Language Agent as Tool-use Decision Maker

Hongru WANG

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

The Chinese University of Hong Kong

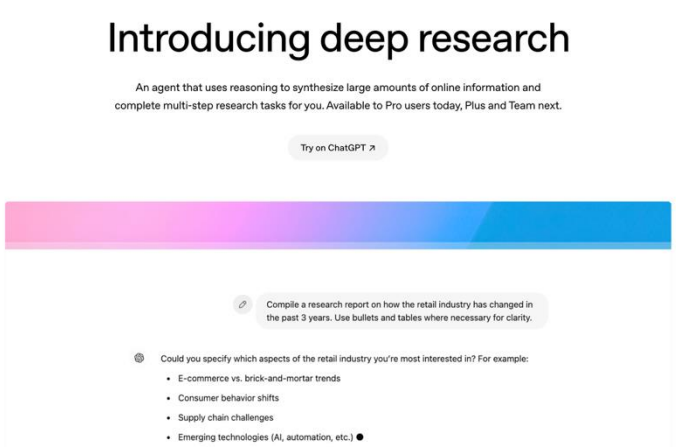
6th July, 2025

Supervisor:	Professor WONG Kam Fai William (CUHK)
Chairman:	Professor YU Jeffrey Xu (CUHK)
Committee Member:	Professor WAI Hoi To (CUHK)
External Examiner:	Professor LIU Zhiyuan (Tsinghua)

Overview

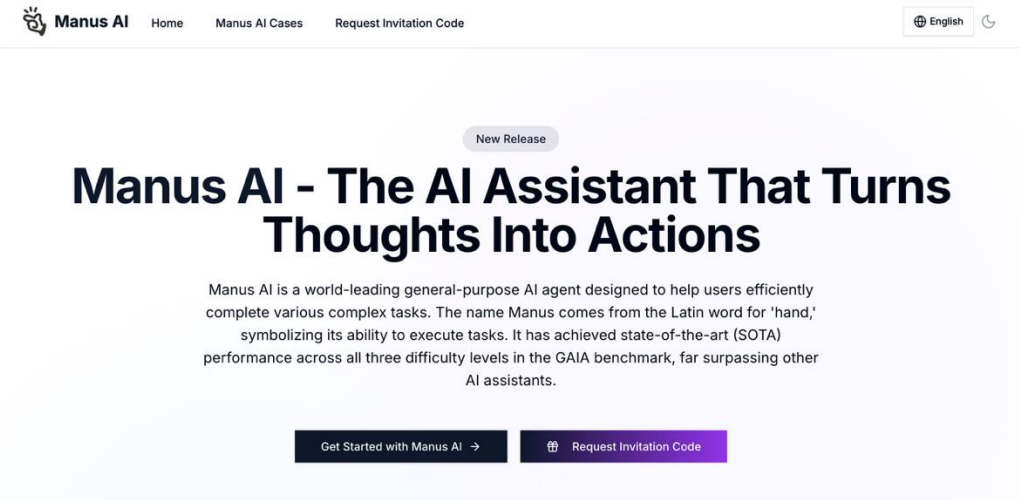
- Introduction
- Methods
  - Building Agents with ***Internal Cognitive*** *Tools*
  - Building Agents with ***External Physical*** *Tools*
  - Building Agents with ***Self-aware Tool Utilization***
- Benchmarks
  - Benchmarking Tool Planning in ***Single-turn*** Interaction
  - Benchmarking Tool Utilization in ***Multi-turn*** Interaction
- Summary and Future Directions

Background

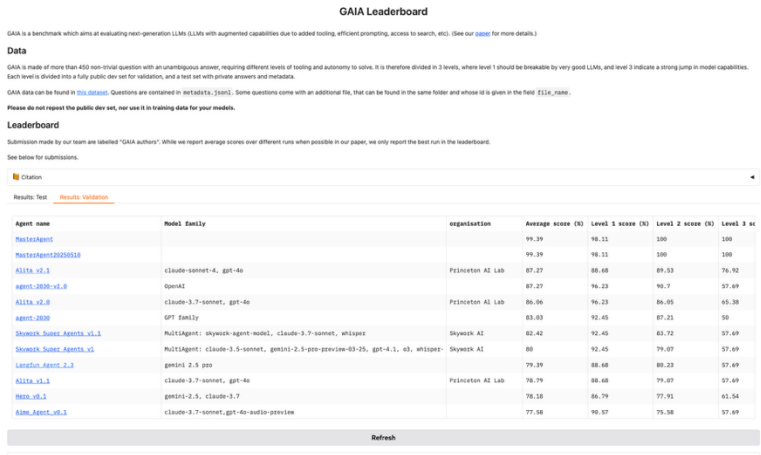


OpenAI Deep Research

Computer-Using Agent

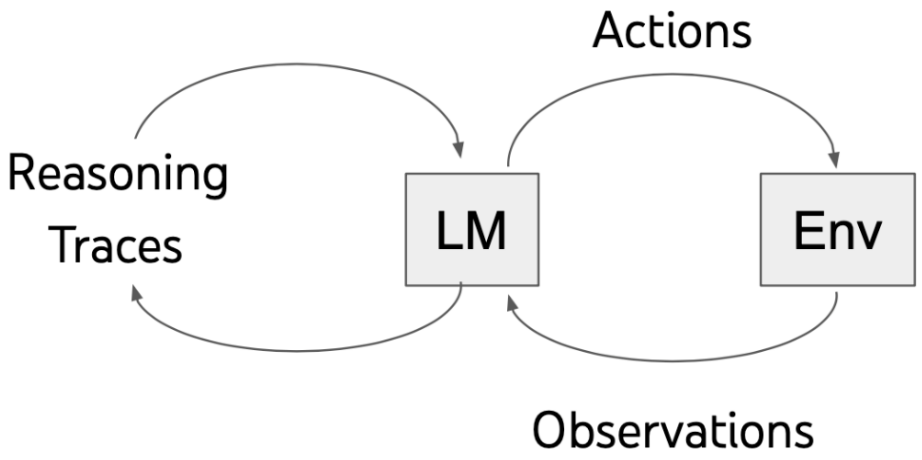


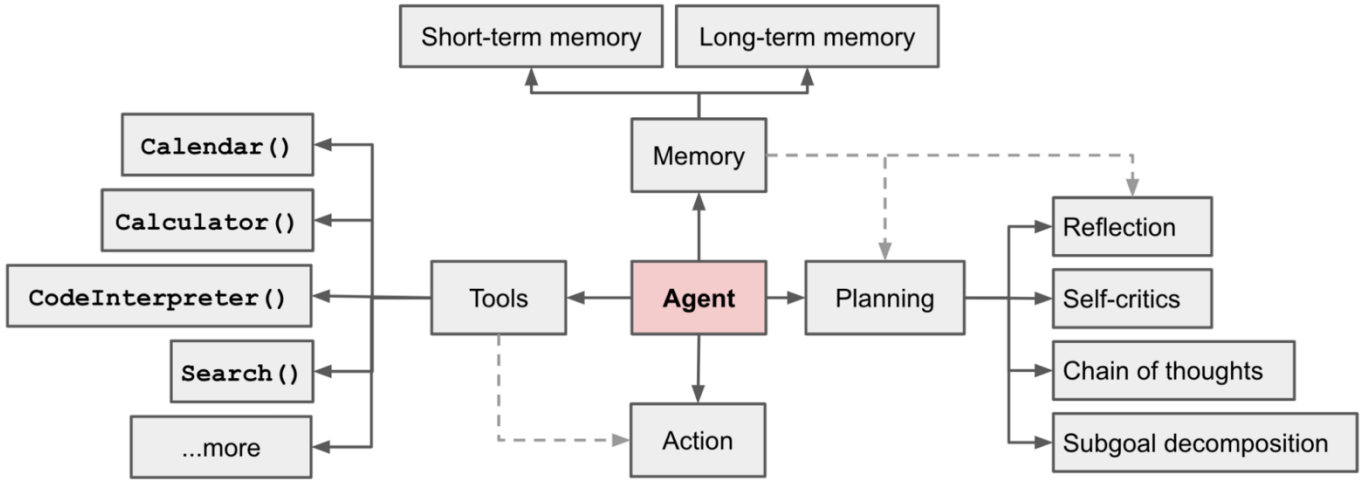
Manus



Our Alita reaches top 1 at GAIA (validation)

Background



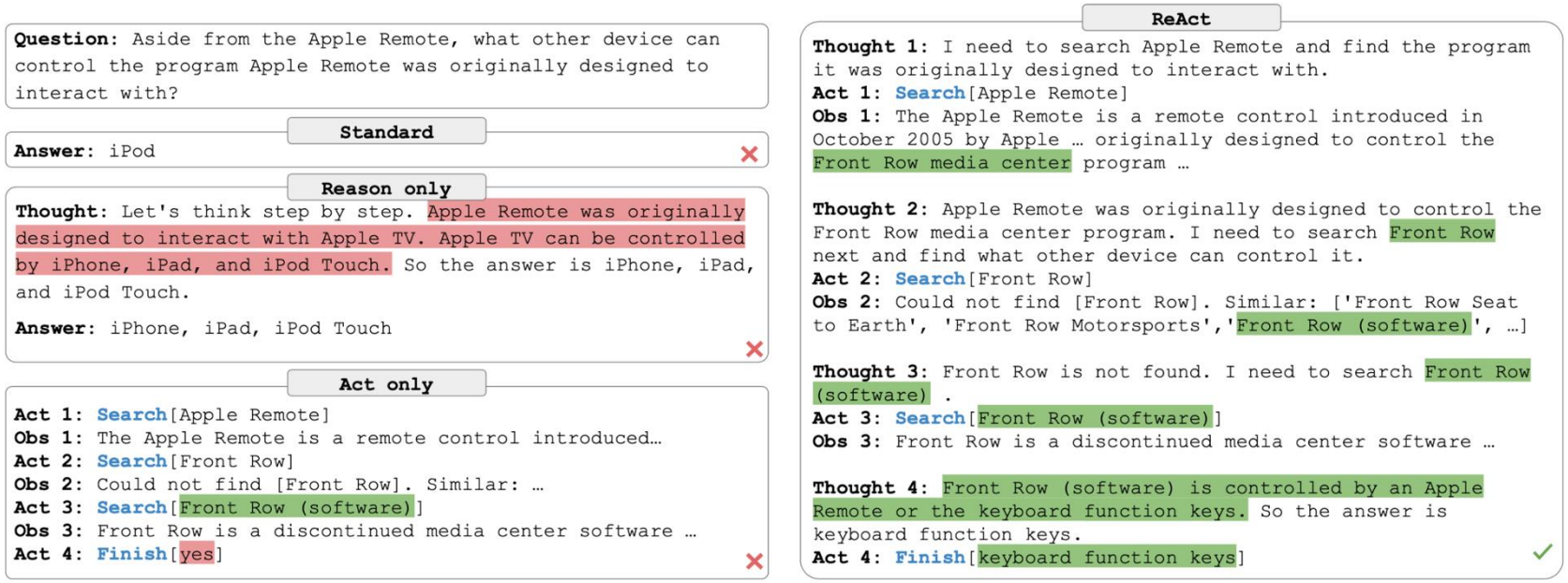


https://react-lm.github.io/

https://lilianweng.github.io/posts/2023-06-23-agent/

Agent = [Reasoning + Acting] * n

Background



ReAct in QA

Thought: The model's understanding about the current state

Act: external APIs to acquire external knowledge

Observation: feedback or results returned by external environment

Background

AlfWorld

You are in the middle of a room. Looking quickly around you, you see a armchair 1, a cabinet 4, … a cabinet 1, a drawer 5, …, a drawer 1, a dresser 1, a garbagecan 1, a safe 1, a shelf 12, …, a shelf 1, a sidetable 1, and a sofa 1.
Your task is to: put two keychain in safe.

(a) ReAct

.....
Obs 16: You open the drawer 4. The drawer 4 is open. In it, you see a keychain 3, and a watch 2.
Act 17: Think [Now I find the first keychain (3). Next, I need to take it. I can find the second keychain (2) later in drawer 4.]
Act 18: take keychain 3 from drawer 4
Obs 18: You pick up the keychain 3 from the drawer 4.
.....
Act 22: put keychain 3 in/on safe 1
Obs 22: You put the keychain 3 in/on the safe 1.
Act 23: Think [Now I put the first keychain in safe. Next, I need to find the second keychain. I can directly go to drawer 4.]
Act 24: go to drawer 4
Obs 24: The drawer 4 is open. In it, you see a watch 2.
Act 25: Think [Now I find the second keychain (2). Next, I need to take it.]
Act 26: take keychain 2 from drawer 4
Obs 26: Nothing happens. ✗

(b) ReAct (Human Thought edit)

.....
Obs 16: You open the drawer 4. The drawer 4 is open. In it, you see a keychain 3, and a watch 2.
Act 17: Think [Now I find the first keychain (3). Next, I need to take it. ~~I can find the second keychain (2) later in drawer 4.~~]
Act 18: take keychain 3 from drawer 4
Obs 18: You pick up the keychain 3 from the drawer 4.
.....
Act 22: put keychain 3 in/on safe 1
Obs 22: You put the keychain 3 in/on the safe 1.
Act 23: Think [Now I put the first keychain in safe. Next, I need to find the second keychain , which is more likely to appear in dresser (1), garbagecan (1), safe (1), sidetable (1), sofa (1), shelf (1-12) .]
Act 24: go to dresser 1
Obs 24: On the dresser 1, you see a box 1, a creditcard 1, a keychain 2, a keychain 1, a laptop 1...
Act 25: Think [Now I find the second keychain (2). Next, I need to take it.]
Act 26: take keychain 2 from dresser 1
Obs 26: You pick up the keychain 2 from the dresser 1.
Act 27: Think [Now I take the second keychain (2). Next, I need to put it in/on safe 1.]
Act 28: go to safe 1
Obs 28: The safe 1 is open. In it, you see a keychain 3, a statue 2, and a watch 3.
Act 29: put keychain 2 in/on safe 1
Obs 29: You put the keychain 2 in/on the safe 1. ✓

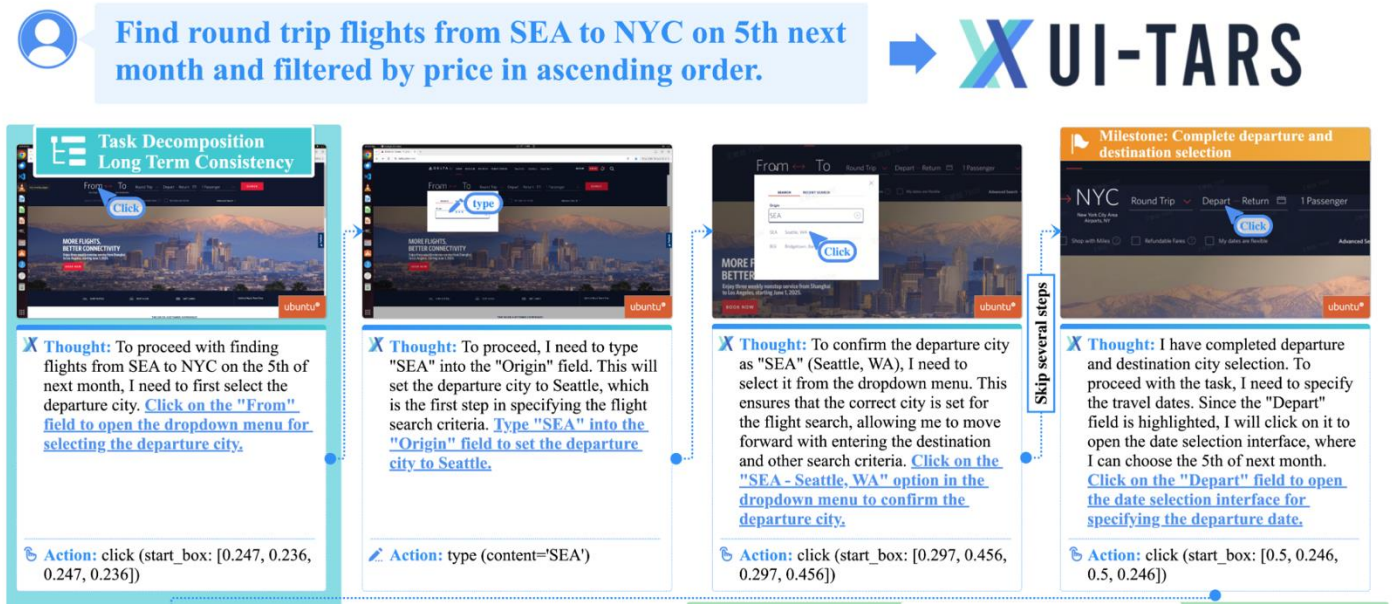
ReAct in Embodied Env

Thought: The model's understanding about the current state

Act: external Actions to acquire external knowledge

Observation: feedback or results returned by external environment

Background



ReAct in Computer-using Env

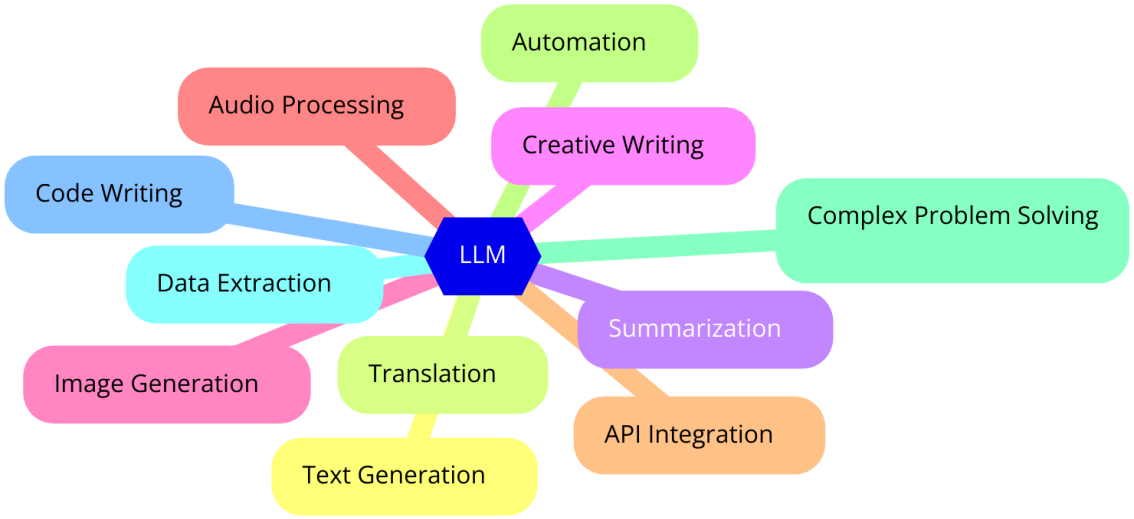
Thought: The model's understanding about the current state

Act: external Actions to acquire external knowledge

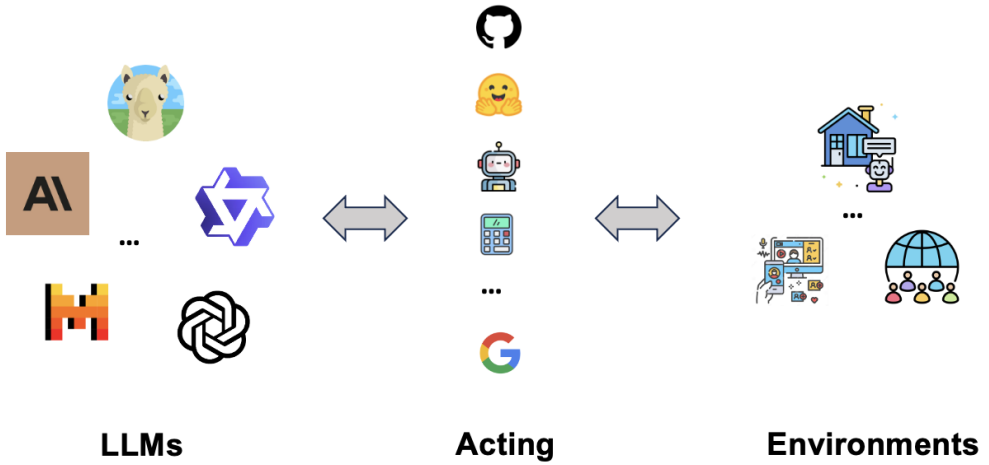
Observation: feedback or results returned by external environment, the next page here

Challenge 1: Unified Framework for Reasoning / Acting

Diverse Tasks (i.e., open-ended question)

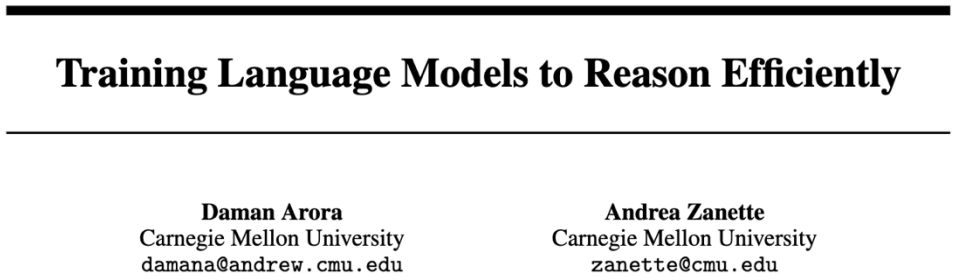


Complex Environments



Challenge 2: Smart Autonomous Agents

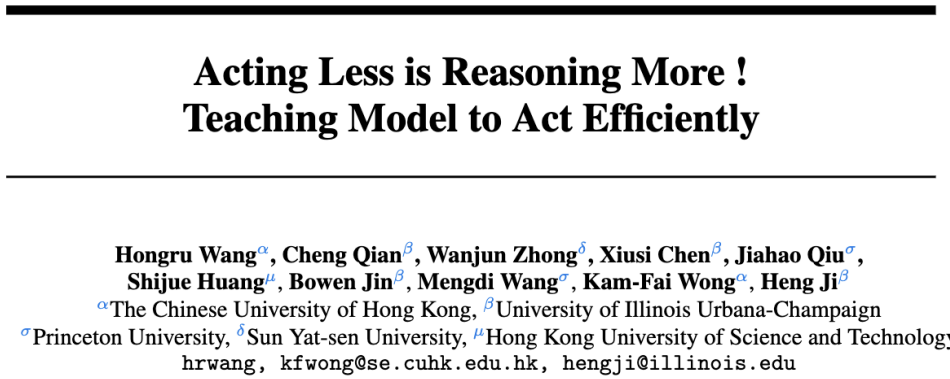
- The **smart** agent must be aware what is capable to do, and what it can not do, and make decisions accordingly. “知之为知之，不知为不知，是知也”
- The **autonomous** agent is designed to **minimize the number of actions** a system needs to take in the real world to learn a task (Yann LeCun 2022)



Training Language Models to Reason Efficiently

Daman Arora
Carnegie Mellon University
damana@andrew.cmu.edu

Andrea Zanette
Carnegie Mellon University
zanette@cmu.edu



Acting Less is Reasoning More !
Teaching Model to Act Efficiently

Hongru Wang[α], **Cheng Qian**[β], **Wanjun Zhong**[δ], **Xiusi Chen**[β], **Jiahao Qiu**[σ], **Shijue Huang**[μ], **Bowen Jin**[β], **Mengdi Wang**[σ], **Kam-Fai Wong**[α], **Heng Ji**[β]
[α]The Chinese University of Hong Kong, [β]University of Illinois Urbana-Champaign
[σ]Princeton University, [δ]Sun Yat-sen University, [μ]Hong Kong University of Science and Technology
hrwang, kfwong@se.cuhk.edu.hk, hengji@illinois.edu

Challenge 3: Real-world Applications and Evaluations

Complex Environments

Siri with App Intents

Siri is more natural, more personal, and more deeply integrated into the system. Apple Intelligence provides Siri with enhanced action capabilities, and developers can take advantage of predefined and pretrained App Intents across a range of domains to not only give Siri the ability to take actions in your app, but to make your app's actions more discoverable in places like Spotlight, the Shortcuts app, Control Center, and more. SiriKit adopters will benefit from Siri's enhanced conversational capabilities with no additional work. And with App Entities, Siri can understand content from your app and provide users with information from your app from anywhere in the system.

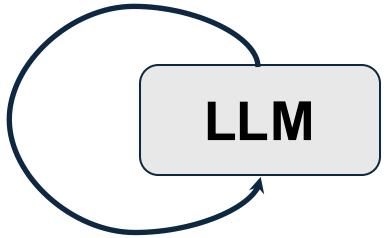
Bring your app to Siri
What's new in App Intents

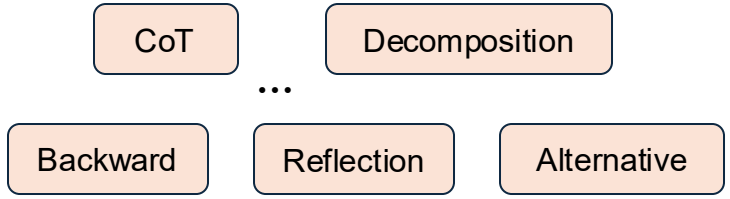
Complex Humans

Apple intelligence is not only a technology, but also represents new ways of interaction.

Reasoning vs Acting vs Planning

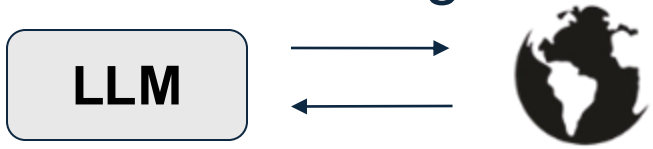
Reasoning

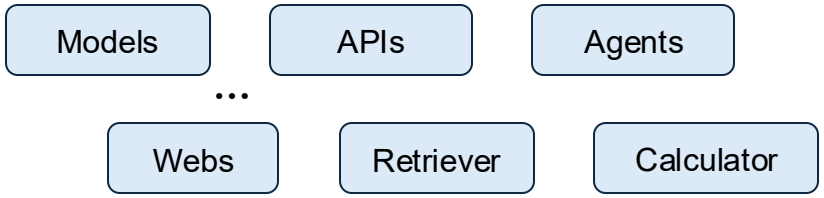




Cognitive mechanism / functions from **Cognitive Science**

Acting

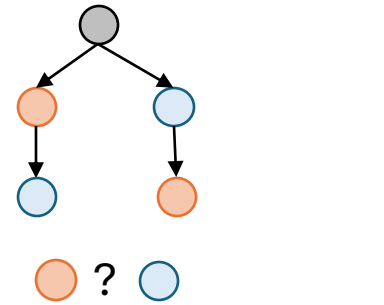




Physical Tools / Actions

Planning

Reason or Act



Decision-making Planning

Reasoning vs Acting vs Planning

reasoning == acting 

If reasoning == acting [Yao et al, …]

Thinking, or reasoning, is a **strange** kind of action

--- The Second Half, Shunyu Yao

https://ysymyth.github.io/The-Second-Half/

Reasoning vs Acting vs Planning

reasoning == acting 

If reasoning == acting [Yao et al, …]


```
{
  "name": "think",
  "description": "Use the tool to think about something. It will not obtain new information or change the 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://www.anthropic.com/engineering/claude-think-tool

Reasoning vs Acting vs Planning

reasoning == acting 

If reasoning == acting [Yao et al, …]

then how should we viem them in an unified manner?

Reasoning and Acting are both Tools

else:

what's real difference between reasoning and acting?

What's Tool?

Claude-3-Haiku Poe

"Tool" is a broad term that can refer to various objects or devices used to perform a specific task or function. In a general sense, a tool is an instrument or implement designed to help a person accomplish a particular goal or activity more efficiently or effectively.

Tools can come in many different forms, such as:

1. Handheld implements: These are tools that are held and manipulated by hand, like hammers, screwdrivers, pliers, saws, etc. They are commonly used in construction, repair, and maintenance work.
2. Powered tools: These tools use some form of power source, such as electricity, batteries, or fuel, to assist or automate a task. Examples include power drills, chainsaws, lathes, and sanders.
3. Measuring tools: These tools are used to quantify or assess various properties, such as length, weight, temperature, or time. Examples include rulers, scales, calipers, and stopwatches.
4. Computer software and applications: In the digital age, software programs and apps on computers, smartphones, and other devices can also be considered tools. They help users perform tasks like writing, analysis, communication, and entertainment.
5. Conceptual tools: Some "tools" are not physical objects but rather mental frameworks, methodologies, or approaches that help people solve problems, make decisions, or accomplish goals. Examples include problem-solving techniques, decision-making models, and management strategies.

Answer from LLMs

Google Scholar cognitive tool

Articles About 5,860,000 results (0.13 sec)

Any time
Since 2025
Since 2024
Since 2021
Custom range...

Sort by relevance
Sort by date

Any type
Review articles

include patents
include citations

Create alert

What are **cognitive tools**?
DH Jonassen - **Cognitive tools** for learning, 1992 - Springer
… **tools tools** that extend the mind This workshop was about **cognitive tools** - computer-based **tools** … Computer-based **cognitive tools** are in effect **cognitive** amplification **tools** that are part of …
☆ Save 99 Cite Cited by 508 Related articles All 5 versions

[PDF] Technology as **cognitive tools**: Learners as designers
DH Jonassen - ITForum Paper, 1994 - tecfa.unige.ch
… **Cognitive tools** are generalizable computer **tools** that … **Cognitive tools** and environments activate **cognitive** learning strategies and critical thinking. They are computationally based **tools** …
☆ Save 99 Cite Cited by 383 Related articles All 4 versions »

[BOOK] Computers as **Cognitive Tools**: 1
SP Lajoie, SJ Derry - 1993 - books.google.com
… are employed, and the forms of "**cognitive tools**" that are embedded within systems to help … computers as **tools** for enhancing learning. Computers as **Cognitive Tools** is appropriate for …
☆ Save 99 Cite Cited by 924 Related articles All 10 versions »

[BOOK] **Cognitive tools** for learning
PAM Kommers, DH Jonassen, JT Mayes - 1992 - research.utwente.nl
… to address the theme of **cognitive tools** as discussed in this book … **tools** and was the main reason that '**cognitive tools**' became … during instruction allows for **cognitive** amplification. Some …
☆ Save 99 Cite Cited by 342 Related articles All 8 versions »

Answer from Scholars

Unification of Reasoning and Acting

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*).

Physical Tools

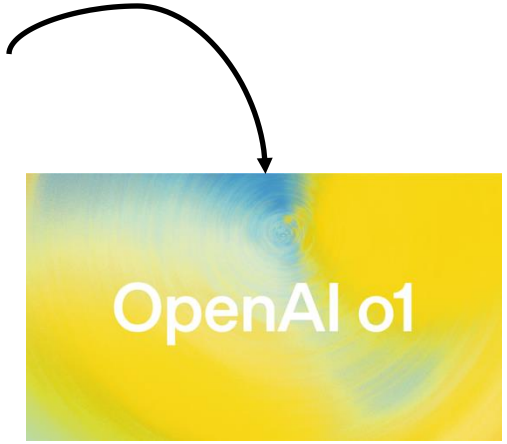
- ✓ Models
- ✓ Retriever
- ✓ Calculator
- ✓ Programs
- ✓ Webs
- ✓ Robots
- ✓ Knowledge Sources
- ✓ ……

external

Cognitive Tools

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

internal



Reasoning ~= Acting (in) Tools

- **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*).

Essence of Tool

- **Useful:** A tool must effectively complete one or multiple tasks. It typically receives inputs and produces outputs.
- **On-demand:** A tool must be used as needed, meaning it is invoked based on the current state.

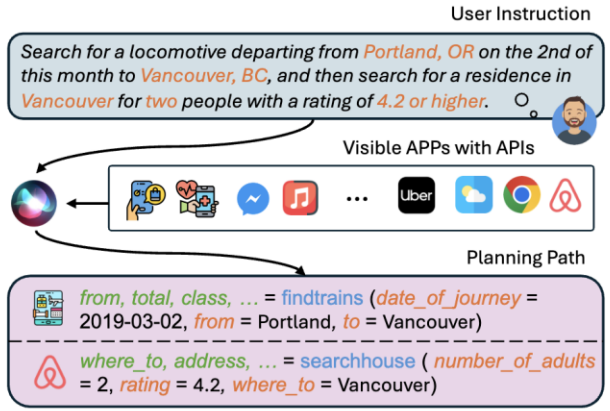
Some Typical Tools

	Useful	On-demand
Chain-of-thoughts (CoT) Reflection Decomposition …	✓	✓
		internal cognitive tools
APIs Actions Search Engine **Seek Human Help** …	✓	✓
		external physical tools

To CoT or not to CoT? Chain-of-thought helps mainly on math and symbolic reasoning

Zayne Sprague♠, Fangcong Yin♠, Juan Diego Rodriguez♠, Dongwei Jiang♢,
Manya Wadhwa♠, Prasann Singhal♠, Xinyu Zhao♠,
Xi Ye♡, Kyle Mahowald♠, Greg Durrett♠

♠The University of Texas at Austin, ♢Johns Hopkins University, ♡Princeton University
zaynesprague@utexas.edu



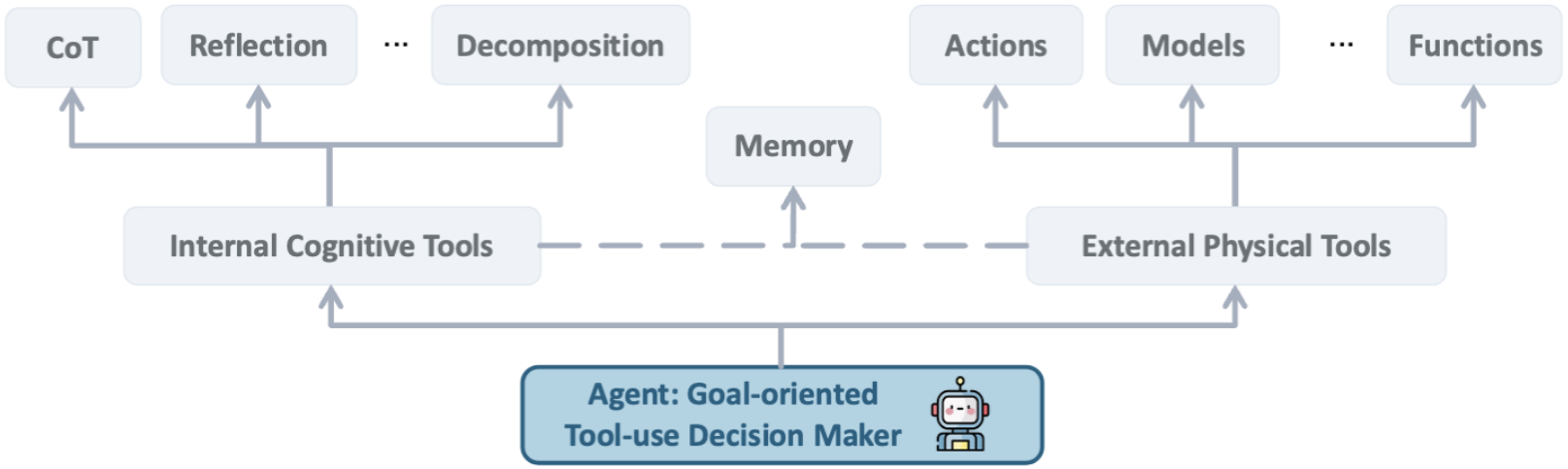
AppBench

These tools effectively **address inherent limitations** of LLMs, such as outdated information, while also **expanding the capabilities to interact with the external environment**.

AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction (EMNLP 2024)

Tool-integrated Agents

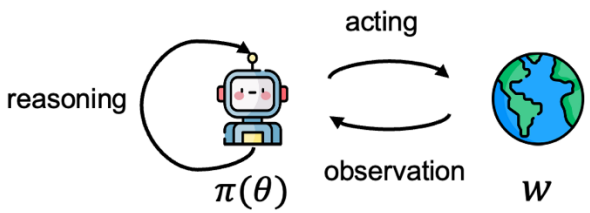
- An agent is an entity that coordinates internal cognitive tools (e.g., reflection) and external physical tools (e.g., function callings) to acquire knowledge in order to achieve a specific goal.



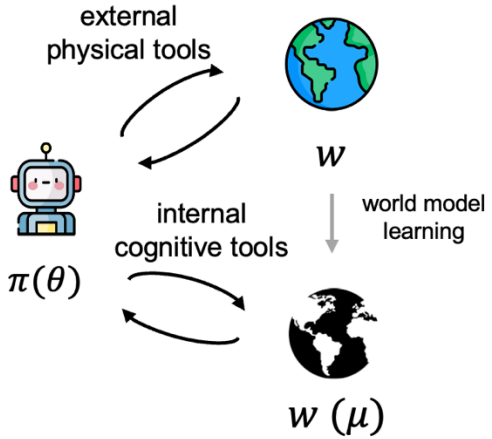
Tool-integrated Agents

- 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.

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



(a) ReAct-based Agent



(b) Tool-integrated Agent

Tool-integrated Agents

- 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.

  - Flexible and Robust
    - It degrade to previous ReAct paradigm if we consider the internal tools and internal knowledge as 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 stands for external part.

    - If we solely consider internal tools, it is proved that simply outcome-based reward can trigger various tool utilization such as reflection and decomposition to solve the problem in Large Reasoning Models (i.e., DeepSeek-R1). Alternatively, simply outcome-based reward also trigger various external tool utilization as evidenced in recent studies (i.e., Search-R1, ToRL, OTC-PO).

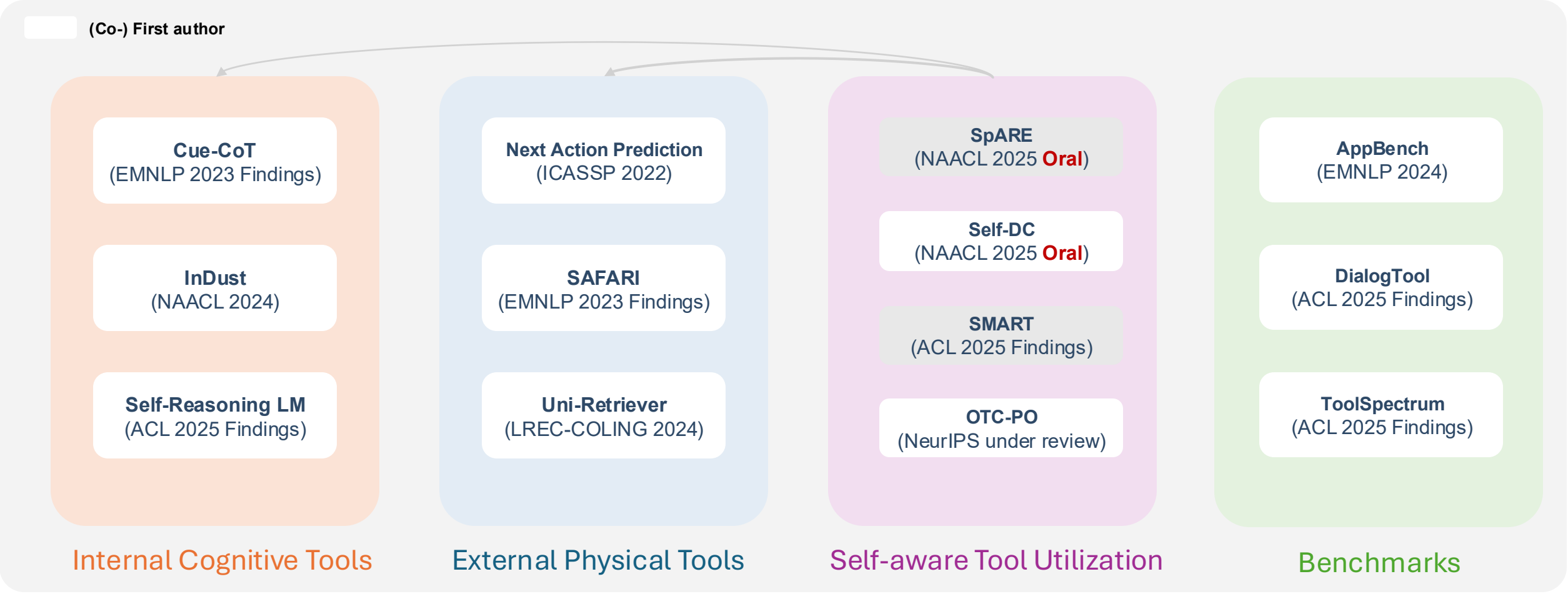
Tool-integrated Agents

- 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.

  - Potential Next Scaling Law
    - *Next Tool 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.

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools



Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

(Co-) First author

Cue-CoT
(EMNLP 2023 Findings)

InDust
(NAACL 2024)

Self-Reasoning LM
(ACL 2025 in Sub)

→

How to elicit the internal reasoning capabilities of LLMs and leverage various cognitive tools?

Internal Cognitive Tools

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

(Co-) First author

Cue-CoT
(EMNLP 2023 Findings)

InDust
(NAACL 2024)

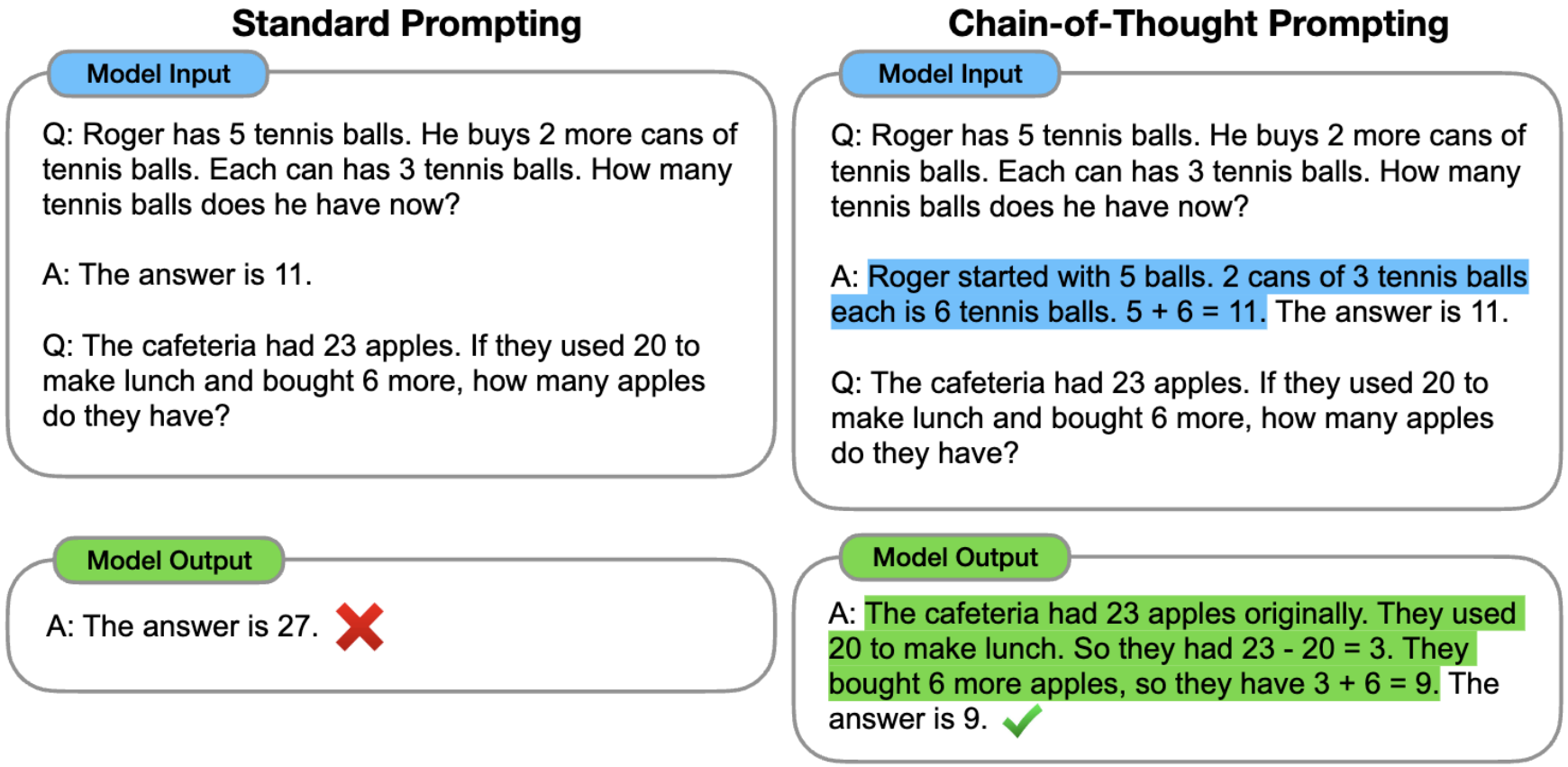
Self-Reasoning LM
(ACL 2025 in Sub)

→

How to elicit the internal reasoning capabilities of LLMs and leverage various cognitive tools?

Internal Cognitive Tools

Chain-of-Thoughts (CoT)



CoT has proved to be an effective mean to elicit reasoning capabilities of LLMs in certain tasks.

Chain-of-Thought Prompting Elicits Reasoning in Large Language Models (NeurIPS 2022)

Chain-of-Cues

Dialogue Context

"What are some things you think you should know when having a baby, but no one tells you?"

"1. Prenatal care tips, such as maintaining a healthy diet, reducing alcohol consumption, ...
2. Possible emergencies during childbirth, such as fetal growth restriction, ...
3. Postpartum care for the mother, such as breastfeeding, ...
4. ..."

"What are some things that can help me prepare for childbirth?"

Linguistic Cues

Emotion: anxiety

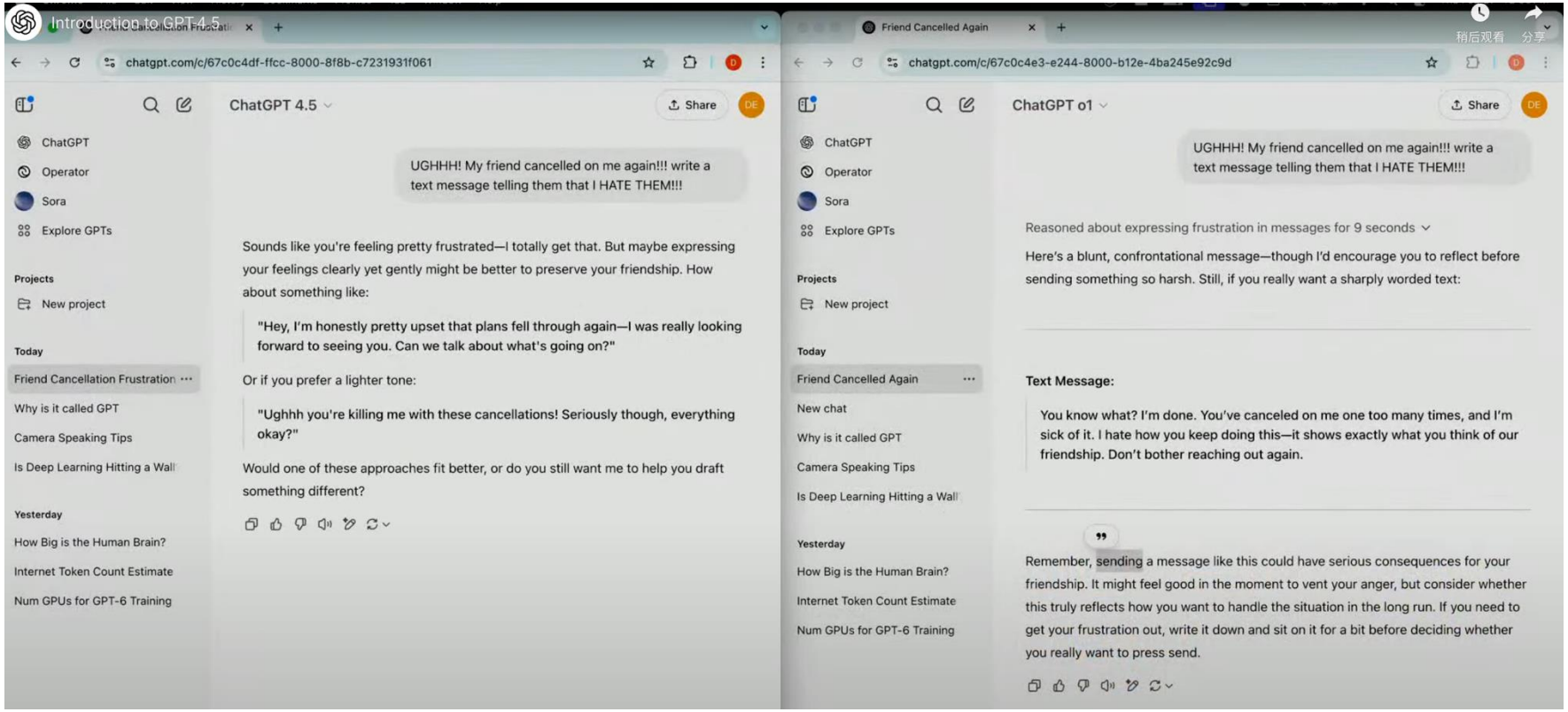
Personality: pay close attention to details and approach with thoughtful consideration, ...

Psychology: feeling uncertain and anxious about the future, need some more specific advice, ...

Lots of linguistic cues underlying dialogue context are effective means as **intermediate reasoning results** (Chain-of-Cues) to generate more personalized and acceptful responses.

Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs (EMNLP 2023 Findings)

Chain-of-Cues



Introduction to GPT-4.5: https://youtu.be/cfRYp0nItZ8

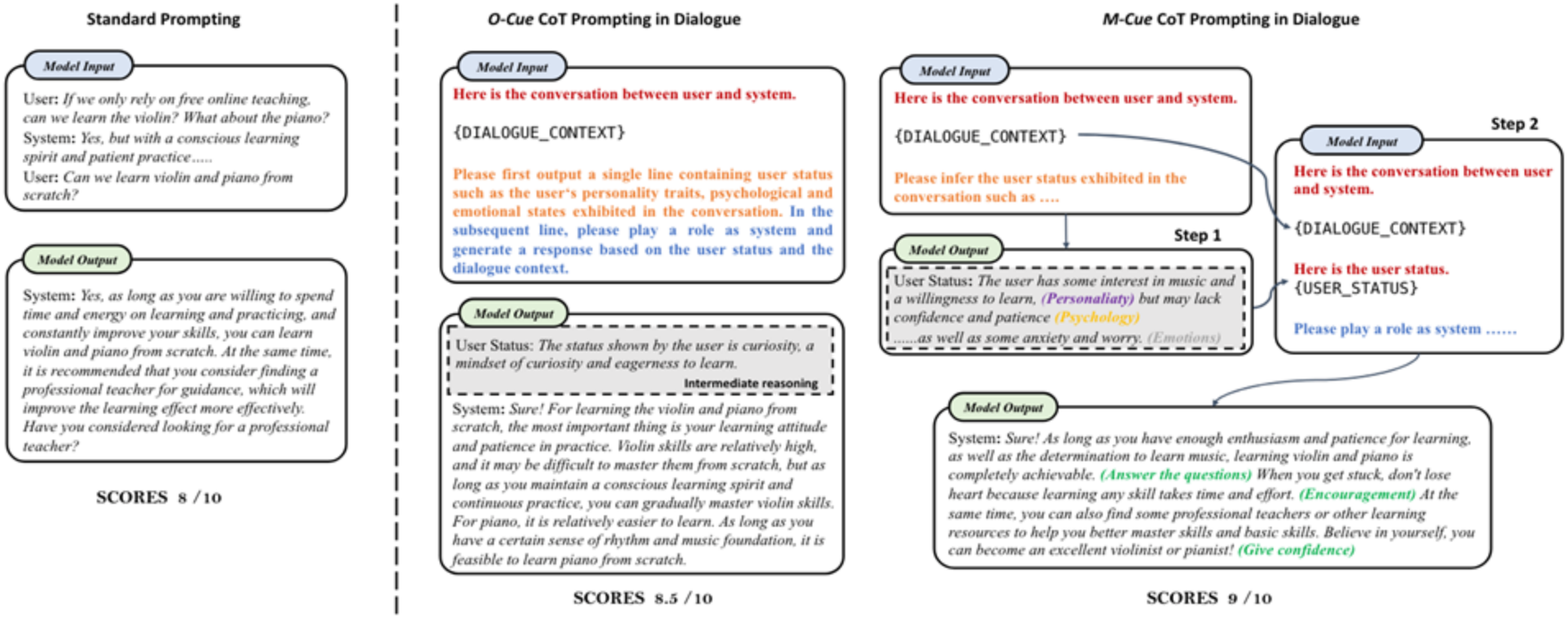
GPT-4.5

Lots of internal cognitive capabilities

Emotional intelligence,
More natural / human,
Creativity,
Warm
....

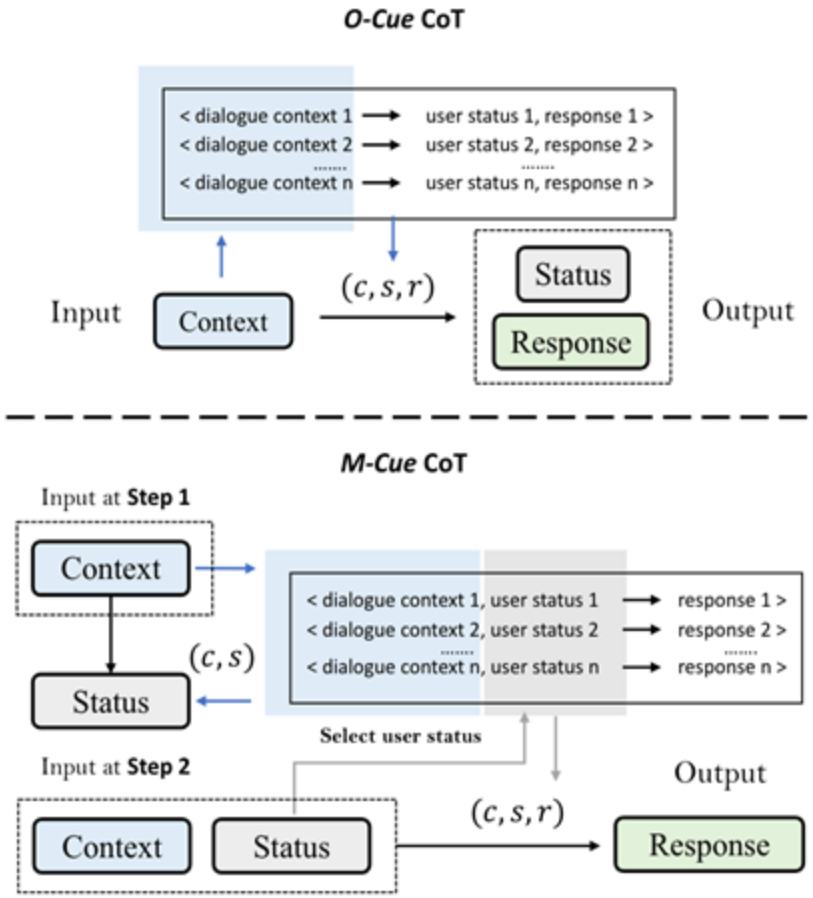
Chain-of-Cues – Two Reasoning Variants

- O-Cue: simply generate everything in one step, like Chain-of-Thoughts;
- M-Cue: generate all reasoning step by step, two major advantages: 1) reduce context length, and 2) utilize intermediate reasoning for in-context learning.



Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs (EMNLP 2023 Findings)

Chain-of-Cues – In-context Learning



O-Cue

select (c, s, r) according to dialogue context to infer cues and response together.

M-Cue

select (c, s) according to dialogue context to infer status, and then select demonstrations (c, s, r) **according to inferred cues**

Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs (EMNLP 2023 Findings)

Chain-of-Cues – Results

Model	Prompt	Helpfulness			Acceptability		
		Zhihu	D4	PsyQA	Zhihu	D4	PsyQA
Zero-shot Setting							
BELLE	*O-Cue*	67.40	76.34	69.31	55.82	52.50	53.43
	M-Cue	81.54	71.60	79.25	60.23	72.41	73.65
CHATGLM	*O-Cue*	48.29	56.68	33.00	32.39	39.19	31.34
	M-Cue	85.02	72.10	83.57	66.67	51.27	55.40
CHATGPT	*O-Cue*	67.91	50.40	61.90	53.14	52.38	58.15
	M-Cue	95.57	87.88	90.34	65.22	61.08	56.12
One-shot Setting							
							random selection
BELLE	*O-Cue*	64.31	50.53	65.15	53.35	40.07	53.81
	M-Cue	83.30	69.59	73.81	73.61	56.14	61.90
CHATGLM	*O-Cue*	-	-	-	-	-	-
	M-Cue	90.28	75.10	91.85	74.55	54.03	64.75
CHATGPT	*O-Cue*	76.47	51.94	65.44	63.86	50.47	56.03
	M-Cue	91.60	86.67	88.96	76.83	58.19	61.41
							top-1 selection
BELLE	*O-Cue*	63.77	57.51	69.92	54.93	41.02	55.87
	M-Cue	82.77	69.94	73.99	74.32	54.38	62.24
CHATGLM	*O-Cue*	-	-	-	-	-	-
	M-Cue	89.25	77.26	91.77	73.43	57.17	58.74
CHATGPT	*O-Cue*	76.86	50.93	55.85	59.63	52.02	57.58
	M-Cue	93.19	88.84	91.77	78.46	56.84	59.48

- As win rate > 50%, it means the responses are better than standard prompting. O-Cue and M-Cue both are better than Standard Prompting, and M-Cue is more effective and robust
- LLMs
  - BELLE: low **long-context understanding ability**; middle **instruction-following ability**
  - ChatGLM: middle long-context understanding ability; low instruction-following ability
  - ……
  - ChatGPT: both high

Self-Reasoning Language Model (SRLM)

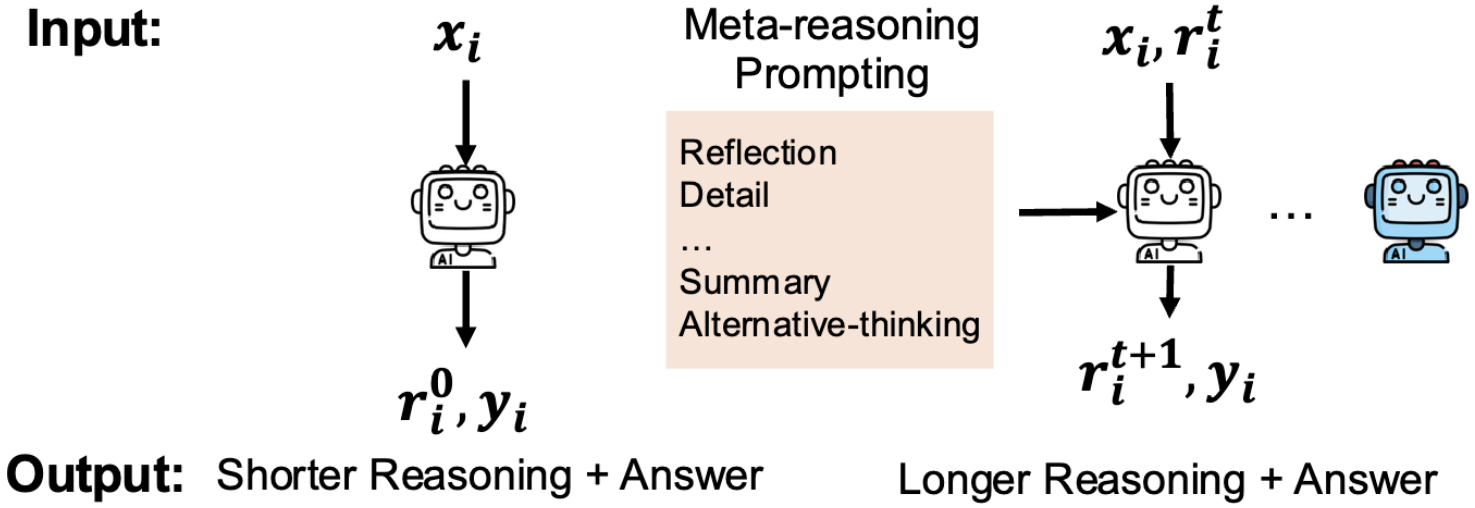
The model need to **understand role of each tool** and **composite them selectively** to solve the problem.

*What if we want to **dynamically call multiple internal cognitive tools** on demand to solve general problems?*

Outcome-based RL (e.g., DeepSeek-R1) is **not appliable**

Self-Reasoning Language Model (SRLM)

- The key lies in generating and obtaining high-quality and longer CoT traces.
  - With longer CoTs during inference, these LLMs could explore more creative and diverse reasoning rationales while assembling various meta-reasoning skills observed in human cognition, i.e., decomposition and reflection.



Self-Reasoning Language Model (SRLM)

- We introduce ***Self-Reasoning Language Model*** (SRLM), where **the model itself can synthesize longer CoT data and iteratively improve performance through self-training**. It is capable to act as both: 1) a response generation model (i.e., *learn to reason*); and 2) reasoning models to refine its own reasoning rationales (i.e., *how to reason*). It is noted that we **do not do continual training.**

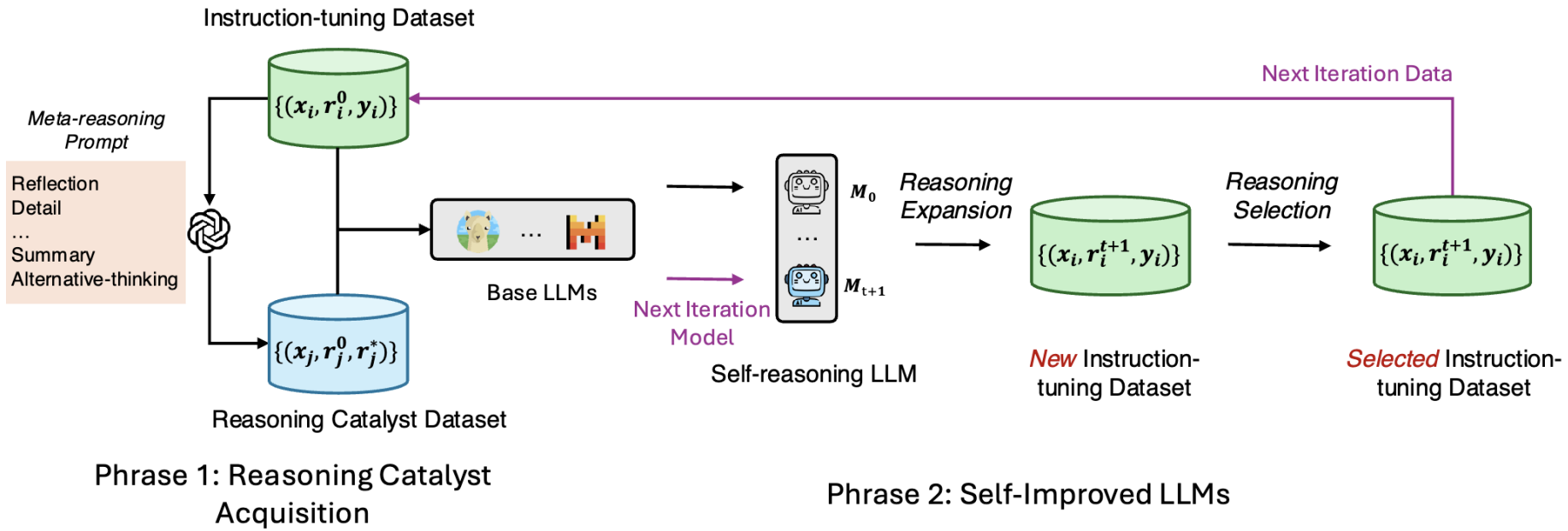


Figure 2: The framework of our proposed Self-Reasoning Language Models, which consists of two phrases.

Phrase-1: Reasoning Catalyst Acquisition

“It’s better to teach a man to fish than to give him fish." 授人以鱼不如授人以渔

You are an expert at meta reasoning theory from cognitive science. Given the question, corresponding summarized reasoning and answer, you can always uncover hidden or unspoken reasoning, even when it isn’t explicitly stated.
You need to add any missing reasoning thoughts that you think it is helpful or may occurs to understand and solve this question based on given summarized reasoning result. Your new reasoning should be more comprehensive, detailed and clear. Your answer should follow the format like <thoughts> your new reasoning here </thoughts>
Inside <thoughts> </thoughts>, you need to explicitly indicate which meta reasoning skill is used, some of them are shown below, but note you can use anything else if you think it is helpful or it naturally occurs when you solve this question.

<decomposition>: breaking down a complex problem into smaller, more manageable parts. Making sure that you also provide answers for all decomposed problems in this section. You can decompose iteratively but should not contain same problem or exceed the max iteration depth which is three.

<backward>: starting with the desired observations at any previous reasoning step and working backward to identify the new reasoning directions.

<detail>: any details including but not limited to logic and reasons for your reasoning in this way, you are encouraged to add this at every unclear or unnatural reasoning step.

<summary>: summarize your reasoning to help future thinking.

<alternatives>: directly thinking in other ways, try to explore different solutions as much as possible to solve given problem.

<reflection>: you are encouraged to regularly reflect on your past reasoning in current response at various levels of detail, from sentence down to individual word. This will help you better understand and think through problems. It’s okay to make mistakes; use them as opportunities to learn and improve.

<analogy>: you are encouraged to regularly consider other analogous problems with the problem you’ve encountered at various reasoning steps, along with their solutions. Reference existing theories or methods that guided your approach to solving these problems. These similar problems can be at various levels of detail - from larger overarching issues down to smaller sub-problems you encountered along the way. The key is to demonstrate a diverse range of problems and solutions, to show how you have approached and resolved challenges that are analogous to the current situation.

<check>: consider different edge cases or test cases carefully.

<other>: other meta reasoning skills you think is helpful or worthy to try to solve the task.

Notice:

1. All tags must be properly invoked and closed, using the format like <reflection> and </reflection>.

2. You should always use first-person perspective.

3. You can add any new meta reasoning skill at any positions except <reflection>. Note <reflection> can not be invoked without any reasoning in current response and it can be invoked at any positions when you already have some reasoning results.

4. You cannot change the original reasoning. However, if you identify any errors or improvements in the reasoning, you can add new reasoning steps using above meta-reasoning skills afterwards to correct or clarify the path, ensuring a better understanding and solution.

5. You can apply the same reasoning skills multiple times or use different skills simultaneously.

6. Your answer should start with <thoughts>, and end with </thoughts>.

Defining lots of reasoning modules as internal cognitive tools

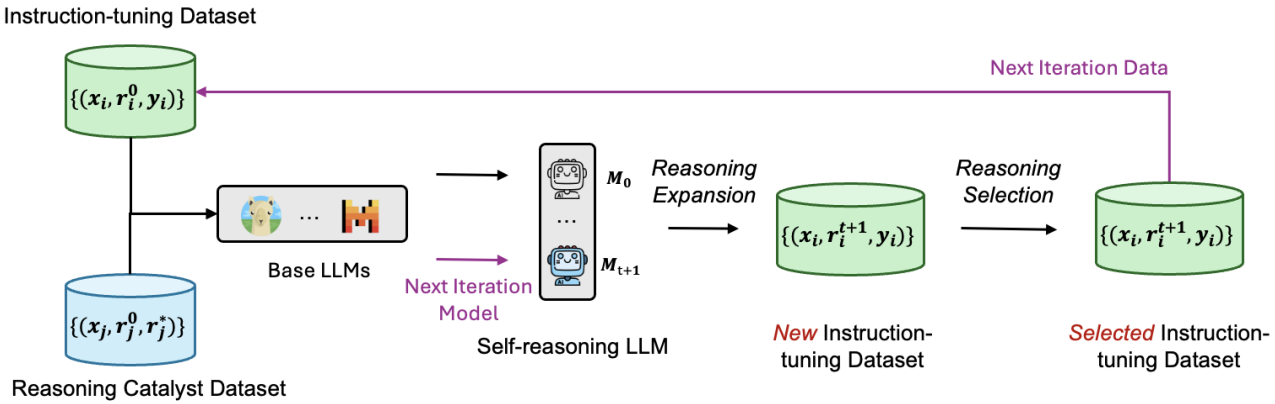
CoT: Let’s think step by step
Reflection: ….
Critic: ……

$$r_j^* = \mathcal{M}_{meta}(x_j, r_j^0)$$

Phrase 2: Self-Improved LLMs

- **Iterative Reasoning Expansion:** Just feed instruction-tuning dataset in last iteration into SRLM;
- **Iterative Reasoning Selection**: Choose the better one from current sample and previous sample.

Type equation here.



$$L = -\log p(r, y|x) \; -\log p(r^*|x, r)$$

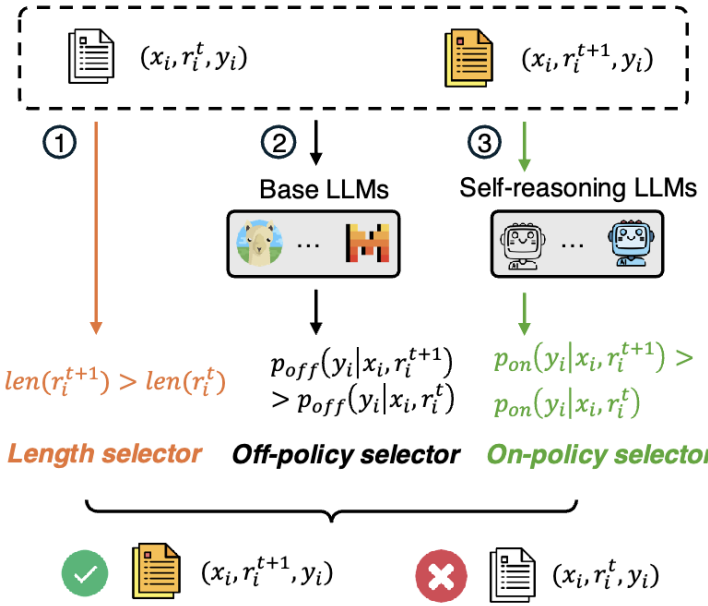
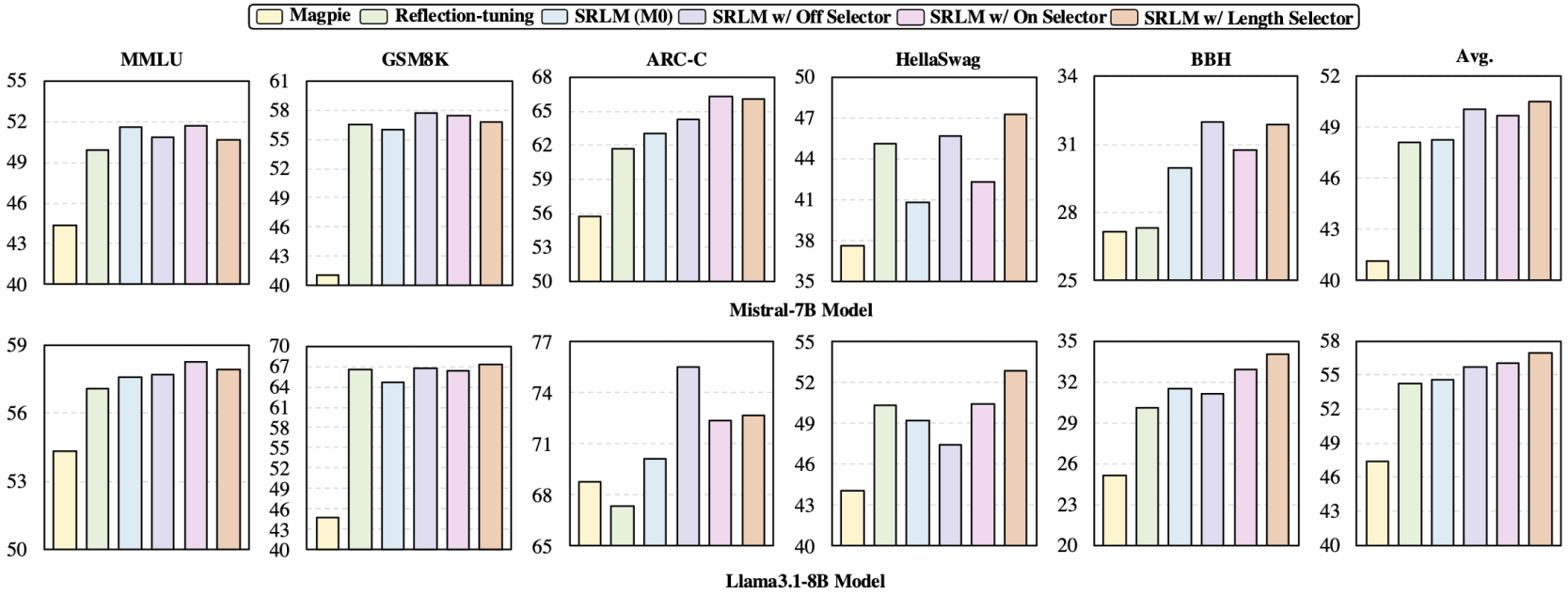


Figure 3: Three different types of reasoning selectors.

Experimental Results

- Incorporating additional reasoning expansion samples into the instruction-tuning dataset acts as a *catalyst*, leading to improved performance. SRLM (M0) vs Reflection-tuning and Magpie
- Small Self-Reasoning Language Models can generate high-quality rationales iteratively. It is evident that SRLM with different selectors at various iterations all leads to improvement compared with the initial SRLM (M0).
- All types of selectors outperform SRLM (M0), with the on-policy selector excelling in MMLU, while the length selector performs better on HellaSwag and BBH.



Analysis

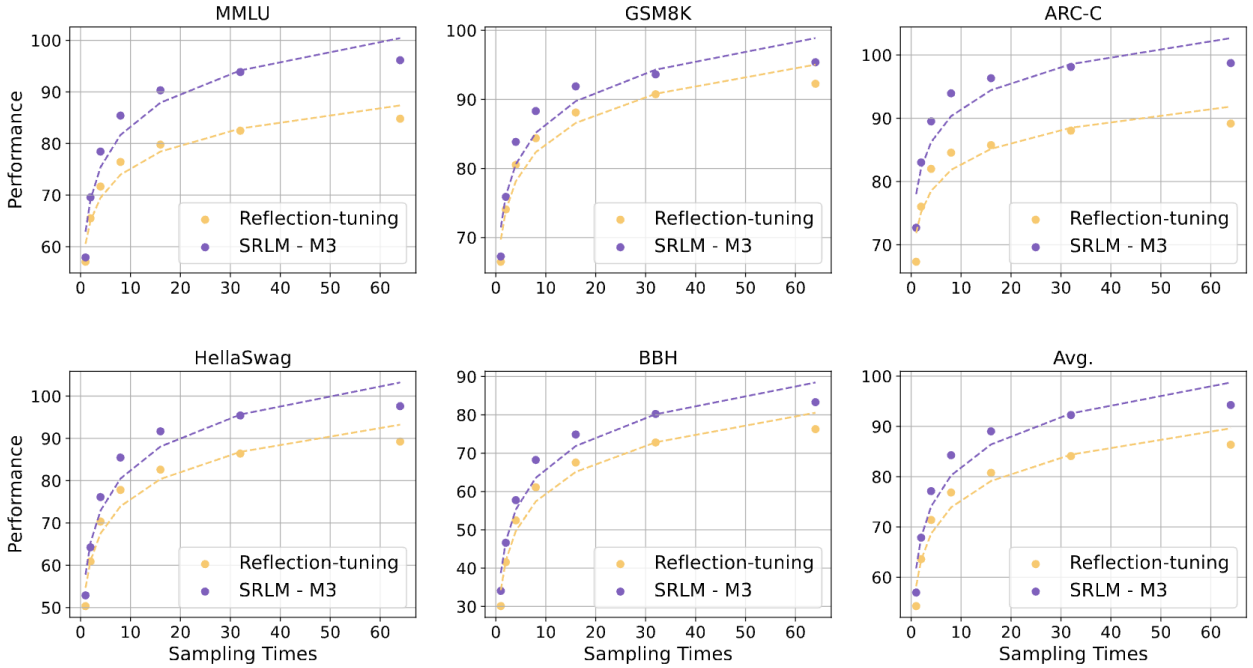


Figure 5: The performance of the baseline and our proposed SRLM ($\mathcal{M}_3$ with length selector) different sampling times ranging from 1 to 64 (i.e., 1, 2, 4, 8, 16, 32, 64) on five various benchmarks and the Avg. performance.

It is believed that our proposed SRLM can explore **more depth, diverse and creative reasoning paths** toward the final correct solution, revealing its promising potential.

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

(Co-) First author

Cue-CoT
(EMNLP 2023 Findings) → Pioneering work to propose **dialogue Chain-of-Thought (CoT)** considering the emotion, personality and psychology of the user (~90 Citations).

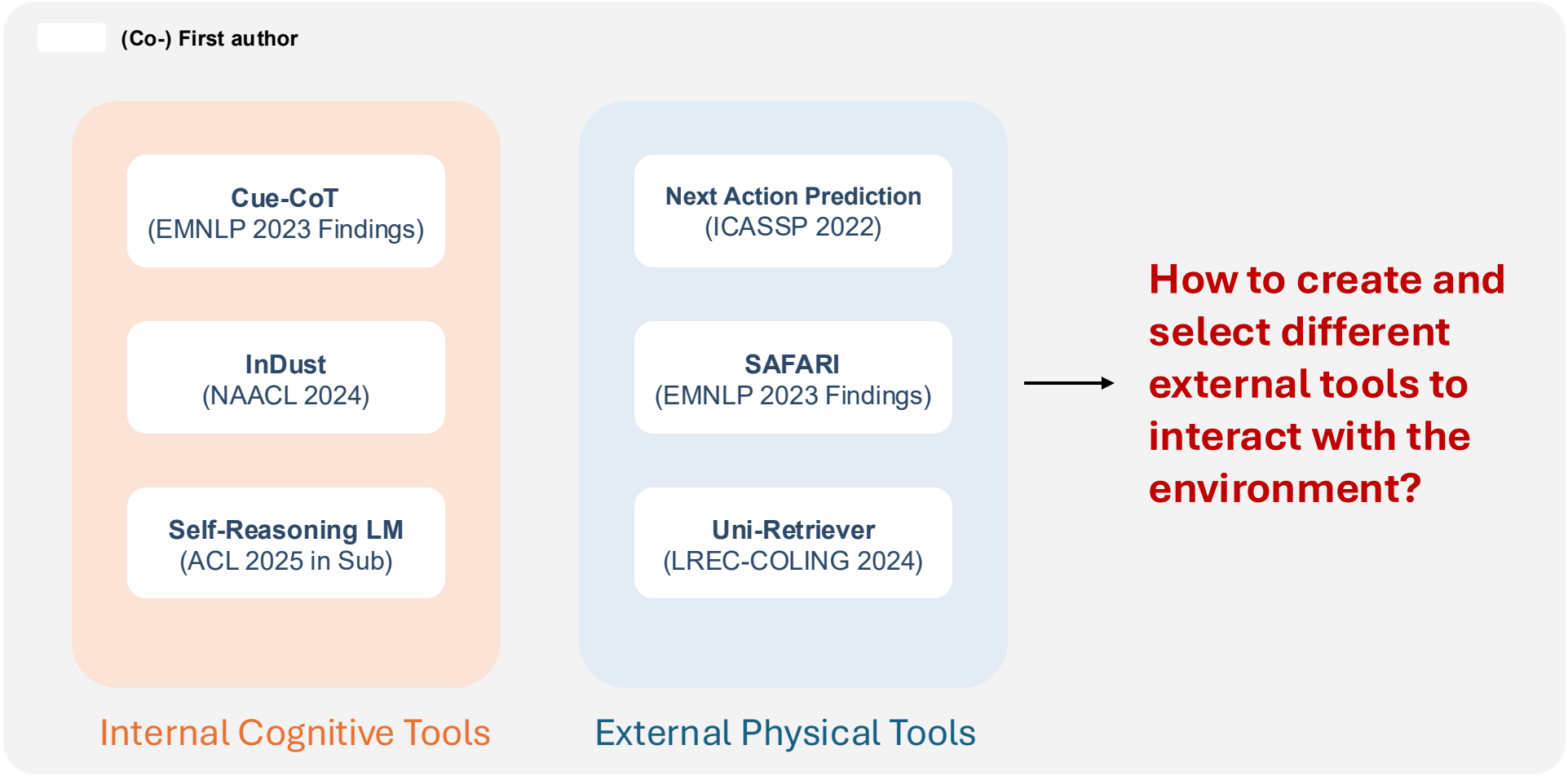
InDust
(NAACL 2024) → First Work to leverage **perspective-thinking** cognitive function to handle inductive instruction for safety.

Self-Reasoning LM
(ACL 2025 in Sub) → First Work to use small LLM itself to **synthesize longer CoT data with lots of cognitive tools** and iteratively improve performance through self training.

Internal Cognitive Tools

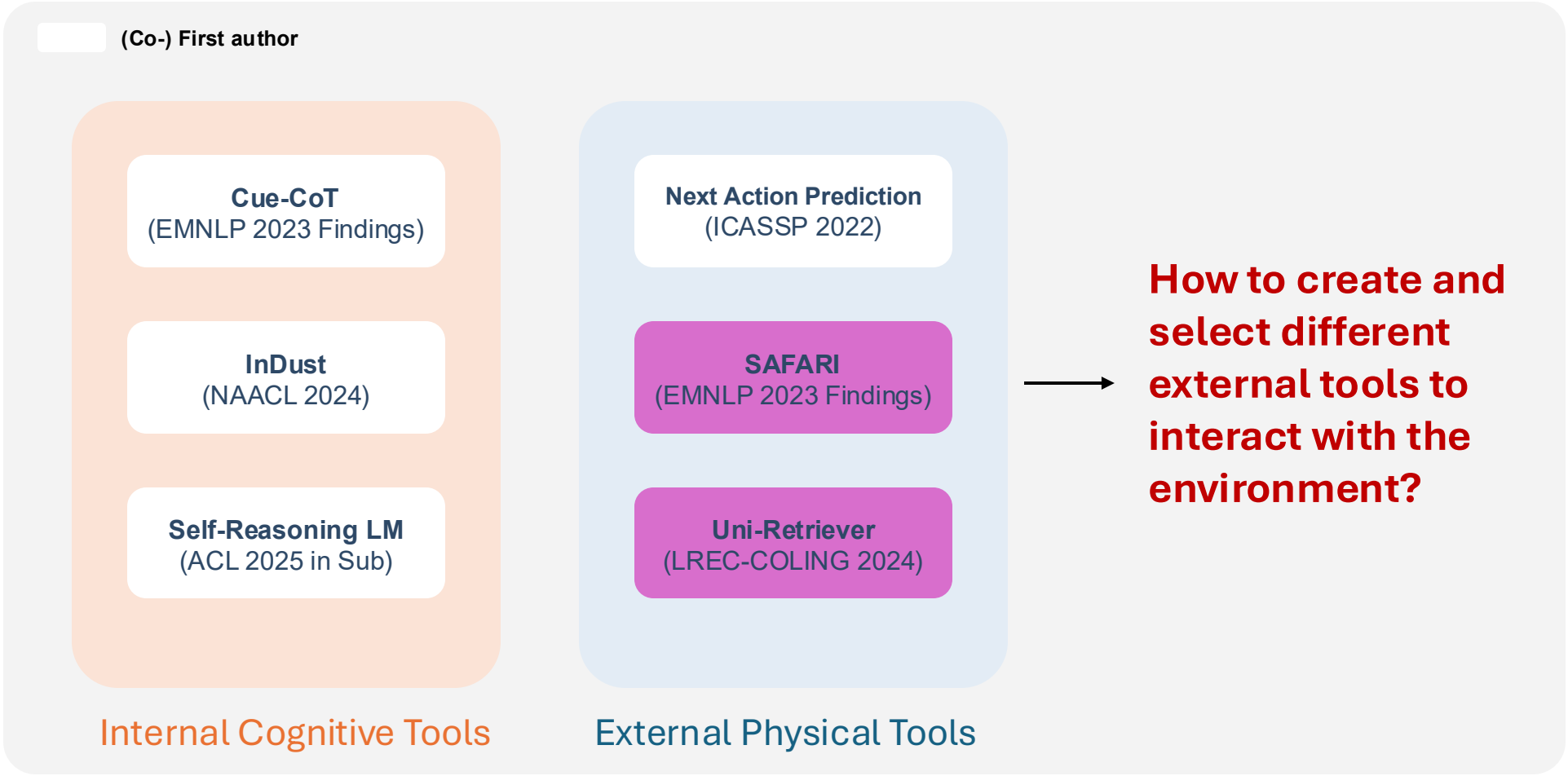
Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools



Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools



SAFARI – External Physical Tools

Dialogue Agent requires access to various external knowledge sources to deliver **reliable, informative, personalized, and helpful response.**

Research Question 1: How to build universal retriever to retrieve various evidences from different knowledge sources?

Research Question 2: How to decide which knowledge source to retrieve and plan call order of multiple knowledge sources if required?

RQ1: Tool Creation – Universal Retriever

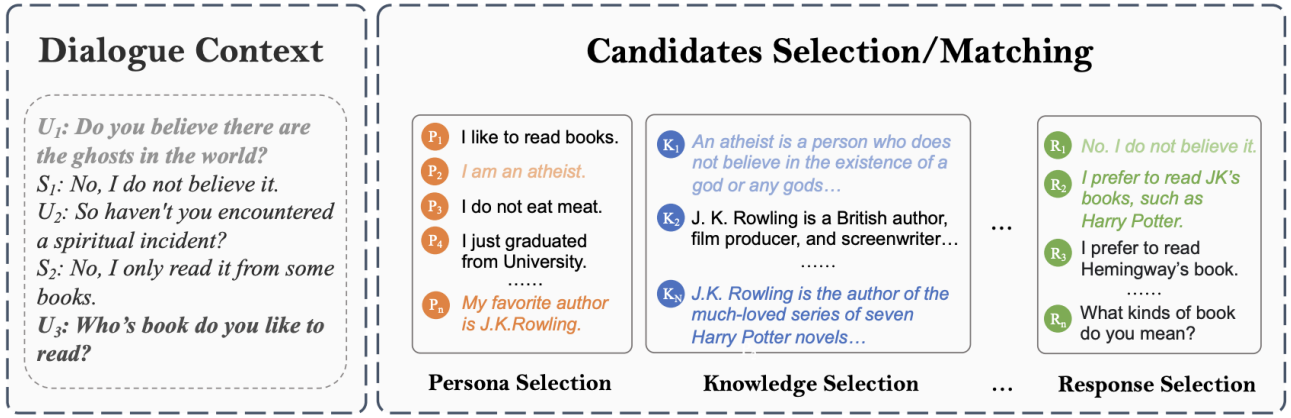
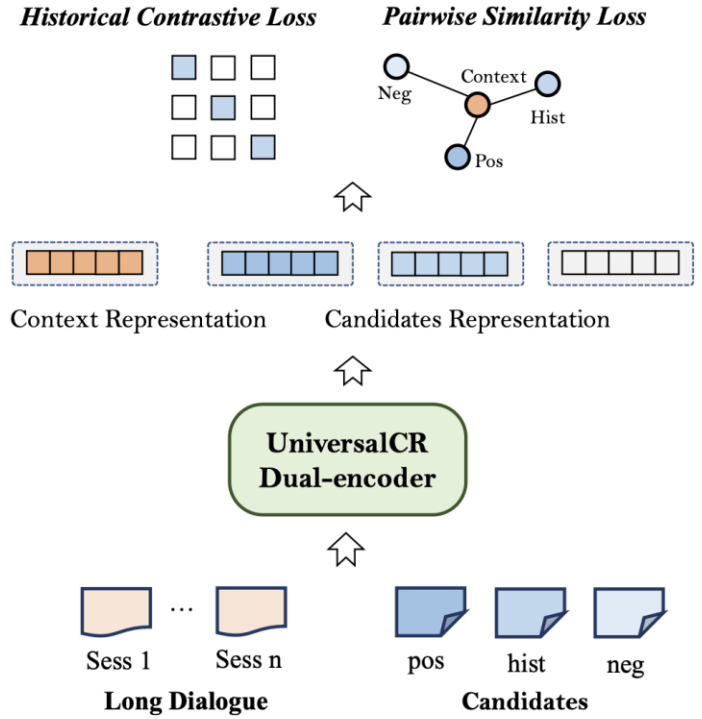


Figure 1: Different candidates selection tasks in a dialogue system: persona selection, knowledge selection, and response selection task. According to u_3 in the dialogue context, it is obvious to select p_n, k_n, and r_2 as target persona, knowledge, and response for the next turn respectively, while the p_2, k_1, and r_1 are historical selected persona, knowledge and response for historical turn u_1.

- **Hard negative mining**: Using historical selected candidates as semi-hard negative samples
- **Two loss constraints**: historical contrastive loss and pairwise similarity loss



Our proposed **UniCR** framework based on dual-encoder.

UniRetriever: Multi-task Candidates Selection for Various Context-Adaptive Conversational Retrieval (LREC-COLING 2024)

RQ2: Knowledge Source as Tools

- **Planning:** make a series of decision to determine whether or not use knowledge, which and when.

$$\mathcal{M} : \boldsymbol{c} \rightarrow K_i, K_j, ..., K_n \quad or \quad \texttt{NULL}, \qquad (1)$$

- **Retrieval:** retrieve *top-n* results from local databases according to the decided used source knowledge

$$\mathcal{R} : K_i, K_j, ..., K_n \rightarrow k_i^j, ..., k_n^m \qquad (2)$$

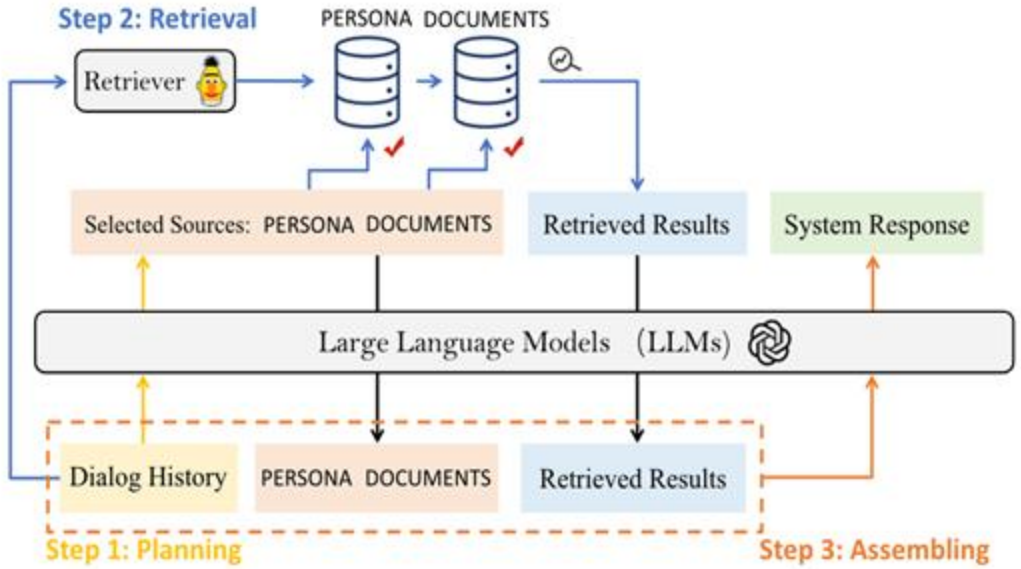
- **Assembling:** incorporate all retrieved middle results into the final response generation

Three Sub-tasks

Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogue (EMNLP 2023 Findings)

RQ2: Knowledge Source as Tools

- **Two Learning Methods**



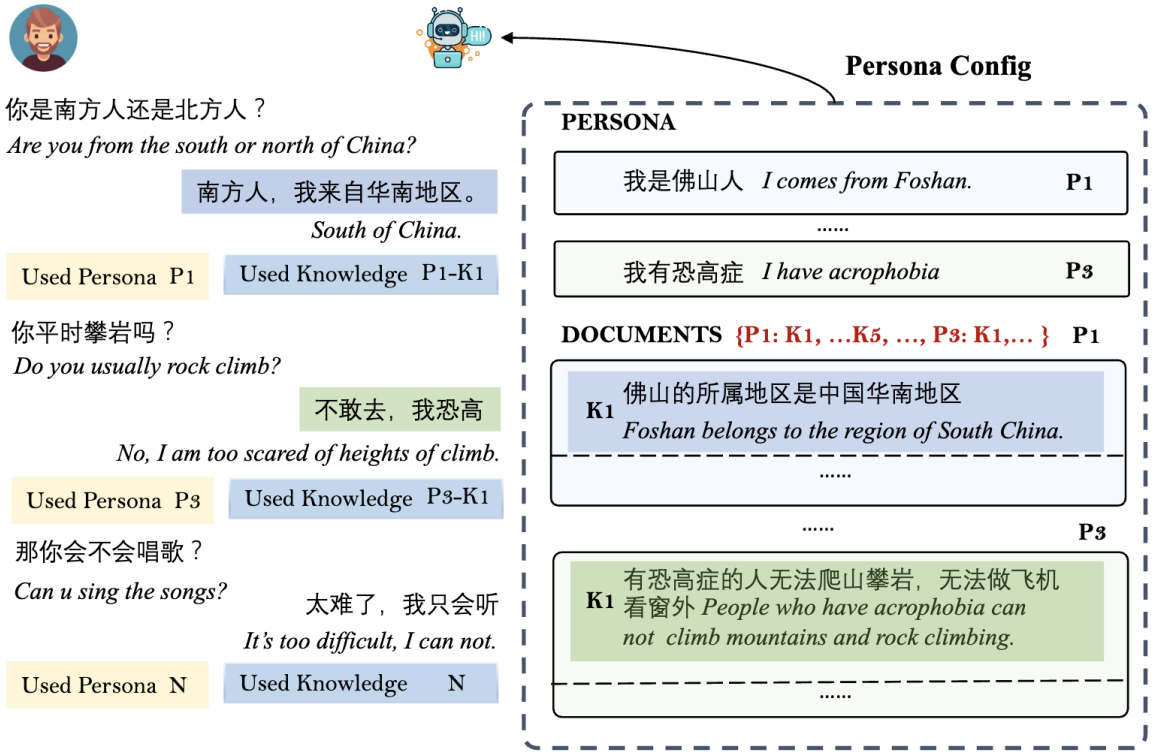
Learn from Data

There are different knowledge bases storing relevant information:
K_1: {K_1_DESC}
K_2: {K_2_DESC}
.....
There exists a dependency between these knowledge bases.
{DEPENDENCY_DESC}
Here is the dialogue between the user and the system: **{DIALOGUE}**
Based on the user's last question, please determine if it requires invoking the corresponding knowledge base. If the invocation is necessary, output the names of the knowledge bases in the order they should be invoked. If no invocation is needed, output **NULL**.

Learn from In-context Demonstrations

Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogue (EMNLP 2023 Findings)

RQ2: Knowledge Source as Tools



(c) An example of KBP dataset

- **Dataset**: KBP, three different knowledge sources
  - Do not need any external knowledge NULL
  - Only require PERSONA source of knowledge PERSONA
  - Require both PERSONA and DOCUMENT sources of knowledge PERSONA DOCUMENT
- **Evaluation**: Planning, Retrieval, Generation
- **Methods**:
  - Supervised finetuning
  - Prompting-based method

Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogue (EMNLP 2023 Findings)

SAFARI -- Results

Model	NULL	Persona	Both
Supervised			
BELLE-LLAMA-7B-2M	42.67 (194)	14.08 (17)	83.77 (1018)
CHATGLM-6B	**47.10** (129)	**31.96** (69)	**86.59** (1031)
Unsupervised			
Zero-shot			
BELLE-LLAMA-7B-2M	**28.55** (940)	8.94 (54)	32.47 (235)
CHATGLM-6B	25.60 (1225)	0.0 (0)	0.43 (4)
CHATGPT	11.45 (116)	**20.67** (233)	**74.88** (880)
In-context			
BELLE-LLAMA-7B-2M	9.22 (36)	18.21 (1193)	0.0 (0)
CHATGLM-6B	25.67 (1190)	1.49 (9)	4.62 (30)
CHATGPT	**27.95** (699)	**23.14** (238)	**41.98** (292)

Table 4: The F1 of different decisions in **Planning** of different LLMs under supervised/unsupervised settings. We also report the frequency of different decisions in the bracket. There are 181 NULL, 125 PERSONA and 923 PERSONA, and DOCUMENTS in the ground planning.

Model	BLEU1	Rouge-L	P.C	K.C
Supervised Setting				
BELLE-LLAMA-7B-2M	**30.48**	**34.61**	75.34	**46.62**
CHATGLM-6B	23.81	26.70	**76.99**	42.39
Unsupervised Setting				
Zero-shot				
BELLE-LLAMA-7B-2M	11.84	19.24	30.59	27.34
CHATGLM-6B	6.18	14.50	14.73	24.73
CHATGPT	12.06	24.44	**73.47**	**38.00**
In-context				
BELLE-LLAMA-7B-2M	**19.51**	22.25	72.98	24.89
CHATGLM-6B	13.74	19.69	16.92	24.89
CHATGPT	16.03	**25.62**	46.38	35.56

Table 6: The performance of **Assembling** under supervised/unsupervised settings.

Model	BLEU1	RougeL	P.C	K.C
CHATGLM-6B	23.81	26.70	76.99	42.39
+ Ground Planning	24.29	27.01	86.16	57.12
+ Ground Retrieval	**25.86**	29.15	79.52	53.95
+ Ground P & R	25.71	**29.43**	**90.56**	**72.99**
- Dependency	23.32	25.53	75.67	38.49
- Documents	23.06	25.34	75.91	36.53
- Planning*	23.51	25.98	72.90	24.89
- Planning**	23.69	26.81	71.60	34.91

Table 7: Ablation study on the impact of different steps and modules in SAFARI.

- Difficulty level of tasks: planning > retrieval > assembling. In-context learning does not help a lot at planning.
- Complex relationship between different knowledge sources (i.e., using the user profile to look up the movie database) remains an open challenge. Simply merge them all in one knowledge source lead to worse performance.

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

(Co-) First author

Cue-CoT
(EMNLP 2023 Findings)

InDust
(NAACL 2024)

Self-Reasoning LM
(ACL 2025 in Sub)

Internal Cognitive Tools

Next Action Prediction
(ICASSP 2022) → Pioneering work to use BERT to provide dense reward signals to guide the selection of dialogue actions via **reward shaping in RL**.

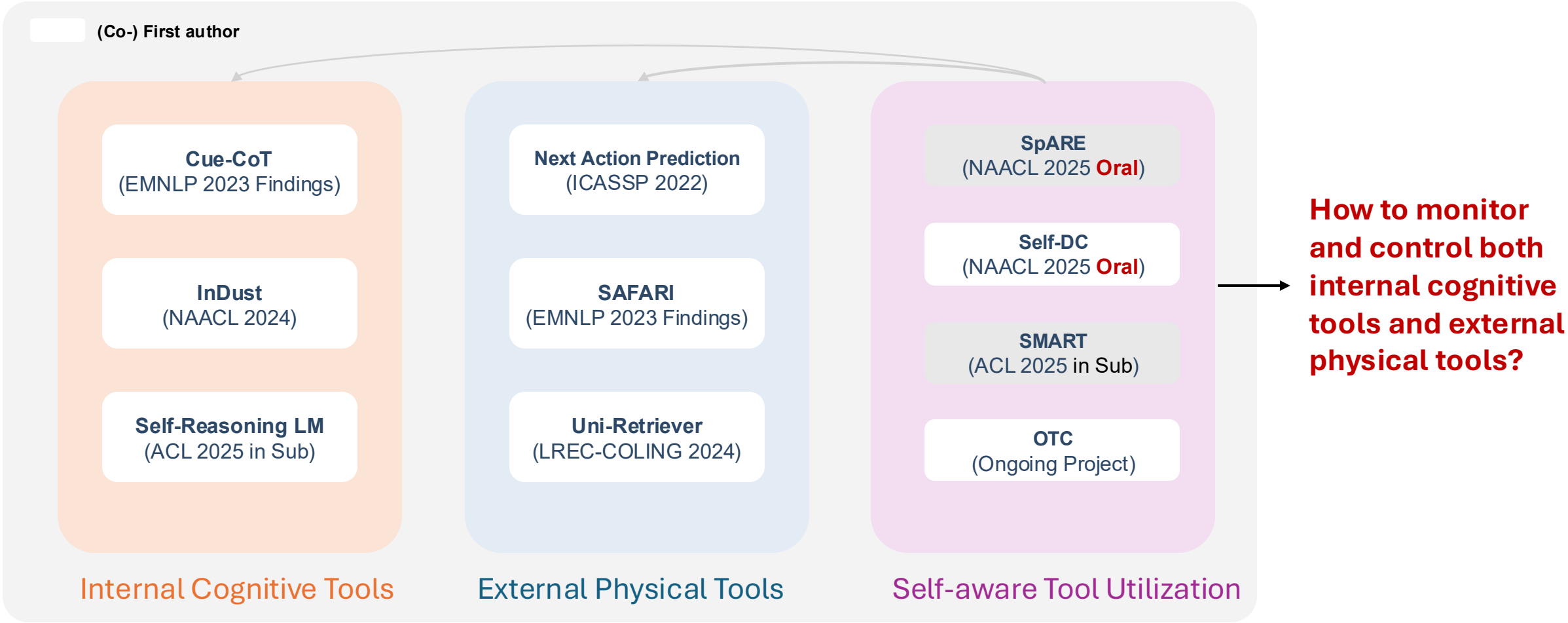
SAFARI
(EMNLP 2023 Findings) → First work to focus on **multi-source retrieval-augmented generation (RAG)** and teach LLMs to select suitable sources for response generation.

Uni-Retriever
(LREC-COLING 2024) → Build a **universal conversational retriever** to address multiple knowledge sources involved in dialogue response generation.

External Physical Tools

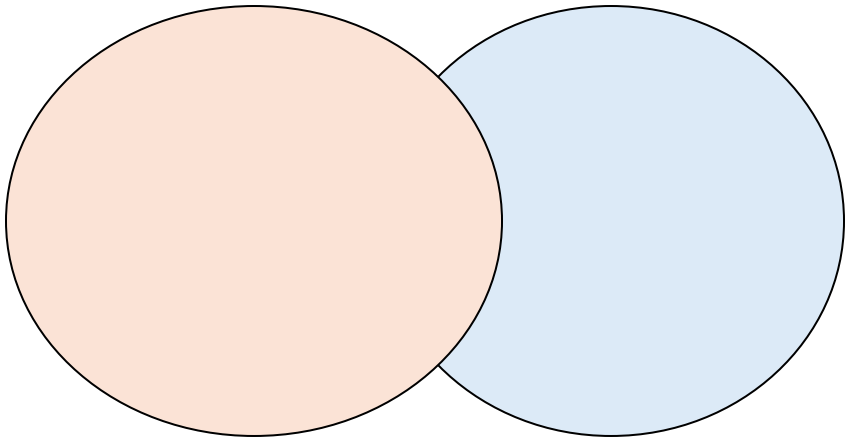
Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools



How to monitor and control both internal cognitive tools and external physical tools?

Can we better monitor and control LLMs' behaviors?



Internal Cognitive Tools

External Physical Tools

Tool Planning: Different tools serve different purposes and are designed to produce distinct types of outputs.

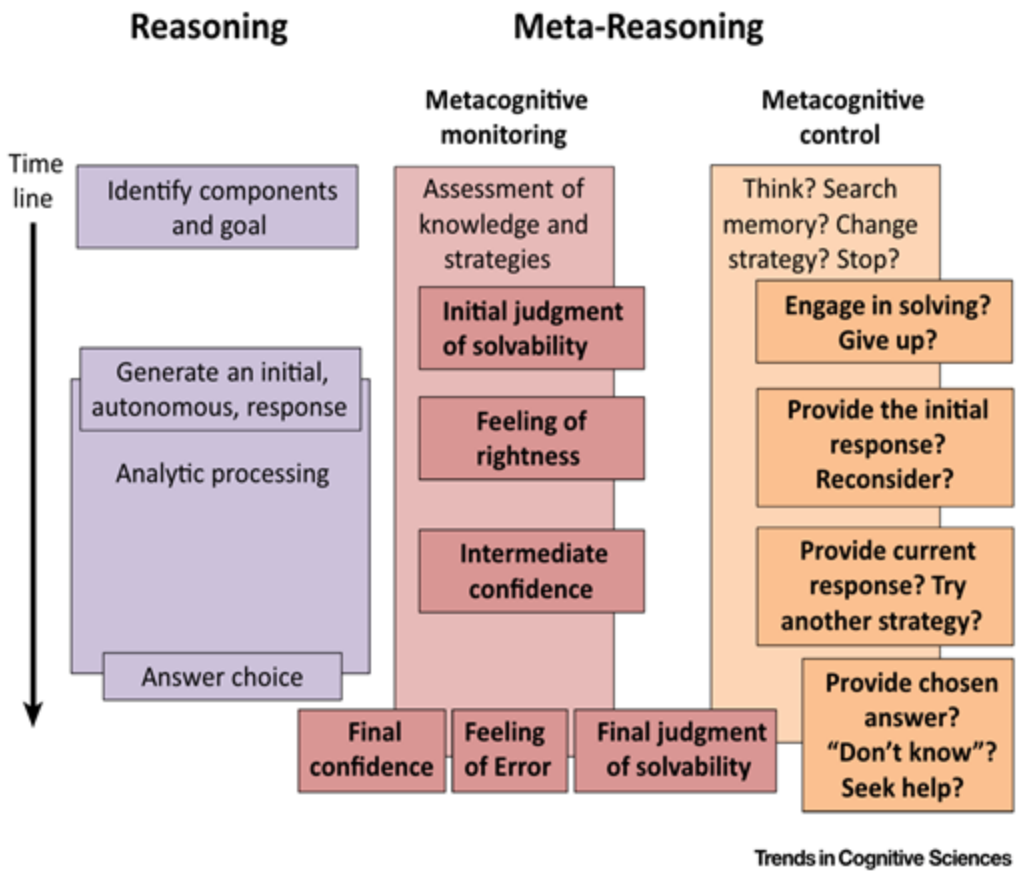
Tool Management: When multiple tools have similar functionality, the choice of which tool to use depends on several factors: personalization, efficiency, and practical limitations.

Tool Conflicts: When the information or results from different tools contradicts, it can be challenging to determine which source to trust.

All of these necessitate monitoring and control of LLMs' behaviors.

Let's recall how human call tools…

Natural Mind



Artificial Mind

Monitoring

Judgement of solvability
Intermediate confidence
Reward model
Uncertainty estimation
…

Control

Cognitive tools
Physical tools
Tool planning
…

Meta-Reasoning: Monitoring and Control of Thinking and Reasoning

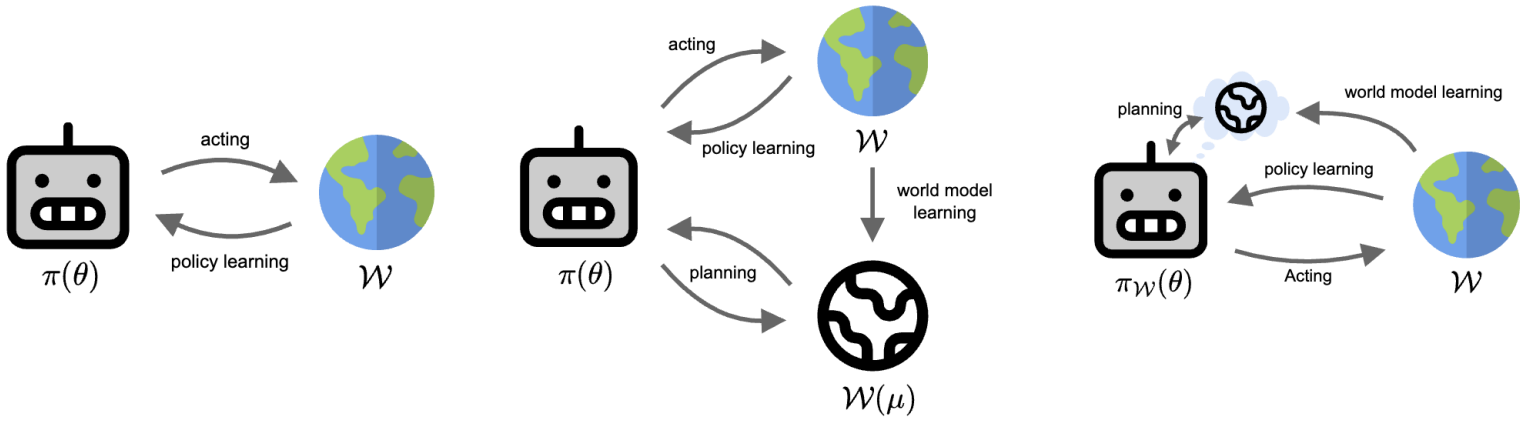
Monitoring and Control – Meta-Reasoning Theory

❖ We want the agent call **internal tools when they know certain knowledge**, while only invoke **external tools when they do not know certain knowledge**.

Why?

*"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."*

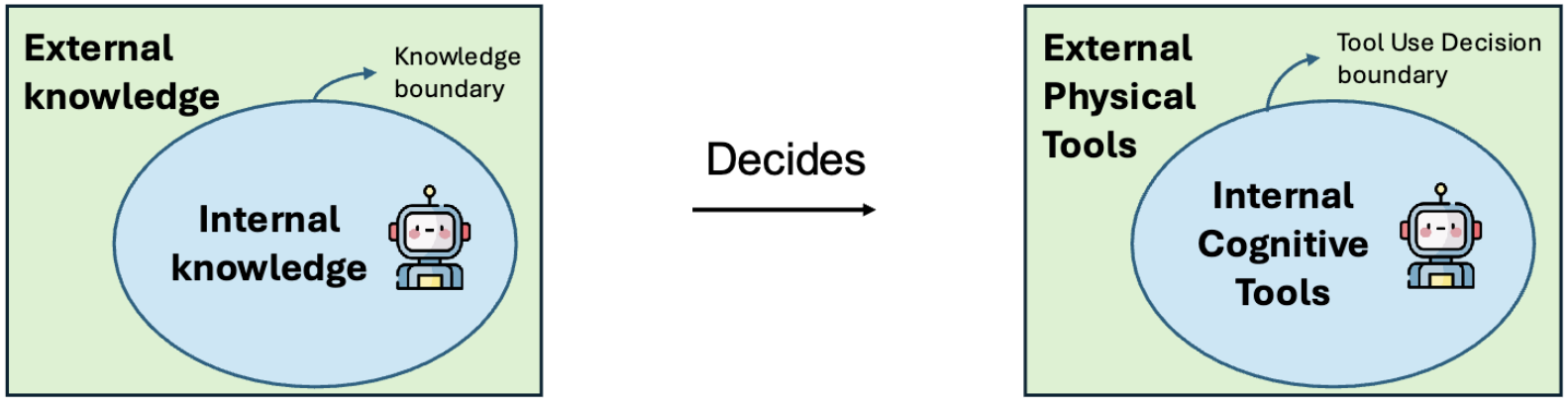
--- Yann Lecun



A Path Towards Autonomous Machine Intelligence
DYNA-THINK: Synergizing Reasoning, Acting, and World Model Simulation in AI Agents

Monitoring and Control – Meta-Reasoning Theory

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. The challenge here is **self-aware tool utilization**



Monitor: Self-aware Knowledge Boundary

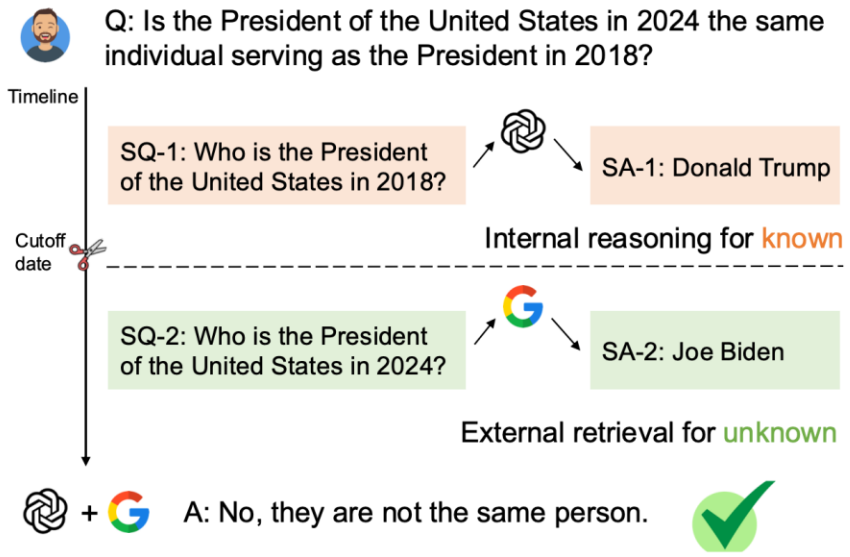
Control: Self-aware Tool Utilization

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

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

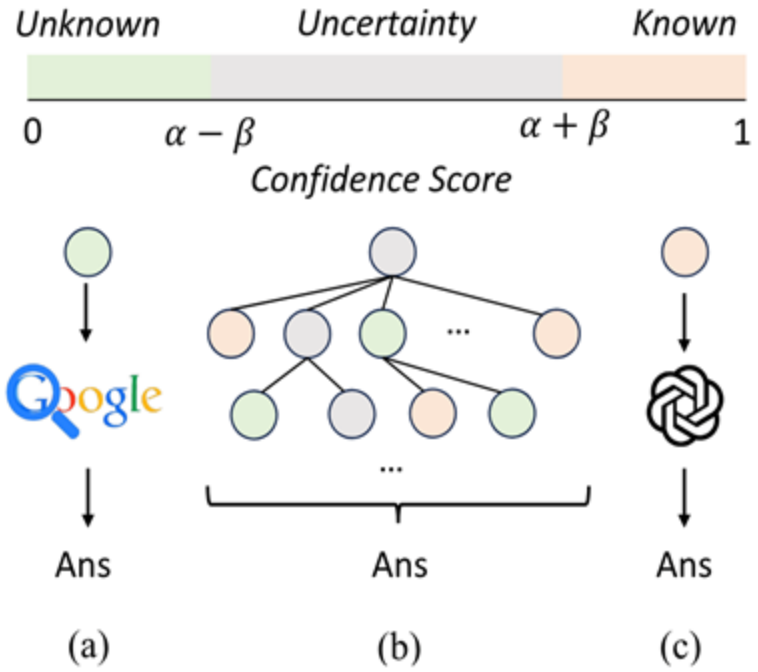
Given a LLM, its knowledge boundary is fixed at time *t*.

Thus, given one LLM and one question, there are four cases.



- ***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*.

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



Our proposed **Self-DC** framework, including a) retrieve-then-read for unknown questions, b) decompose-and-combination for uncertain questions; and c) generate-then-read for known questions.

Since different LLMs have different knowledge boundaries, we design a two-step prompting strategy:

Step1: knowledge boundary assessment for different LLMs, i.e., uncertainty estimation such as prompting LLMs to generate confidence scores or multiple sampling. (***monitor***)

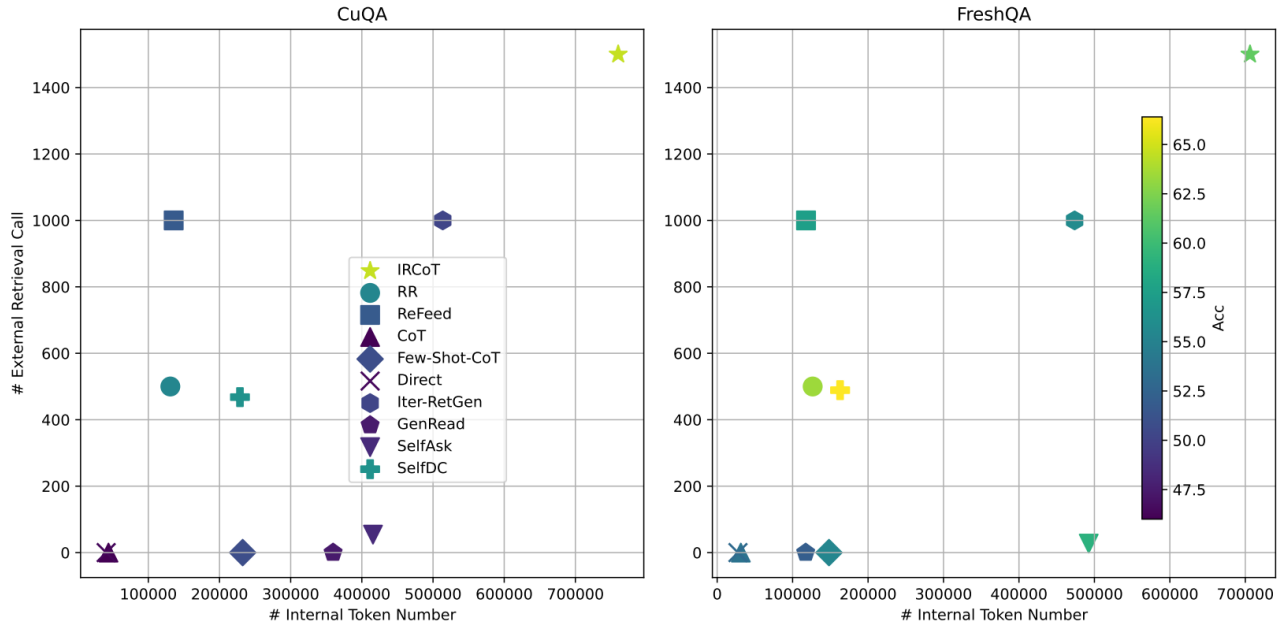
Step2: divide-and-conquer (***control***)

This is **the first work to consider the relationship between reasoning and acting** in terms of trade-off between effectiveness and efficiency.

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

- Self-DC achieves **better trade-off between efficiency and effectiveness** than retrieval-based methods.

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**



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

Agentic RL – OTC-PO

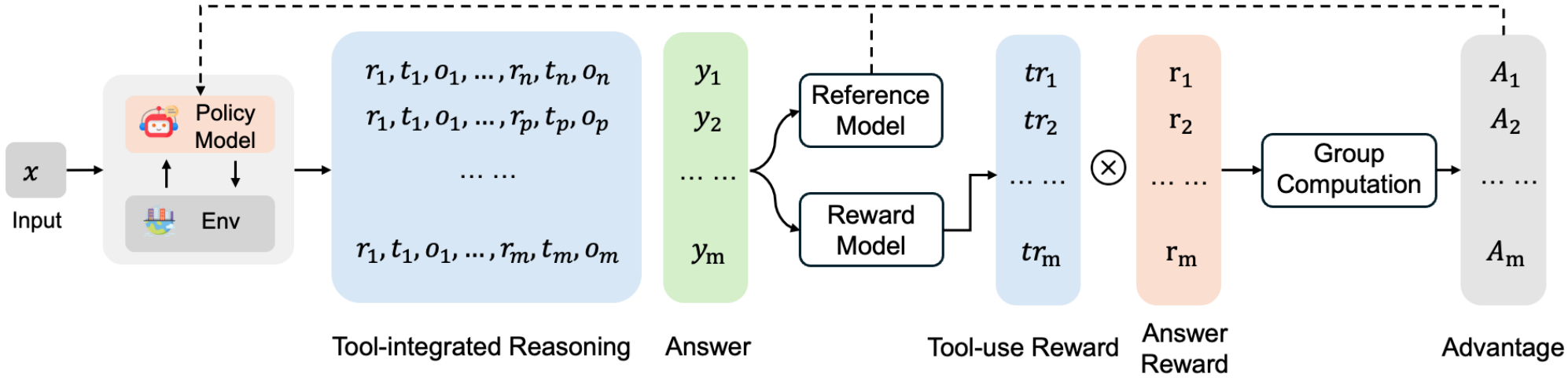
Can we effectively align an agent's tool use boundary to its knowledge boundary via RL, so that smarter tool use could be achieved from experience?

Agentic RL – OTC-PO

We start from one fundamental assumption that given one problem and one LLM, there exist an **optimal number of external tools required**, defined as **minimal number** of tool calls to solve the problem correctly.

Solution: add tool-use reward as a **coefficient** of (outcome reward + format reward)

Why tool-use reward? → Tool overuse and underuse brings serious efficiency issues, especially considering the cost of various tool calls in terms of time, money and computation.



Agentic RL – OTC-PO

- We are **the first** to define this problem as follows: Here is a tool-integrated reasoning trajectory:

$$\tau_k = (r_0, tc_0, o_0), (r_1, tc_1, o_1), ...(r_k, tc_k, o_k),$$

where r_i, tc_i, o_i denotes the reasoning, tool call and returned observation respectively. The objective of task is to provide the correct answer with minimal cost of tools given the question q and model M.

$$\arg\min_{\tau} \ \text{Cost}(\tau) \quad \text{subject to} \quad \mathcal{M}(q, \tau) = \hat{a},$$

- We are **the first** to define tool productivity (TP) as the fraction between benefits and cost.

$$\text{TP} = \frac{\sum_{i=1}^{N} \mathbb{I}\{y_i = \hat{y}_i\}}{\sum_{i=1}^{N} tc_i}$$

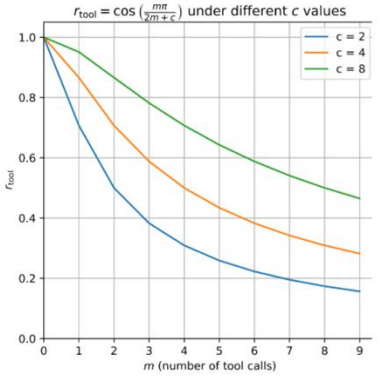
where I is the indicator function which equals 1 if the generated answer is the ground truth answer.

Reward Design -- OTC-PO

The key idea is to penalize trajectories that involve excessive use of tool calls.

❖ **OTC-PPO**

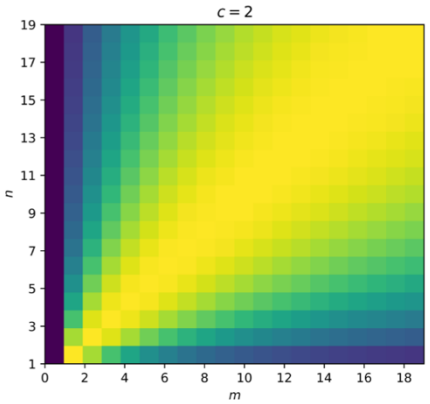
$$r_{tool} = \cos\left(\frac{m * \pi}{2m + c}\right)$$



OTC-PPO

❖ **OTC-GRPO**

$$r_{tool} = \begin{cases} 1 & \text{if } f(m,n) = n = 0 \\ \cos\left(\frac{m*\pi}{2m+c}\right) & \text{if } n = 0 \\ \sin\left(\frac{f(m,n)*\pi}{2n}\right) & \text{otherwise} \end{cases} \qquad f(m,n) = \begin{cases} 0, & \text{if } m = 0 \text{ and } n = 0 \\ m, & \text{if } n = 0 \\ \frac{2nm}{m+n}, & \text{otherwise} \end{cases}$$

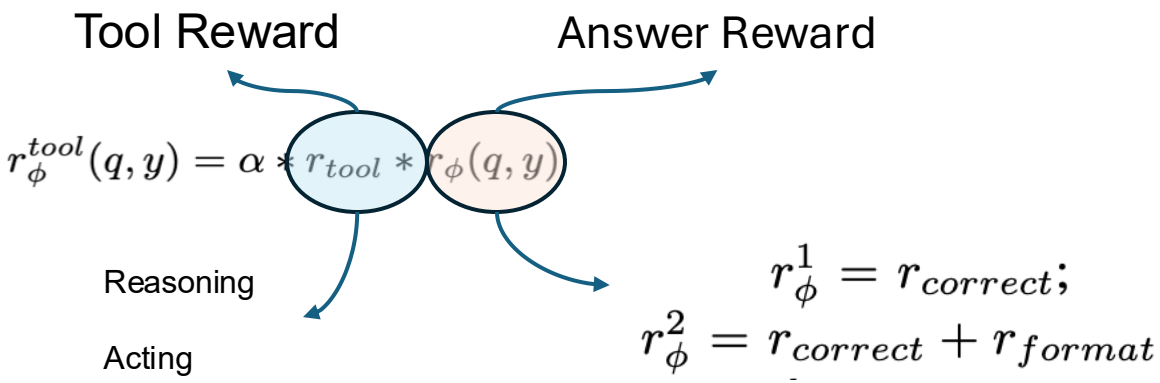


OTC-GRPO

m is the *number of tool calls for current trajectory of the sample* while n stands for the *minimal number of tool calls for the sample till now*.

Reward Design -- OTC-PO

❖ **Unified Tool-integrated Reward Function**



- Maximally preserves overall accuracy;
- Mitigates the risk of reward hacking compared to additive forms;
- Generalizes well to different formulations of r_{tool} (including both reasoning and acting) and $r_{\emptyset}$ (appliable to different agentic tasks)

Agentic RL – OTC-PO

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%)

Models	AIME24			AIME25		
	EM (↑)	TC (↓)	TP (↑)	EM (↑)	TC (↓)	TP (↑)
Qwen2.5-Math-1.5B(-Base)						
Qwen2.5-Math-1.5B-Ins	10.0	-	-	10.0	-	-
Qwen2.5-Math-1.5B-Ins-TIR	13.3	1.1	12.1	13.3	1.4	9.5
ToRL-GRPO	23.3	2.2	10.6	23.3	2.3	10.1
OTC-GRPO	20.0	1.1 (▼ 50.0%)	18.2 (▲ 71.7%)	20.0	1.1 (▼ 41.2%)	18.2 (▲ 80.2%)
Qwen2.5-Math-7B(-Base)						
Qwen2.5-Math-7B-Ins	10.0	-	-	16.7	-	-
Qwen2.5-Math-7B-Ins-TIR	26.7	1.6	16.4	16.7	1.4	12.2
Base-RL	33.3	-	-	6.7	-	-
ToRL-GRPO	36.7	2.1	17.5	26.7	2.1	12.7
OTC-GRPO	36.7	0.7 (▼ 66.7%)	52.4 (▲ 199.4%)	23.3	0.8 (▼ 61.9%)	29.1 (▲ 129.1%)

- Simply optimize answer correctness via RL bring serious over-optimization problem such as **tool overuse issue**.
- OTC-PO significantly **reduce the tool calls and improve the tool productivity**, especially when the size of model increases.

Agentic RL – OTC-PO

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)**

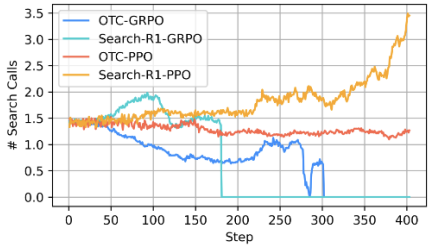
Undesirable tool use behavior

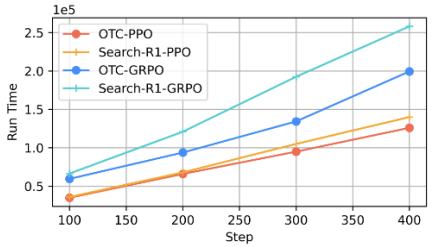
OTC-PPO: <think> First, I need to determine if Giuseppe Verdi and Ambroise Thomas are both opera composers. </think>
<search> Giuseppe Verdi and Ambroise Thomas both Opera composers ***</search>***
<observation> ... </observation>
<think> Based on the search results, Giuseppe Verdi and Ambroise Thomas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

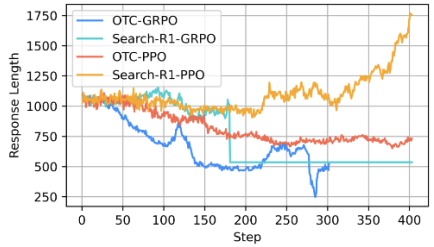
OTC-GRPO: To answer the question of whether Giuseppe Verdi and Ambroise Thomas are both Opera composers, let's first reason through the information:
1. Giuseppe Verdi is a well-known Italian opera composer. He is famous for his numerous operas, including "Rigoletto," "La Traviata," and "Aida."
2. Ambroise 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)**

- **Cognitive offloading** appears more often in larger LLMs.
- **(Minimizing external tool calls = maximizing internal reasoning) = Smart Agent.**

Agentic RL – OTC-PO







Models	TriviaQA		PopQA		2Wiki		Musique		Bamboogle	
	EM (↑)	TC (↓)	EM (↑)	TC (↓)	EM (↑)	TC (↓)	EM (↑)	TC (↓)	EM (↑)	TC (↓)
Qwen2.5-3B(-Base)										
Search-R1-PPO	0.566	1.580	0.425	1.631	0.258	1.675	0.051	1.922	0.063	1.766
Search-R1-GRPO	0.587	1.455	0.345	1.542	0.257	1.991	0.084	2.263	0.203	1.859
OTC-PPO	0.551	**1.008**	0.409	**1.009**	0.235	**1.050**	0.045	1.051	0.063	**1.016**
OTC-GRPO	**0.608**	1.046	**0.441**	1.030	**0.341**	1.561	**0.124**	1.734	**0.266**	1.547
Qwen2.5-7B(-Base)										
Search-R1-PPO	0.596	3.353	0.420	3.315	0.326	4.116	0.135	4.294	0.375	3.641
Search-R1-GRPO	0.578	1.704	0.411	1.754	0.340	2.521	0.130	2.616	0.203	1.859
OTC-PPO	**0.623**	1.066	0.425	1.083	**0.363**	1.868	**0.152**	1.942	**0.391**	1.828
OTC-GRPO	0.597	**0.430**	**0.431**	**0.739**	0.311	**0.938**	0.130	**1.224**	0.250	**0.781**

Simple

Faster

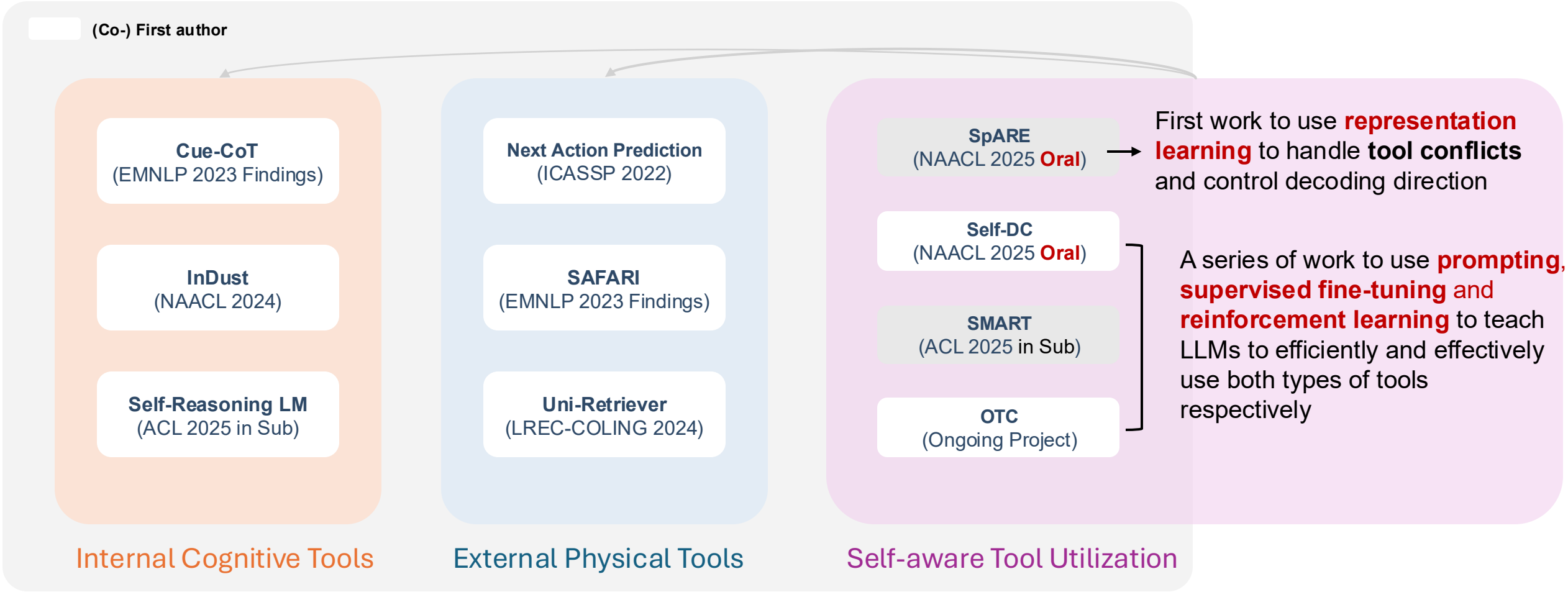
Generalizable

Scalable

Acting Less is Reasoning More! Teaching Model to Act Efficiently

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools



Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

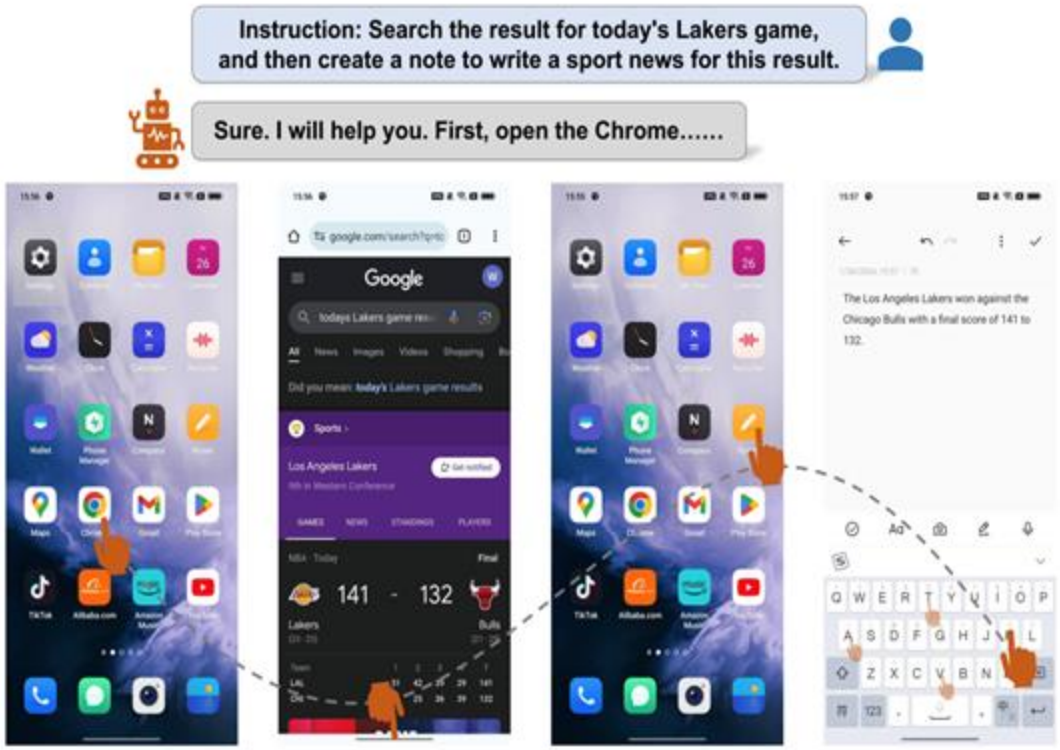
How to evaluate the performance of existing LLMs or Agent under complex real-world applications? ←

(Co-) First author

- **AppBench** (EMNLP 2024)
- **DialogTool** (ACL 2025 Findings)
- **ToolSpectrum** (ACL 2025 Findings)

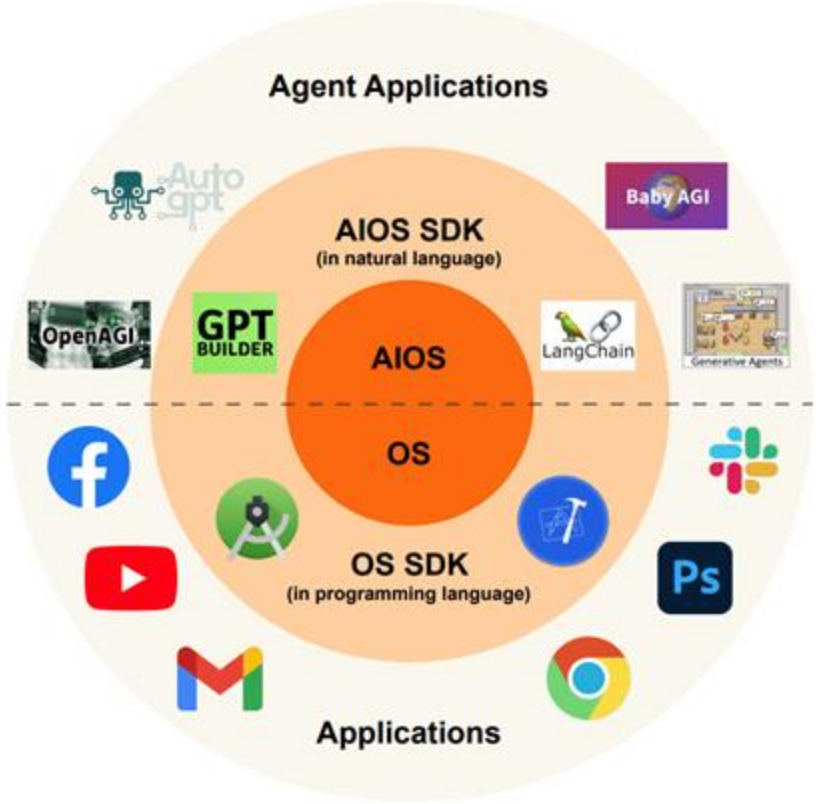
Benchmarks

Benchmarks



Human-computer Interaction
GUI, Websites, Apps, …

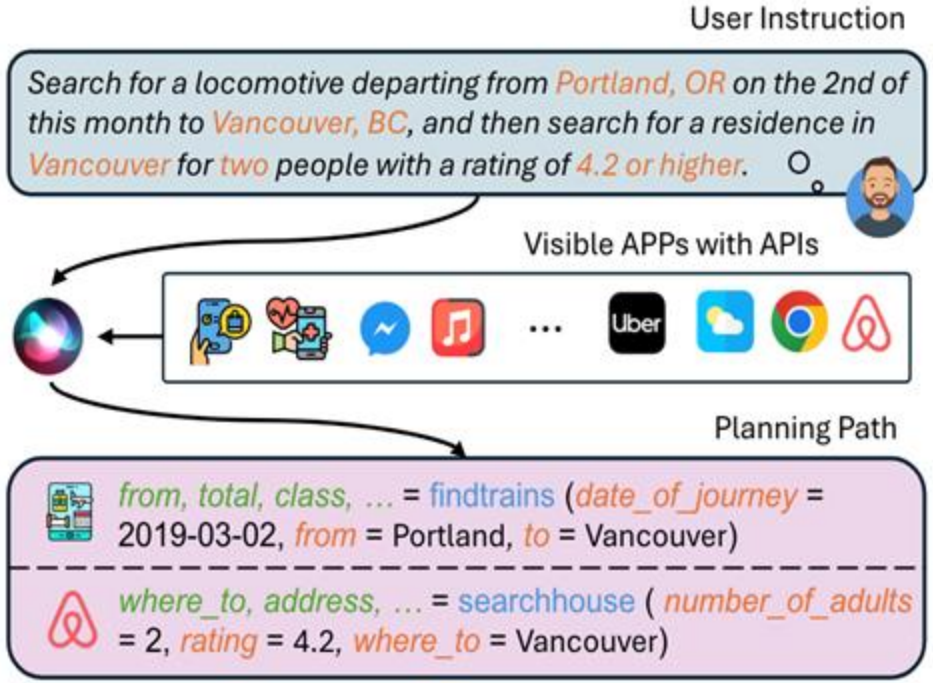
How **human** interact with the world.



Model-computer Interaction
LLM as planner / controller,
Language Agent, …

How **model** interact with the user / world.

Benchmarking Tool Planning in *Single-turn* Interaction



Single-turn Interaction

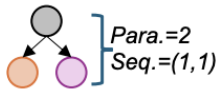
✓ Tool Selection ✓ Tool Planning

✓ Tool Management

- Two Key Challenges
  - ❑ **Graph structure**: some APIs can be executed independently while others need to be executed one by one, resulting in graph-like execution order;
  - ❑ **Permission constraints**: which source is authorized to execute the API call.
- Four Data Types
  - **Single App Single API (SS):** common cases
  - **Single App Multiple API (SM):** mostly *sequential*
  - **Multiple App Single API (MS)**: mostly *parallel*
  - **Multiple App Multiple API (MM):** both *sequential* and *parallel*

AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction (EMNLP 2024)

Benchmarking Tool Planning in *Single-turn* Interaction

Data Type	Example	Structure
SS	Instruction: *Find a house with a rating of 4.6 or higher for a trip to Delhi for two people, inquire about laundry service availability* Output: House: address, phone_number, total_price, has_laundry_service, … = searchhouse(number_of_adults='2', rating='4.60', where_to='Delhi')	
SM	Instruction: *Please book a Hatchback car with insurance to be picked up from Warsaw Chopin Airport on March 7th at 1:30 pm, and returned on March 13th in Warsaw.* Output: Rents: pickup_location, price_per_day, ... = getcarsavailable(car_type='Hatchback', city='Warsaw', end_date='2019-03-13', pickup_time='13:30', start_date='2019-03-07') Rents: car_type, car_name, ... = reservecar(add_insurance='True', car_type=car_type, end_date=end_date, pickup_location=#**pickup_location**, pickup_time=pickup_time, start_date=start_date)	
MS	Instruction: *Search for a locomotive departing from Portland, OR on the 2nd of this month to Vancouver, BC, and then search for a residence in Vancouver for two people with a rating of 4.2 or higher.* Output: Train: from, total, class, … = findtrains (date_of_journey = 2019-03-02, from = Portland, to = Vancouver) House: address, phone_number, total_price, has_laundry_service, … = searchhouse(number_of_adults='2', rating='4.2', where_to= Vancouver')	
MM	Instruction: *Please make a reservation for 3 people at one Korean restaurant in San Francisco at 1:30 pm on March 12th, and also book a Luxury taxi for 3 to 4 Embarcadero Center.* Output: Restaurant: restaurant_name, has_vegetarian_options, phone_number, rating, address, price_range, category, ... = findrestaurants (category ='Korean', has_seating_outdoors='True', location='San Francisco') Restaurant: date, time, location, ... = reserverestaurant (date='2019-03-12', location=location, number_of_seats='3', restaurant_name = #**restaurant_name**, time='13:30') Rents: destination, ride_type, ride_fare, wait_time, number_of_seats = getride(destination='4 Embarcadero Center', number_of_seats='3', ride_type='Luxury')	

AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction (EMNLP 2024)

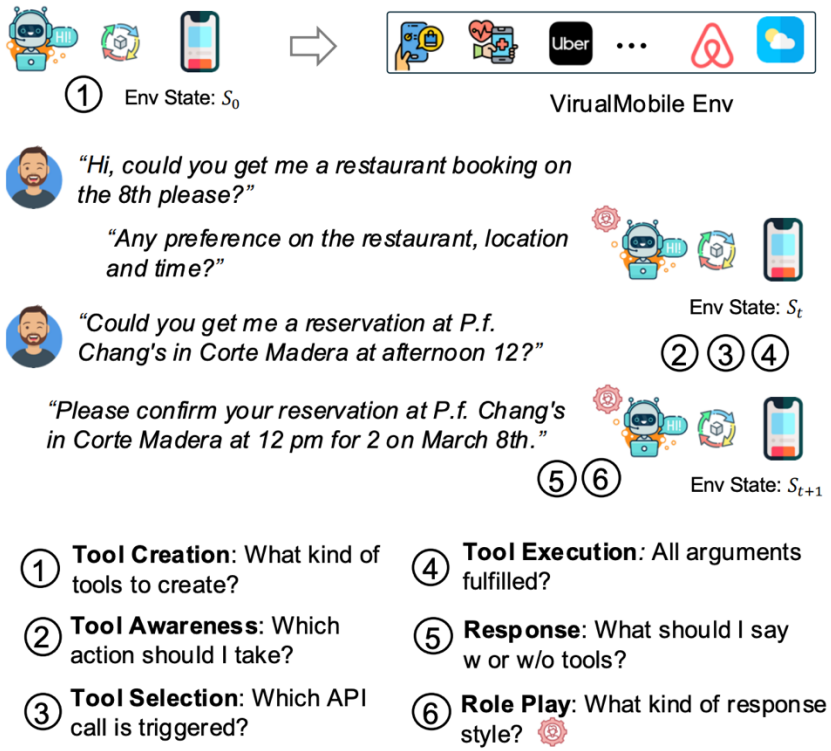
Benchmarking Tool Planning in *Single-turn* Interaction

Models	SS			SM			MS			MM		
	$F1_{app}$	$F1_{api}$	Succ	$F1_{app}$	$F1_{api}$	Succ	$F1_{app}$	$F1_{api}$	Succ	$F1_{app}$	$F1_{api}$	Succ
Mistral-7B	55.97	16.31	0.51	36.59	15.09	0.50	33.72	6.42	0.00	28.92	7.56	0.00
Vicuna-13B	43.20	3.70	2.00	34.71	4.63	0.50	20.43	3.10	0.00	21.05	2.52	0.00
LLaMA3-8B	63.04	42.67	23.23	37.20	25.33	0.50	30.65	19.52	0.10	26.39	17.80	0.05
LLaMA3-70B	71.20	70.00	50.00	46.48	46.96	10.50	32.61	32.96	2.50	28.97	28.53	0.50
QWen1.5-7B	48.14	19.54	0.00	30.13	16.71	0.00	23.24	10.11	0.00	23.76	11.55	0.00
QWen1.5-14B	72.89	28.41	10.10	41.89	25.51	1.50	42.22	21.98	0.80	32.36	15.07	0.00
QWen1.5-72B	81.23	24.28	12.50	**51.89**	25.27	1.00	**45.94**	13.42	0.62	**38.53**	11.51	0.00
GPT-3.5	63.60	57.95	30.81	41.49	43.65	6.50	33.17	34.53	7.00	27.79	28.09	1.00
GPT-4o	**88.31**	**86.87**	**70.92**	50.83	**50.57**	**20.50**	39.39	**39.14**	**11.00**	32.62	**32.35**	**2.00**

- Overall, GPT-4o achieves the best overall performance, while LLaMA3-70B sometimes outperforms GPT-3.5, mostly in scenarios only involving single APP.
- As the size of the model increases, the performance can get further improved regardless of the type of instructions and the improvement becomes less significant with multiple APPs.
- The complexity of planning highly impacts the performance of these models. MM is the most challenging tasks, followed by MS, SM and SS.

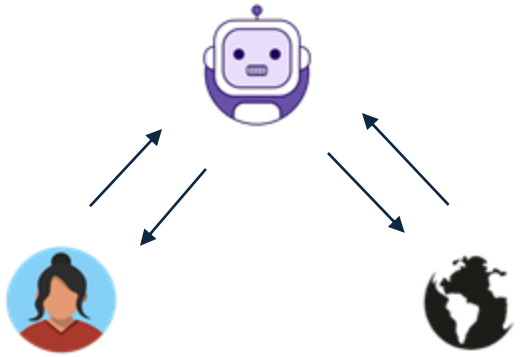
AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction (EMNLP 2024)

Benchmarking Tool Utilization in *Multi-turn* Interaction



Multi-turn Interaction

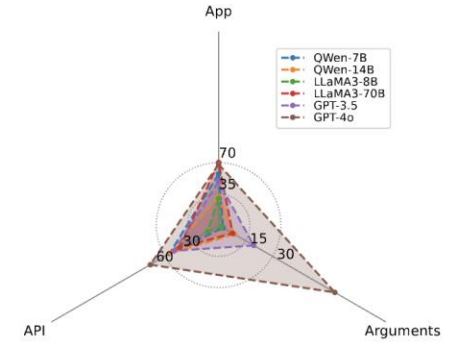
Benchmark	Tool Learning						Evaluation			
	Apps	APIs	Argu.	C. S. E.	States	Awareness	Role	Hierarchical	Resp.	Multi-turn
APIBench (Patil et al., 2023)	3	1,715	(1.5/1.0)	✗|✓|✓	✗	✗	✗	✗	✗	✗
API-Bank (Li et al., 2023)	8	53	(2.5/1.0)	✗|✓|✓	✗	✗	✗	✗	✓	✓
ToolBench (Qin et al., 2023c)	49	16,464	(1.0/1.0)	✗|✓|✓	✗	✗	✗	✓	✗	✓
ToolQA (Zhuang et al., 2023)	6	13	(1.0/1.0)	✗|✓|✓	✗	✗	✗	✗	✓	✗
GAIA (Mialon et al., 2023)	-	-	-	✗|✓|✓	✗	✗	✗	✗	✓	✗
UltraTool (Huang et al., 2024)	22	2032	(4.1/1.6)	✓|✓|✓	✗	✗	✗	✗	✗	✗
AgentBench (Liu et al., 2023)	8	-	-	✗|✓|✓	✗	✗	✗	✗	✗	✗
MINT (Wang et al., 2024c)	8	-	-	✗|✓|✓	✗	✗	✗	✗	✓	✗
AgentBoard (Ma et al., 2024)	9	-	-	✗|✓|✓	✓	✗	✗	✗	✗	✓
DialogTool	16	31♡	(**4.2/7.5**)	✓|✓|✓	✓	✓	✓	✓	✓	✓



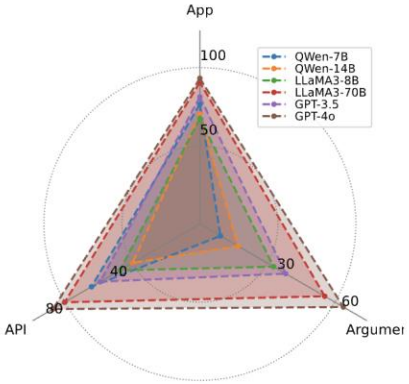
User-Agent-Environment Interactions

Benchmarking Tool Utilization in *Multi-turn* Interaction

Models	Tool Creation	Tool Utilization			Role-consistent Responses			
		Awareness	Selection	Execution	BLEU	R.L	Role	Human
ChatGLM3-6B	31.5	58.9	32.8	6.8	7.8	7.5	4.8	1.64
LLaMA2-7B	33.2	63.5	27.4	7.0	6.8	5.7	6.2	1.25
QWen1.5-7B	21.9	68.9	54.7	11.3	8.0	7.4	7.0	2.82
Mistral-7B	11.4	42.5	51.8	22.6	8.0	7.1	6.7	2.28
LLaMA3-8B	62.2	46.3	61.4	45.6	8.3	7.7	7.0	2.69
LLaMA2-13B	48.8	47.1	51.1	11.7	7.7	6.4	6.5	2.17
Vicuna-13B	-	64.5	62.9	12.3	10.1	11.5	6.0	2.59
QWen1.5-14B	27.9	51.7	55.6	21.8	9.3	10.9	7.5	2.44
QWen1.5-72B	49.7	**75.5**	71.9	49.3	10.8	**15.3**	7.4	3.37
LLaMA2-70B	23.0	34.7	57.8	32.6	8.5	10.7	6.2	2.56
LLaMA3-70B	**69.7**	40.2	57.1	68.1	9.0	11.3	7.7	2.98
GPT-3.5	63.3	67.9	50.0	42.6	10.2	11.9	6.7	3.42
GPT-4o	66.7	63.5	**77.8**	**68.7**	**11.4**	14.5	**8.3**	**3.56**



(a) Tool Selection under *Flat* Setting



(b) Tool Selection under *Hierarchical* Setting

- No LLMs achieve an accuracy exceeding 80% at the tool creation and utilization, and most LLMs performed poorly in **tool creation and execution tasks** compared to their performance in **awareness and selection** tasks, revealing the complexity and challenges of our dataset and environment.
- We can find that models likely achieves much higher performance under hierarchical setting (select App first and then API) instead of flat one.

Tool-integrated Agents

ROADMAP: Building Agent with Self-aware Tool Utilization with both Internal Cognitive Tools and External Physical Tools

(Co-) First author

Pioneer work to consider the **graph dependency structure** and **permission constraints** of various APIs of different Apps in mobile device.

AppBench
(EMNLP 2024)

Pioneer work to consider the **stateful API calls** of various Apps, and evaluate the whole lifecycle of tool utilization: tool creation, awareness, selection, execution and response generation.

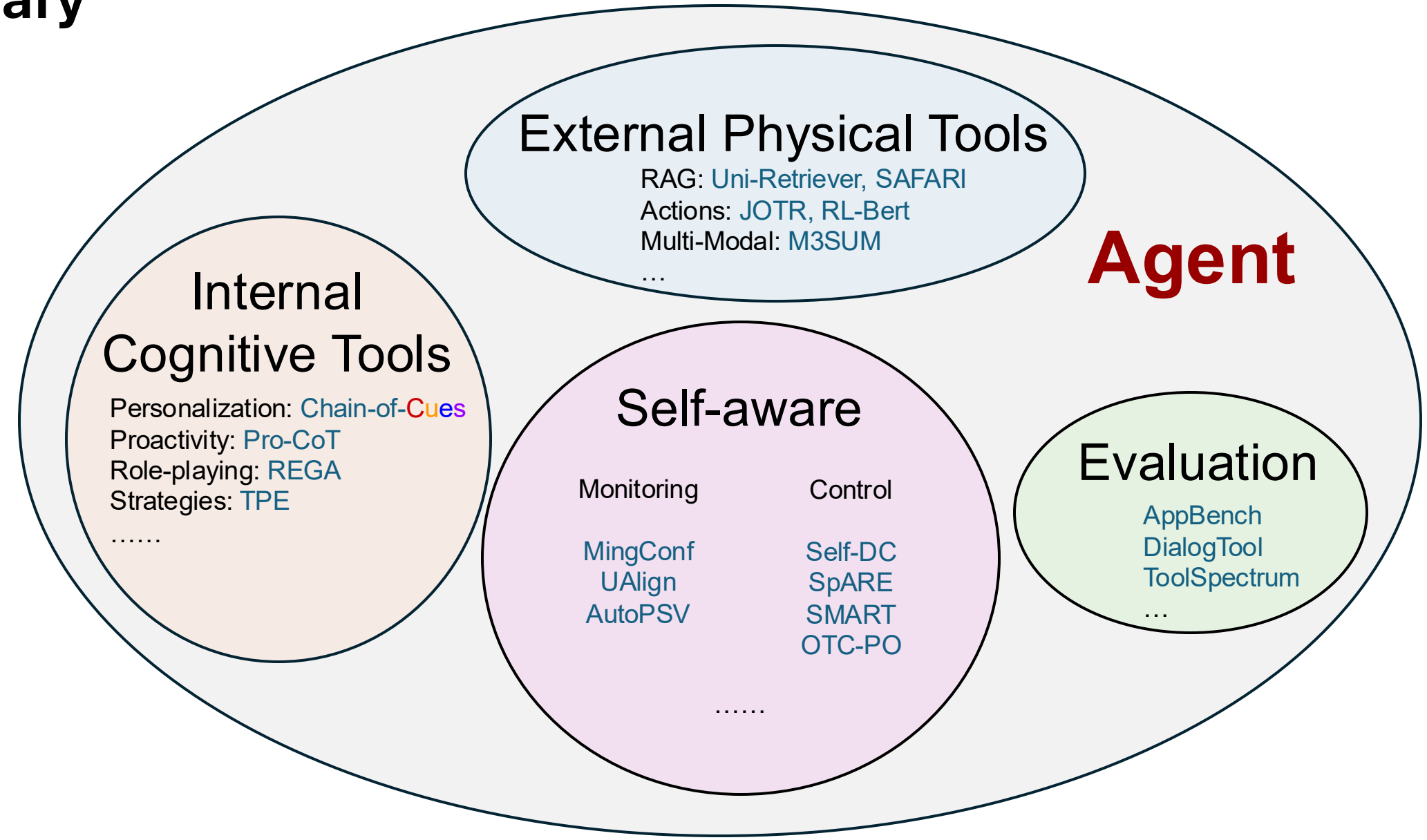
DialogTool
(ACL 2025 Findings)

First work to consider the **personalized tool utilization.**

ToolSpectrum
(ACL 2025 Findings)

Benchmarks

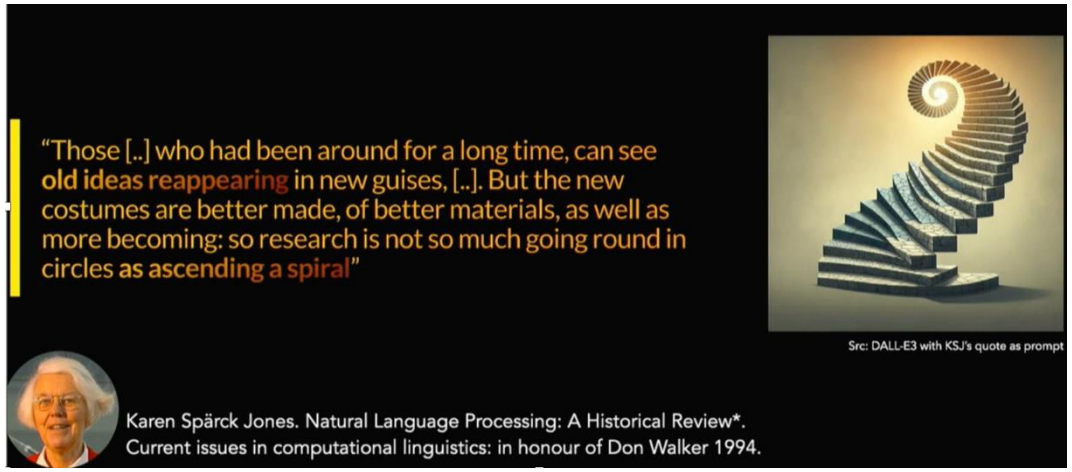
Summary


External Physical Tools
RAG: Uni-Retriever, SAFARI
Actions: JOTR, RL-Bert
Multi-Modal: M3SUM
…
Agent
Internal Cognitive Tools
Personalization: Chain-of-Cues
Proactivity: Pro-CoT
Role-playing: REGA
Strategies: TPE
……
Self-aware
Monitoring
MingConf
UAlign
AutoPSV
Control
Self-DC
SpARE
SMART
OTC-PO
……
Evaluation
AppBench
DialogTool
ToolSpectrum
…

Thesis Statement

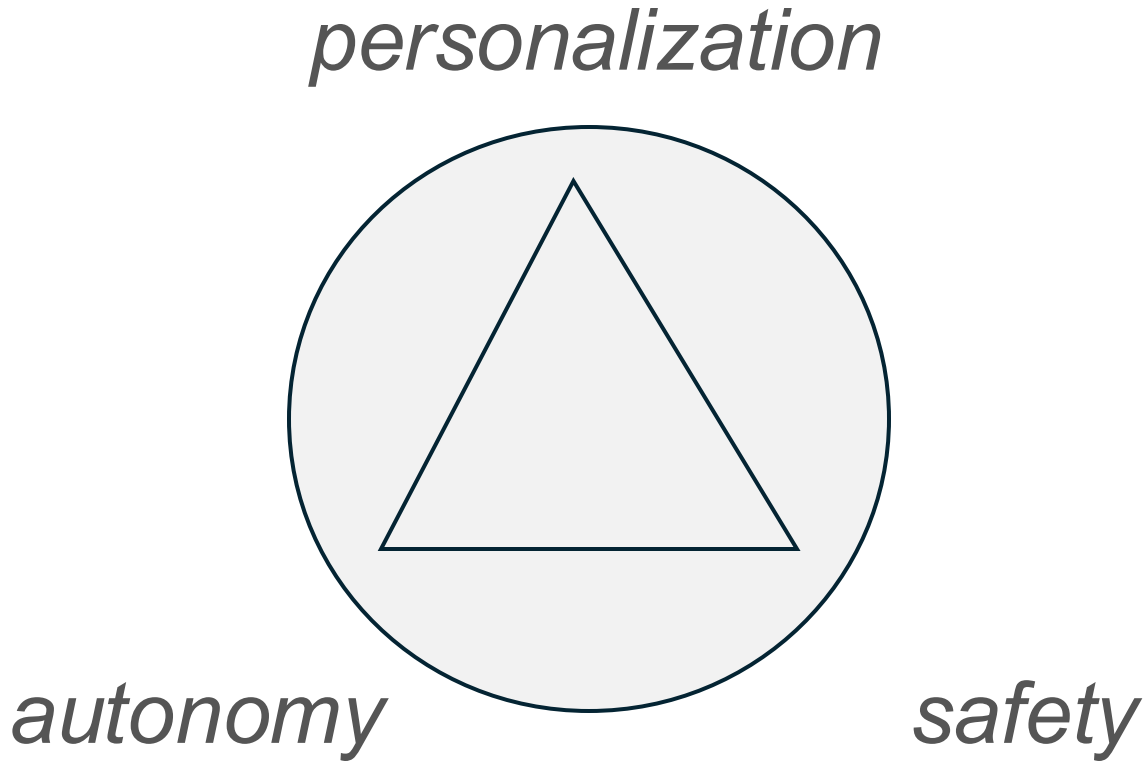
This thesis explores the design of self-aware agents capable of coordinating both internal cognitive tools and external physical tools. It introduces a series of frameworks and methodologies to examine the interaction between these tools, the mechanisms for monitoring and control, and the development of optimal tool-use behaviors.

Future Direction



Time

- Task-oriented Dialogue System (DS)
  - NLU / DST / DPL / NLG
- Open-domain Dialogue System (DS)
  - RAG / Search
- Unified DS / LLM-based DS
- Tool Learning
- Language Agent (*new* Task-oriented DS)
  - Such as Tau-bench / AppBench
- ……



Impossible Triangle in Agent

Publication List (*Co*-first author)

19 (Co-)first author papers

1. **Hongru WANG**, Wai-Chung Kwan, Min Li, Zimo Zhou and Kam-Fai Wong, "KddRES: A Multi-level Knowledge-driven Dialogue .. Towards Customized Dialogue System", Computer Speech & Language
2. Wai-Chung Kwan*, **Hongru WANG***, Huimin Wang and Kam-Fai Wong, "A Survey on Recent Advances and Challenges in Reinforcement Learning Methods", Machine Intelligence Research
3. **Hongru Wang**, Lingzhi Wang, Yiming Du, Liang Chen, Jingyan Zhou, Yufei Wang, Kam-Fai Wong, "A Survey of the Evolution of Language Model-based Dialogue Systems", ACM Computing Survey (Major Revision)
4. **Hongru WANG**, Deng Cai, Wanjun Zhong, Shijue Huang, Jeff Z. Pan, Zeming Liu, Kam-Fai Wong, "Self-Reasoning Language Models: Unfold Hidden Reasoning Chains with Few Reasoning Catalyst", ACL 2025 Findings
5. **Hongru WANG**, Wenyu Huang, et al., Zeming Liu, Jeff Z. Pan, Kam-Fai Wong, "Rethinking Stateful Tool Use in Multi-Turn Dialogues: Benchmarks and Challenges", ACL 2025 Findings
6. Zihao Cheng*, **Hongru WANG***, Zeming Liu, Yuhang Guo, et al., Yunhong Wang, Haifeng Wang, "ToolSpectrum: Towards Personalized Tool Utilization for Large Language Models, ACL 2025 Findings
7. Boyang XUE*, **Hongru WANG***, Rui Wang, et al., Wenxuan Zhang, Kam-Fai Wong, "MlingConf: A Comprehensive Study of Multilingual Confidence Estimation on LLMs", ACL 2025 Findings
8. Yuheng Lu*, Qian Yu*, **Hongru WANG***, Zeming Liu, Wei Su, et al., Yunhong Wang, Haifeng Wang, "TransBench: Breaking Barriers for Transferable GUI Agents in Dynamic Digital Environments", ACL 2025 Findings
9. **Hongru WANG**, Boyang Xue, Baohang Zhou, et. al, Kam-Fai Wong, "Self-DC: When to Reason and When to Act? Self Divide-and-Conquer for Compositional Unknown Questions", NAACL 2025 Oral

Publication List (*Co*-first author)

19 (Co-)first author papers

10. **Hongru WANG**, Rui Wang, Boyang XUE, et. al, Jeff Z. Pan, Kam-Fai Wong, "AppBench: Planning of Multiple APIs from Various APPs for Complex User Instruction", EMNLP 2024
11. **Hongru WANG**, Huimin Wang, Lingzhi Wang, et. al, Kam-Fai Wong, "TPE: Towards Compositional Reasoning over Conceptual Tools with Multi-persona Collaboration", NLPCC 2024
12. Rui Wang*, **Hongru WANG***, Fei Mi, Boyang XUE, Yi Chen, Kam-Fai Wong, Ruifeng Xu, "Enhancing Large Language Models Against Inductive Instructions with Dual-critique Prompting", NAACL 2024
13. **Hongru WANG**, Yujia Qin, Yankai Lin, Jeff Z. Pan, Kam-fai Wong, "Empowering Large Language Models: Tool Learning for Real-World Interaction", SIGIR 2024 Tutorial
14. **Hongru WANG**, Boyang Xue, Baohang Zhou, et. al, Kam-Fai Wong, "UniRetriever: Multi-task Candidates Selection for Various Context-Adaptive Conversational Retrieval", LREC-COLING 2024
15. **Hongru WANG**, Baohang Zhou, Zhengkun Zhang, Yiming Du, David Ho, Kam-Fai Wong, "M3Sum: A Novel Unsupervised Language-guided Video Summarization", ICASSP 2024
16. **Hongru WANG**, Minda Hu, Yang Deng, et. al, Irwin King, Kam-Fai Wong, "Large Language Models as Source Planner for Personalized Knowledge-grounded Dialogue", EMNLP 2023 Findings
17. **Hongru WANG**, Rui Wang, Fei Mi, et. al, Ruifeng Xu and Kam-Fai Wong, "Cue-CoT: Chain-of-thought Prompting for Responding to In-depth Dialogue Questions with LLMs", EMNLP 2023 Findings
18. **Hongru WANG**, Zezhong Wang, Wai-Chung Kwan, Kam-Fai Wong, "MCML: A Novel Memory-based Contrastive Meta-Learning Method for Few Shot Slot Tagging", IJCNLP-AACL 2023
19. **Hongru WANG**, Huimin Wang, Zezhong Wang and Kam-Fai Wong, "Integrating Pretrained Language Model for Dialogue Policy Learning", ICASSP 2022

Publication List (*Co*- author)

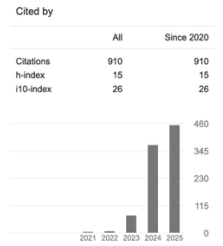
22 (Co-)author papers

1. Nan Hu, Jiaoyan Chen, Yike Wu, Guilin Qi, **Hongru WANG**, et al., Jeff Z. Pan, "Can LLMs Evaluate Complex Attribution in QA? Automatic Benchmarking using Knowledge Graphs", ACL 2025
2. Yan Yang, Yixia Li, **Hongru WANG**, Xuetao Wei, James Jianqiao Yu, Yun Chen, Guanhua Chen, "ImPart: Importance-Aware Delta-Sparsification for Improved Model Compression and Merging …", ACL 2025
3. Yu Zhao, et al., **Hongru WANG**, Xuanli He, Kam-Fai Wong, Pasquale Minervini, "Steering Knowledge Selection Behaviours in LLMs via SAE-Based Representation Engineering", NAACL 2025 Oral
4. Yan Yang, Zeguan Xiao, Xin Lu, **Hongru WANG**, et al., Guanhua Chen, Yun Chen, "SeqAR: Jailbreak LLMs with Sequential Auto-Generated Characters", NAACL 2025
5. Jianqiao Lu, Zhiyang Dou, **Hongru WANG**, et. al, Zhijiang Guo, "AutoPSV: Automated Process-Supervised Verifier", NeurIPS 2024
6. Rongwu Xu, Zehan Qi, Zhijiang Guo, Cunxiang Wang, **Hongru WANG**, Yue Zhang, Wei Xu, "Knowledge Conflicts for LLMs: A Survey", EMNLP 2024
7. Jingtao Cao, Zhang Zheng, **Hongru WANG**, Kam-Fai Wong, "VLEU: a Method for Automatic Evaluation for Generalizability of Text-to-Image Models", EMNLP 2024
8. Zezhong Wang, Fangkai Yang, Lu Wang, Pu Zhao, **Hongru WANG**, et. al, Kam-Fai Wong, "SELF-GUARD: Empower the LLM to Safeguard Itself", NAACL 2024
9. Rui Wang, Jianzhu Bao, Fei Mi, Yi Chen, **Hongru WANG**, et. al, Kam-Fai Wong, Ruifeng Xu, "Retrieval-free Knowledge Injection through Multi-Document Traversal for Dialogue Models", ACL 2023

……

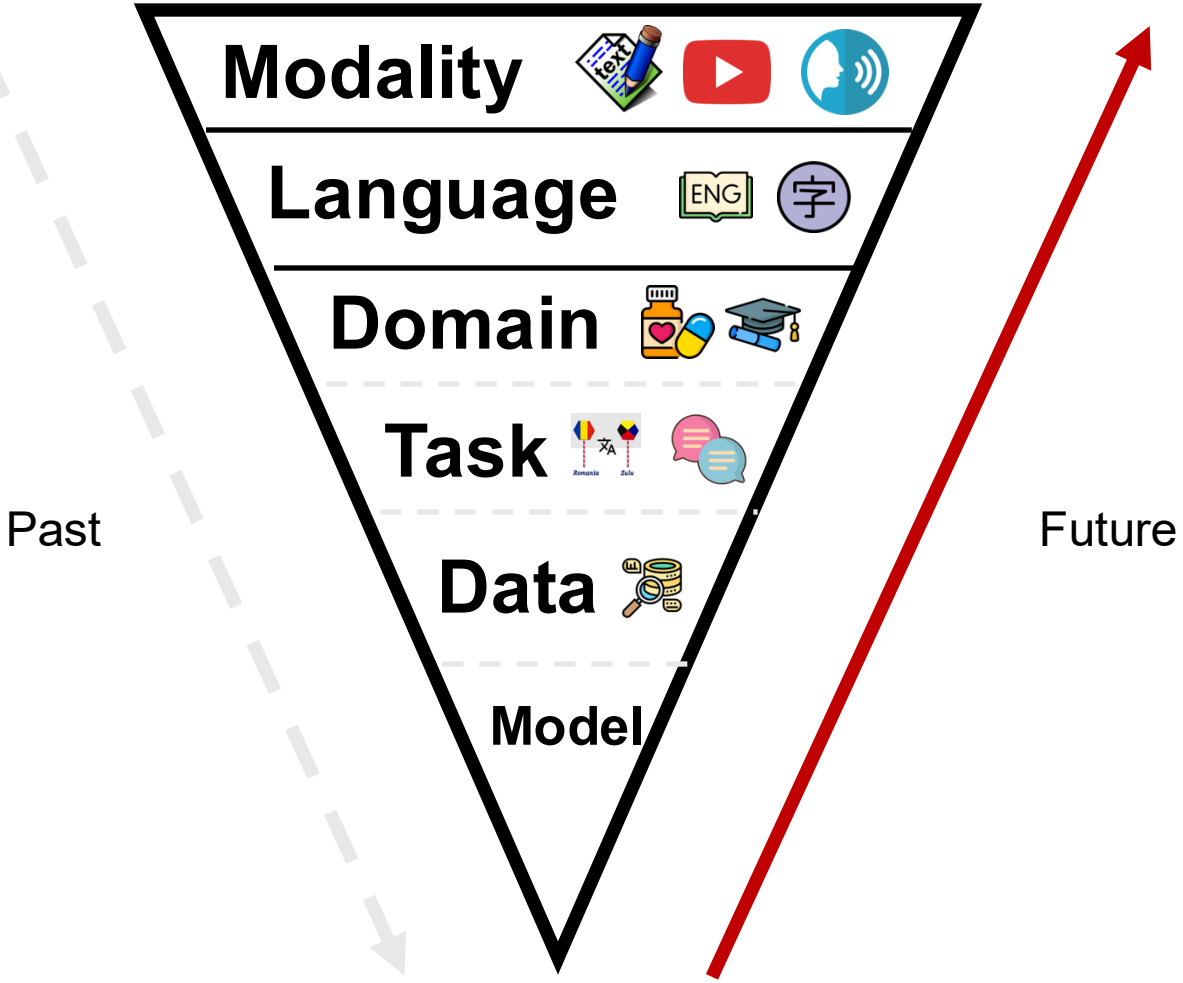
Others

Hongru WANG
Ph.D. Candidate
Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong



- **Google Scholar:** Citations: 900+, h10-index: 26
- **Best Paper Awards:** SIGHAN @ACL 2024, International Doctoral Forum
- **Community**: Area Chair@ NeurIPS 2025, Reviewer@ ARR, IJCAI, ……
- **Publications**: 40+ papers, including 19 (Co)First author papers @ACL, EMNLP, NAACL, COLING, Computer Speech & Language and 22 Co-author papers
- **Representative works:** Cue-CoT, SAFARI, Self-DC, AppBench, OTC, Theory of Agent
- **First Tool Learning Tutorial @ SIGIR 2024**
- **First Cantonese-based Task-oriented Dialogue System (KddRES)**
- **Champion on WWW2024 Online Safety Prize Challenge**
- **Funding and Grants:** TBF22ENG004 and OSCP2023-2024
- **Interns and Visiting**: BlenderLab@UIUC, EdinburghNLP, ByteDance-Seed
- **Co-founder and Committee Member**: NLP Academic Exchange Platform (NICE)
  - Homepage: https://nice-nlp.github.io/
  - Fans: 15w+; Talks: 60+; Invited Speakers: 150+; Views: 15w+
  - Committee: Senior Members: 8 professors; Members: 13 PhD students
- **Research Interests**: Dialogue System, Tool Learning, Language Agent

Others



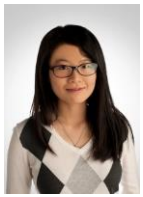
- Multi-modal: VLEU (EMNLP 2024), OSPC (WWW 2024), ...
- Multi-lingual: KddRES (CSL Journal), MlingConf (ACL 2025 Findings), Self-Denote, ...
- Explainable AI: ReadPrompt (EMNLP 2023), K-Dial (EMNLP 2023), ...
- .....

Acknowledgement

Prof. Kam-Fai Wong (CUHK)

Prof. Heng Ji (UIUC)

Prof. Jeff Z. Pan (University of Edinburgh)



Prof. Mendi Wang (Princeton University)

.....

Big thanks to my supervisor and collaborators!

Dr. Huimin Wang

Dr. Lingzhi Wang

Dr. Wai-chung Kwan

Dr. Xiusi Chen

Zezhong Wang

Jingtao Cao

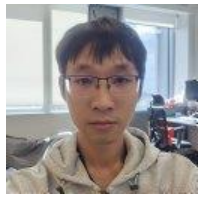
Yiming Du

Liang Chen

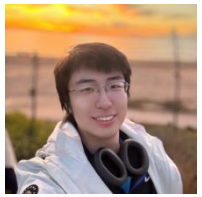
Boyang Xue

Rui Wang

Minda Hu



Yu Zhao



Cheng Qian

Jiahao Qiu

.....

Thank You!

(Q & A)