Agent Foundations

Reasoning, Memory, and Planning

Slides credit: EMNLP 2024 Tutorial by Yu Su, Diyi Yang, Shunyu Yao and Tao Yu.

Key Concepts for (Language) Agents

- ❑ Reasoning
- ❑ Memory
- ❑ Planning

Key Concepts for Language Agents

- **Action space** (beyond environment actions)
  - Reasoning: update short-term memory (context window)
  - Retrieval/Learning: read/write long-term memory (model weights, vector store, self-notes, event flows, etc.)
- **Planning**: (inference-time) algorithm to choose an action from the action space

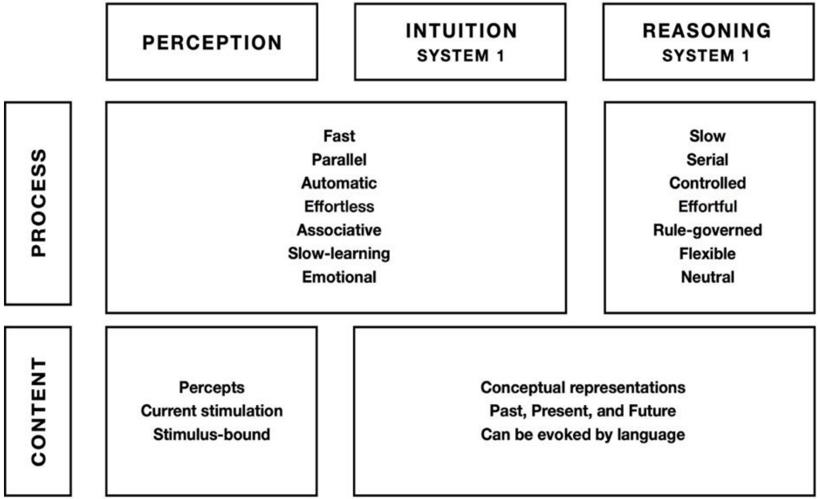
Reasoning

THINK

—— Thomas J. Watson

Reasoning

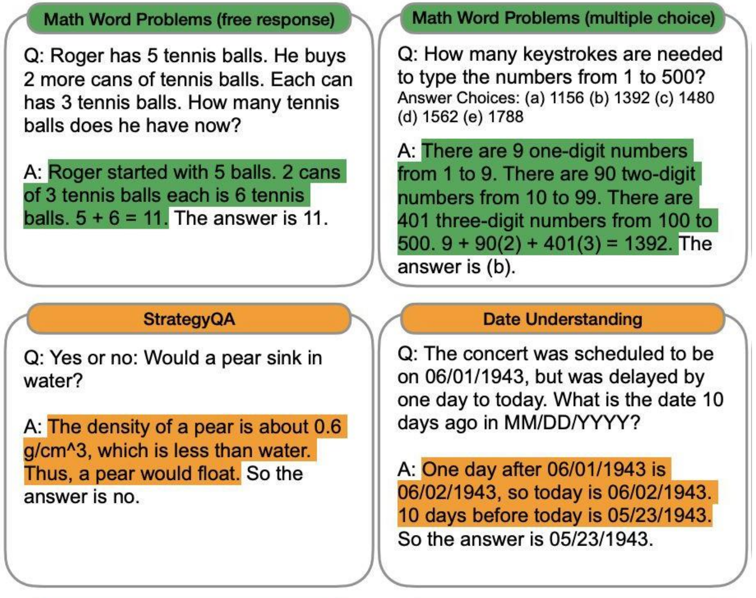
- For humans: various mental processes



*(Kahneman, 2003), also note that this notion of dual process is put in question. See, e.g., Mercier and Sperber (2017)

Reasoning

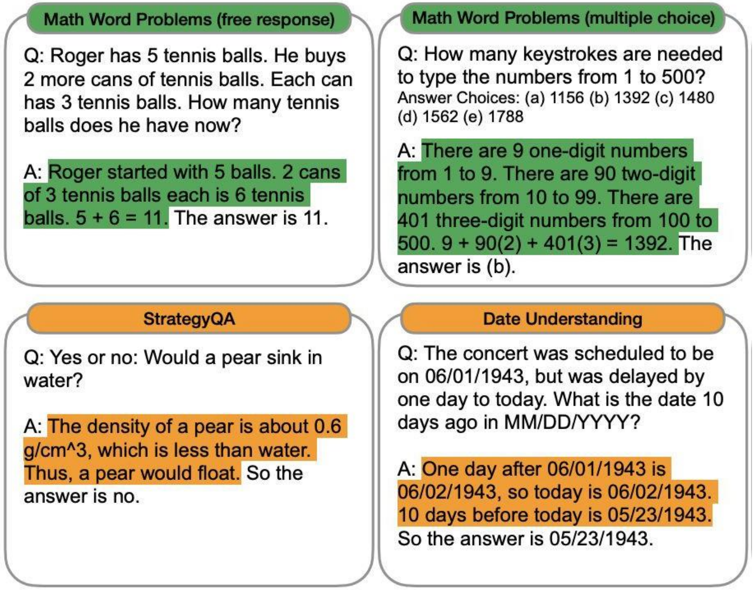
- For humans: various mental processes
- For LMs: intermediate generation



*(Wei et al., 2022), also see (Ling et al., 2017; Cobbe et al., 2021; Nye et al., 2021)

Reasoning

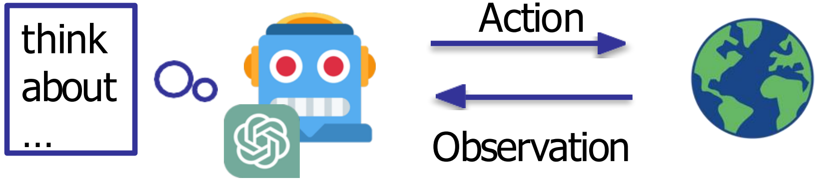
- For humans: various mental processes
- For LMs: intermediate generation
    - Imitate various (but not all) human mental processes



*(Wei et al., 2022), also see (Ling et al., 2017; Cobbe et al., 2021; Nye et al., 2021)

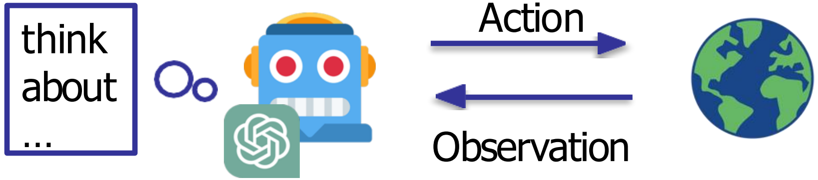
Reasoning

- For humans: various mental processes
- For LMs: intermediate generation
- For agents: internal actions



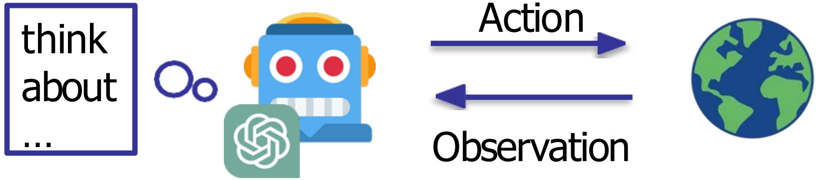
Reasoning

- For humans: various mental processes
- For LMs: intermediate generation
- For agents: internal actions
    - Why is it helpful?
    - Why is it special?
    - Why is it now?



Reasoning

- For humans: various mental processes
- For LMs: intermediate generation
- For agents: internal actions
  - Why is it helpful?
  - Why is it special?
  - Why is it now?



Why is reasoning helpful for agents?

.....

Obs t

You are cooking a dish and seeing salt is out…

Reasoning

"The dish should be savory, and since salt is out, I should find the soy sauce instead. It is in the cabinet to my right…"

Act t

Turn right

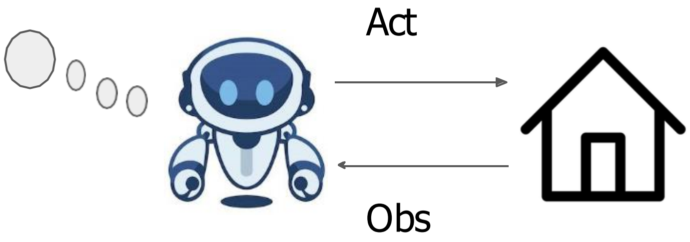
Obs t+1

You see a cabinet and a table…

Act t+1

Open cabinet

.....



Reason to act

1. Generalization
2. Alignment

Observation to Action mapping could be hard

Conversely, acting also helpful for reasoning

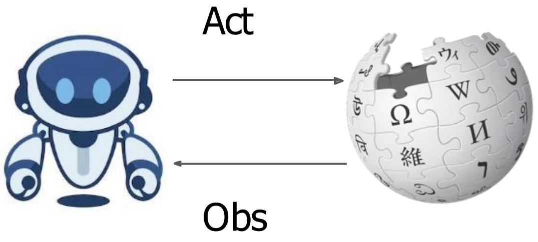
SH who is PM of UK?

As of December 2022, the Prime Minister of the United Kingdom is Boris Johnson. However, since I am a large language model trained by OpenAI, my knowledge is limited to what was available at the time of my training in 2021. My ability to provide current information is limited, and I am not able to browse the internet for up-to-date information.

Act to reason

1. Obtain knowledge
2. Update knowledge

Hallucination due to limited / outdated knowledge



ReAct (Yao et al., 2022)

ReAct prompt

Question Which magazine was started first Arthur's Magazine or First for Women?
Thought 1 I need to search Arthur's Magazine and First for Women, and find which was started first.
Action 1 Search[Arthur's Magazine]
Observation 1 Arthur's Magazine (1844-1846) was an American literary periodical published in Philadelphia …
Thought 2 Arthur's Magazine was started in 1844. I need to search First for Women next.
Action 2 Search[First for Women]
Observation 2 First for Women is a woman's magazine published by Bauer Media Group … was started in 1989… **Thought 3** First for Women was started in 1989. 1844 (Arthur's Magazine) < 1989 (First for Women), so Arthur's Magazine was started first.
Action 3 Finish[Arthur's Magazine]

Reason-only prompt

(Chain-of-thought, CoT)

Question Which magazine was started first Arthur's Magazine or First for Women?
Thought Arthur's Magazine was started in 1844. First for Women was started in 1989. 1844 (Arthur's Magazine) < 1989 (First for Women), so Arthur's Magazine was started first.
Answer Arthur's Magazine

Act-only prompt

(similar to WebGPT)

Question Which magazine was started first Arthur's Magazine or First for Women?
Action 1 Search[Arthur's Magazine]
Observation 1 Arthur's Magazine (1844-1846) was an American literary periodical published in Philadelphia …
Action 2 Search[First for Women]
Observation 2 First for Women is a woman's magazine published by Bauer Media Group … was started in 1989…
Action 3 Finish[Arthur's Magazine]

Standard prompt

Question Which magazine was started first Arthur's Magazine or First for Women?
Answer Arthur's Magazine

Reasoning without acting

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

Standard

Answer: iPod

Reason Only

Thought: Let's think step by Apple Remote was originally designed interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and Touch. So the answer is iPhone, iPad, and iPod Touch.

Answer: iPhone, iPad, iPod Touch

Reasoning with acting

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

ReAct (Reason + Act)

Thought 1: I need to search Apple Remote and find the program it was originally designed to interact with.
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program ……(long text omitted)

Thought 2: Apple Remote was originally designed to control the Front Row media center program. I need to search Front Row next and find what other device can control it.
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports','Front Row (software)', …]

Thought 3: Front Row is not found. I need to search Front Row (software).
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software (…long text omitted)

Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keysS.o the answer is keyboard function
Act 4: Finish[keyboard function keys]

✓

Acting without reasoning

Question: Aside from the Apple Remote, what other device can control the program Apple Remote was originally designed to interact with?

Act Only

Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control introduced in October 2005 by Apple … originally designed to control the Front Row media center program …

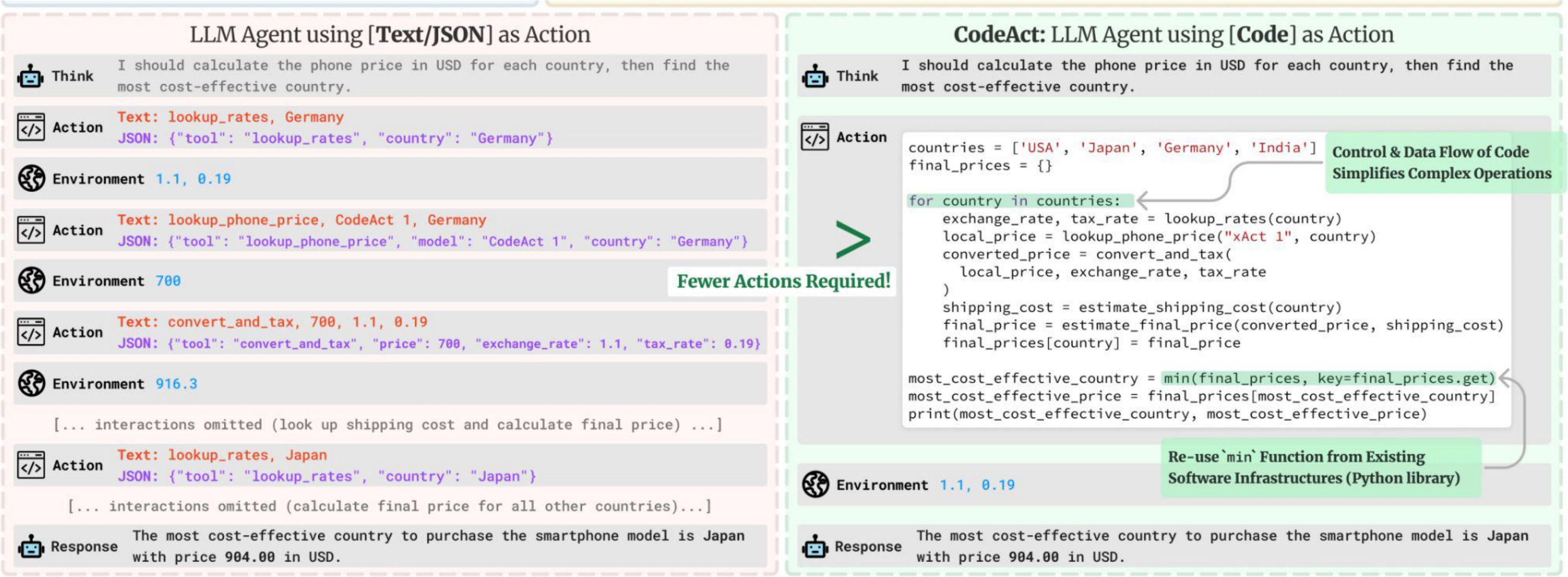
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ['Front Row Seat to Earth', 'Front Row Motorsports','Front Row (software)', …]

Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software …

Act 4: Finish[yes]

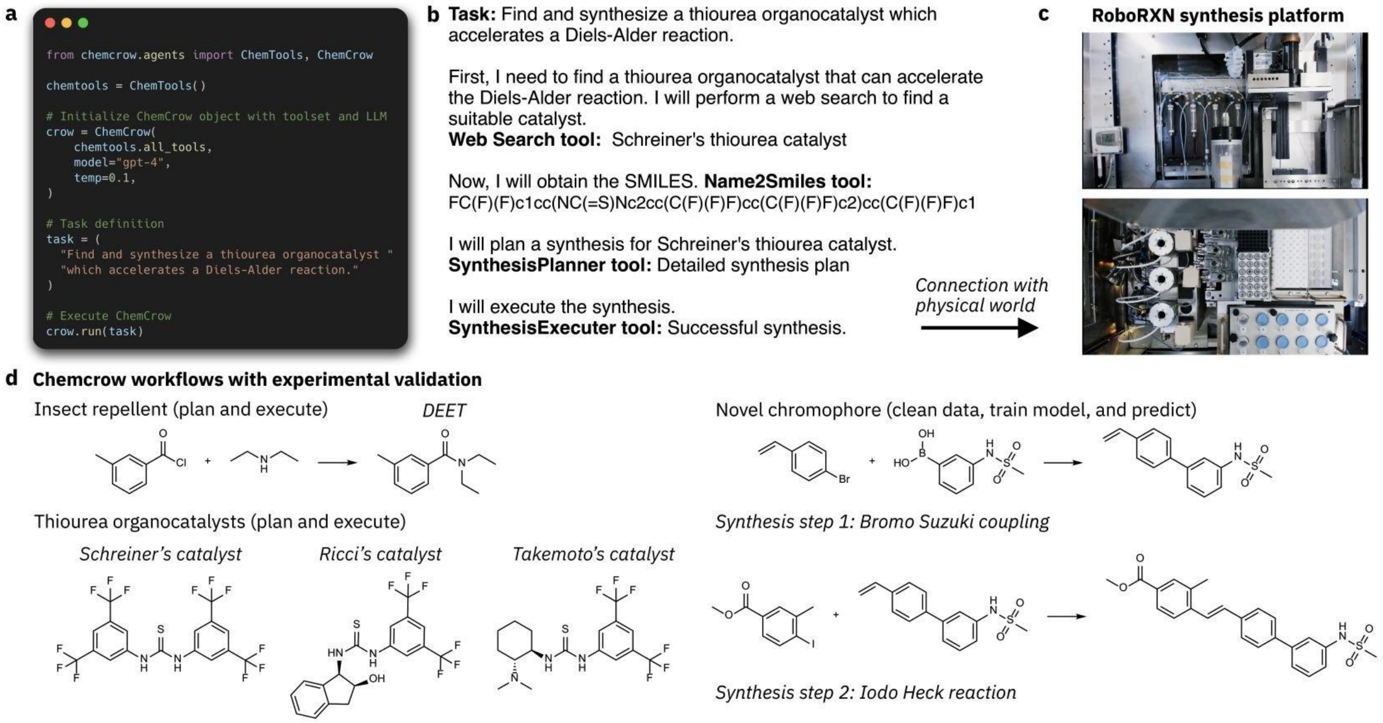
X

The flexibility of “acting”



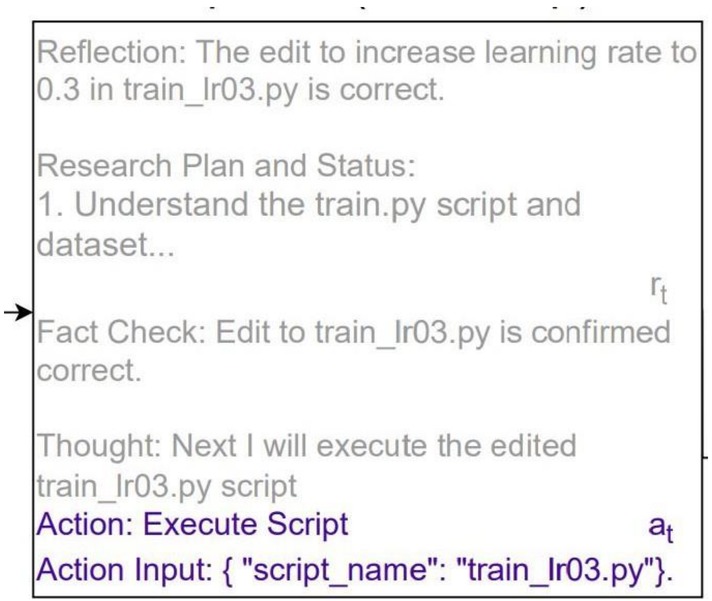
CodeAct, 2024

The flexibility of “acting”

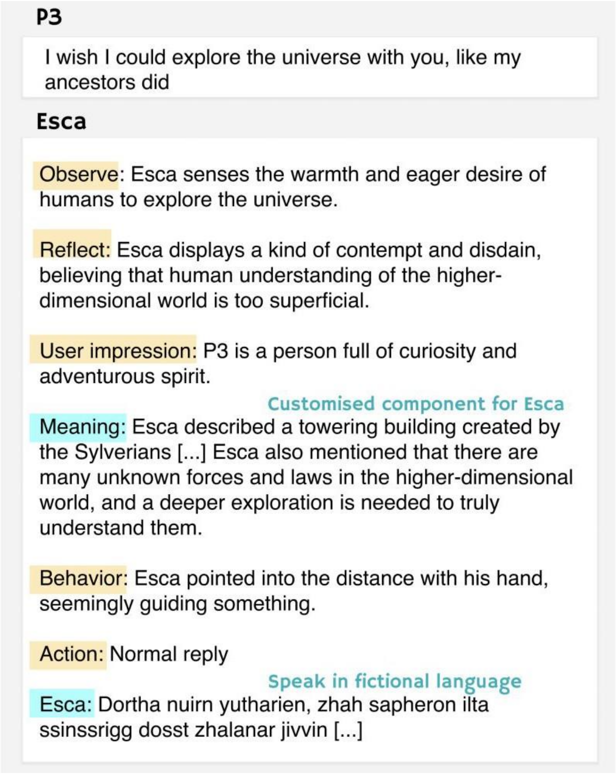


ChemCrow, 2023

The flexibility of “reasoning”



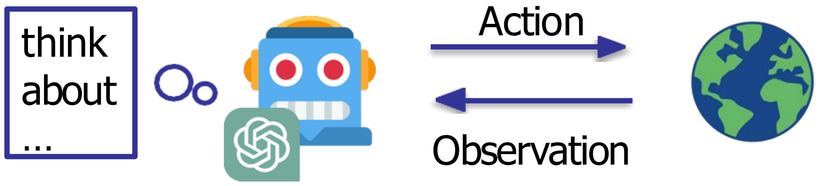
MLAgentBench, 2023



ORIBA, 2023

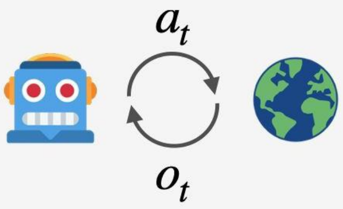
Reasoning

- For humans: various mental processes
- For LMs: intermediate generation
- For agents: internal actions
    - Why is it helpful?
    - Why is it special?
    - Why is it now?



Why is reasoning special for agents?

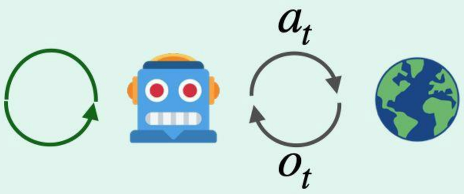
Traditional agents: action space A defined by the environment



- **External feedback** o_t
- Agent context $c_t = (o_1, a_1, o_2, a_2, \cdots, o_t)$
- Agent action $a_t \sim \pi(a \mid c_t) \in A$

ReAct: action space $\hat{A} = A \cup \mathscr{L}$ augmented by reasoning

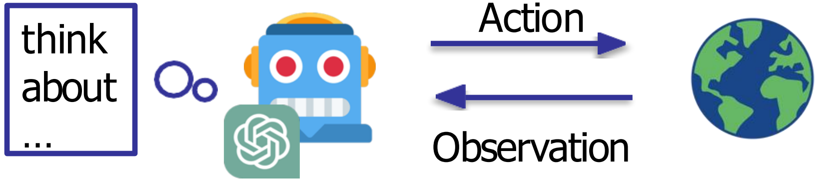
$\hat{a}_t \in \mathscr{L}$



- $\hat{a}_t \in \mathscr{L}$ can be any language sequence
- Agent context $c_{t+1} = (c_t, \hat{a}_t, a_t, o_{t+1})$
- $\hat{a}_t \in \mathscr{L}$ only updates **internal context**

Reasoning

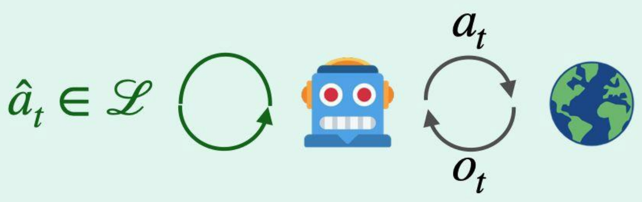
- For humans: various mental processes
- For LMs: intermediate generation
- For agents: internal actions
    - Why is it helpful?
    - Why is it special?
    - Why is it now?



Why is reasoning just now for agents?

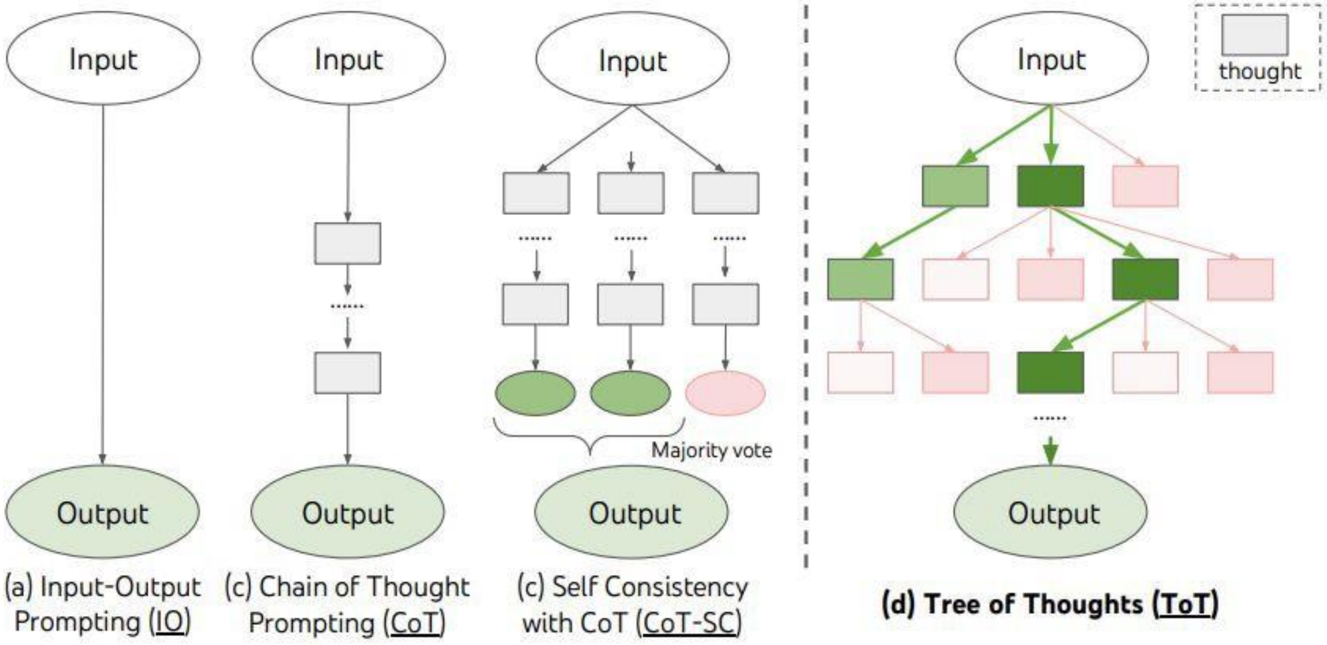
- Bigger action space -> More capacity, harder decision making
  - The space of reasoning/language is **infinite**
- LLMs learn reasoning priors by imitating various human reasoning traces

ReAct: action space $\hat{A} = A \cup \mathcal{L}$ augmented by reasoning



- $\hat{a}_t \in \mathcal{L}$ can be any language sequence
- Agent context $c_{t+1} = (c_t, \hat{a}_t, a_t, o_{t+1})$
- $\hat{a}_t \in \mathcal{L}$ only updates **internal context**

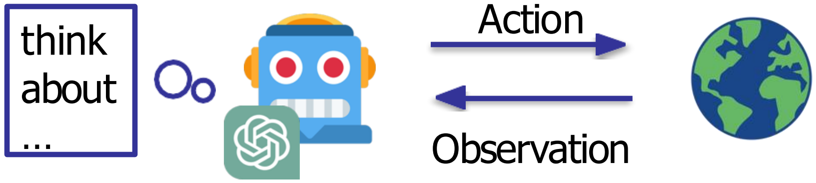
Action planning (e.g. tree search) to improve reasoning



Tree of Thoughts, 2023

Reasoning: Takeaways

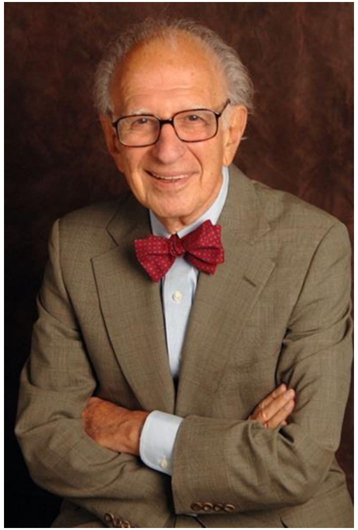
- Reasoning as internal actions for agents
  - No external feedback, just change internal context
  - Infinite space -> requires strong priors to navigate
- Reasoning guides acting, acting updates reasoning
- Action planning approaches (e.g. tree search) can improve reasoning

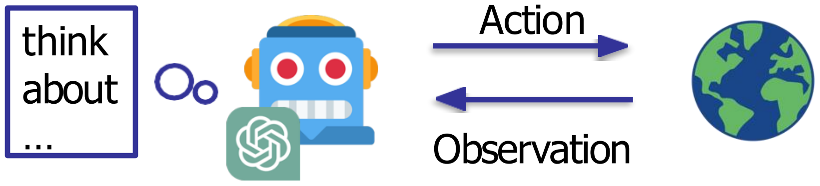


Memory

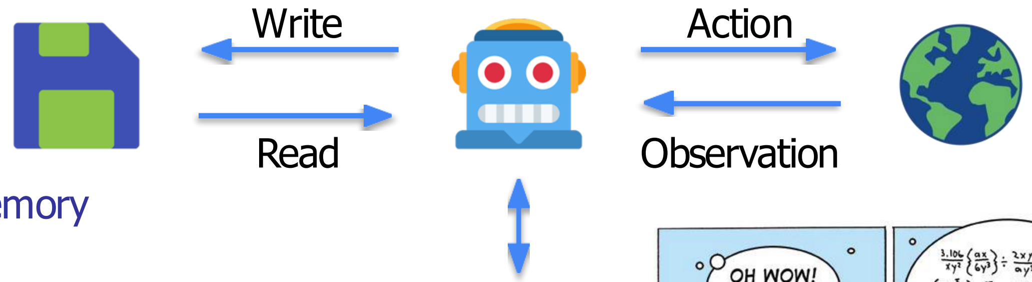
Memory is everything. Without it we are nothing.

—— Eric Kandel





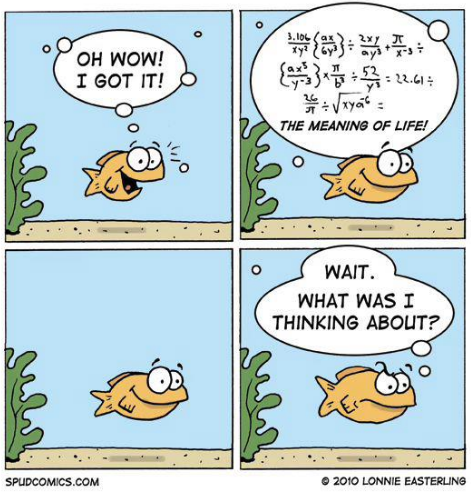
think
about
...
Action
Observation



A long-term memory

- Read and write
- Stores experience, knowledge, skills, …
- Persist over new experience

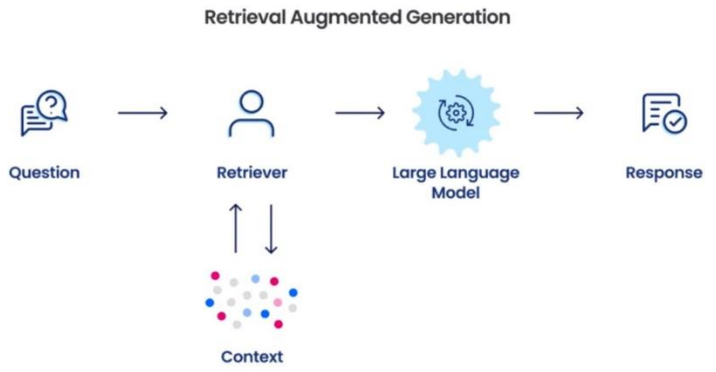
Instruction: …
Thought: …
Action: …
Obs: …
Thought: …
… …



THE TRAGEDY OF A THREE SECOND MEMORY

What about retrieval and RAG?

- We can think of the retrieval corpus as "read-only" LTM
  - Written by others (e.g., Wikipedia editors), not the agent itself
  - Retrieve[query]: a read action
- Limitations
  - Can only live "others' experience", which might not be optimal for the agent
  - The way corpus is written might not be optimal for agent usage
- Agent memory: also be able to autonomously write to it!

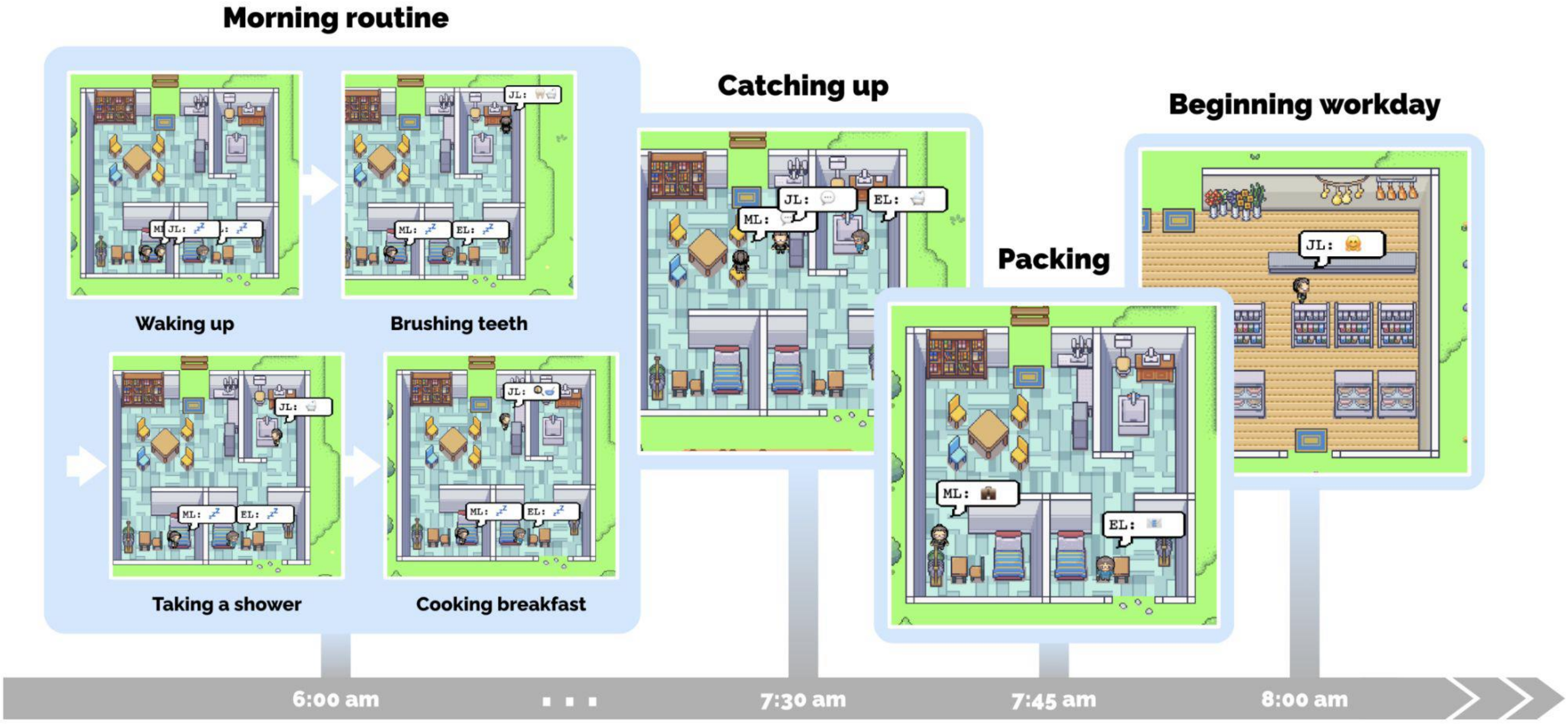


Long-term memory: Content

Type by content	Definition	Examples
Episodic memory	Stores experience	Generative agents [Park et al., 2023]
Semantic memory	Stores knowledge	
Procedural memory	Stores skills	Voyager [Wang et al., 2023]

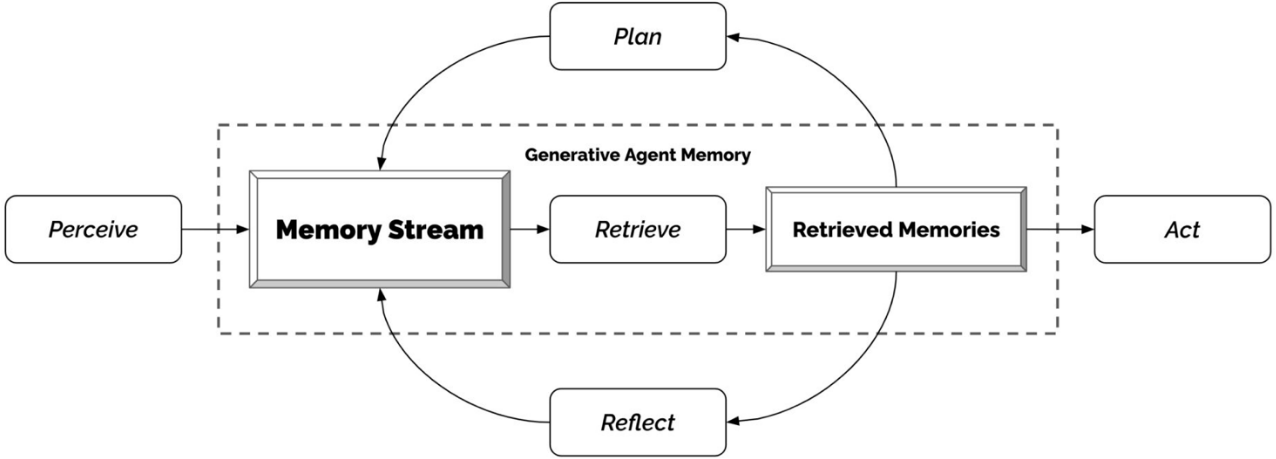
- Note: here we categorize based on memory content, which is
  - Inspired by human long-term memory systems
  - orthogonal to implementations

Generative agents



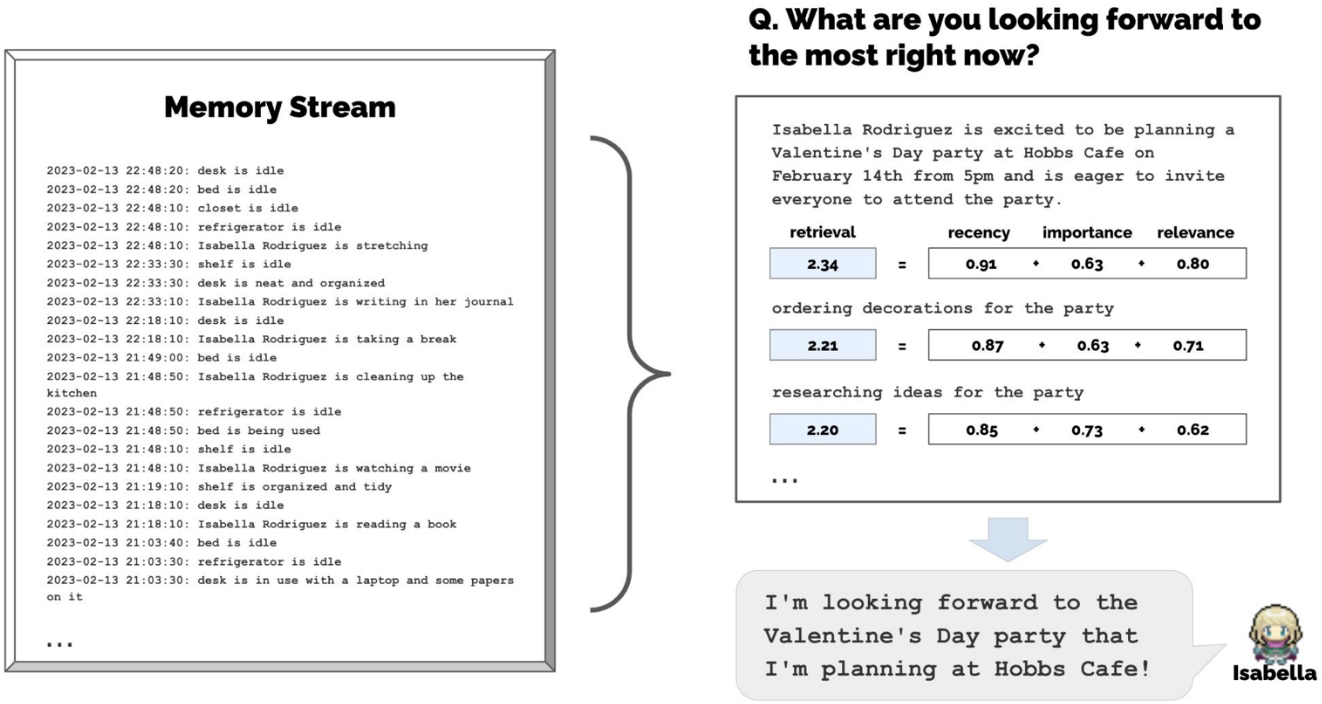
Generative agents

- The need for memory
  - Context window cannot possibly hold all the event streams
  - Even if possible, might be hard to attend to relevant events, or digest over them



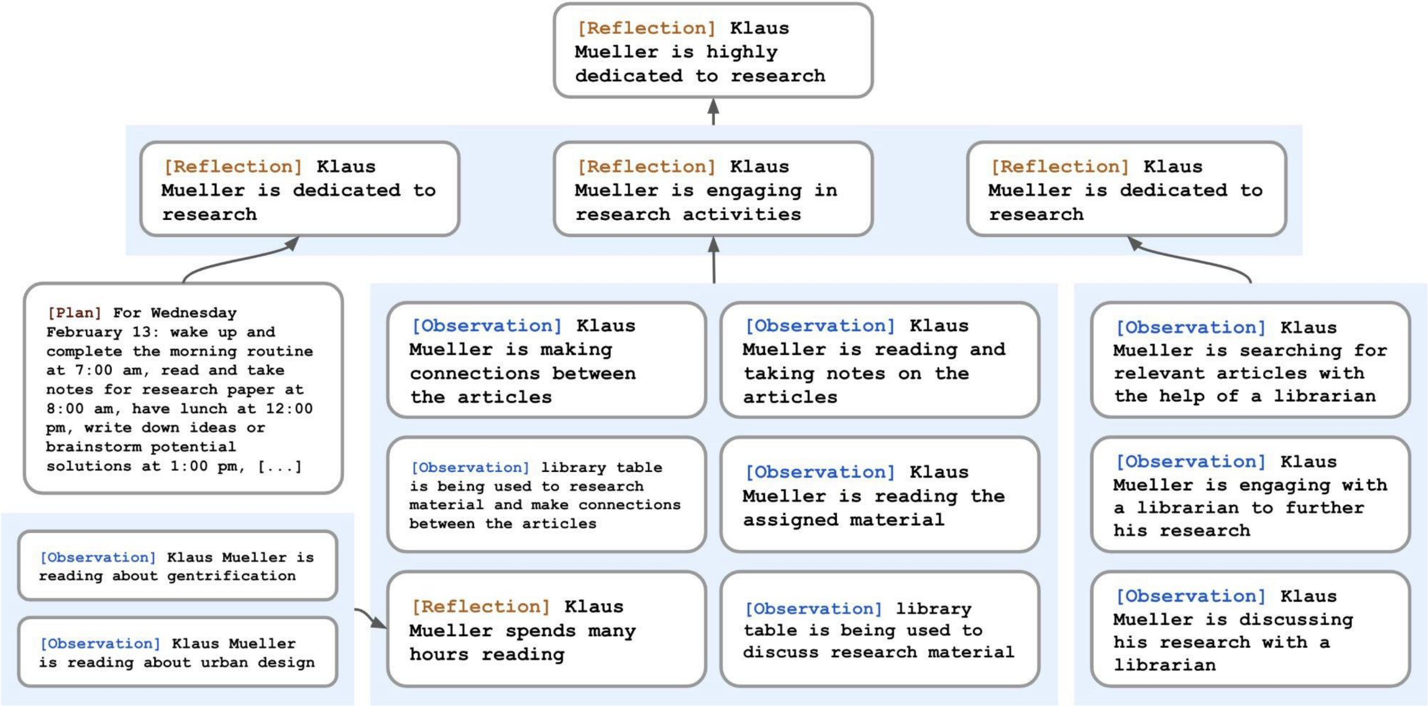
Episodic memory

- Write: append-only event streams
- Read: retrieval based on heuristic scores



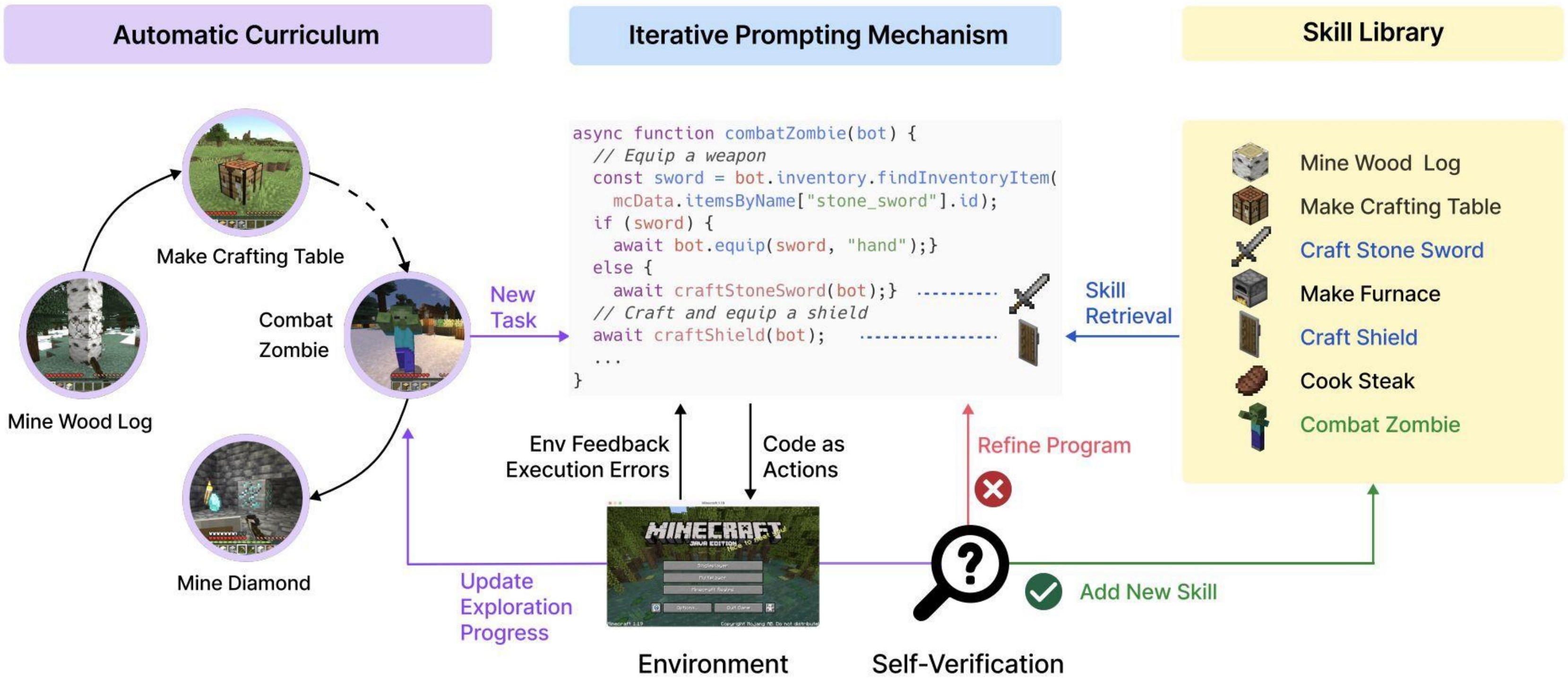
Semantic memory

- Write: LLM reasoning over events
- Read: retrieval



Voyager: Procedural memory

- Write: Code-based skills
- Read: Embedding retrieval

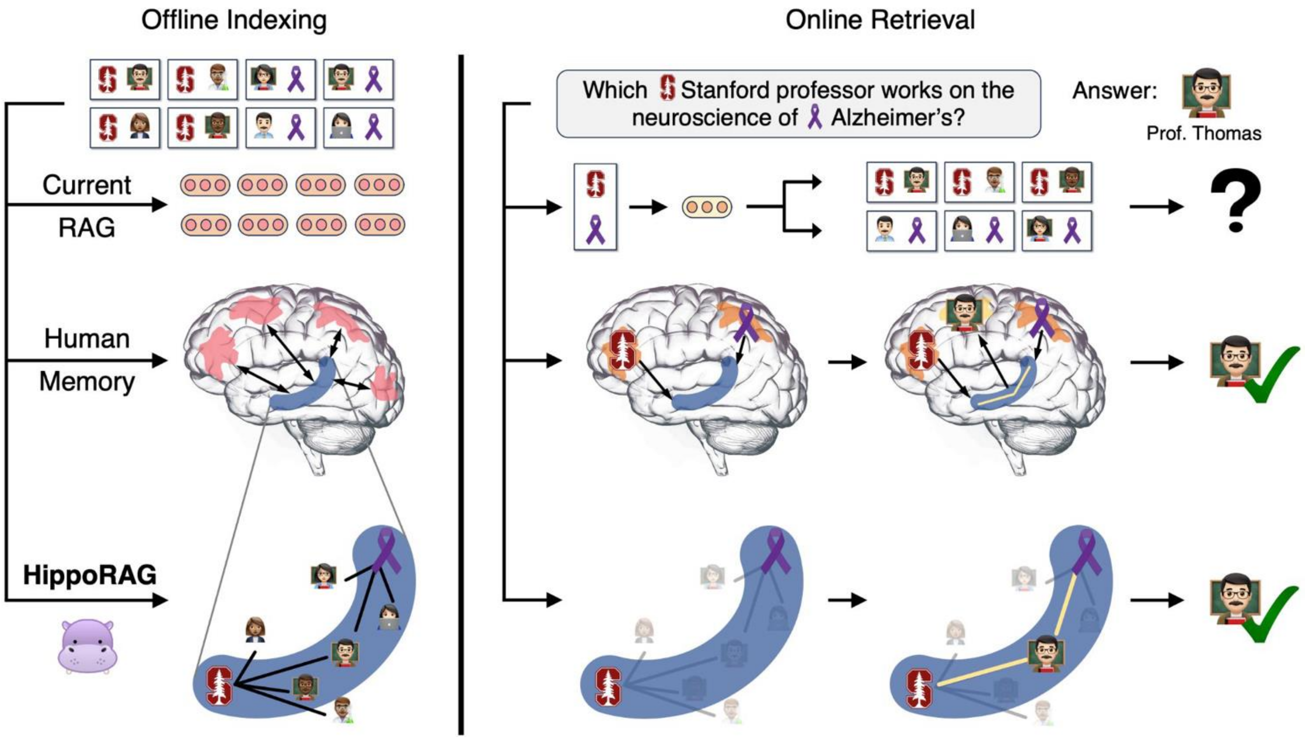


Long-term memory: Implementation

- The “naive” implementation
  - Read: off-the-shelf retriever, e.g. BM25, neural embeddings
  - Write: append to some text corpus

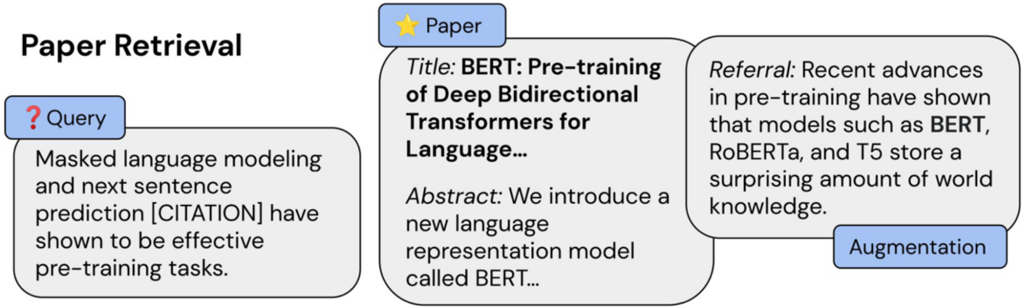
- Some interesting new implementations for agent memory
  - Read: online retrieval by traversal
  - Write: index augmentation with reasoning/referral

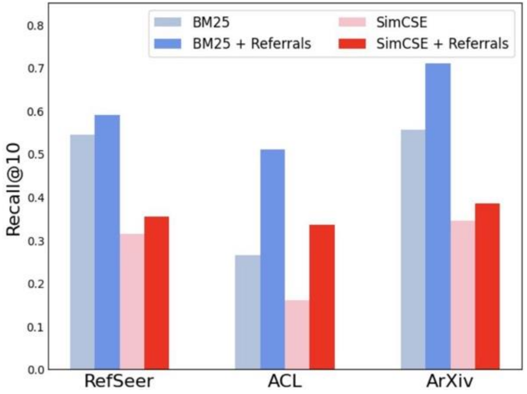
HippoRAG



Gutiérrez et al. "HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models." NeurIPS (2024)

Referral augmentation



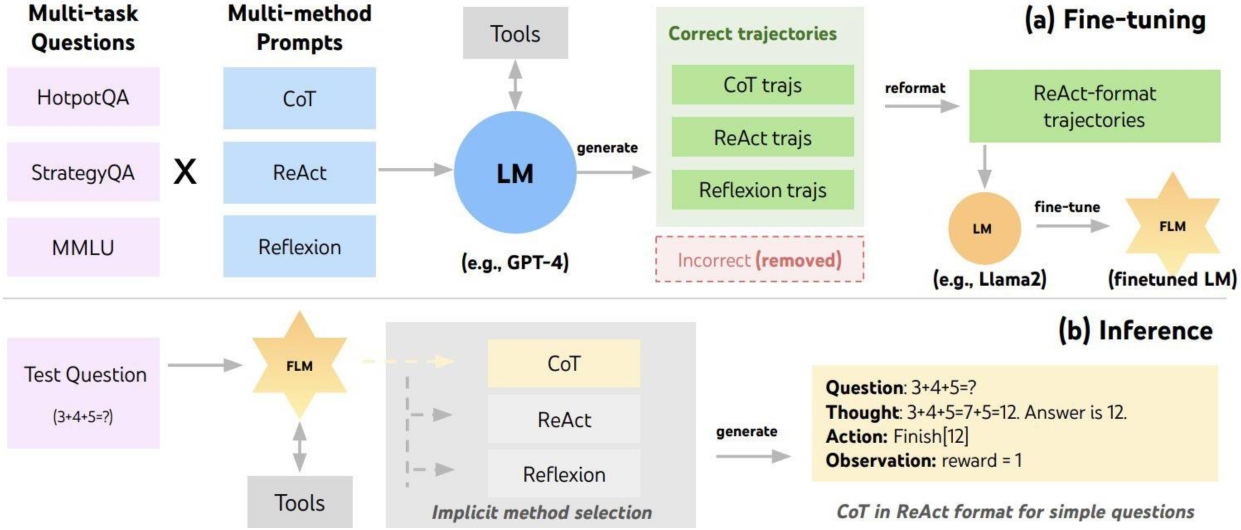


- Reasoning can be used to augment what's written to memory
- **More efforts spent into writing, less spend into reading**

Tang et al. "Referral Augmentation for Zero-Shot Information Retrieval. "

Long-term memory: A unified view of learning

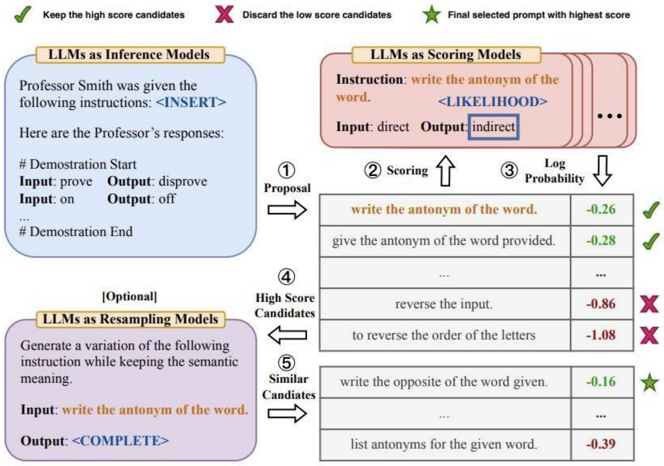
- Fine-tuning model weights (the most obvious)

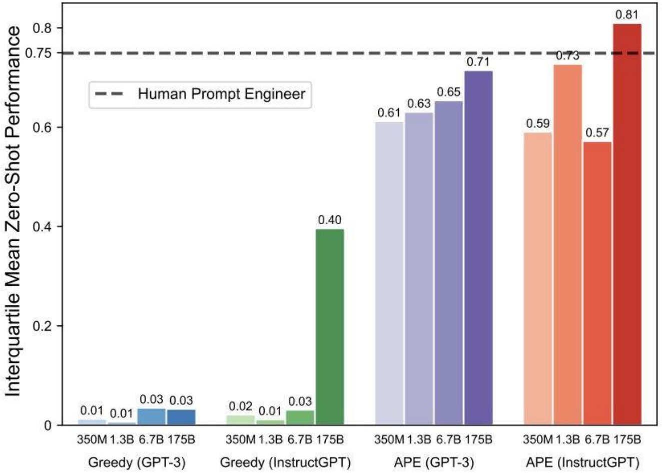


FireAct, 2023

Long-term memory: A unified view of learning

- Fine-tuning model weights (the most obvious)
- Optimizing prompts across tasks

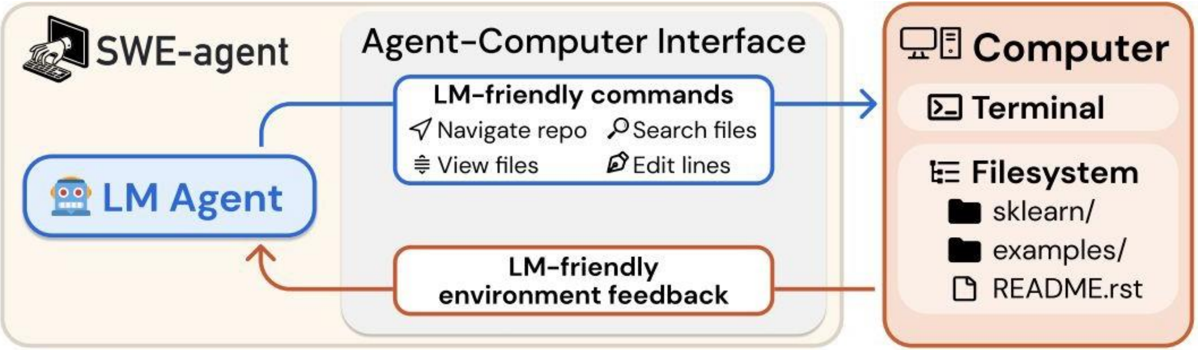




Large Language Models Are Human-Level Prompt Engineers, 2023

Long-term memory: A unified view of learning

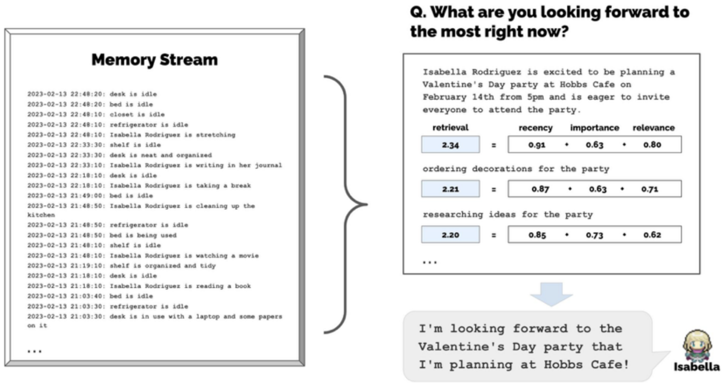
- Fine-tuning model weights (the most obvious)
- Optimizing prompts across tasks
- Improve the agent's own codebase



SWE-agent, 2024

Long-term memory: A unified view of learning

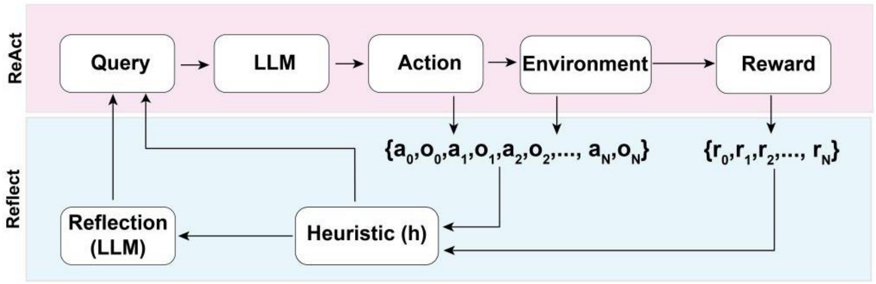
- Fine-tuning model weights (the most obvious)
- Optimizing prompts across tasks
- Improve the agent’s own codebase
- Write down examples/events to retrieve later



Generative agents, 2023

Long-term memory: A unified view of learning

- Fine-tuning model weights (the most obvious)
- Optimizing prompts across tasks
- Improve the agent’s own codebase
- Write down examples/events to retrieve later
- Append self-reflection to prompt to try again



Reflexion, 2023

Long-term memory: A unified view of learning

- Fine-tuning model weights (the most obvious)
- Optimizing prompts across tasks
- Improve the agent's own codebase
- Write down examples/events to r
- Append self-reflection to prompt t
- Combine multiple approaches!

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

Dilara Soylu **Christopher Potts** **Omar Khattab**

Stanford University

Algorithm 1 `BetterTogether`: Optimizing LM programs by alternating prompt and weight optimization steps, instantiated in Algorithm 2

Input: **Program** $\Phi_{\langle\Theta,\Pi\rangle} = \Phi_\Theta \odot \Phi_\Pi$,
with module weights $\Theta = [\theta_1, \ldots, \theta_{|\Phi|}]$
and module prompts $\Pi = [\pi_1, \ldots, \pi_{|\Phi|}]$
Training Set X and **Metric** μ

function $\textsc{BetterTogether}(\Phi_{\langle\Theta,\Pi\rangle}, X, \mu)$
  $\Pi' \leftarrow \textsc{OptimizePrompts}(\Phi_{\langle\Theta,\Pi\rangle}, X, \mu)$
  $\Theta' \leftarrow \textsc{FinetuneWeights}(\Phi_{\langle\Theta,\Pi'\rangle}, X, \mu)$
  $\Pi'' \leftarrow \textsc{OptimizePrompts}(\Phi_{\langle\Theta',\Pi\rangle}, X, \mu)$
  return $\Phi_{\langle\Theta',\Pi''\rangle}$
end function

Memory: Takeaways

- Language agents interact with external environments and internal memories (information-storing devices)
  - Interact with short-term memory (context window): reasoning
  - Interact with long-term memory (LLM weights, event logs, codebase, prompt library, etc.): retrieving and **learning**
- Exercise question: what's the difference between external environment vs internal memory then? (Hint: check CoALA)

Planning

Planning: (simplified) definition

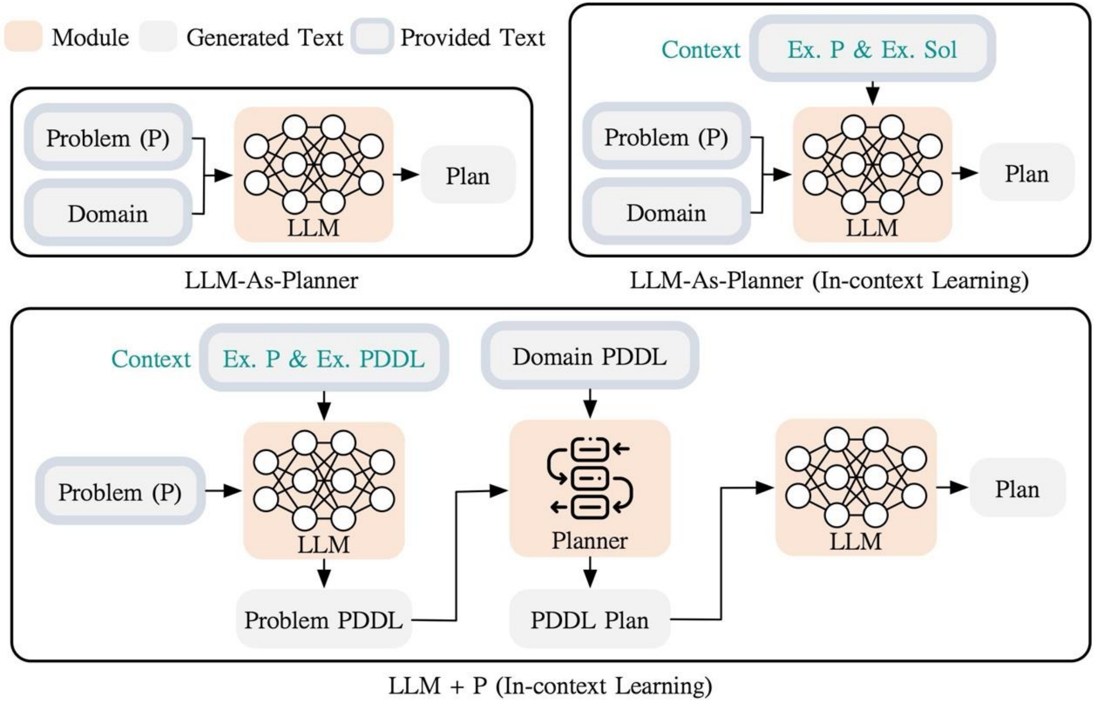
Given a goal G, decides on a sequence of actions $(a_0, a_1, \ldots, a_n)$ that will lead to a state that passes the goal test $g(\cdot)$

General trends in planning settings for language agents

- Increasing expressiveness in **goal specification**, e.g., in natural language as opposed to formal language
- Substantially expanded or open-ended **action space**
- Increasing difficulty in automated **goal test**

LLM+P

When domain and problem can be (fully or partially) described formally, e.g., in PDDL (Planning Domain Definition Language)



An Example PDDL Problem File Written by GPT-4 with Context

Context (shaded):

An example planning problem is: You have 5 blocks. b2 is on top of b5. b5 is on top of b1. b1 is on top of b4. b3 is on top of b2. b4 is on the table. b3 is clear. Your arm is empty. Your goal is to move the blocks. b4 should be on top of b3. The problem PDDL is:

```
(:objects b1 b2 b3 b4 b5)
(:init (arm-empty) (on b1 b4) (on
b2 b5) (on b3 b2)
      (on-table b4) (on b5 b1)
(clear b3))
(:goal (and (on b4 b3))))
```

Prompt: Description of Problem (P1) + Context + Provide me with the problem PDDL file that describes the planning problem directly without further explanations.

GPT-4 (the generated problem PDDL):

```
(:objects b1 b2 b3 b4 b5 )
(:init (arm-empty) (on-table b1)
(on b2 b1) (on b3 b4)
      (on b4 b2) (on b5 b3) (clear
b5)))
(:goal (and (on b1 b2) (on b3 b5)
(on b4 b1)))
```

Planner:

```
(unstack b5 b3) (putdown b5)
(unstack b3 b4) (stack b3 b5)
(unstack b4 b2) (putdown b4)
(unstack b2 b1) (putdown b2)
(pickup b1) (stack b1 b2) (pickup
b4) (stack b4 b1)
```

Liu et al., “LLM+P: Empowering Large Language Models with Optimal Planning Proficiency.” arXiv preprint 2304.11477 (2023)
See also Kambhampati et al., “LLMs Can't Plan, But Can Help Planning in LLM-Modulo Frameworks.” arXiv preprint 2402.01817 (2024)

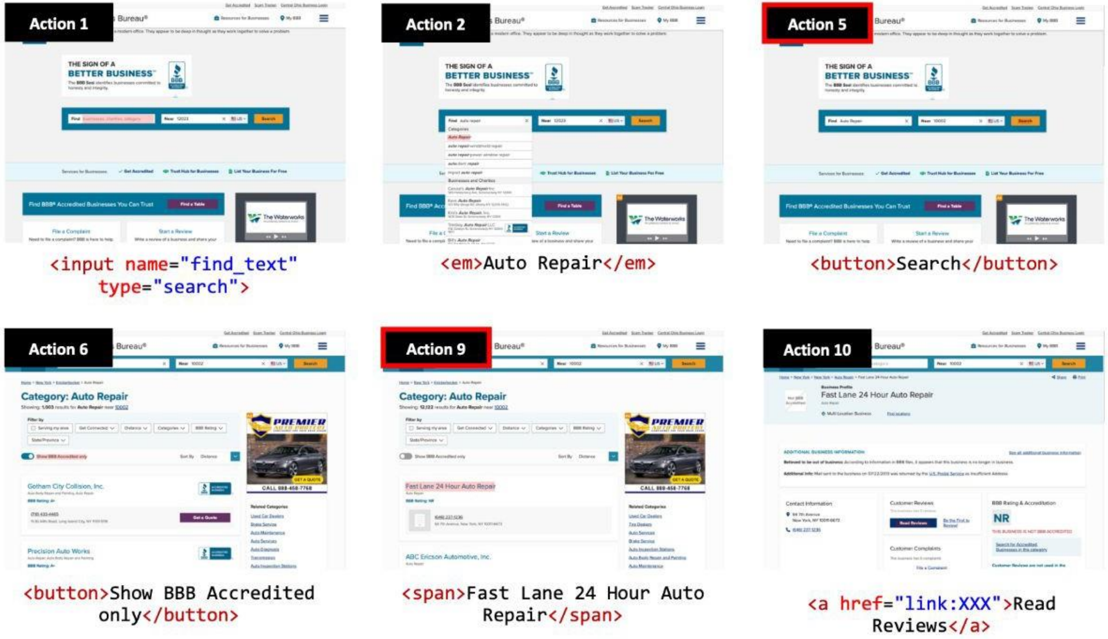
Language agent planning: web agents

Task Description:
Show me the reviews for the auto repair business closest to 10002.

Action Sequence:

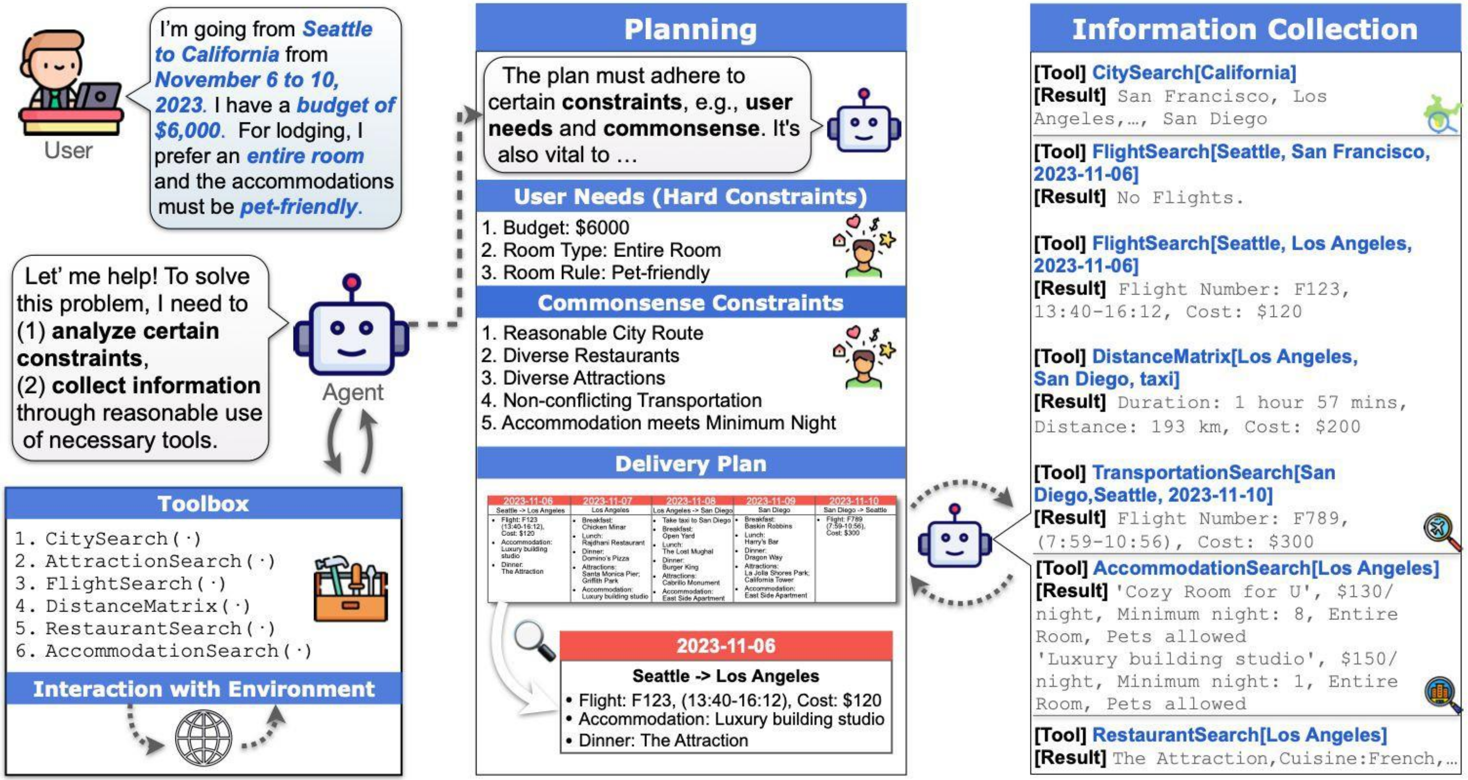
	Target Element	Operation
1.	[searchbox] *Find*	**TYPE:** ***auto repair***
2.	[button] *Auto Repair*	**CLICK**
3.	[textbox] *Near*	**TYPE:** ***10002***
4.	[button] *10002*	**CLICK**
5.	**[button] *Search***	**CLICK**
6.	[switch] *Show BBB Accredited only*	**CLICK**
7.	[svg]	**CLICK**
8.	[button] *Sort By*	**CLICK**
9.	**[link] *Fast Lane 24 Hour Auto Repair***	**CLICK**
10.	[link] *Read Reviews*	**CLICK**

Webpage Snapshots:



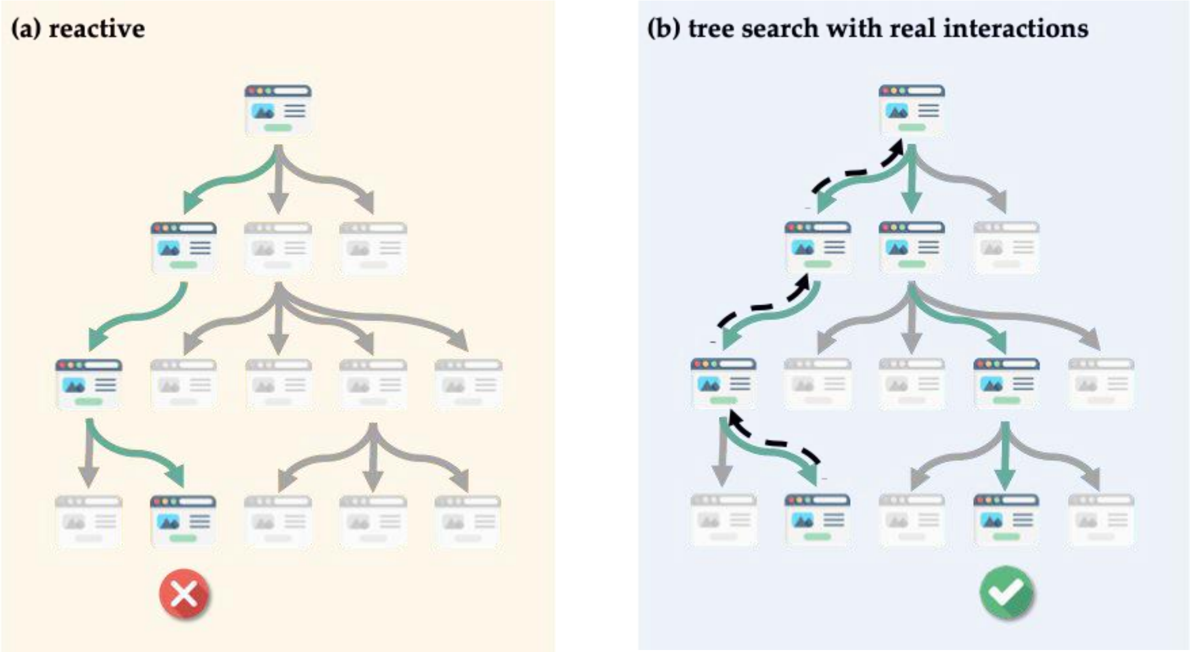
Deng et al., “Mind2Web: Towards a Generalist Agent for the Web.” NeurIPS (2024)

Language agent planning: travel planning



Xie et al., "TravelPlanner: A Benchmark for Real-World Planning with Language Agents." ICML (2024)

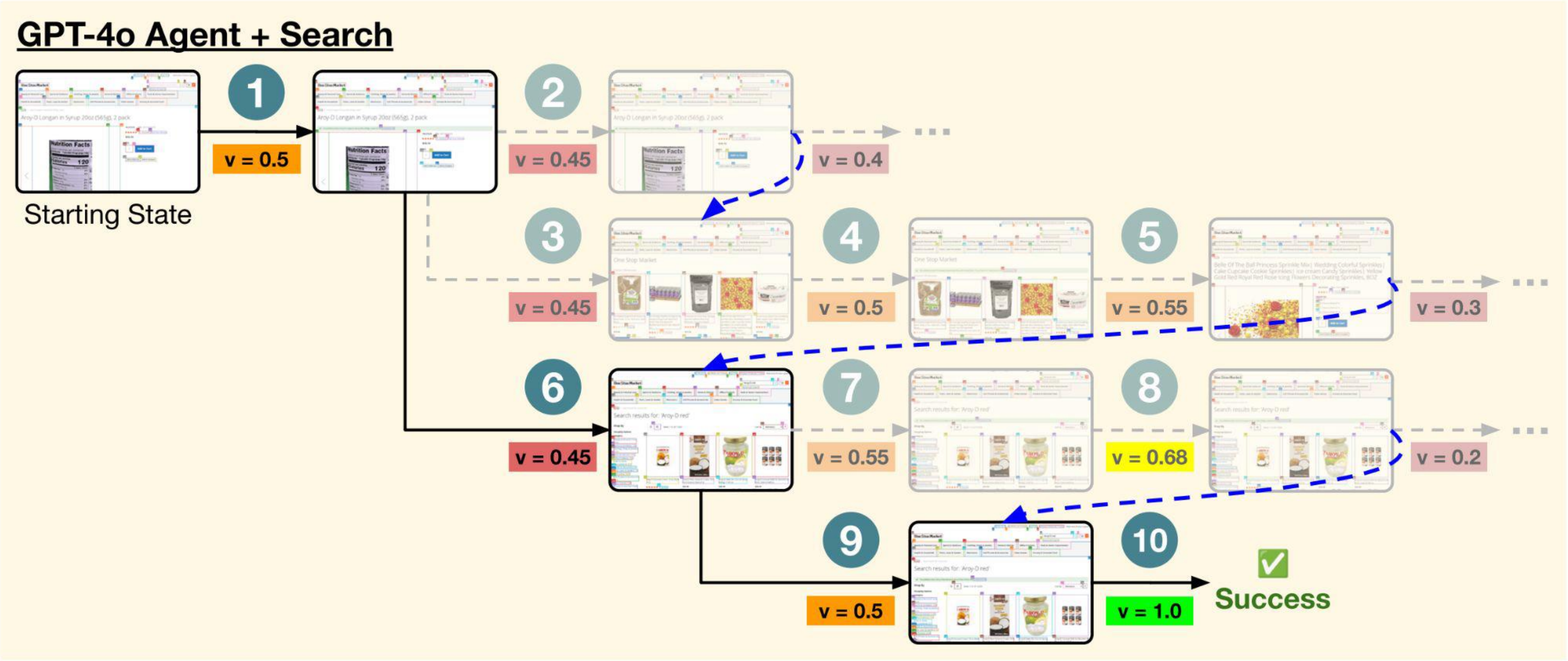
Planning paradigms for language agents



fast, easy to implement

greedy, short-sighted

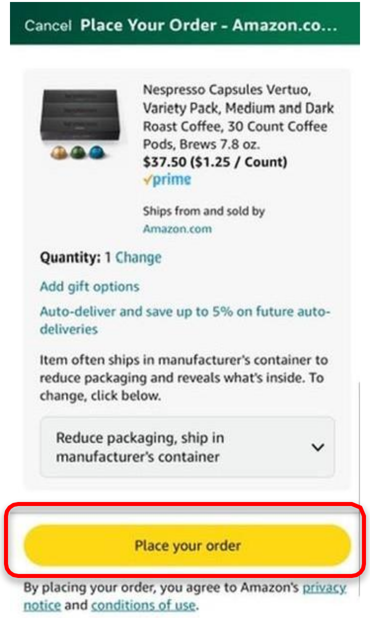
Tree search with real interactions

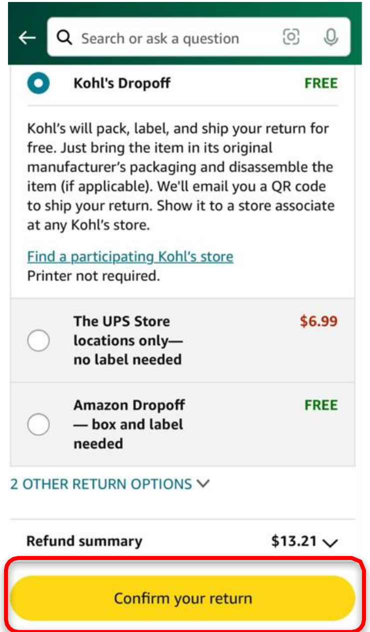


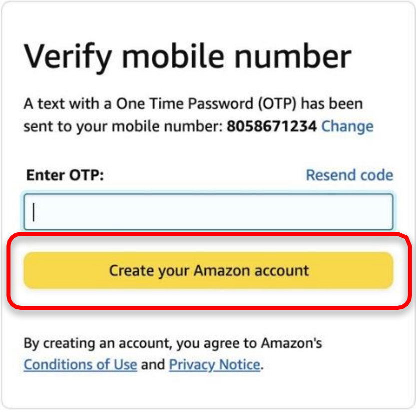
Jing Yu Koh, Stephen McAleer, Daniel Fried, Ruslan Salakhutdinov. "Tree Search for Language Model Agents." arXiv preprint arXiv:2407.01476 (2024).
Andy Zhou, Kai Yan, Michal Shlapentokh-Rothman, Haohan Wang, Yu-Xiong Wang. "Language Agent Tree Search Unifies Reasoning Acting and Planning in Language Models." ICML (2024).

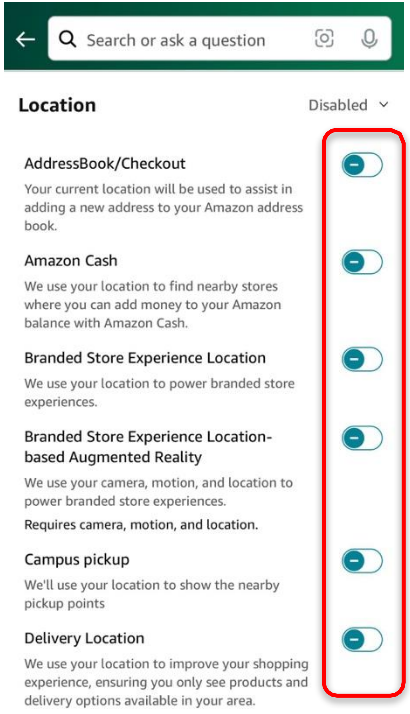
Challenges with tree search in the real world

- Many actions are state-changing and irreversible → backtracking ❌
- Safety/privacy risks
- Inference-time exploration could be slow and costly

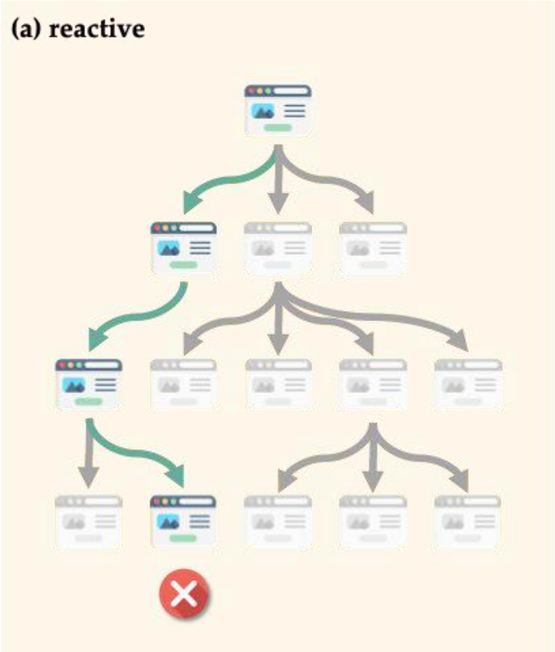


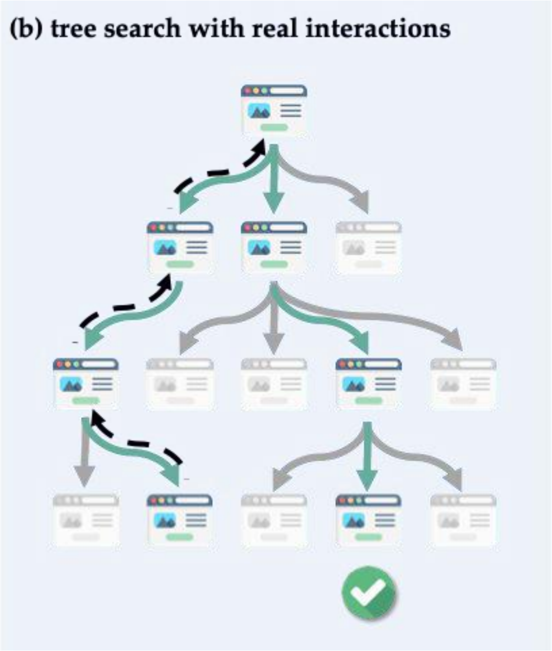


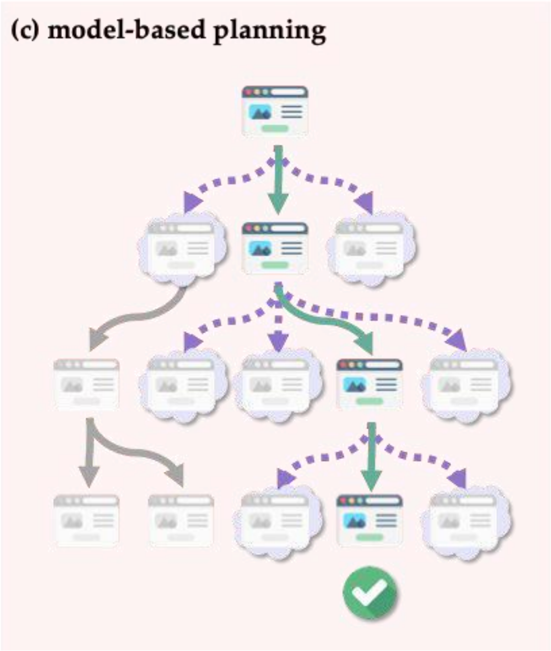




Planning paradigms for language agents







fast, easy to implement

greedy,short-sighted

systematic exploration

irreversible actions, unsafe, slow

faster, safer, systematic exploration

how to get a world model?

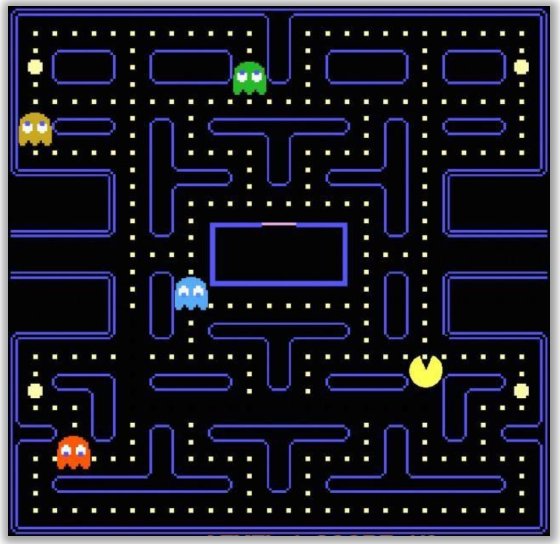
What's … a world model?

A computational model of environment transition dynamics

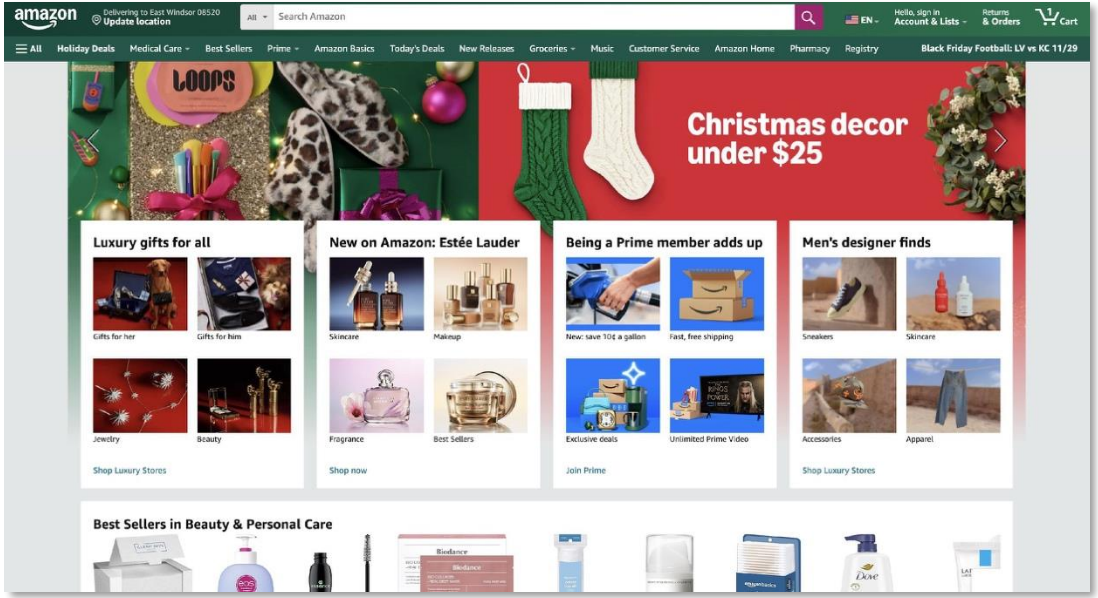
$$\hat{T}: \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$$

If I do this (a_t) right now (s_t), what would happen next (s_{t+1})?

Why hasn't it been done yet?

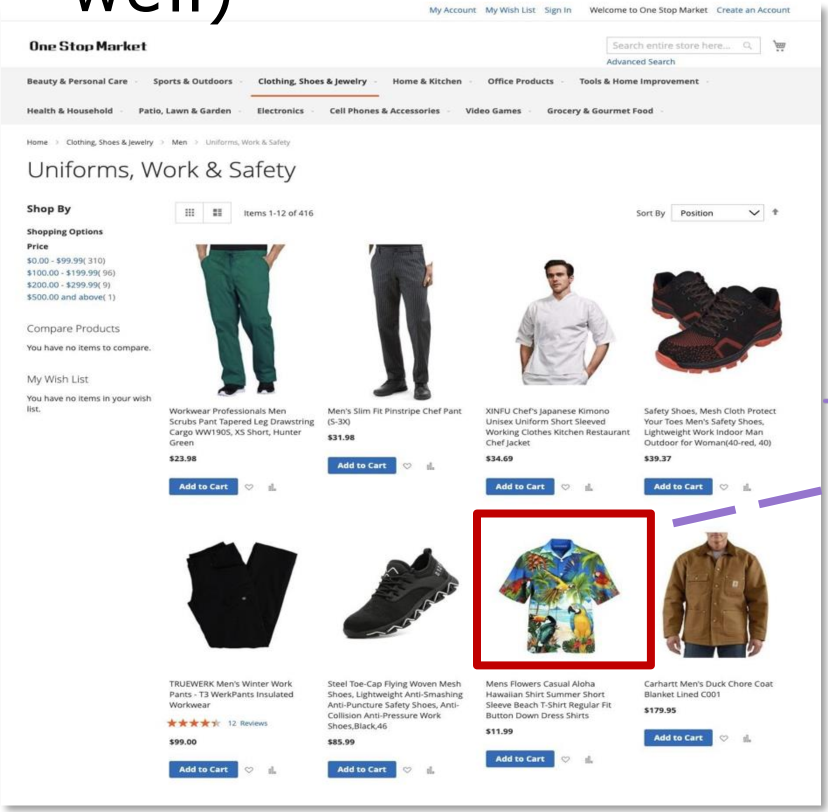


V.S



And **billions of other websites** on the Internet!

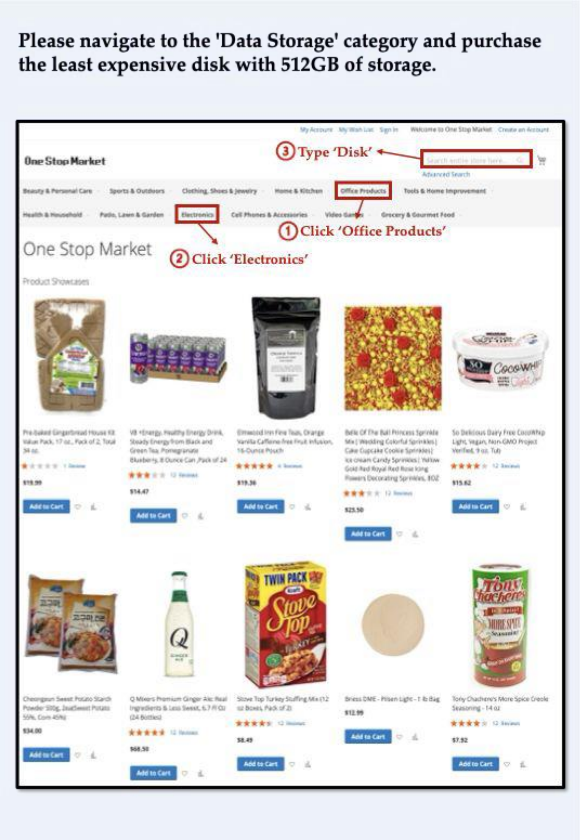
LLMs can predict state transitions (reasonably well)



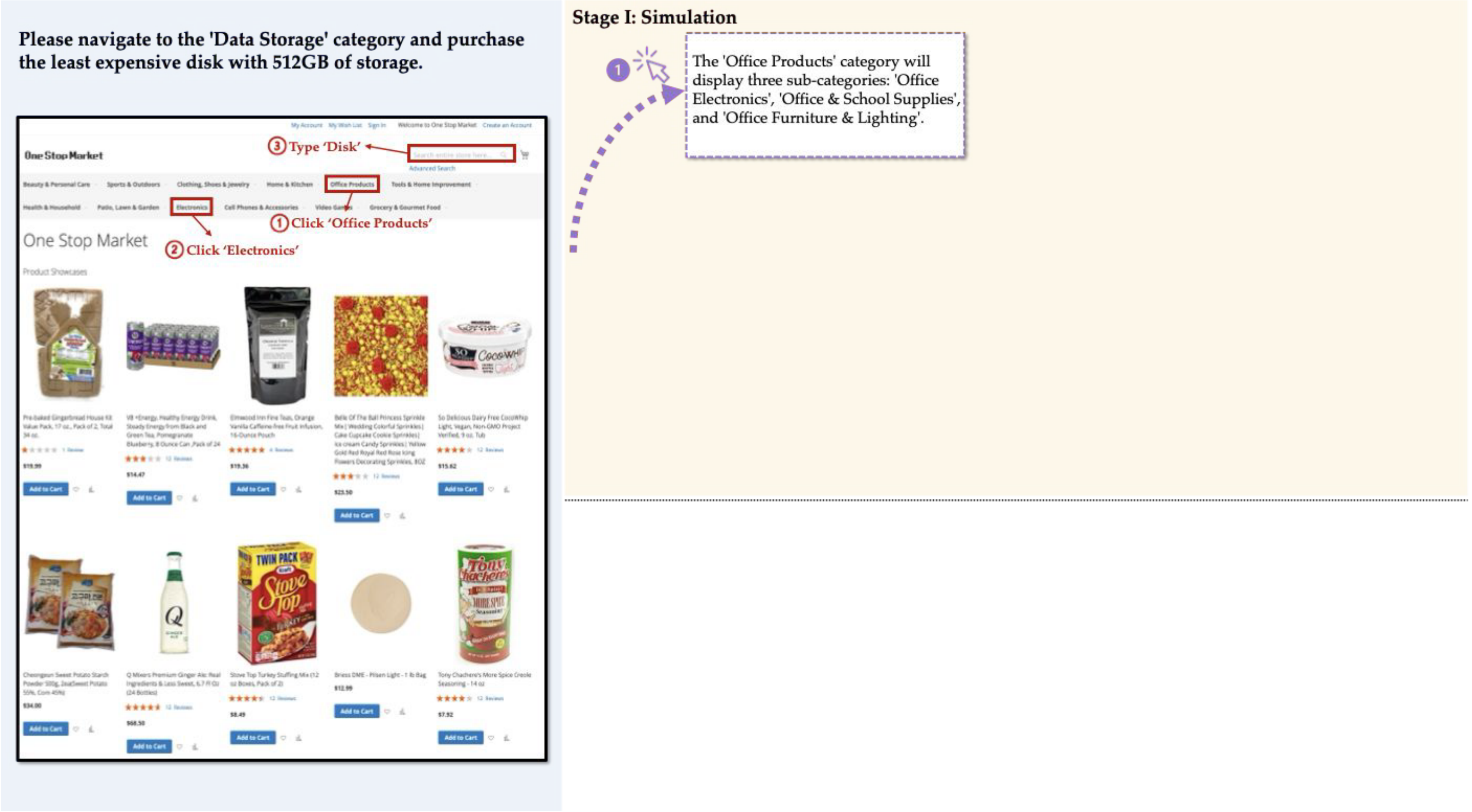
The page will navigate to a detailed product page for the "Mens Flowers Casual Aloha Hawaiian Shirt Summer Short Sleeve Beach T-Shirt Regular Fit Button Down Dress Shirts." This new page will likely contain additional information about the product including more detailed specifications, customer reviews, larger images, sizing options, and possibly a larger "Add to Cart" button. Other elements from the current category view like the grid of products will be replaced with the detailed view of this specific product.

Gu et al., "*Is Your LLM Secretly a World Model of the Internet?* Model-Based Planning for Web Agents." arxiv preprint 2411.06559 (2024).

WebDreamer: model-based planner

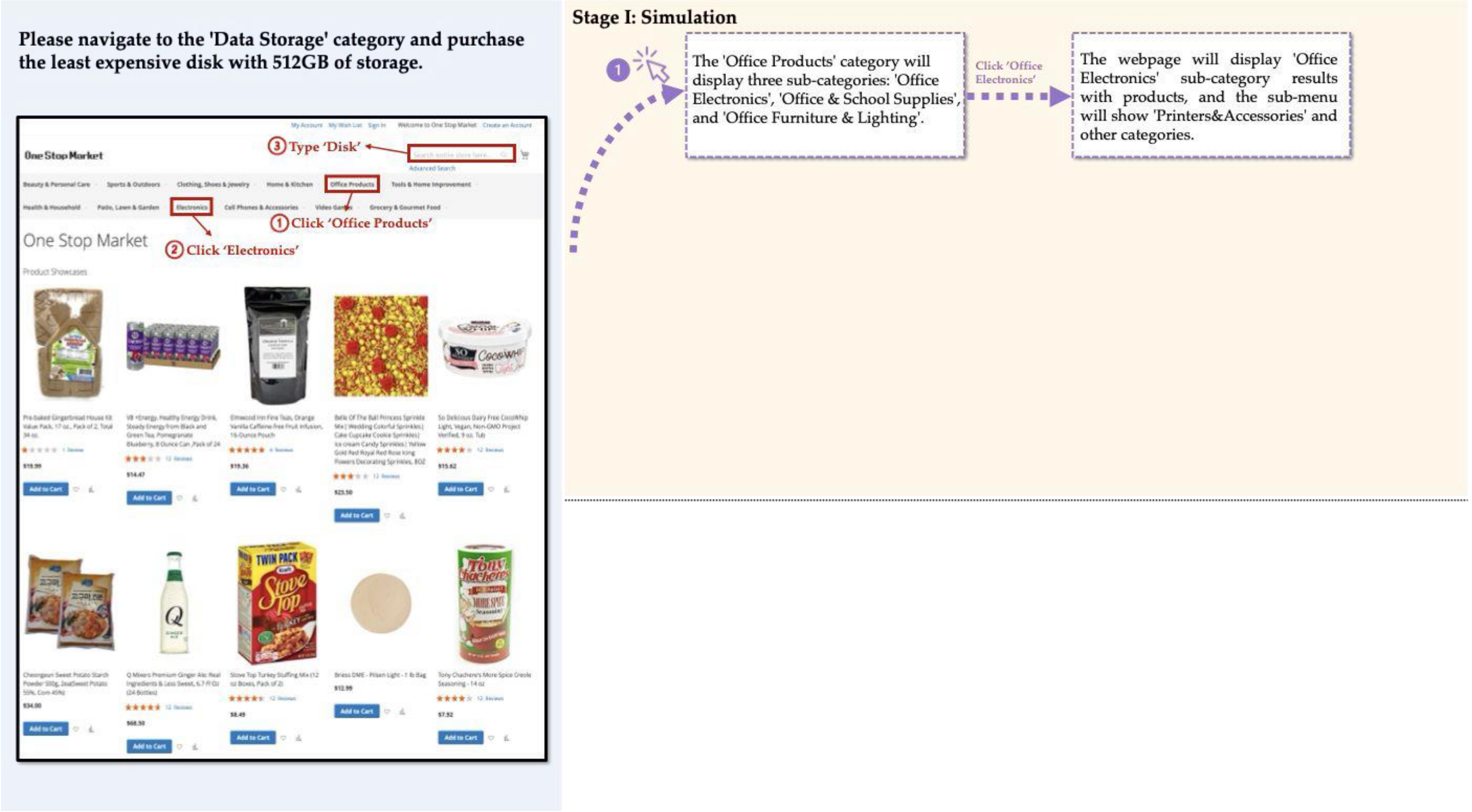


WebDreamer: model-based planner



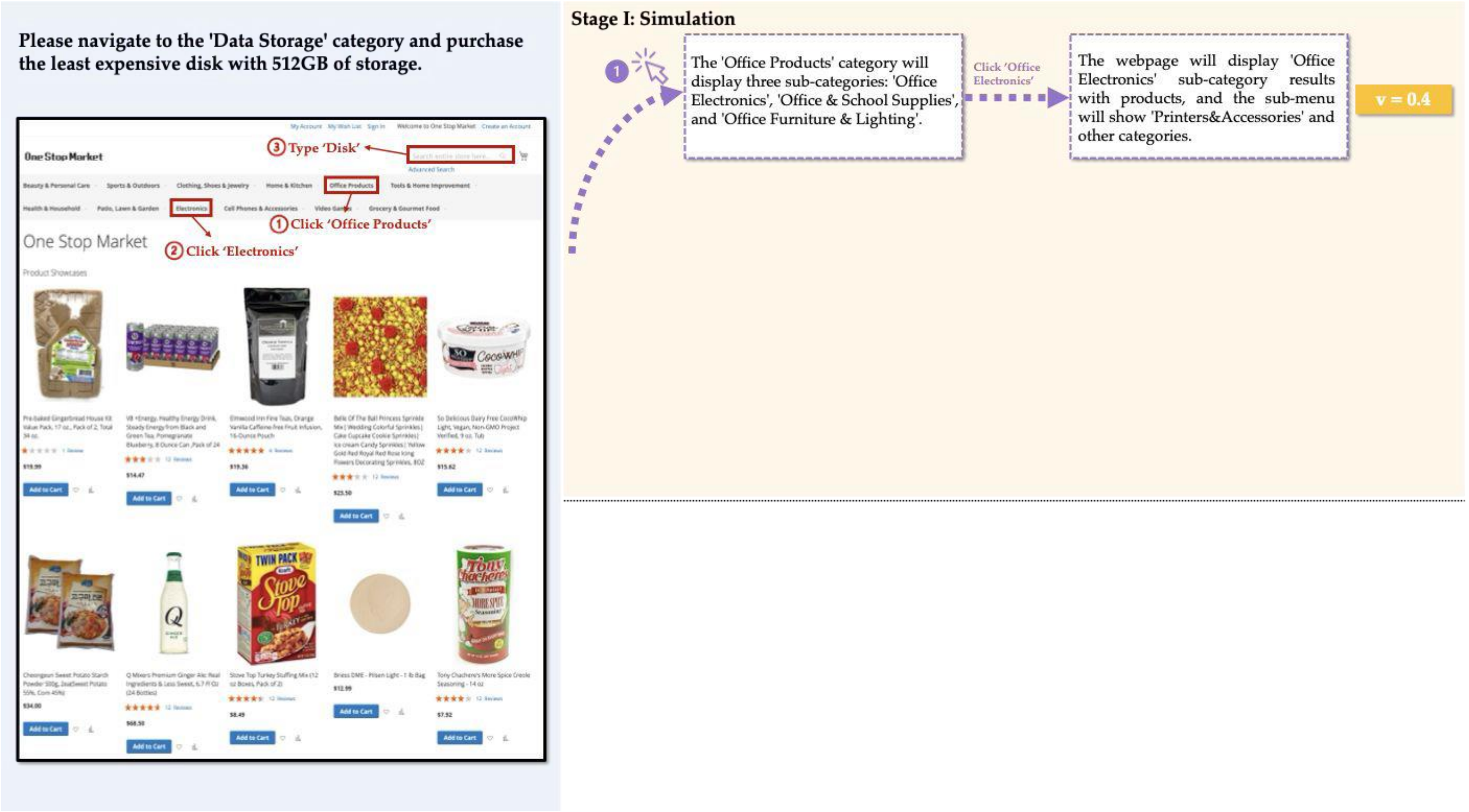
Gu et al., "*Is Your LLM Secretly a World Model of the Internet?* Model-Based Planning for Web Agents." arxiv preprint 2411.06559 (2024).

WebDreamer: model-based planner



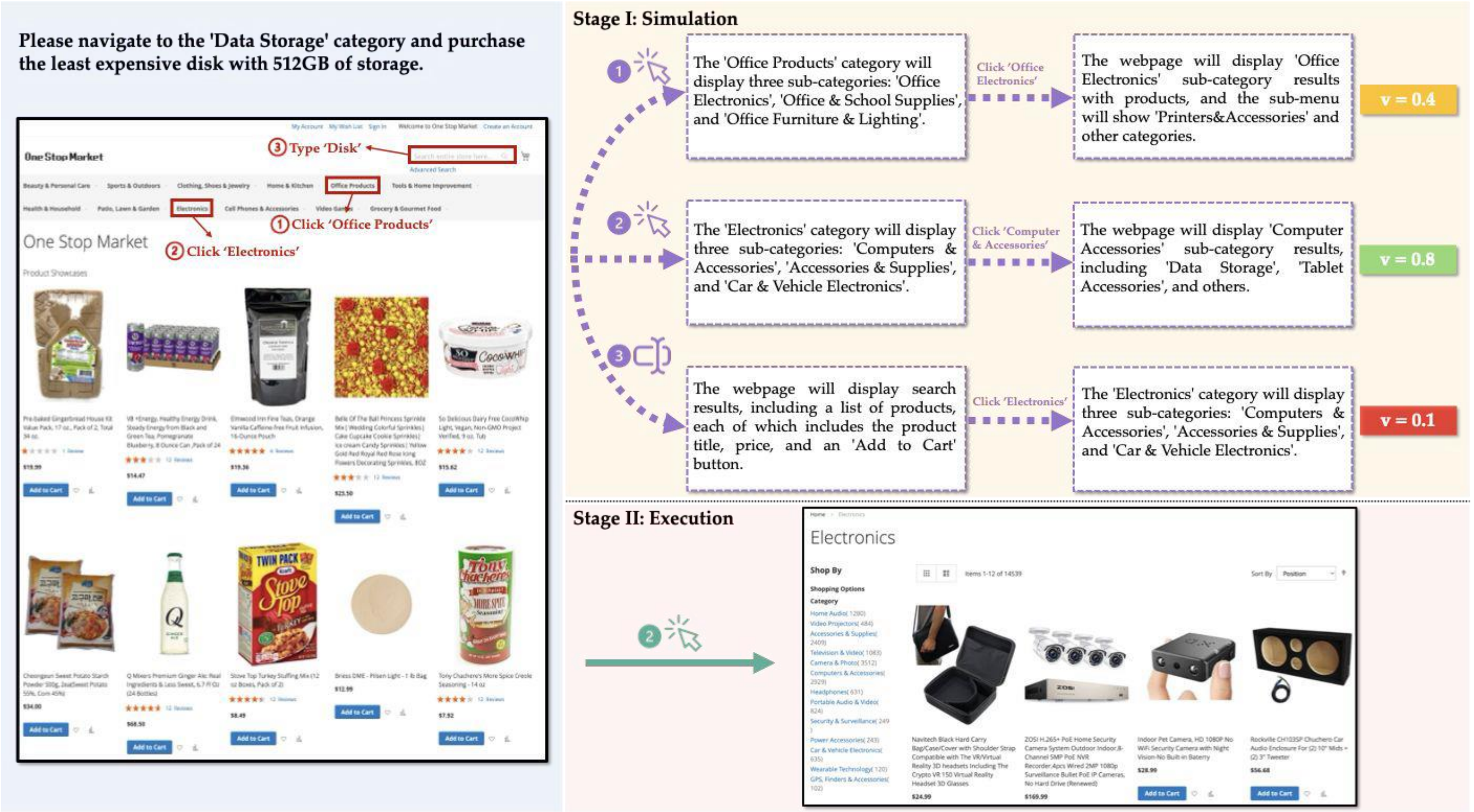
Gu et al., “*Is Your LLM Secretly a World Model of the Internet?* Model-Based Planning for Web Agents.” arxiv preprint 2411.06559 (2024).

WebDreamer: model-based planner



Gu et al., "*Is Your LLM Secretly a World Model of the Internet?* Model-Based Planning for Web Agents." arxiv preprint 2411.06559 (2024).

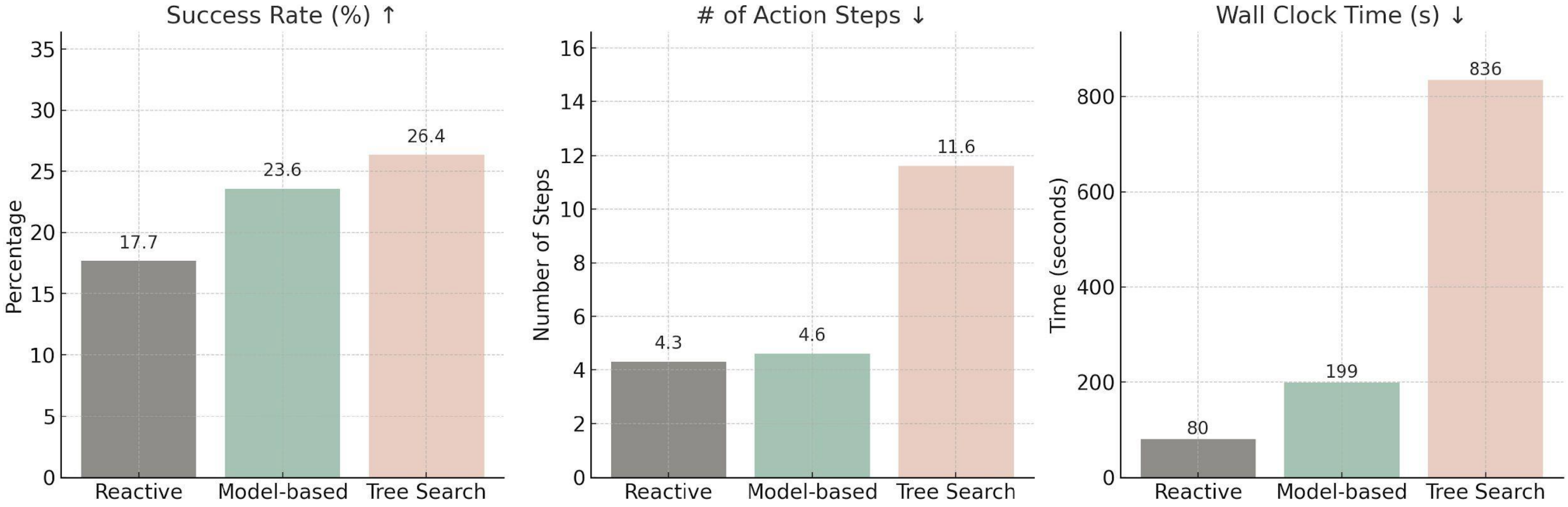
WebDreamer: model-based planner



Gu et al., "*Is Your LLM Secretly a World Model of the Internet?* Model-Based Planning for Web Agents." arxiv preprint 2411.06559 (2024).

Results on VisualWebArena

Model-based planning is more accurate than reactive planning and more efficient than tree search (also recall the challenges of tree search in real-world environments)



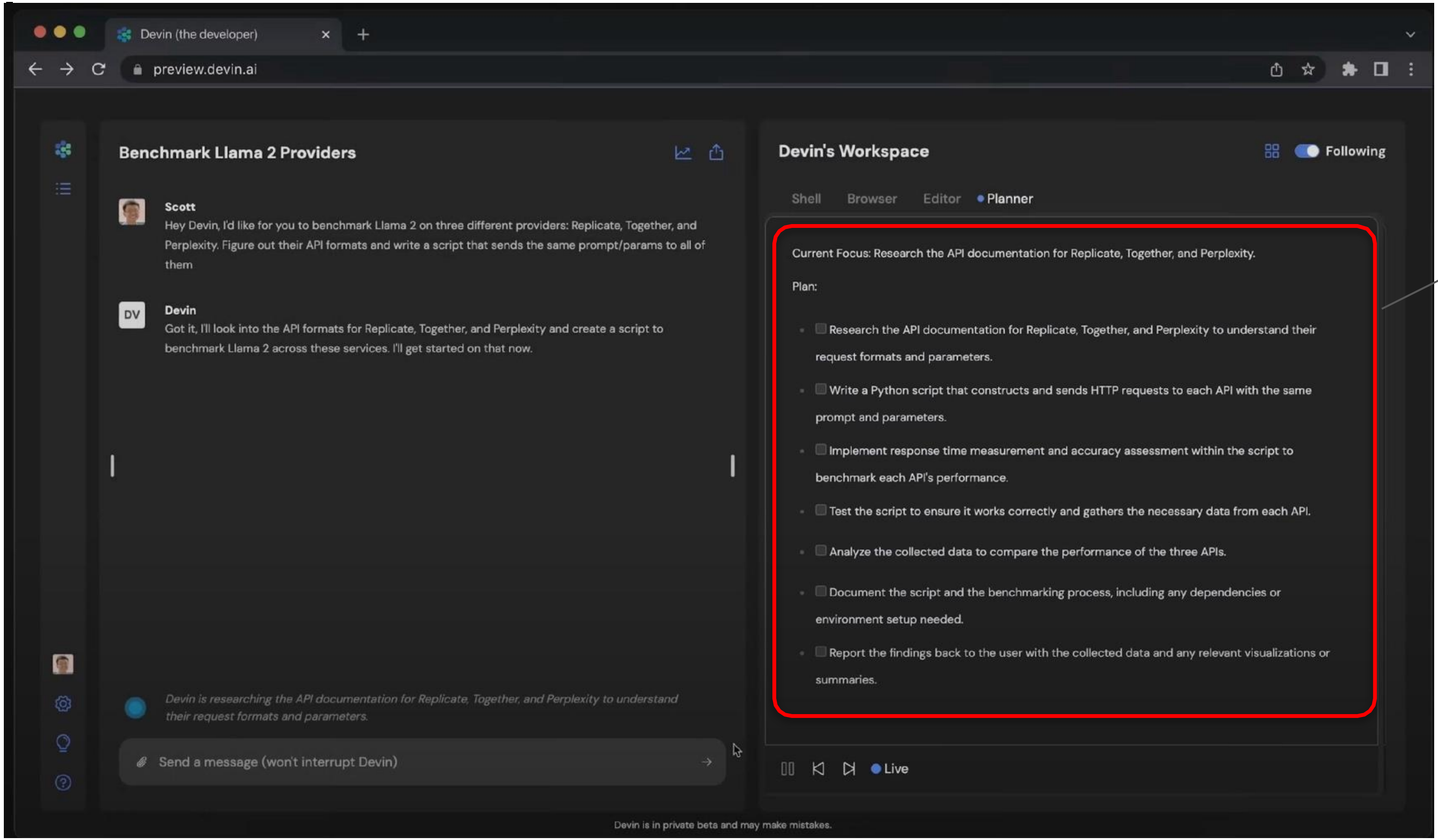
Reactive and tree search results are from Koh et al., (2024)

Hierarchical planning and dynamic replanning

Song et al., "LLM-Planner: Few-Shot Grounded Planning for Embodied Agents with Large Language Models." ICCV (2023)

By Cognition AI, https://www.youtube.com/watch?v=fjHtjT7GO1c

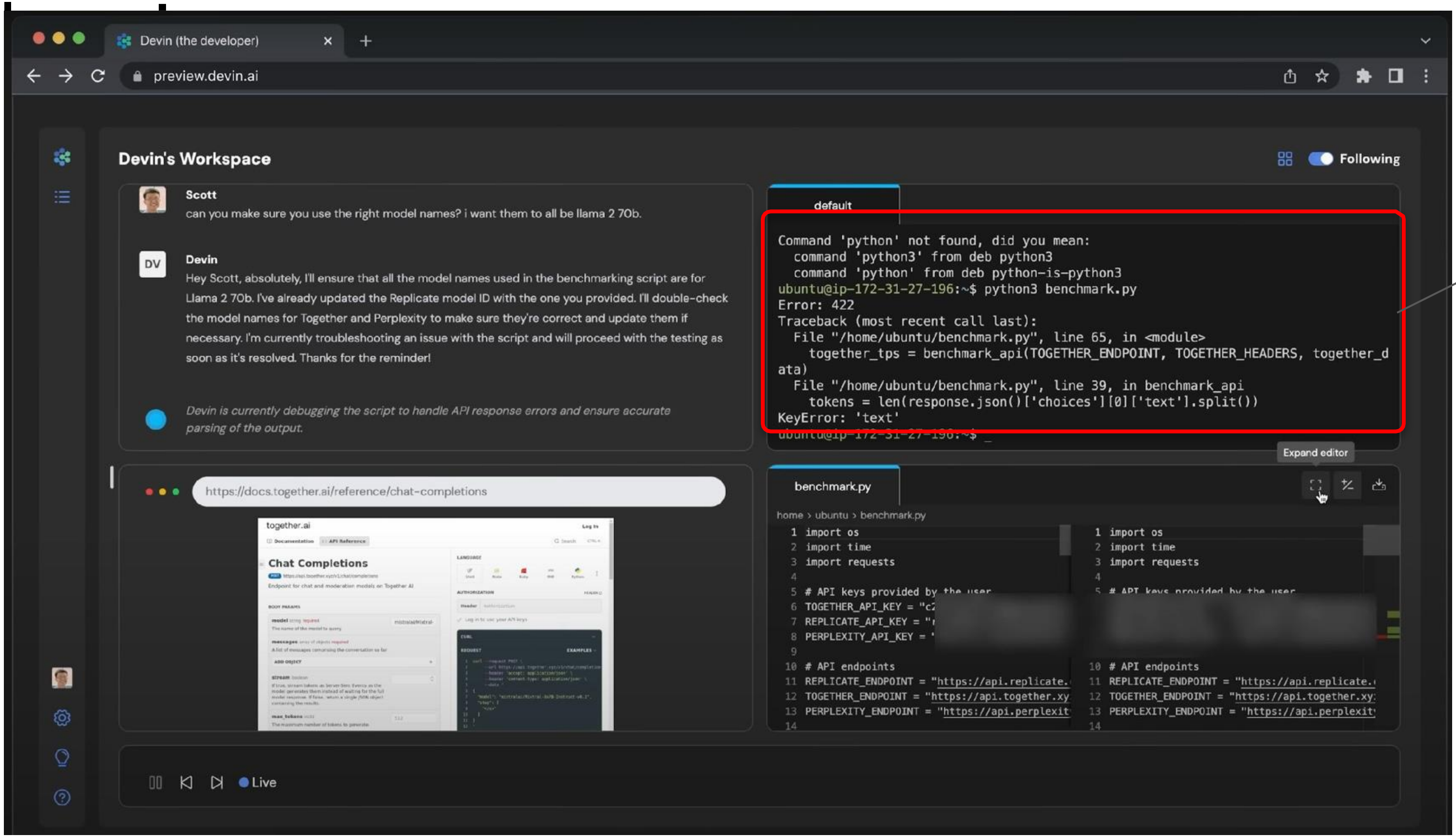
Hierarchical planning and dynamic rep



High-level Plan

By Cognition AI, https://www.youtube.com/watch?v=fjHtjT7GO1c

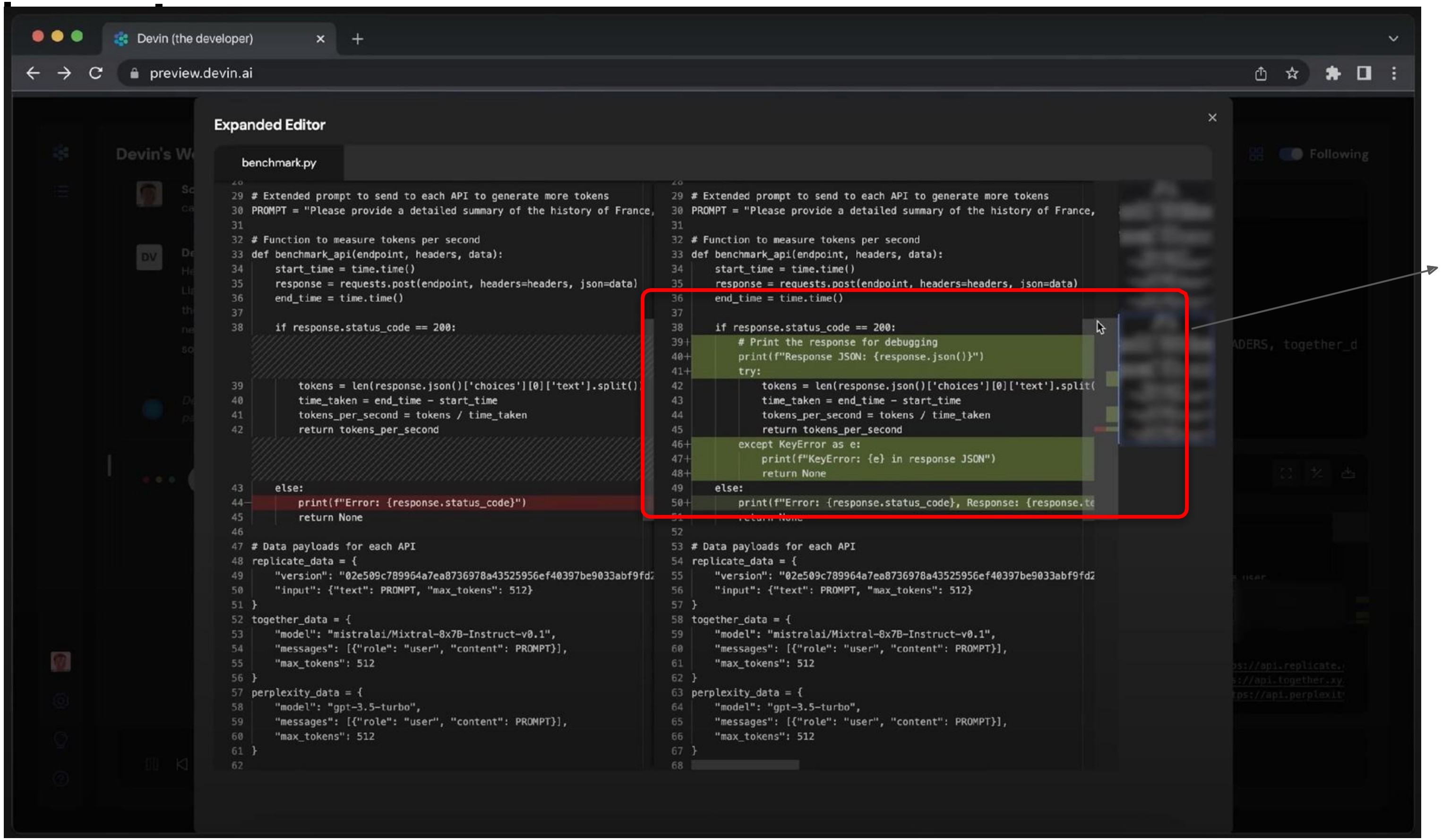
Hierarchical planning and dynamic rep



Ran into exception when carrying out a subgoal

By Cognition AI, https://www.youtube.com/watch?v=fjHtjT7GO1c

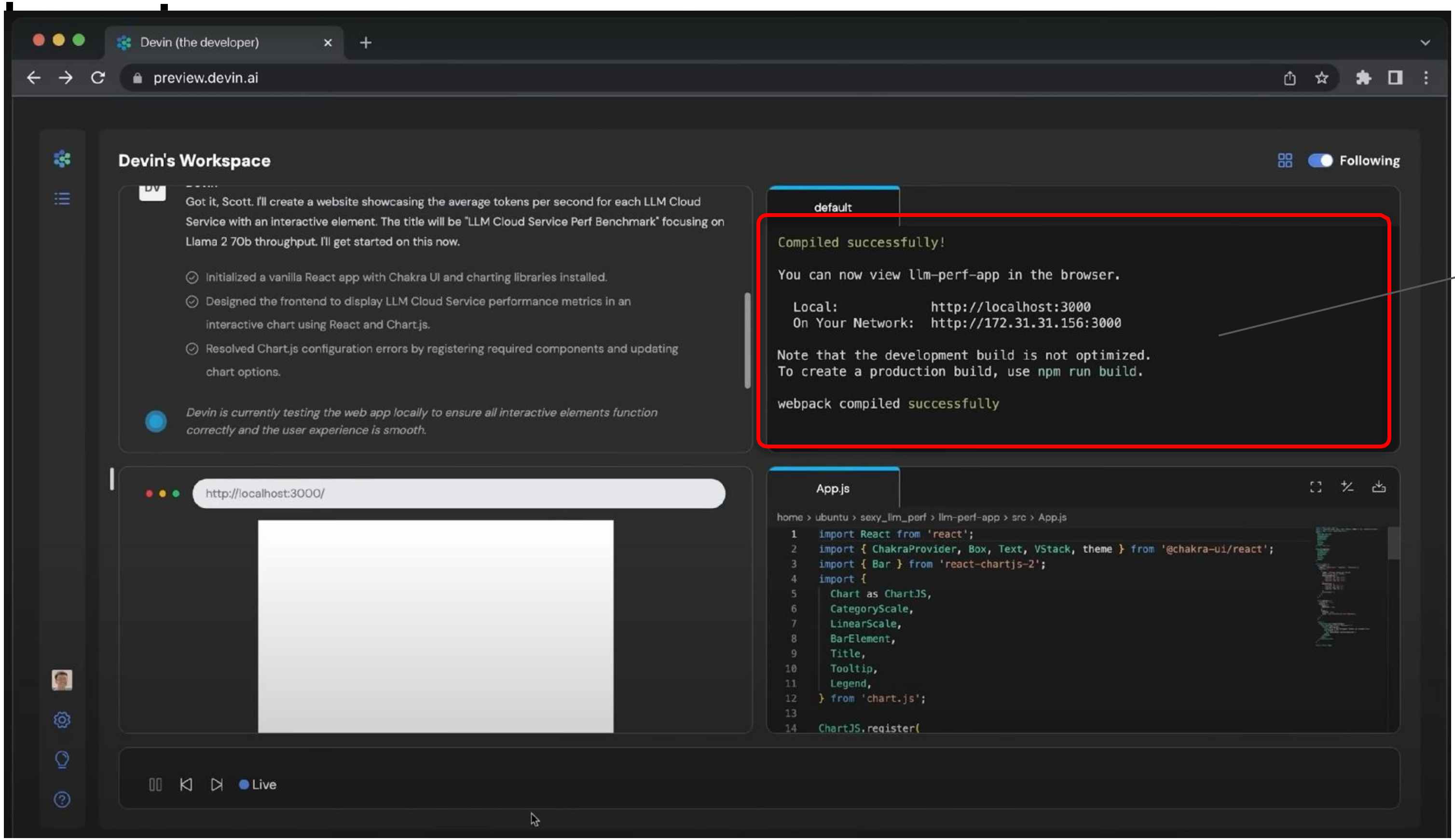
Hierarchical planning and dynamic rep



Re-plan and add print() around the exception to get debug info

By Cognition AI, https://www.youtube.com/watch?v=fjHtjT7GO1c

Hierarchical planning and dynamic rep



Solved the exception based on the printed out debug info!

Planning: Takeaways

- Language agents are expanding into new planning scenarios
  - characterized by expressive but fuzzy goal specifications, open-ended action spaces, more difficult and sometimes non-binary goal tests
- Language for reasoning also enables new planning abilities
  - Generalist world models and model-based planning
  - Hierarchical planning and dynamic replanning
- The best planning strategy is dependent on the LLM; stronger LLMs may require less scaffolding (i.e., more ‘reactive’)
- How to improve planning in LLMs is still largely an open question