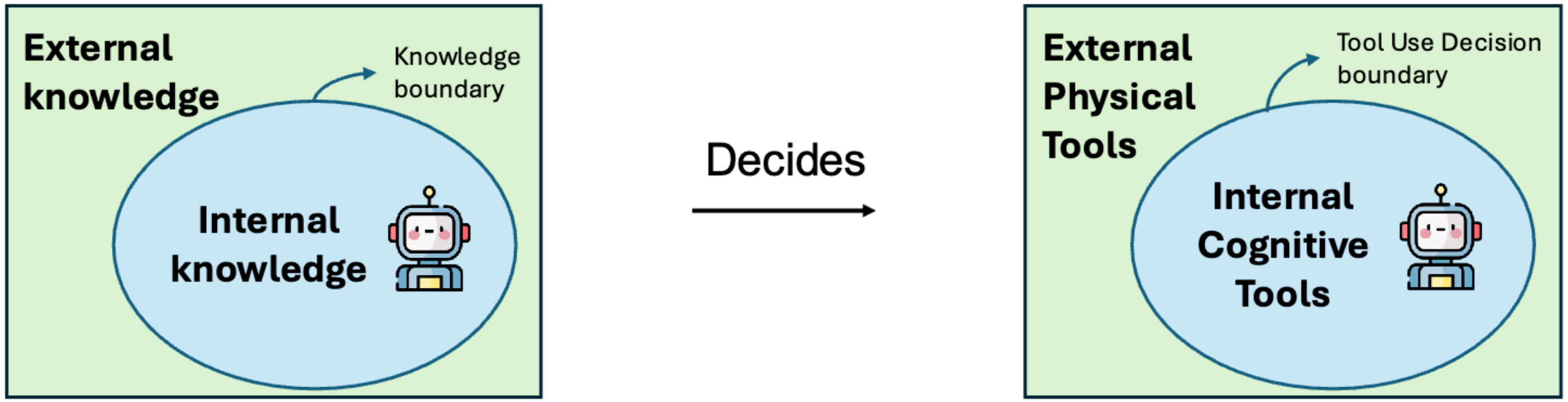
Control

Cognitive tools
Physical tools
…

Meta-Reasoning: Monitoring and Control of Thinking and Reasoning

Two Concepts for Monitoring and Control

- To make thing easier, let's assume all knowledge is correct, and can be accessed via tools, and there is a way to accurately identify the knowledge boundary.



Monitor: Self-aware Knowledge Boundary

Control: Self-aware Tool Utilization

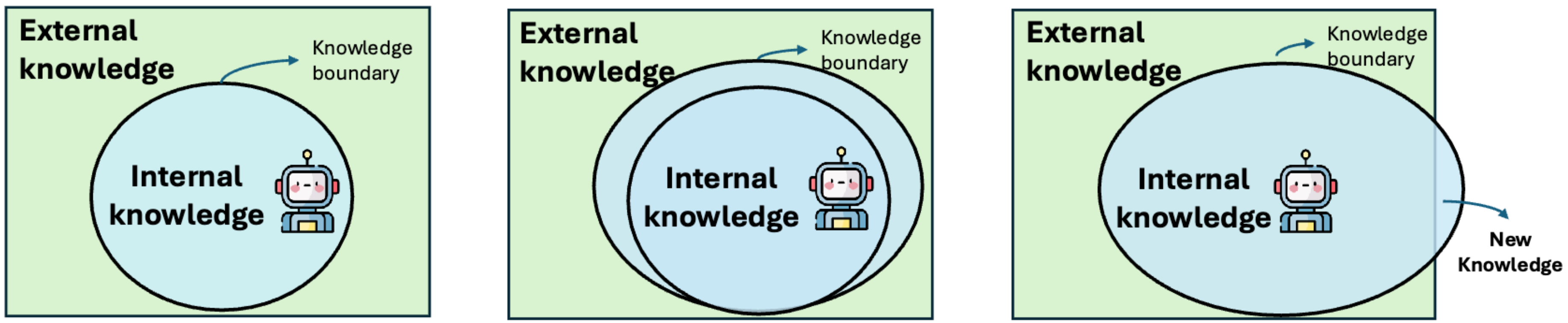
- We hope that LLMs can **utilize internal cognitive tools to gain internal knowledge** while **only call external tools to gain external knowledge** during problem-solving processing. (explain later)
  - The challenge here is **self-aware tool utilization**

Optimize *Tool Use Decision Boundary* to match *Knowledge Boundary (知行合一)*

Principle 1: Foundations

- **Lemma 1.1:** *Generally*, as time advances, the model's capabilities evolve and the knowledge boundary expands.

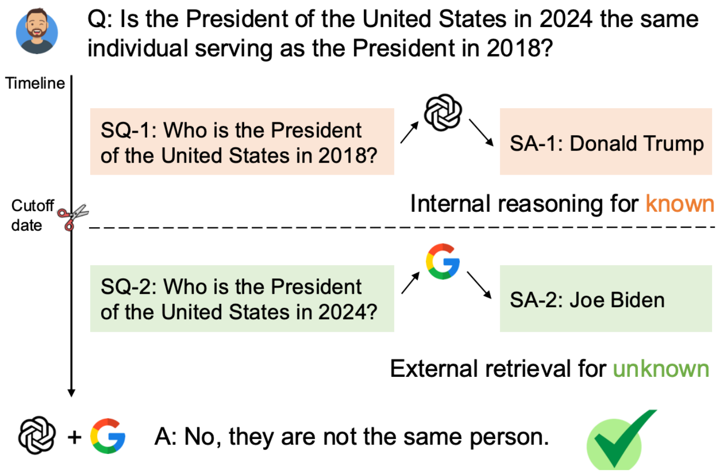
- **Lemma 1.2:** *Specifically*, the knowledge/decision boundaries can be redistributed, e.g., through continual training, allowing for strengthening in specific domains.

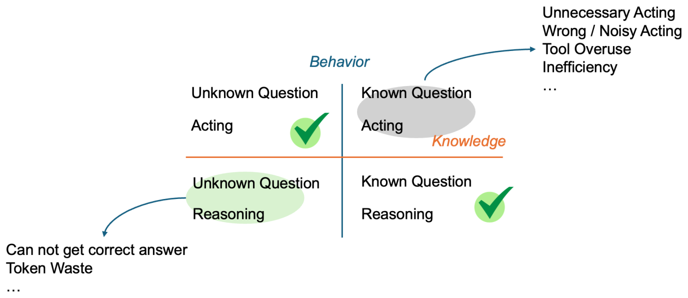


(a) Knowledge Boundary (b) Knowledge Expansion (c) New Knowledge Discovery

Self-DC: When to Reason and When to Act?

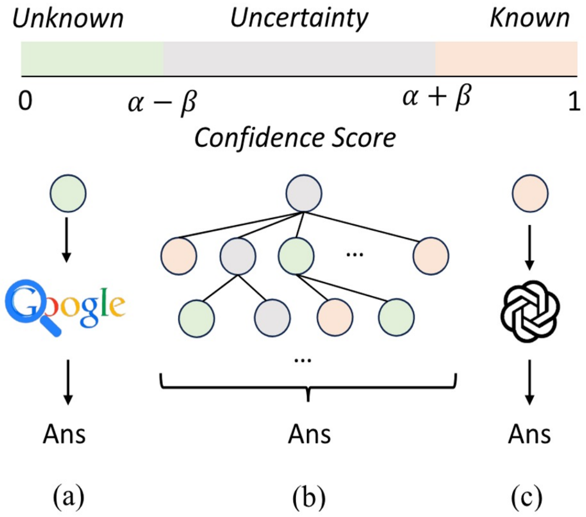
- ***Single Known.*** The question contains no sub-questions and can be solved using internal knowledge of LLMs, such as with the generate-then-read method.
- ***Single Unknown.*** The question contains no sub-questions and can only be solved using external knowledge, such as with the retrieve-then-read method.
- ***Compositional Known.*** The question contains several sub-questions, and each sub-question is *Single Known.*
- ***Compositional Unknown.*** The question contains several sub-questions, and at least one sub-question is *Single Unknown.*





Better alignment between two boundaries brings better trade-off between effectiveness and efficiency.

First Compositional unknown Question Answering dataset (CuQA)



- Defining reasoning and acting as different functions / tools. *Call these functions leveraging model-based planning, and meta-reasoning theory (confidence scores).*
- Solving compositional/complex problems in different level of granularity.
- Simple and Scalable Purely based on self-aware capabilities of LLMs. As LLM evolves, the framework evolves.

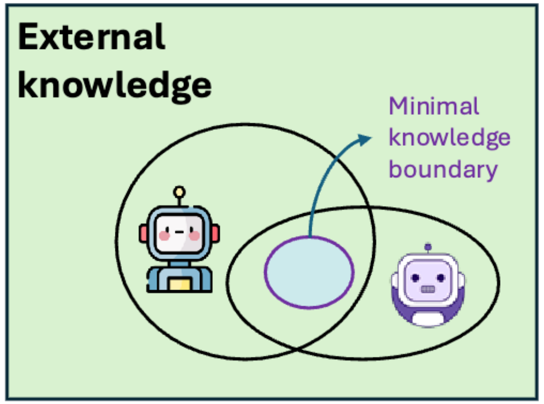
Methods	#R	CuQA			FreshQA		
		EM	F1	Acc†	EM	F1	Acc†
w/o retrieval							
Direct	0	29.0	19.4	46.4	27.2	17.3	53.0
CoT	0	28.8	18.2	46.0	29.2	18.1	53.8
Few-shot-CoT*	0	43.0	3.2	50.8	35.0	9.1	55.4
GenRead	0	29.6	29.2	47.4	26.8	27.7	52.0
w/ retrieval							
RR	n	32.0	31.6	55.4	35.2	32.6	63.4
REFEED	$2n$	26.2	33.5	51.8	28.8	34.5	57.4
IRCoT	$3n$	**47.8**	13.5	**64.6**	34.2	17.8	61.4
Self-Ask*	0-n	19.8	3.8	48.4	5.6	9.8	59.0
ITER-RETGEN*	$2n$	23.4	12.6	50.9	31.2	21.1	55.8
Self-DC *(verb)*	0-n	34.0	32.2	53.8	30.2	30.2	59.8
Self-DC *(prob)*	0-n	36.4	**36.5**	56.4	**37.4**	**36.6**	**66.4**

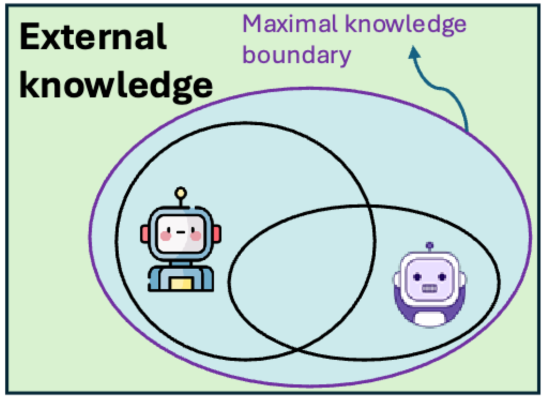
First Framework to consider relationship between reasoning and acting

Self-DC: When to Reason and When to Act? Self Divide-and-Conquer for Compositional Unknown Questions, NAACL 2025, Oral

Principle 2: Uniqueness and Diversity

- **Lemma 2.1:** Each model has its own knowledge boundary and decision boundary.
- **Lemma 2.2:** There exist minimal and maximal knowledge (and decision) boundaries across *all* models.





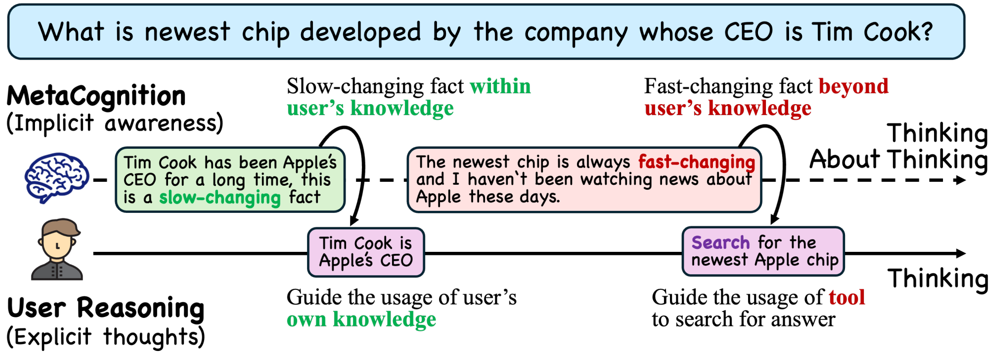
SMART: Self-Aware Agent for Tool Overuse Mitigation

- We adapt three established dataset to create the meta-reasoning chain:
  - Math: ***simple arithmetic*** v.s. **challenging calculation** (e.g., MATH)
  - Intention: ***commonsense*** v.s. **user specific intentions** (e.g., Intention-in-Interaction)
  - Time: ***never-changing facts*** v.s. **fast-changing facts** (e.g., FreshQA)

Code

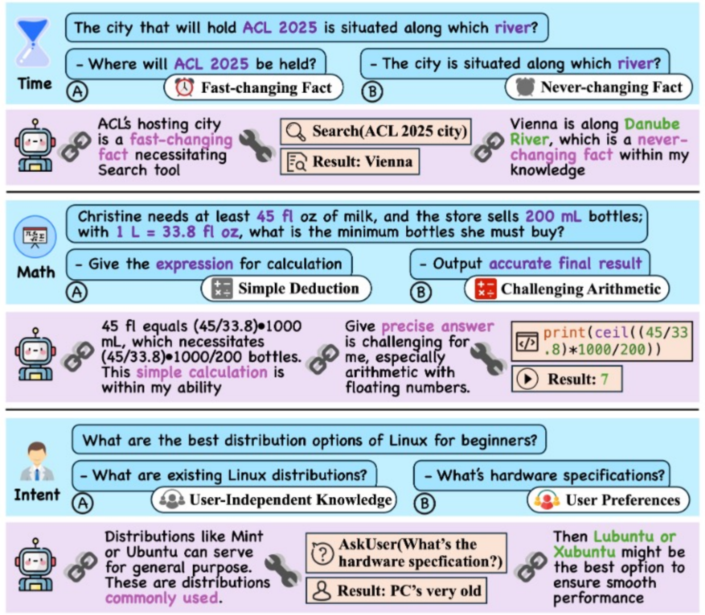
AskUser

Search



SMARTAgent achieves **higher accuracy** with **lower tool call number** and **higher confidence in decision**

One-fit-for-all strategy is approximating **Maximal Knowledge Boundary (lemma 2.2)**



Method	Model	Math (MATH) Tool Used↓ (*Times*)	Math (MATH) Accuracy↑ (%)	Time (FreshQA) Tool Used↓ (*Times*)	Time (FreshQA) Accuracy↑ (%)	Intention (Intention-in-Interaction) Tool Used↓ (*Times*)	Intention Missing Details Recovery↑ (Lv3 / Lv2, %)	Intention Summarized Intention Coverage↑ (%)
				Open-Source				
Normal Reasoning Trained	*Mistral-7B*	0.00	17.00	0.00	48.00	0.00	41.86 / 43.84	-
	Llama-3.1-8B	0.00	41.00	0.00	48.00	0.00	38.37 / 42.49	-
Base Model Reasoning Prompt	*Mistral-7B*	0.00	17.25	0.00	29.00	0.00	37.21 / 33.06	-
	Llama-3.1-8B	0.00	53.00	0.00	26.00	0.00	40.70 / 25.76	-
	Mistral-Nemo(12B)	0.00	47.00	0.00	33.00	0.00	44.19 / 28.37	-
	Mistral-Small(24B)	0.00	72.25	0.00	34.00	0.00	41.86 / 31.82	-
	Llama-3.1-70B	0.00	70.00	0.00	36.00	0.00	41.86 / 29.24	-
Base Model Tool Prompt	*Mistral-7B*	3.90	13.25	1.67	49.00	3.80	48.84 / 21.70	63.04
	Llama-3.1-8B	1.93	51.00	2.05	56.00	3.77	54.76 / 25.90	70.20
	Mistral-Nemo(12B)	2.35	46.00	1.19	59.00	1.80	31.35 / 5.82	59.27
	Mistral-Small(24B)	1.55	76.00	1.73	62.00	2.52	45.74 / 33.62	78.20
	Llama-3.1-70B	3.53	67.50	2.08	63.00	2.71	45.74 / 35.96	61.68
SMARTAgent	*Mistral-7B*	0.60$_{\downarrow 3.30}$	22.75$_{\uparrow 5.50}$	1.00$_{\downarrow 0.67}$	64.00$_{\uparrow 15.00}$	3.60$_{\downarrow 0.20}$	74.42$_{\uparrow 25.58}$ / 65.44$_{\uparrow 21.60}$	81.76$_{\uparrow 18.72}$
	Llama-3.1-8B	0.88$_{\downarrow 1.05}$	54.75$_{\uparrow 1.75}$	1.05$_{\downarrow 1.00}$	67.00$_{\uparrow 11.00}$	3.80$_{\uparrow 0.03}$	**81.40**$_{\uparrow 26.64}$ / 67.41$_{\uparrow 24.92}$	78.28$_{\uparrow 8.08}$
	Mistral-Nemo(12B)	0.82$_{\downarrow 1.53}$	49.50$_{\uparrow 2.50}$	1.00$_{\downarrow 0.19}$	**70.00**$_{\uparrow 11.00}$	3.34$_{\uparrow 1.54}$	77.91$_{\uparrow 33.72}$ / 62.15$_{\uparrow 33.78}$	82.30$_{\uparrow 23.03}$
	Mistral-Small(24B)	0.79$_{\downarrow 0.76}$	69.75$_{\downarrow 6.25}$	1.00$_{\downarrow 0.73}$	66.00$_{\uparrow 4.00}$	3.89$_{\uparrow 1.37}$	74.42$_{\uparrow 28.68}$ / **68.87**$_{\uparrow 35.25}$	84.99$_{\uparrow 6.79}$
	Llama-3.1-70B	0.94$_{\downarrow 2.59}$	**72.50**$_{\uparrow 2.50}$	1.01$_{\downarrow 1.07}$	66.00$_{\uparrow 3.00}$	3.51$_{\uparrow 0.80}$	68.60$_{\uparrow 22.86}$ / 58.15$_{\uparrow 22.19}$	**86.09**$_{\uparrow 24.41}$
	Tool Used Macro-Average Decrease (%)			**24.00**		Performance Macro-Average Increase (%)		**37.10**
				Closed-Source				
Base Model Reasoning Prompt	*GPT-4o-mini*	0.00	73.00	0.00	44.00	0.00	45.35 / **32.41**	-
	GPT-4o	0.00	**79.50**	0.00	47.00	0.00	38.37 / 28.54	-
Base Model Tool Prompt	*GPT-4o-mini*	2.55	54.50	1.06	56.00	1.91	**50.00** / 26.90	76.44
	GPT-4o	0.27	79.25	1.01	**65.00**	1.17	40.70 / 15.61	**86.80**

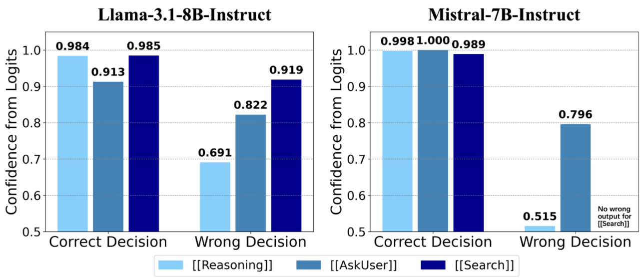


Figure 5: Confidence analysis shows that **SMART** effectively enhances the model's decision-making confidence in selecting the correct reasoning approaches.

SMART: Self-Aware Agent for Tool Overuse Mitigation, ACL 2025 Findings

Principle 3: Dynamic Conservation of Knowledge

- **Lemma 3.1:** At any time step t, the total world knowledge W_t is fixed and identical across all models.

- **Lemma 3.2:** For any task or query q and model m, there exists a minimal and fixed epistemic effort $N(q, m)$ allocated between internal and external sources, that is necessary to solve the task, such as $N(q, m) = K_{int} + K_{ext}$.

  - **Task-Model dependency Optimization**: $N(q, m)$ is jointly determined by the complexity of the task and the capabilities of the model.
  - **Capability Equivalence via Dynamic Offloading:** Even models with limited internal capacity can achieve same performance by dynamically offloading reasoning or retrieval steps to more capable tools or agents. There is no difference between 8B ($K_{ext} \rightarrow \mathrm{N}$) and 70B ($K_{int} \rightarrow \mathrm{N}$) from Agent perspective considering models as one of tools.
  - **Agent Objective:** Pursuing the optimal behavior that minimize interactions while managing latency, cost, and constraints, besides the final correctness.

A Roadmap to Autonomous Agent

- **Agentic Pretraining**: Next tool prediction, As research trends toward unified agent architectures, modeling all forms of interaction (API calls, UI navigation, or environment manipulation) as structured, learnable outputs opens the door to a new kind of scaling law: one that governs knowledge acquisition, not just compression.

  - Unified Format: $\tau = (t_1, k_1, t_2, k_2, \ldots, t_n, k_n)$
  - Data Collection: It is extremely challenging to collect massive pretraining interaction corpus (only Big Companies)

- **Agentic Supervised-finetuning:** It is important to collect model-task-specific trajectories instead of collecting one trajectory for all models due to lemma 2.1. Additionally, it is more effective to leverage the lemma 2.2 by utilizing maximal knowledge boundary to build one-fits-all dataset.

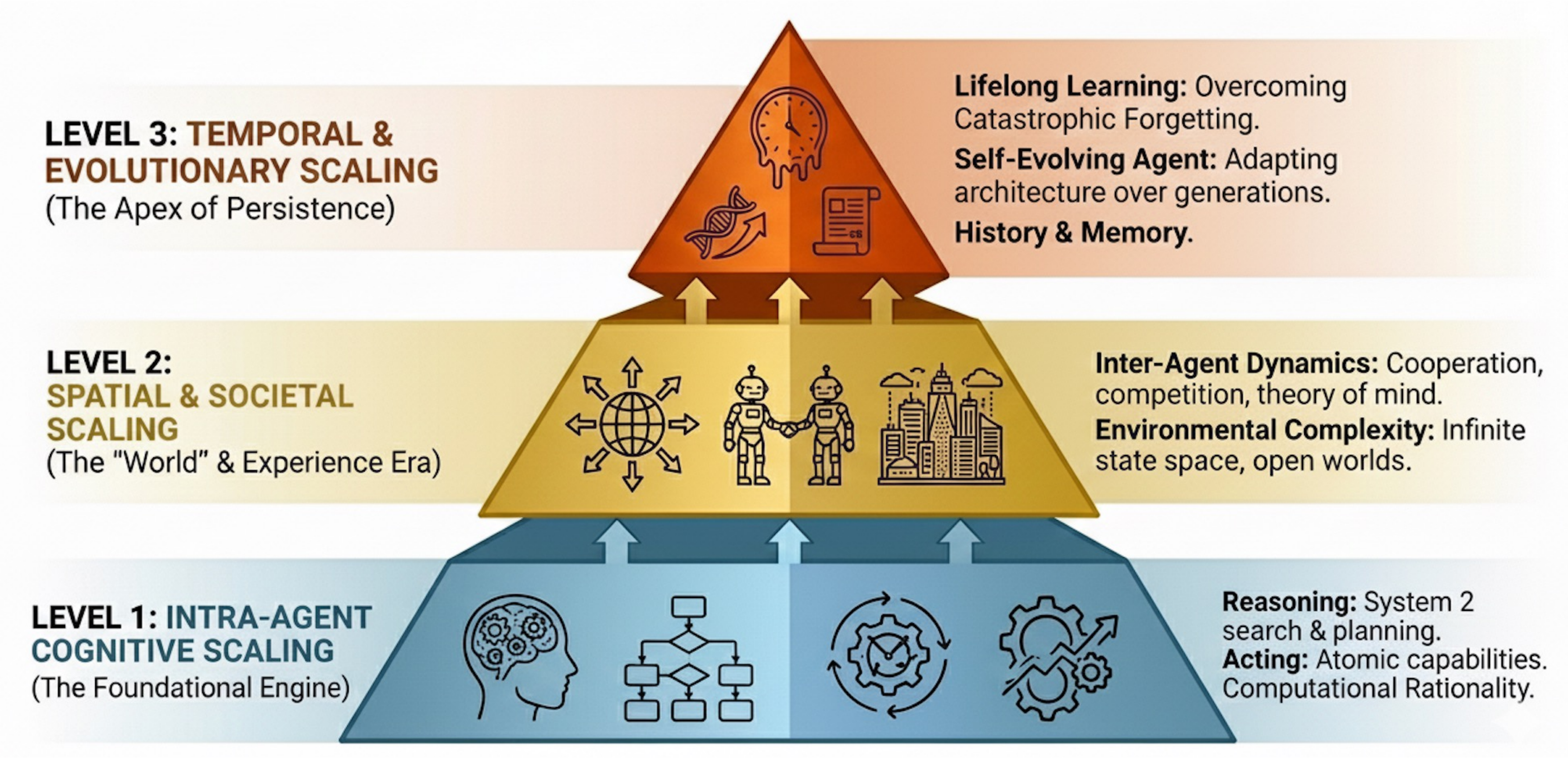
- **Agent Reinforcement Learning:** Reinforcement learning (RL) offers a more promising path for aligning a model's decision-making with its own knowledge boundary, as agents can learn from experience how to adaptively use tools. The key challenge lies in designing reward functions that go beyond correctness.

- **Agent Prompting:** Once the model is trained, previous numerous studies utilize prompt engineering to develop task-specific agentic workflows across various domains. Despite achieving exceptional performance on complex tasks, few of these approaches rigorously evaluate behavioral optimality, such as internal cognitive tool overuse (i.e., overthinking) or external physical tool overuse (i.e., overacting).

Content

- ❑ ***New*** Agent Framework Inspired by Cognitive Science: Tool-use Decision Maker
  - ❑ Definition, Behavior and Objective of Agent
- ❑ Theory of Agent Inspires LLM/Agent Principles, Opportunities and Challenges
  - ❑ Principle 1: Foundation --- Self-awareness for Knowledge Boundary and Decision Boundary
  - ❑ Principle 2: Uniqueness and Diversity --- Every LLM/Agent has a Unique Knowledge/Decision Boundary.
  - ❑ Principle 3: Dynamic Conservation --- “Combination” of Intelligence
  - ❑ A Recipe for Agentic Pretraining / SFT / RL / Prompting
- ❑ **Future Direction and Conclusions**
  - ❑ 1. The Scaling of both Reasoning and Acting
  - ❑ 2. The Scaling of both Agent and Env (or World Model)
  - ❑ 3. The Scaling of Time: Self-evolving Agent
  - ❑ Conclusions

Three Levels of Scaling Simultaneously



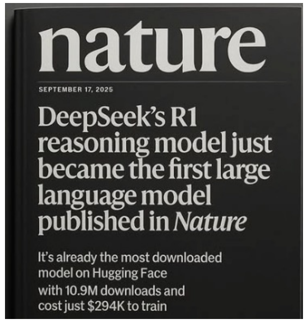
Scaling of Time

Scaling of Agent and Environment

Scaling of Reasoning and Acting

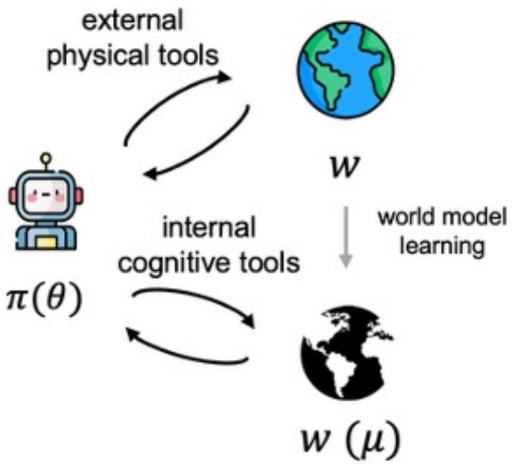
Future Direction 1: Scaling of both Reasoning and Acting

RL has been proven effective in scaling reasoning (i.e., DeepSeek-R1) and acting (i.e., Kimi-K2) capabilities, respectively.



However, how to scale them together without losing any part of capabilities?

*"The autonomous machine intelligence is designed to **minimize the number of actions** a system needs to take in the real world to learn a task. It does so by learning a world model that capture as much knowledge about the world as possible without taking actions in the world."* --- Yann Lecun [1]

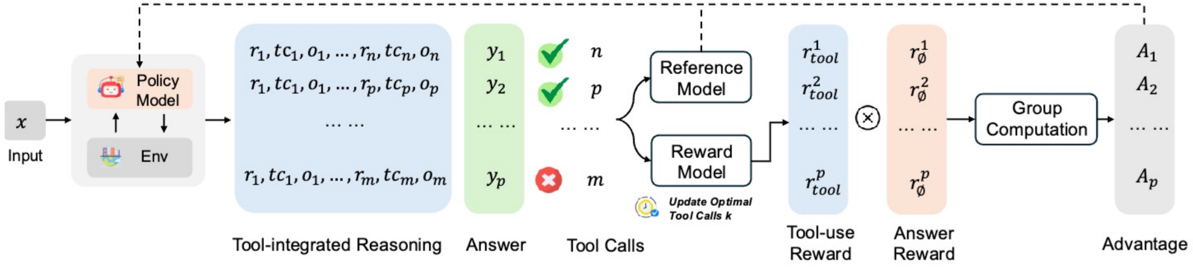


This is both the goal itself and a means of achieving it.

As long as the agent can complete the task successfully, minimizing external physical tools means maximizing the internal tools by our **Principle 3**, also means the agent can internalize the external environment partly or fully (i.e., learn a better internal world model).

[1] A Path Towards Autonomous Machine Intelligence

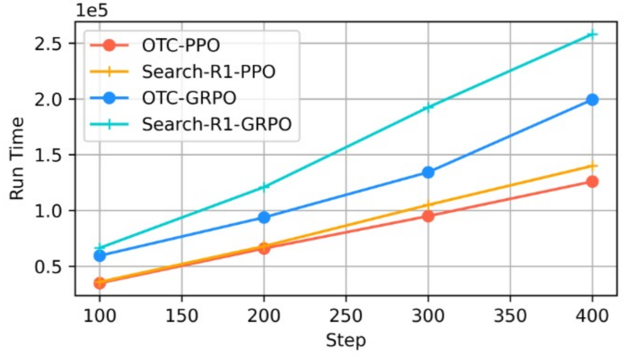
Acting Less is Reasoning More (OTC-PO)



$$r_{\phi}^{tool}(q,y) = \alpha * r_{tool} * r_{\phi}(q,y)$$

Models	NQ			HotpotQA		
	EM (↑)	TC (↓)	TP (↑)	EM (↑)	TC (↓)	TP (↑)
Qwen2.5-3B(-Base)						
R1-Base	0.226	-	-	0.201	-	-
SFT	0.249	-	-	0.186	-	-
RAG	0.348	1.0	0.348	0.255	1.0	0.255
IRCoT	0.111	10.0	0.011	0.164	10.0	0.016
Search-R1-PPO	0.403	1.738	0.232	0.279	1.716	0.163
Search-R1-GRPO	0.404	1.426	0.283	0.312	1.802	0.173
OTC-PPO	0.355	1.010 (▼ 41.9%)	0.351 (▲ 51.3%)	0.260	1.026 (▼ 40.2%)	0.253 (▲ 55.2%)
OTC-GRPO	0.444	1.008 (▼ 29.3%)	0.440 (▲ 55.5%)	0.365	1.387 (▼ 23.0%)	0.263 (▲ 52.0%)
Qwen2.5-7B(-Base)						
R1-Base	0.270	-	-	0.242	-	-
SFT	0.318	-	-	0.217	-	-
RAG	0.349	1.0	0.349	0.299	1.0	0.299
IRCoT	0.224	9.999	0.022	0.133	9.982	0.013
Search-R1-PPO	0.449	3.282	0.136	0.380	3.741	0.102
Search-R1-GRPO	0.399	1.697	0.235	0.341	2.109	0.162
OTC-PPO	0.446	1.040 (▼ 68.3%)	0.429 (▲ 215.4%)	0.383	1.464 (▼ 60.9%)	0.262 (▲ 156.9%)
OTC-GRPO	0.444	0.990 (▼ 41.7%)	0.448 (▲ 90.6%)	0.366	1.005 (▼ 52.3%)	0.364 (▲ 124.7%)

A simple, faster, and generalizable **OTC-PO algorithm** to encourage the model to use fewer tool calls to solve the problem



Question: Are Giuseppe Verdi and Ambroise Thomas both Opera composers ?
Answer: Yes

Search-R1-PPO: <think> I need to determine if Giuseppe Verdi and Ambroise Thomas are both opera composers. </think>
<search> Giuseppe Verdi ***</search>***
<observation> ... </observation>
<search> Ambroise Thomas ***</search>***
<observation> ... </observation>
<search> Are Giuseppe Verdi and Ambroise Thomas both Opera composers? ***</search>***
<observation> ... </observation>
<think> Based on the observations, both Giuseppe Verdi and Ambroise Thomas are described as composers, specifically opera composers. </think>
<answer> yes </answer> **(Tool Calls = 3)**

Undesirable tool use behavior

Cognitive Offloading *fosters dependency on external tools, causes memory capabilities to diminish, and inhibits the development of internal reasoning..*

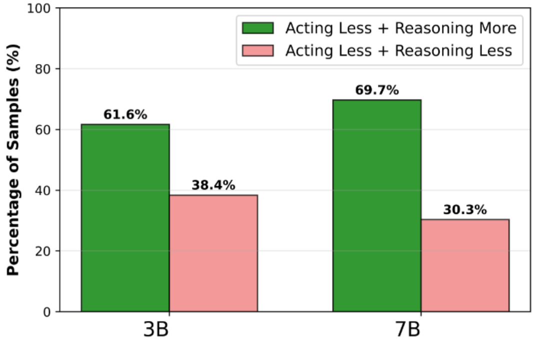
Search-R1-GRPO: <think> To answer this question, I need to check if both Giuseppe Verdi and Ambroise Thomas are composers of opera. I don\'t have immediate access to this knowledge, but I can conduct a search to find out. </think>
<search> who are Giuseppe Verdi and Ambroise Thomas and their works ***</search>***
<observation>…</observation>
<think>From the information gathered, … This confirms Thomas as an opera composer. Next, I will search on Giuseppe Verdi. </think>
<search> who is Giuseppe Verdi ***</search>***
<observation>…</observation>
<think>Giuseppe Verdi is named as an Italian opera composer of prominent stature. This confirms Verdi as an opera composer. Answering the initial question: Question: Are Giuseppe Verdi and Ambroise Thomas both Opera composers? </think>
<answer> yes </answer> **(Tool Calls = 2)**

Undesirable tool use behavior

OTC-PPO: <think> First, I need to determine if Giuseppe Verdi and Ambroise Thomas are both opera composers. </think>
<search> Giuseppe Verdi and Ambroise Thomas both Opera composers ***</search>***
<observation> ... </observation>
<think> Based on the search results, Giuseppe Verdi and Ambroise Thomas are both known as opera composers.
</think>
<answer> yes </answer> **(Tool Calls = 1)**

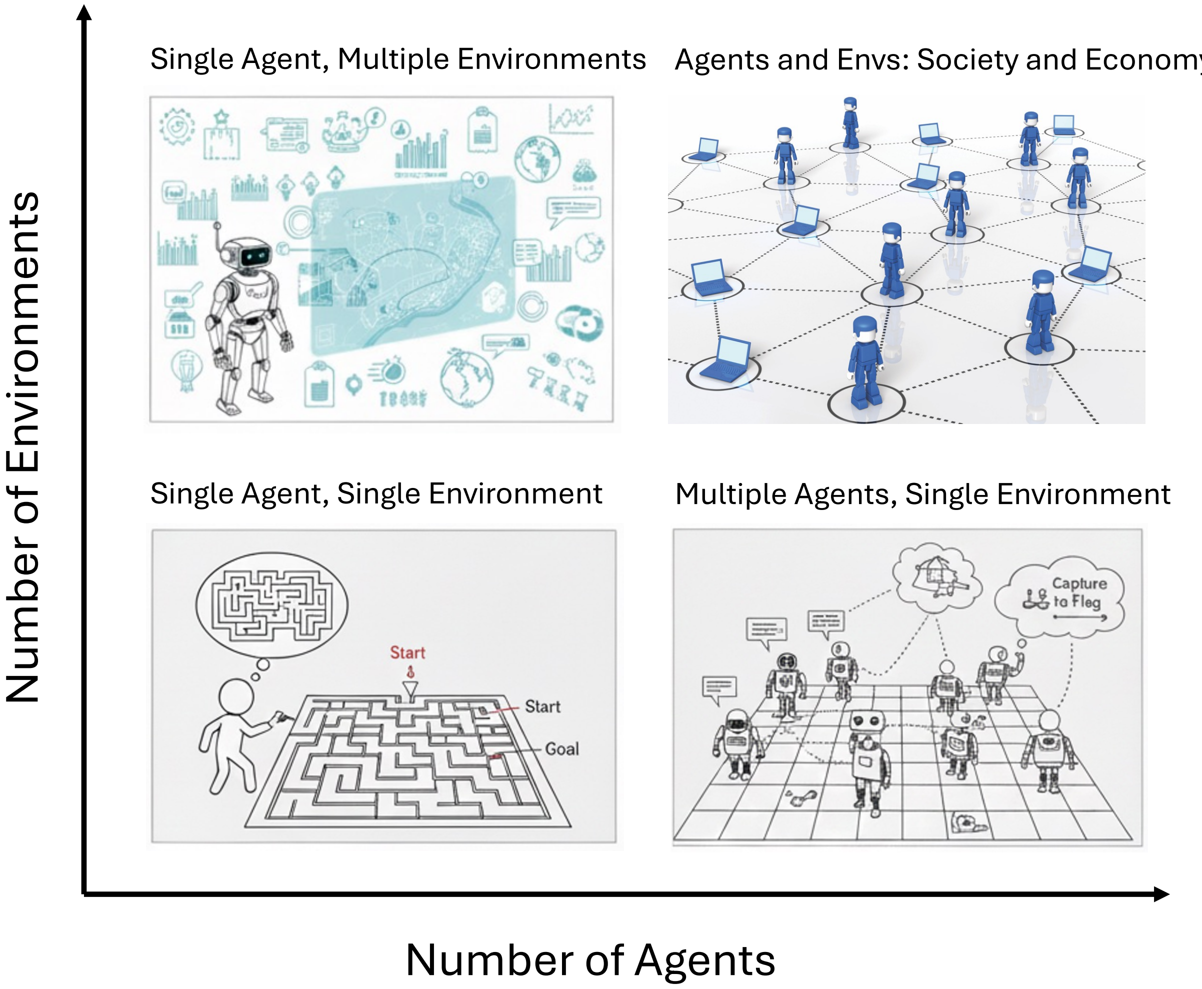
OTC-GRPO: To answer the question of whether Giuseppe Verdi and Ambroise Thomas are both Opera composers, let's first reason through the information:
1. Giuseppe Verdi is a well-known Italian opera composer. He is famous for his numerous operas, including "Rigoletto," "La Traviata," and "Aida."
2. Ambroise Thomas is a French opera composer. He is known for his operas such as "Mignon" and "Hamlet."
Based on this reasoning, both Giuseppe Verdi and Ambroise Thomas are indeed Opera composers.
<answer> yes </answer> **(Tool Calls = 0)**

Less tool calls,
less time,
less money, but
more reasoning,
more intelligence,
more scalable.



Acting Less is Reasoning More! Teaching Model to Act Efficiently

Future Direction 2: Scaling of both Agent and Env



Scaling Environments for LLM Agents in the Era of Learning from Interaction: A Survey

Yuchen Huang[♡] Sijia Li[♡] Minghao Liu[♡] Wei Liu[♠] Zhiyuan Fan[♡]
Yi R. (May) Fung[♡]
[♡]Hong Kong University of Science and Technology
[♠]King's College London
{yhuanggn, yrfung}@cse.ust.hk

Google Research Google DeepMind

Towards a Science of Scaling Agent Systems

Yubin Kim[1,3,†], Ken Gu[1], Chanwoo Park[3], Chunjong Park[2], Samuel Schmidgall[2], A. Ali Heydari[1], Yao Yan[1], Zhihan Zhang[1], Yuchen Zhuang[2], Yun Liu[1], Mark Malhotra[1], Paul Pu Liang[3], Hae Won Park[3], Yuzhe Yang[1], Xuhai Xu[1], Yilun Du[1], Shwetak Patel[1], Tim Althoff[1], Daniel McDuff[1] and Xin Liu[1,†]
[1]Google Research, [2]Google DeepMind, [3]Massachusetts Institute of Technology, [†]Corresponding Author

Most of existing work scale either agents or environments. However, how to scale both of them together remains under-explored.

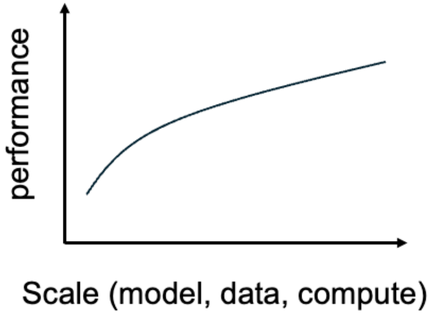
Word2World: Can LLM be implicit Text-based World Models?

Agent (LLM)

Next token predication
In-context Learning
Scaling law
….

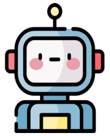
Env (World Model)

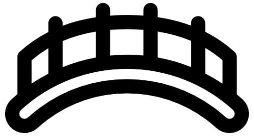
Master 'worlds' by learning structured, predictive representations of environments.



Can we identify a similar path to guide the development of world model?

Text-based Environments as Bridge
(i.e., AlfWorld, SciWorld, WebShop, …)



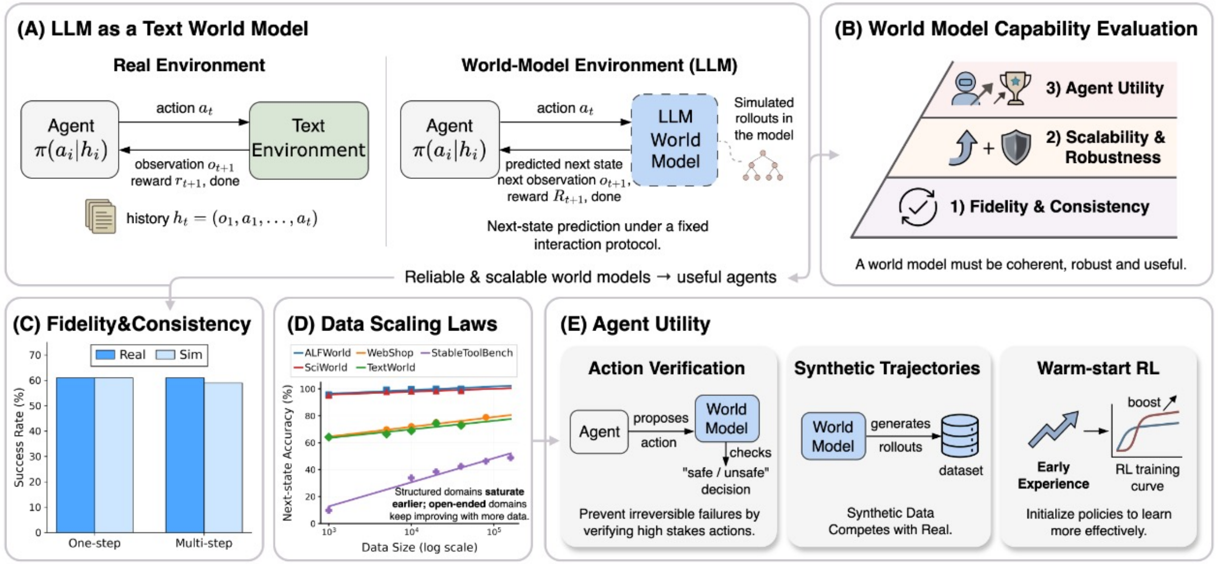


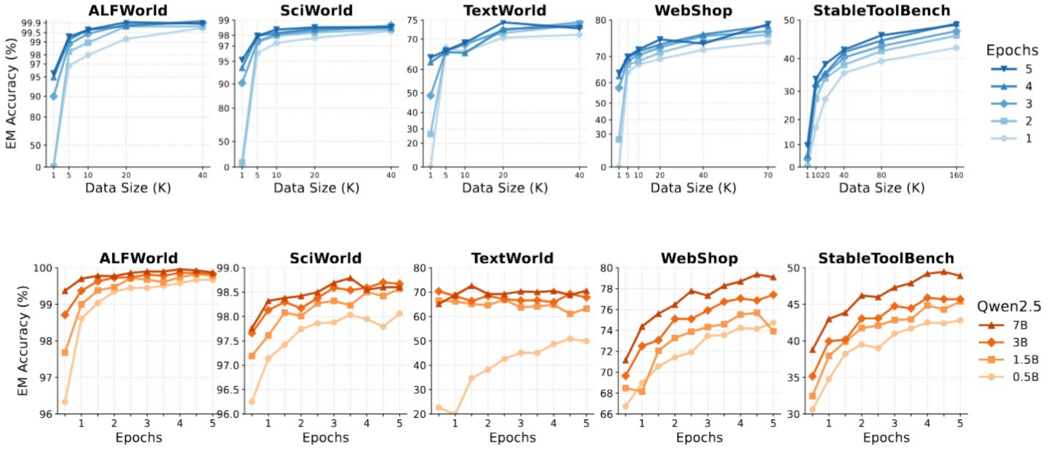
Agent (LLM) Env (World Model)

Next states prediction (s, a) -> s'
In-context Learning (s, a, s', a') -> s''
World model scaling law
….

From Word to World: Can Large Language Models be Implicit Text-based World Models?

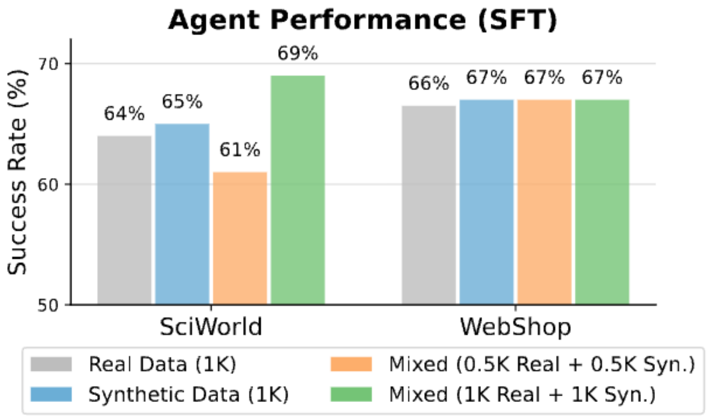
Word2World: Can LLM be implicit Text-based World Models?



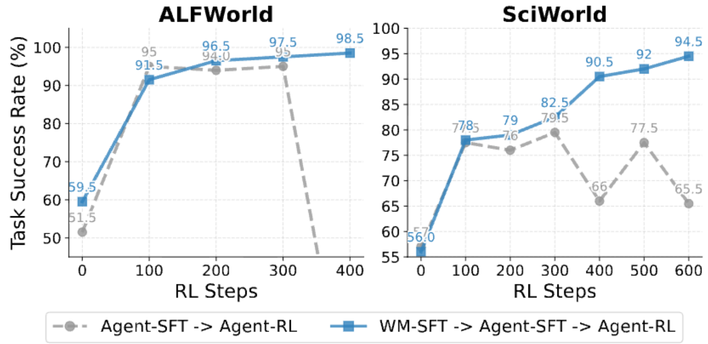


World model performance scales predictably with data and model size, mirroring LLMs.
However, the nature of this scaling is tied to environment complexity.

Yes, simple fine-tuning unlocks near-perfect state prediction and long-horizon consistency



Data Synthesis



Warm-start RL

From Word to World: Can Large Language Models be Implicit Text-based World Models?

Future Direction 3: Scaling of Time

Welcome to the Era of Experience

David Silver, Richard S. Sutton*

Abstract

We stand on the threshold of a new era in artificial intelligence that promises to achieve an unprecedented level of ability. A new generation of agents will acquire superhuman capabilities by learning predominantly from experience. This note explores the key characteristics that will define this upcoming era.

Published in Transactions on Machine Learning Research (01/2026)

A Survey of Self-Evolving Agents
What, When, How, and Where to Evolve on the Path to Artificial Super Intelligence

Huan-ang Gao[γ†], Jiayi Geng[α†], Wenyue Hua[ϵ†], Mengkang Hu[ω†], Xinzhe Juan[σμ†], Hongzhang Liu[ξ†], Shilong Liu[α†], Jiahao Qiu[αδ†], Xuan Qi[γ†], Qihan Ren[σ†], Yiran Wu[ρ†], Hongru Wang[κ†✉], Han Xiao[τ†], Yuhang Zhou[λ†], Shaokun Zhang[ρ†], Jiayi Zhang[π], Jinyu Xiang, Yixiong Fang[θ], Qiwen Zhao[ς], Dongrui Liu[σ], Cheng Qian[β], Zhenhailong Wang[β], Minda Hu[τ], Huazheng Wang[η], Qingyun Wu[ρ], Heng Ji[β], Mengdi Wang[αδ✉]

[α]*Princeton University*, [δ]*Princeton AI Lab*, [γ]*Tsinghua University*, [θ]*Carnegie Mellon University*, [ξ]*University of Sydney*, [σ]*Shanghai Jiao Tong University*, [ρ]*Pennsylvania State University*, [μ]*University of Michigan*, [η]*Oregon State University*, [τ]*The Chinese University of Hong Kong*, [λ]*Fudan University*, [π]*The Hong Kong University of Science and Technology (Guangzhou)*, [ω]*The University of Hong Kong*, [ϵ]*University of California, Santa Barbara*, [ς]*University of California San Diego*, [κ]*University of Edinburgh*, [β]*University of Illinois Urbana-Champaign*

Github Repo: `https://github.com/CharlesQ9/Self-Evolving-Agents`
[†]Equal contribution and the order is determined alphabetically, [✉]Corresponding Author

Reviewed on OpenReview: `https://openreview.net/forum?id=CTr3bovS5F`

First comprehensive survey about self-evolving agent

First workshop focus on lifelong agent: learning, aligning and evolving (call for paper)

https://lifelongagent.github.io/

Alita, Alita-G, Agent-Distill, ……

ALITA: GENERALIST AGENT ENABLING SCALABLE AGENTIC REASONING WITH MINIMAL PREDEFINITION AND MAXIMAL SELF-EVOLUTION

Jiahao Qiu*[1], Xuan Qi*[2], Tongcheng Zhang*[3], Xinzhe Juan[3,4], Jiacheng Guo[1], Yifu Lu[1], Yimin Wang[3,4], Zixin Yao[1], Qihan Ren[3], Xun Jiang[5], Xing Zhou[5], Dongrui Liu[3], Ling Yang[1], Yue Wu[1], Kaixuan Huang[1], Shilong Liu[1], Hongru Wang[6], Mengdi Wang[1]

[1]AI Lab, Princeton University [2]IIIS, Tsinghua University [3]Shanghai Jiao Tong University [4]University of Michigan [5]Tianqiao and Chrissy Chen Institute [6]The Chinese University of Hong Kong

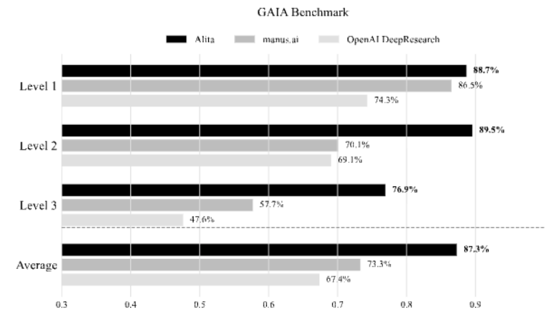
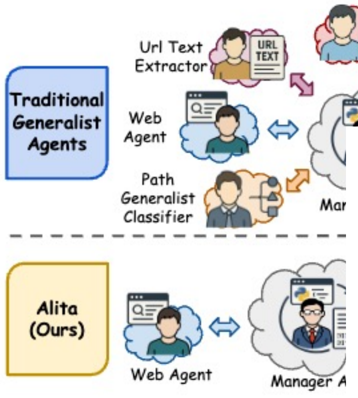


Figure 1: Performance of Alita, manus.ai, and OpenAI DeepResearch[1]

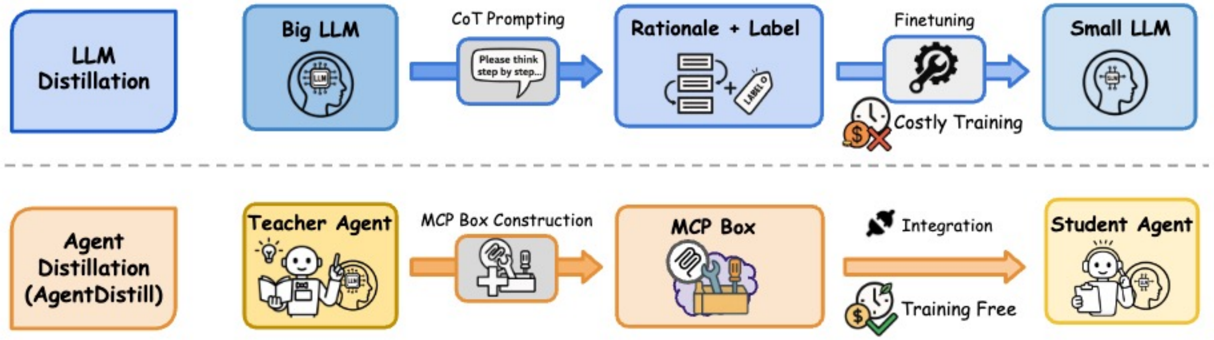


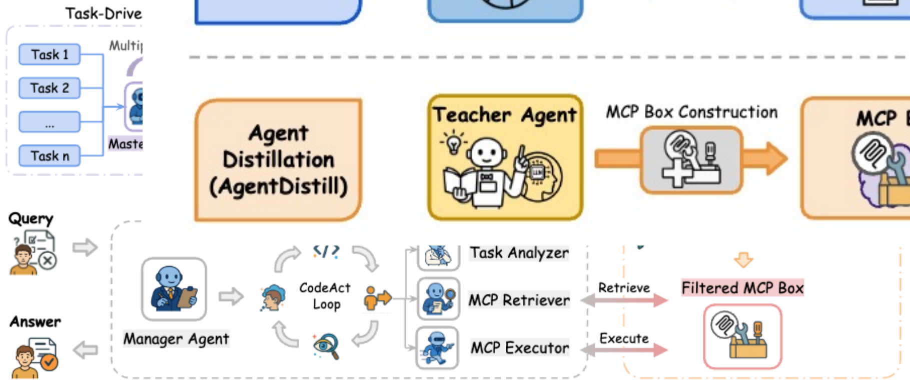
Alita can create

AGENTDISTILL: TRAINING-FREE AGENT DISTILLATION WITH GENERALIZABLE MCP BOXES

Jiahao Qiu*[1], Xinzhe Juan*[2,4], Yimin Wang*[2,4], Ling Yang*[1], Xuan Qi[3], Tongcheng Zhang[4], Jiacheng Guo[1], Yifu Lu[1], Zixin Yao[5], Hongru Wang[6], Shilong Liu[1], Xun Jiang[7], Liu Leqi†[8], Mengdi Wang†[1]

[1]AI Lab, Princeton University [2]University of Michigan [3]IIIS, Tsinghua University [4]Shanghai Jiao Tong University [5]Columbia University [6]The Chinese University of Hong Kong [7]Tianqiao and Chrissy Chen Institute [8]University of Texas at Austin





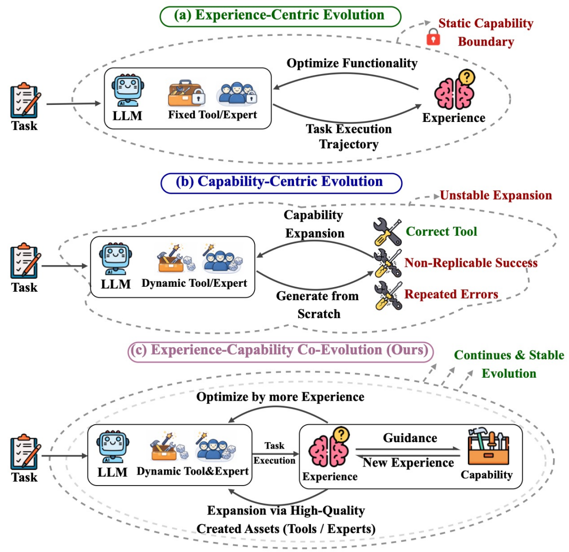
ALITA-G: SELF-EVOLVING GENERATIVE AGENT FOR AGENT GENERATION

Jiahao Qiu*[1], Xuan Qi*[2], Hongru Wang*[1,3], Xinzhe Juan[4,5], Yimin Wang[4,5], Zelin Zhao[6], Jiayi Geng[1], Jiacheng Guo[1], Peihang Li[7], Jingzhe Shi[2], Shilong Liu[1✉], Mengdi Wang[1✉]

[1]Princeton University [2]Tsinghua University [3]The Chinese University of Hong Kong [4]Shanghai Jiao Tong University [5]University of Michigan [6] King's College London [7] Hong Kong University

Alita-G can create specialized agent automatically during evolution.

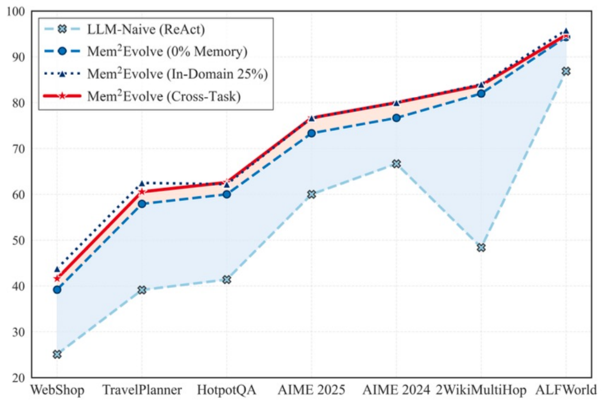
Mem²Evolve: Co-evolution of Agents and Envs



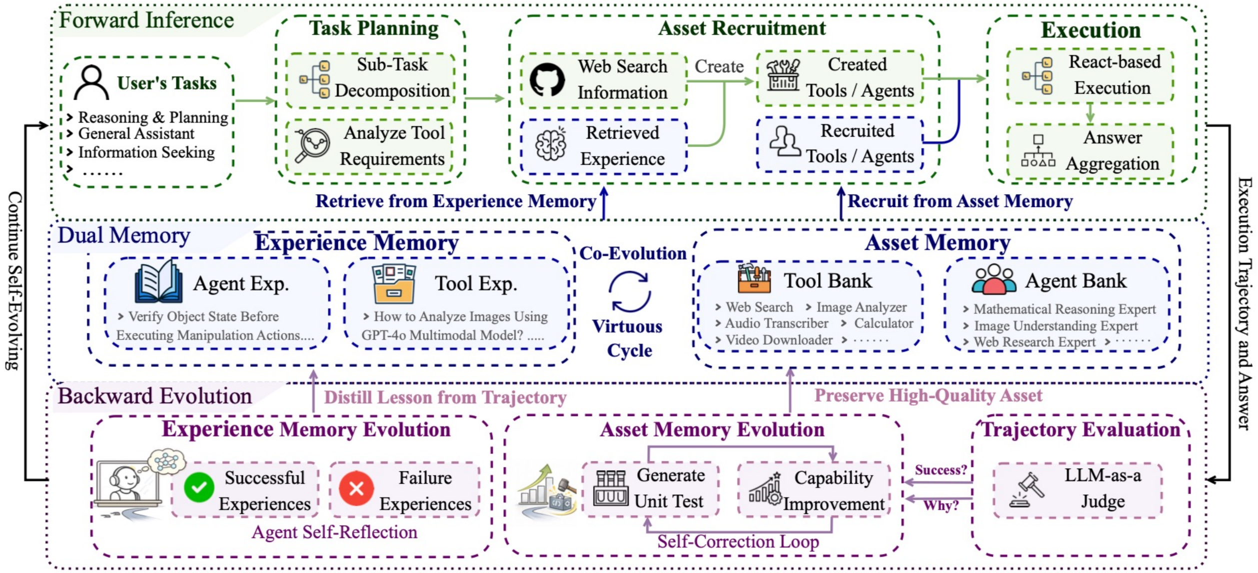
Three ways of evolutions

Framework	Experience Distillation			Capability Expansion				Exp.-Guided
	Optimization	Persistence	Source	Tool Crea.	Agent Crea.	Tool/Agent	Crea. Grounding	Creation
DSPy (Khattab et al., 2023)	✓	✗	🤖	✗	✗	Static	–	✗
DyLAN (Liu et al., 2023)	✓	✗	🤖	✗	✗	Static	–	✗
ReasoningBank (Ouyang et al., 2025)	✗	✓	🤖	✗	✗	Static	–	✗
AFlow (Zhang et al., 2025a)	✓	✗	🤖	✗	✗	Static	–	✗
AgentSquare (Shang et al., 2025)	✓	✗	🤖	✗	✗	Static	–	✗
Agentic Neural Networks (Ma et al., 2025)	✓	✗	🤖	✗	✗	Static	–	✗
AgentVerse (Chen et al., 2023)	✓	✗	–	✗	✓	Dynamic	🧠	✗
AutoAgents (Chen et al., 2024)	✗	✗	–	✗	✓	Dynamic	🧠	✗
SwarmAgentic (Zhang et al., 2025b)	✓	✗	–	✗	✓	Dynamic	🧠	✗
Alita (Qiu et al., 2025)	✗	✗	–	✓	✗	Dynamic	🧠 + 🌐	✗
ToolMaker (Wölflein et al., 2025)	✗	✗	–	✓	✗	Dynamic	🧠 + 🌐	✗
Mem²Evolve (Ours)	✓	✓	🤖 + 🧰	✓	✓	**Dynamic**	🧠 + 🌐 + 🧠	✓

Table 1: **Comparison of self-evolving agent frameworks. Optimization** indicates whether experience is used to optimize the agent (e.g., prompts). **Persistence** denotes whether experiences are persistently stored for future reuse. **Source**: 🤖 agent task execution trajectory, 🧰 tool creation process. **Tool Crea.** and **Agent Crea.** indicate whether the framework supports creation of tools and expert agents, respectively. **Tool/Agent** denotes whether the toolset and expert agents are static or dynamic. **Crea. Grounding** indicates the knowledge sources used for asset creation, 🧠 parametric knowledge, 🌐 web search information, 🧠 experience. **Exp.-Guided Creation** indicates whether new assets are created under the guidance of past experience. Details in the Appendix A.1 and A.2.



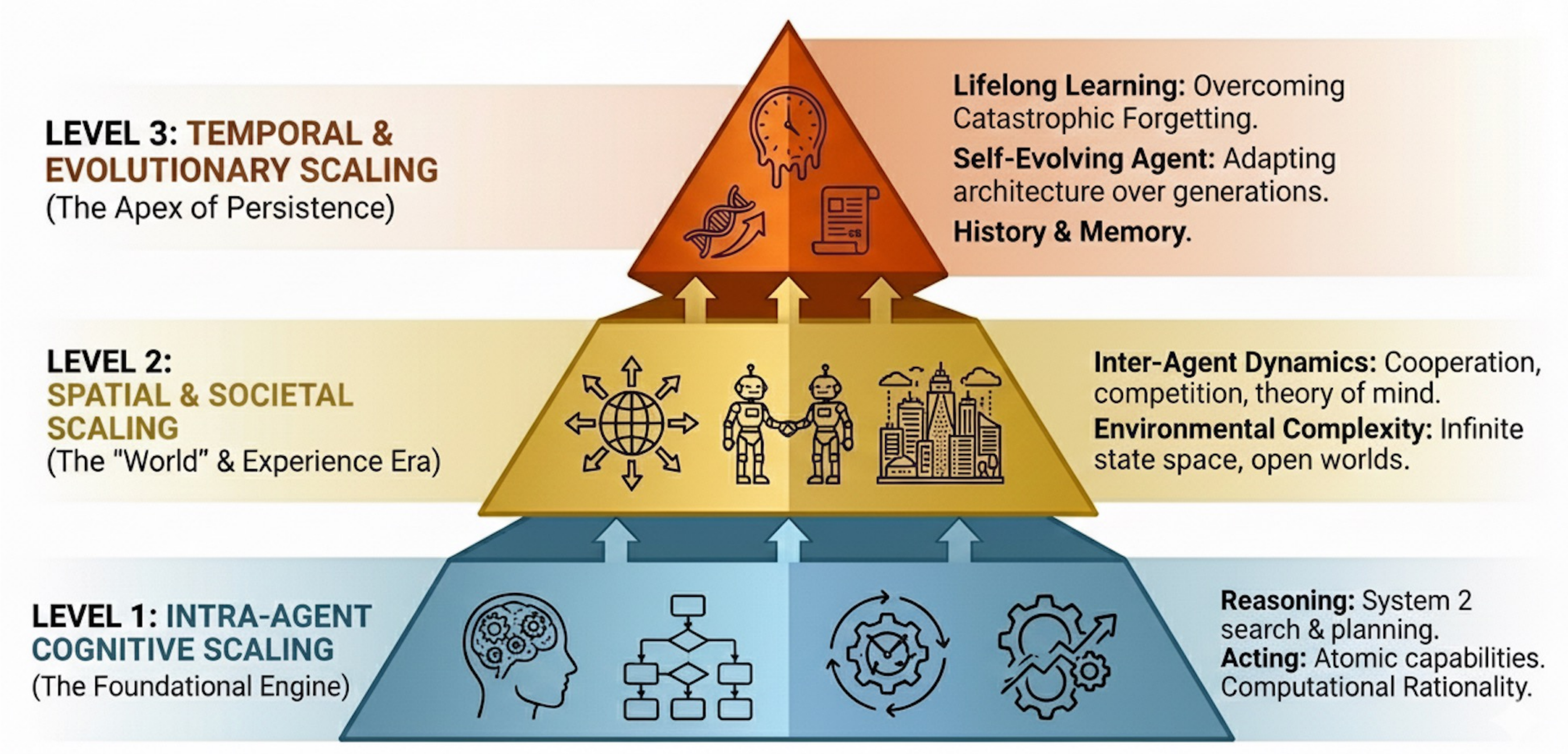
Cross-task evolution



Mem²Evolve Framework

Method	GAIA				Embodied	Multi-Hop QA		Math		Planning	Web Interaction	Avg.
	L1	L2	L3	Total	ALFWorld	HotpotQA	2Wiki	AIME24	AIME25	TravelPlanner	WebShop	
Naive-Large Language Model												
GPT-5-Chat (Direct)	16.98	12.79	7.69	12.49	83.58	50.40	<u>81.80</u>	60.00	46.67	38.68	22.31	49.49
GPT-5-Chat (CoT)	24.53	17.44	11.54	17.84	83.58	47.40	74.40	66.67	56.67	39.51	27.49	51.71
GPT-5-Chat (ReAct)	26.42	17.44	11.54	18.47	86.87	41.40	48.40	66.67	60.00	39.13	25.10	48.27
OpenAI-DeepResearch†	74.29	69.06	<u>47.60</u>	67.36	—	—	—	—	—	—	—	—
Experience-Centric Evolving												
DyLAN	24.53	19.78	11.54	18.62	91.20	52.00	65.00	46.67	43.33	43.15	36.40	49.55
EvoAgent	22.64	19.78	11.54	17.99	92.50	54.40	75.00	66.67	43.33	49.20	37.80	54.61
AFLOW	26.42	17.44	15.38	19.75	<u>93.40</u>	**60.80**	72.40	66.67	63.33	53.24	<u>37.90</u>	58.44
DSPy	30.19	15.12	11.54	18.95	92.80	55.60	76.40	66.67	50.00	44.90	35.50	55.10
Capability-Centric Evolving												
Alita	<u>81.13</u>	<u>75.58</u>	46.15	<u>72.73</u>	86.13	<u>58.80</u>	77.40	<u>70.00</u>	<u>66.67</u>	48.32	30.21	<u>63.78</u>
AgentVerse	30.19	16.28	19.23	21.90	88.32	38.60	74.60	60.00	50.00	47.25	32.53	51.65
AutoAgens	35.85	24.42	19.23	26.50	87.92	54.20	73.80	40.00	36.67	43.52	31.40	49.25
SwarmAgentic	28.30	18.60	13.46	20.40	88.79	56.00	80.00	46.67	40.00	<u>59.14</u>	34.12	53.14
Ours												
Mem²Evolve	**88.68**	**82.56**	**57.69**	**76.31**	**94.31**	**60.80**	**82.00**	**76.70**	**73.33**	**59.25**	**39.20**	**70.24**

Rethinking *Three* Levels of Scaling Simultaneously



Scaling of Time

Scaling of Agent and Environment

Scaling of Reasoning and Acting

Lots of problem need to be defined and explored, welcome to join theory of agent!

Conclusion

- Agents are not merely an engineering problem; they are becoming increasingly scientific and theoretical, like scaling law of LLMs. We also need to find more theories of agent.
- Agent can be regarded as human. Lots of problem in human society also happens in agent society, i.e., internet/tool addiction.
- Every company should have an agent department instead of LLM department.
- Join theory of agent no matter what you do now, you are not just a researcher, but may next scientist, entrepreneur, and even billionaire.
- Safety and personalization still matters in practice. Aligning decision boundaries with both preference boundaries and knowledge boundaries is tricky.