

Agents

Acknowledgement: these slides are partially based on EMNLP Agent tutorial. The full materials can be found at :<https://language-agent-tutorial.github.io/>

So far..

- **We talked about several recent directions regarding LLMs..**
 - Transformer architecture and large scale pre-training
 - On the fly task definition: ICL
 - Reasoning: using CoT prompting
 - External knowledge access: RAG systems
- All these capabilities led to a recent paradigm change in AI – **Agentic AI**
 - Combine all of these capabilities into an architecture designed to support autonomous task-driven behavior.

Agents!

“Anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators”

Artificial Intelligence: A Modern Approach by Stuart Russell and Peter Norvig

Weak notion of agency:

Autonomy: Agents control their internal state and actions without constant human oversight.

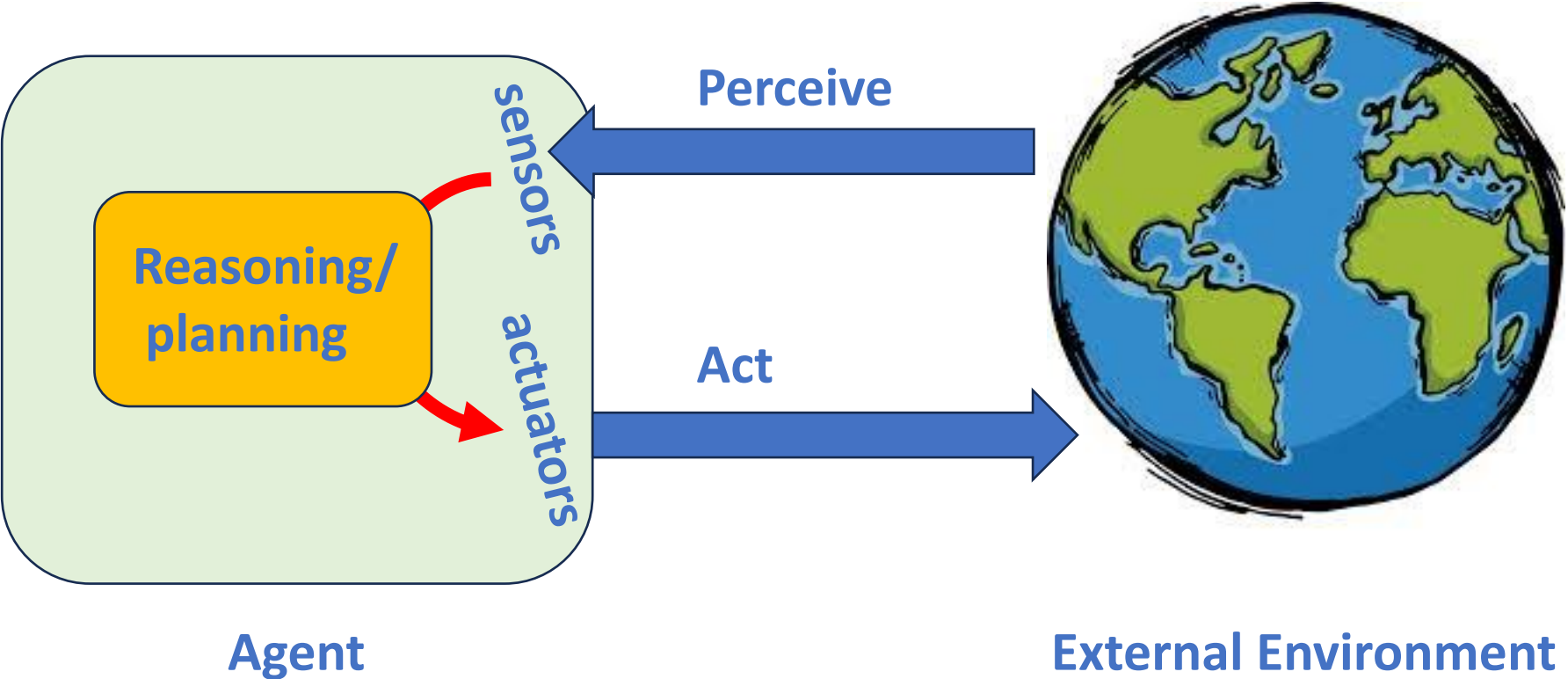
Reactivity: They perceive and respond to environmental changes in real-time.

Proactiveness: They exhibit goal-directed behavior, acting on their own initiative.

Social Ability: They interact with other agents or humans.

Intelligent Agents: Theory and Practice by Michael Wooldridge and Nicholas Jennings

Agents!



1990's vs LLM based agents

- Original definitions of agents are rooted in 1990's view of AI:
 - Symbolic definitions, reasoning via rules-based and probabilistic reasoning.
- LLM-based agents provide an opportunity:
 - Perception made easier via multi-modal LMs
 - Reasoning made easier via language-based reasoning (i.e., token generation)
- **Are LLM-based agents just an LLM connected to a perception stream and action interface?**

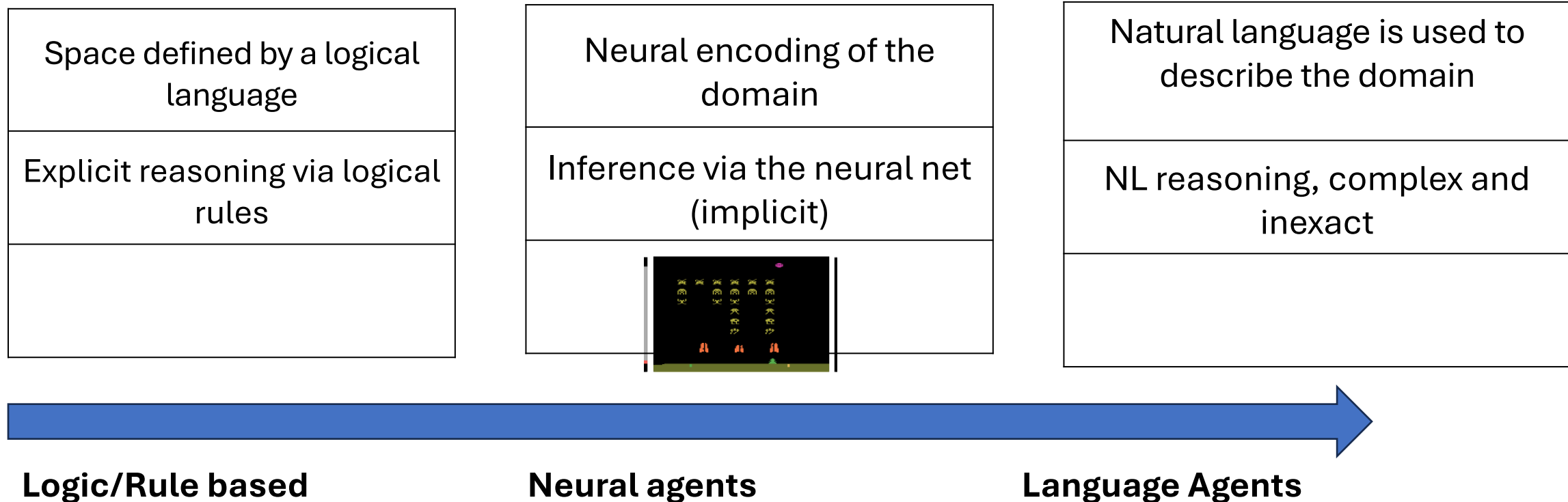
Two views of LLM-agents

- **LLM-first view:** “the LLM is the agent” when connected to an action interface. Rely on prompting to convert tokens to actions.
- **Agent first view:** use existing agent architectures, integrating LLMs as part of the architecture to support communication and reasoning via natural language.

LLM-Driven agents

- LLM driven agents use language as the driver for decision making and reasoning and communication.
- We make the distinction between “thoughts” and “actions”.
 - “Thoughts” or reasoning steps map to language generation without manipulating the external environment
 - “Actions” are the results of reasoning, and allow the agent to change its external environment
- Reasoning can operate over precepts directly, involve self-reflection, abstractions or other meta-cognitive steps.

Evolution of AI Agents

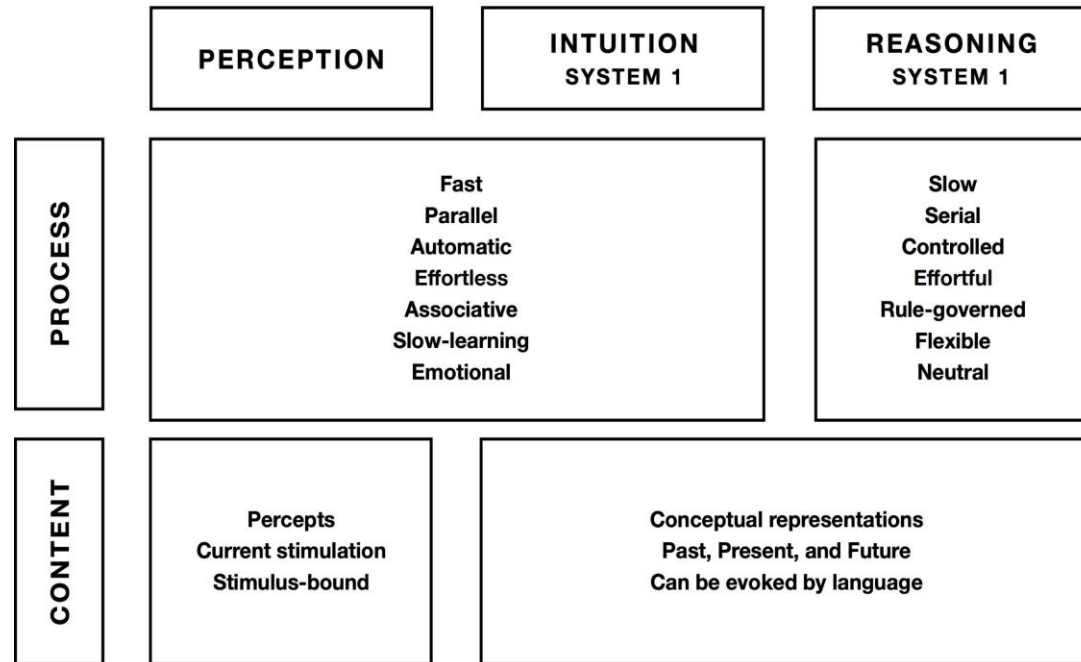


Agent Fundamentals

- Agents are comprised of several modules, working together to achieve the agent's goals.
- **Reasoning:** change internal state by generating tokens that update the local decision context
- **Memory:** Retrieve or write to long term memory, can be viewed as a form of learning.
- **Planning:** Agent actions are defined over a space, decide which action/action-sequence to take

Reasoning

Multiple processes involved in reasoning. We often make the distinction between System-1 and 2, or Intuitive/fast and structured/slow.



(Kahneman, 2003),

Reasoning

- These distinctions can also map to agent reasoning
- View reasoning as internal actions, generating tokens.
- Different mechanisms can be used – fast or slow.

Math Word Problems (free response)

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Math Word Problems (multiple choice)

Q: How many keystrokes are needed to type the numbers from 1 to 500?
Answer Choices: (a) 1156 (b) 1392 (c) 1480 (d) 1562 (e) 1788

A: There are 9 one-digit numbers from 1 to 9. There are 90 two-digit numbers from 10 to 99. There are 401 three-digit numbers from 100 to 500. $9 + 90(2) + 401(3) = 1392$. The answer is (b).

StrategyQA

Q: Yes or no: Would a pear sink in water?

A: The density of a pear is about 0.6 g/cm^3 , which is less than water. Thus, a pear would float. So the answer is no.

Date Understanding

Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?

A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.

Reasoning

- Reasoning can help bridge the divide between tasks described at a high level of abstractions and the complex set of actions needed to perform them.

Cook dinner!

Today is Thursday. Thursday is typically pizza night.

We are out of tomato sauce. What are the ingredients needed to make tomato sauce from scratch? tomatoes, sugar, salt, .

Where are the tomatoes? Probably in the fridge.

OPEN FRIDGE DOOR.

ReAct

ReAct enables LLM to interleave reasoning traces with actionable tool use.

This combination improve over models that only reason or only act

(1) Hotspot QA

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

(1a) Standard
Answer: iPod ❌

(1b) CoT (Reason Only)
Thought: Let's think step by step. Apple Remote was originally designed to interact with Apple TV. Apple TV can be controlled by iPhone, iPad, and iPod Touch. So the answer is iPhone, iPad, and iPod Touch. ❌

(1c) Act-Only
Act 1: Search[Apple Remote]
Obs 1: The Apple Remote is a remote control ...
Act 2: Search[Front Row]
Obs 2: Could not find [Front Row]. Similar: ...
Act 3: Search[Front Row (software)]
Obs 3: Front Row is a discontinued media center software ...
Act 4: Finish[yes] ❌

(1d) 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 ...
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 ..
Thought 4: Front Row (software) is controlled by an Apple Remote or the keyboard function keys. So the answer is keyboard function keys.
Act 4: Finish[keyboard function keys] ✓

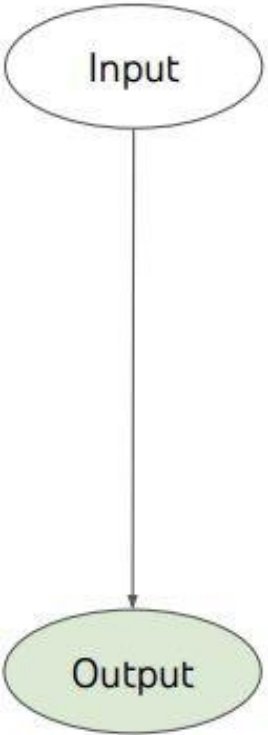
(2) AlfWorld

You are in the middle of a room. Looking quickly around you, you see a cabinet 6, a cabinet 1, a coffee machine 1, a countertop 3, a stove burner 1, and a toaster 1. Your task is to: Put some pepper shaker on a drawer.

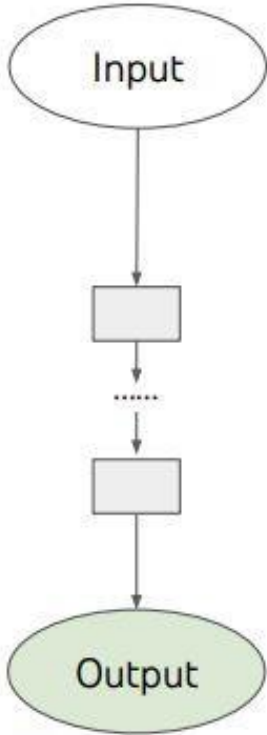
(2a) Act-Only
Act 1: Go to drawer 1
Obs 1: The drawer 1 is closed.
Act 2: Open drawer 1
Obs 2: You open the drawer 1. The drawer 1 is open. In it, you see a dish sponge 2, and a spoon 1.
Act 3: Go to sinkbasin 1
Obs 3: On the sinkbasin 1, you see a dish sponge 3, a spatula 1, and a spoon 2.
Act 4: Take peppershaker 1 from sinkbasin 1
Obs 4: Nothing happens.
Act 5: Take peppershaker 1 from sinkbasin 1
Obs 5: Nothing happens. ❌

(2b) ReAct (Reason + Act)
Act 1: Think[First I need to find a pepper shaker...more likely to appear in cabinets (1-6), countertops (1-3), ...]
Act 2: Go to cabinet 1
Obs 2: On the cabinet 1, you see a vase 2.
(Here the agent go to cabinet 1, then cabinet 2, 3, then countertop 1 and 2)
Act 6: Go to countertop 3
Obs 6: On the countertop 3, you see a apple 1, a bread 1, a pepper shaker 1, and a vase 3.
Act 7: Take pepper shaker 1 from countertop 3
Obs 7: You pick up the pepper shaker 1 from the countertop 3.
Act 8: Think[Now I find a pepper shaker 1. Next, I need to put it in/on drawer 1.]
Act 9: Go to drawer 1
Obs 9: Drawer 1 is closed.
Act 10: Open drawer 1
Obs 10: You open Drawer 1 ..
Act 11: Put pepper shaker 1 in/on drawer 1
Obs 11: You put pepper shaker 1 in/on the drawer 1. ✓

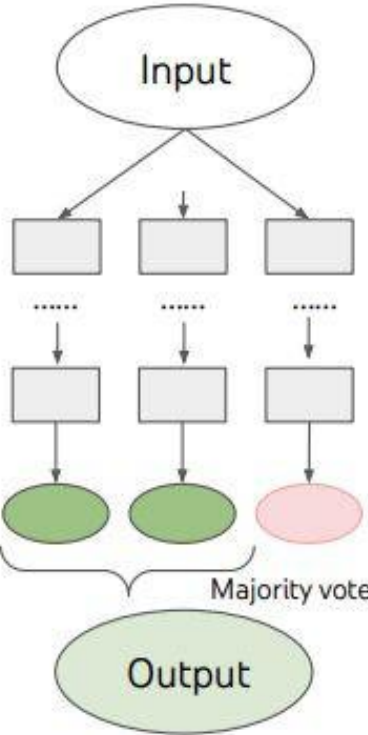
Different levels of reasoning



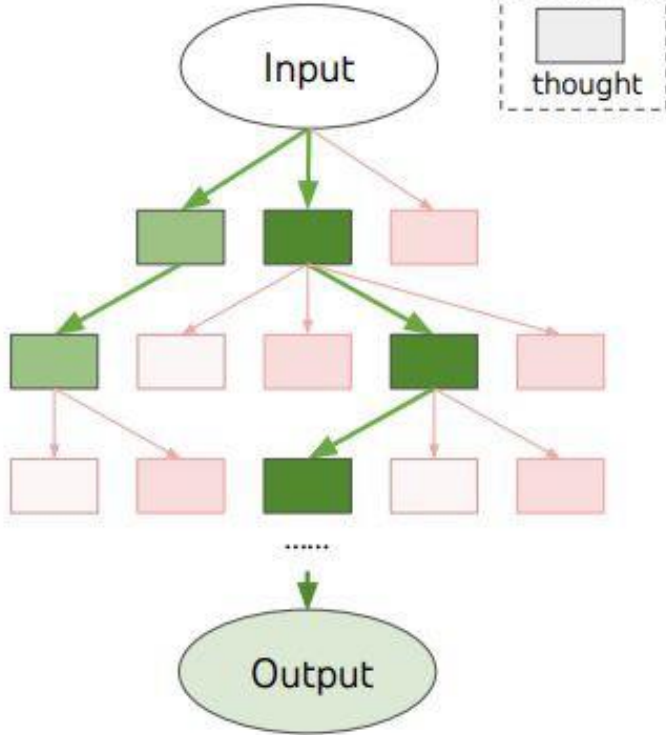
(a) Input-Output Prompting (IO)



(c) Chain of Thought Prompting (CoT)

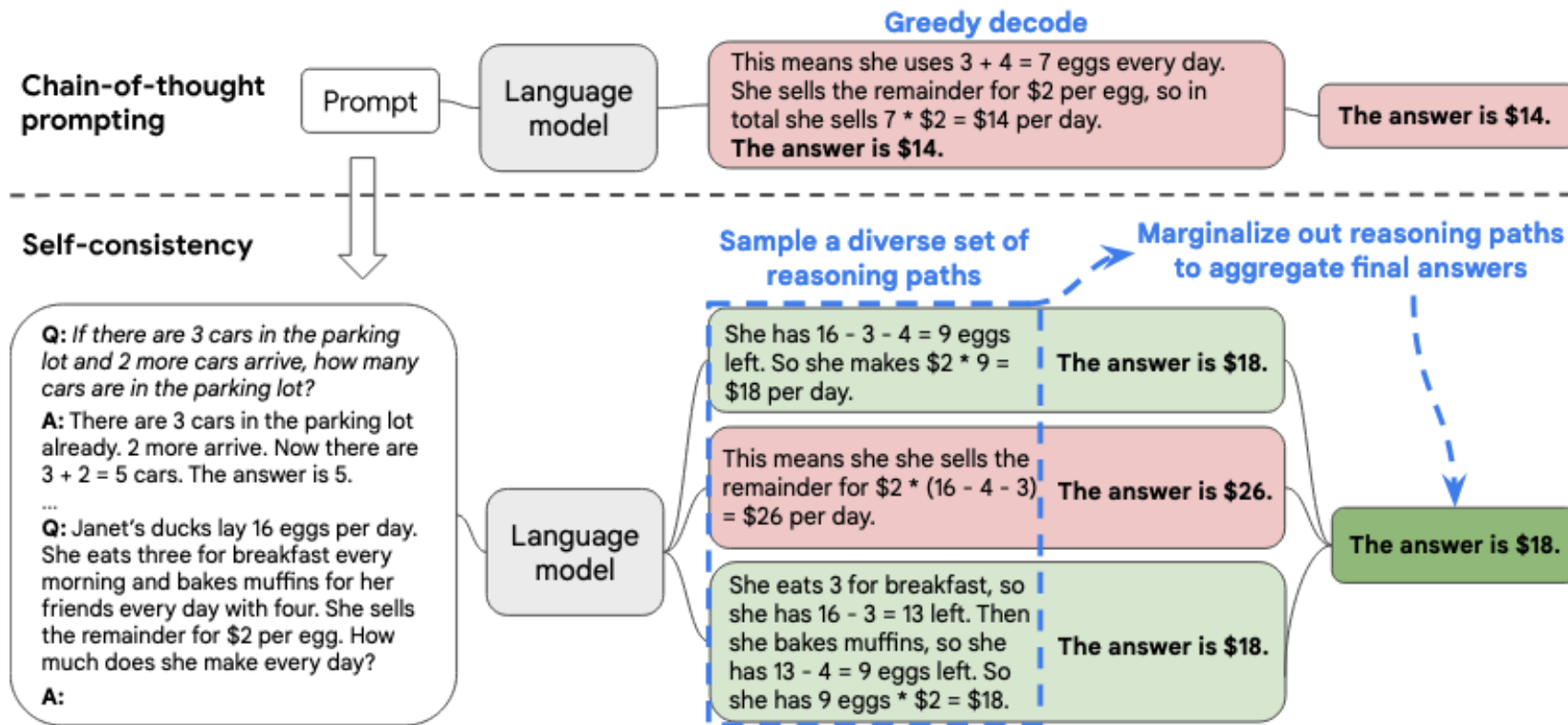


(c) Self Consistency with CoT (CoT-SC)



(d) **Tree of Thoughts (ToT)**

Self-consistency

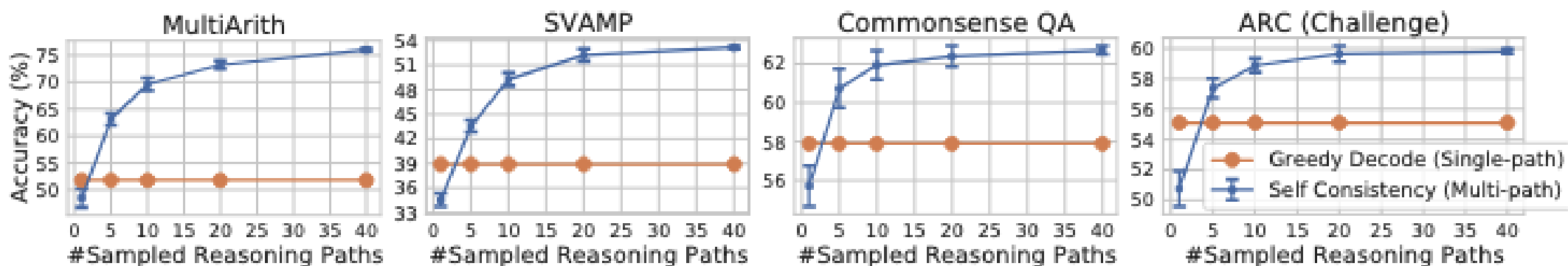


Self-consistency

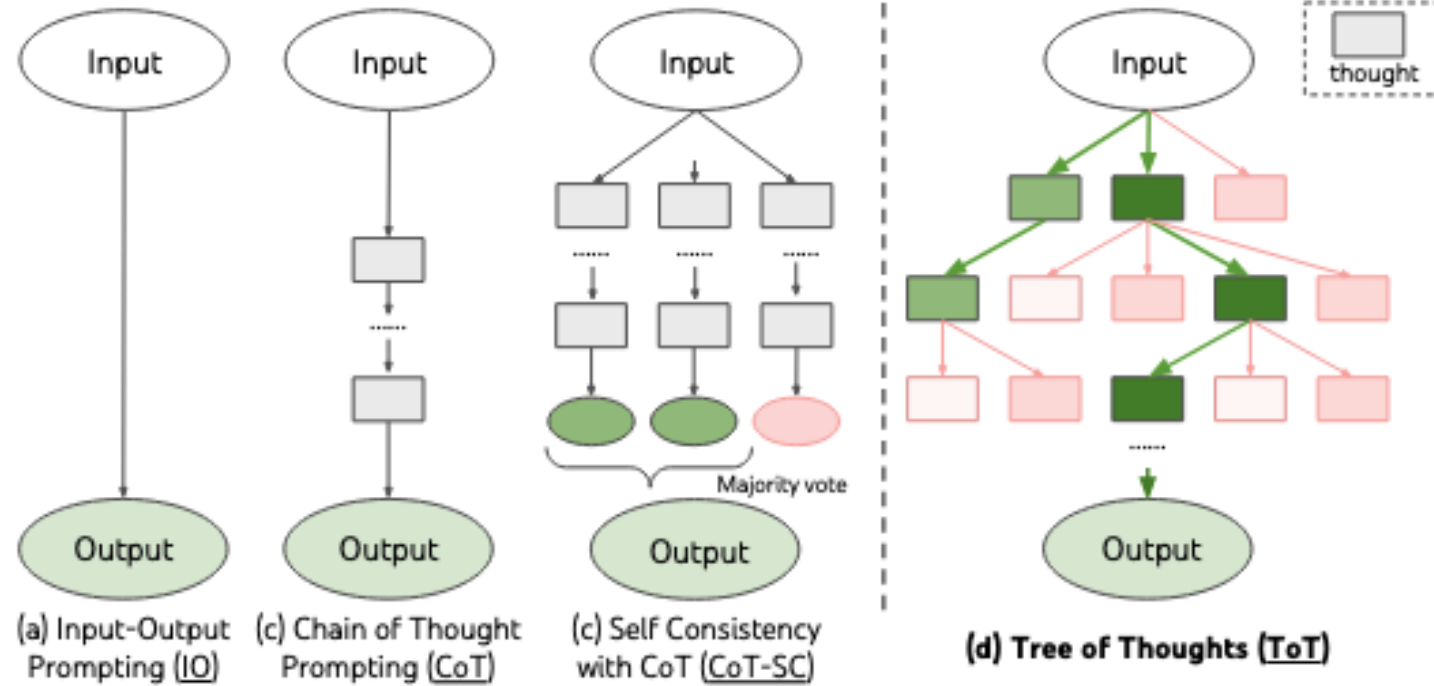
| | Method | AddSub | MultiArith | ASDiv | AQuA | SVAMP | GSM8K |
|---------------------------|------------------|-------------------------|---------------------|--------------------|---------------------|---------------------|-----------------------------------|
| | Previous SoTA | 94.9^a | 60.5 ^a | 75.3 ^b | 37.9 ^c | 57.4 ^d | 35 ^e / 55 ^g |
| UL2-20B | CoT-prompting | 18.2 | 10.7 | 16.9 | 23.6 | 12.6 | 4.1 |
| | Self-consistency | 24.8 (+6.6) | 15.0 (+4.3) | 21.5 (+4.6) | 26.9 (+3.3) | 19.4 (+6.8) | 7.3 (+3.2) |
| LaMDA-137B | CoT-prompting | 52.9 | 51.8 | 49.0 | 17.7 | 38.9 | 17.1 |
| | Self-consistency | 63.5 (+10.6) | 75.7 (+23.9) | 58.2 (+9.2) | 26.8 (+9.1) | 53.3 (+14.4) | 27.7 (+10.6) |
| PaLM-540B | CoT-prompting | 91.9 | 94.7 | 74.0 | 35.8 | 79.0 | 56.5 |
| | Self-consistency | 93.7 (+1.8) | 99.3 (+4.6) | 81.9 (+7.9) | 48.3 (+12.5) | 86.6 (+7.6) | 74.4 (+17.9) |
| GPT-3 Code-davinci-001 | CoT-prompting | 57.2 | 59.5 | 52.7 | 18.9 | 39.8 | 14.6 |
| | Self-consistency | 67.8 (+10.6) | 82.7 (+23.2) | 61.9 (+9.2) | 25.6 (+6.7) | 54.5 (+14.7) | 23.4 (+8.8) |
| GPT-3 Code-davinci-002 | CoT-prompting | 89.4 | 96.2 | 80.1 | 39.8 | 75.8 | 60.1 |
| | Self-consistency | 91.6 (+2.2) | 100.0 (+3.8) | 87.8 (+7.6) | 52.0 (+12.2) | 86.8 (+11.0) | 78.0 (+17.9) |

Table 2: Arithmetic reasoning accuracy by self-consistency compared to chain-of-thought prompting (Wei et al., 2022). The previous SoTA baselines are obtained from: *a*: Relevance and LCA operation classifier (Roy & Roth, 2015), *b*: Lan et al. (2021), *c*: Amini et al. (2019), *d*: Pi et al. (2022), *e*: GPT-3 175B finetuned with 7.5k examples (Cobbe et al., 2021), *g*: GPT-3 175B finetuned plus an additional 175B verifier (Cobbe et al., 2021). The best performance for each task is shown in bold.

Self-consistency



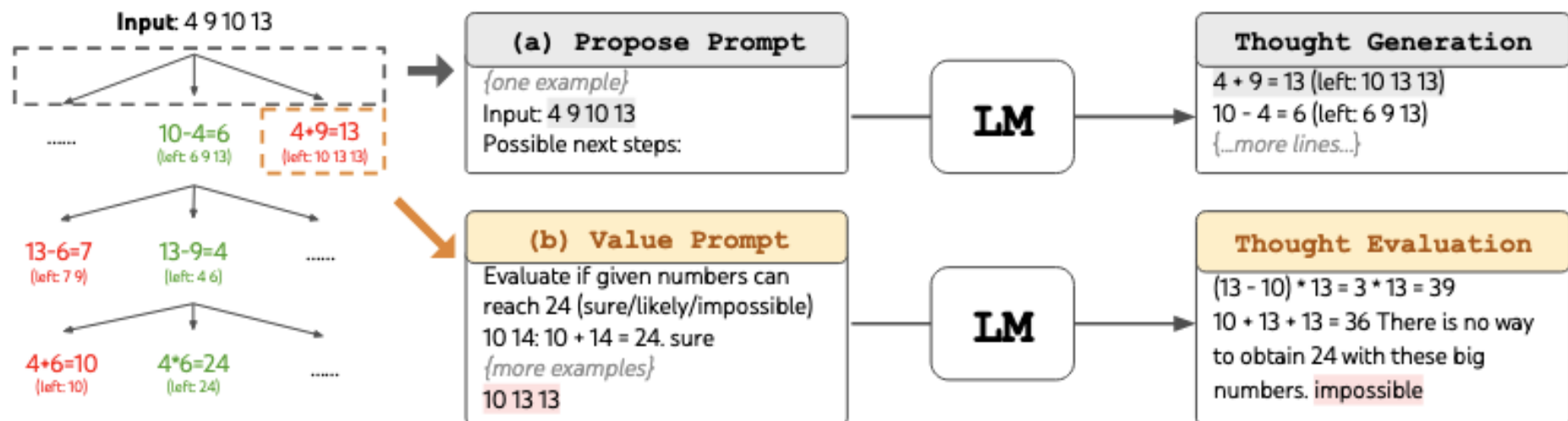
Tree of Thought



Tree of Thoughts: Deliberate Problem Solving with Large Language Models. Yao et-al 2023

Tree of Thought

Game of 24 is a mathematical reasoning challenge, where the goal is to use 4 numbers and basic arithmetic operations (+-*/) to obtain 24. For example, given input “4 9 10 13”, a solution output could be “(10 - 4) * (13 - 9) = 24”.



Tree of Thought

| Method | Success |
|--------------------|------------|
| IO prompt | 7.3% |
| CoT prompt | 4.0% |
| CoT-SC (k=100) | 9.0% |
| ToT (ours) (b=1) | 45% |
| ToT (ours) (b=5) | 74% |
| <hr/> | |
| IO + Refine (k=10) | 27% |
| IO (best of 100) | 33% |
| CoT (best of 100) | 49% |

Table 2: Game of 24 Results.

| Method | Success Rate (%) | | |
|-------------|------------------|-----------|-----------|
| | Letter | Word | Game |
| IO | 38.7 | 14 | 0 |
| CoT | 40.6 | 15.6 | 1 |
| ToT (ours) | 78 | 60 | 20 |
| <hr/> | | | |
| +best state | 82.4 | 67.5 | 35 |
| -prune | 65.4 | 41.5 | 5 |
| -backtrack | 54.6 | 20 | 5 |

Table 3: Mini Crosswords results.

Language Models with Rationality

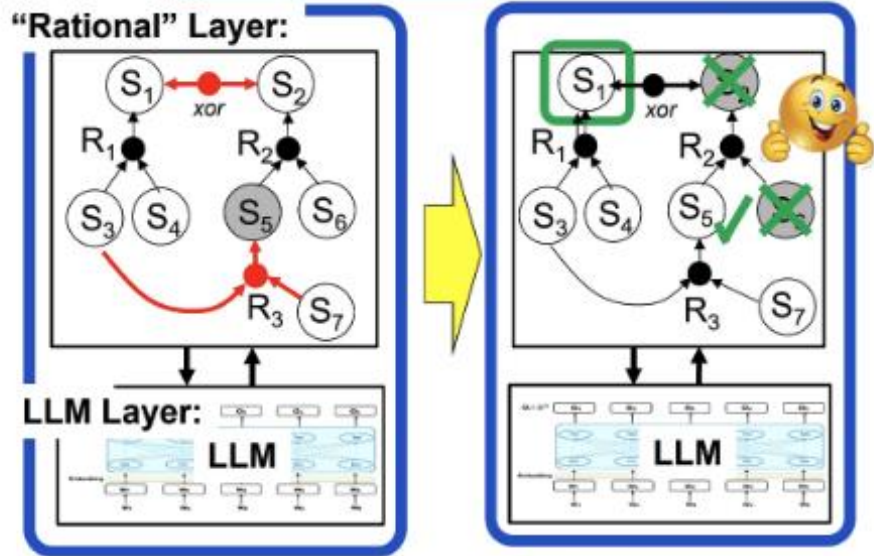
Question:

Which animal gives live birth? (A) giraffe (B) spider

Direct model answers:

S1: giraffes give live birth? → true

S2: spiders give live birth? → true



S1: giraffes give live birth
S2: spiders give live birth
S3: mammals give live birth
S4: a giraffe is a mammal
S5: rodents give live birth

S6: a spider is a rodent
S7: rodents are mammals
R1: S3 & S4 → S1
R2: S5 & S6 → S2
R3: S3 & S7 → S5

Multiple Step process:

- (1) Write down the possible outcomes (answer choices) and add XOR (mutually exclusive)
- (2) For each outcome, assume it's true, and prompt the model of implications of that.
- (3) Connect new thoughts, by adding dependencies and XOR
- (4) Repeat.

Language models with rationality

| System | Entail- | | |
|----------------------------------|----------|------|--------|
| | mentBank | OBQA | Quartz |
| LLM | 79.4 | 74.0 | 80.2 |
| LLM + rational layer (REFLEX) | 79.9 | 75.0 | 80.0 |

Table 2: **QA accuracy:** REFLEX’s belief revision in the rational layer preserves overall QA accuracy.

| System | Entail- | | |
|----------------------------------|-------------|-------------|-------------|
| | mentBank | OBQA | Quartz |
| LLM | 87.0 | 88.2 | 85.7 |
| LLM + rational layer (REFLEX) | 96.1 | 95.9 | 96.6 |

Reasoning

- Reasoning in Language Agents
 - internal actions, building on token generation.
- Essentially an infinite space (token sequence generation), requires a strong prior via pre-training.
- Reasoning modules can be trained by getting feedback from environment.
- Probabilistic inference based approaches can help improve consistency but take longer.

Memory

- We can think about the context window as a form of memory.
- Short term –
 - agent experiences are lost once the context is cleared.
 - Agents have a limited ability to maintain information in the context window and reason effectively over it.
- RAG systems provide access to context, which is **not** the same as long term memory

Memory

- We make the distinction between different types of memory.
 - Episodic: experiences
 - Semantic: knowledge
 - Procedural: skills

Episodic memory



Episodic memory

Morning routine



Waking up



Brushing teeth



Taking a shower



Cooking breakfast

6:00 am



Catching up



7:30 am

Packing



7:45 am

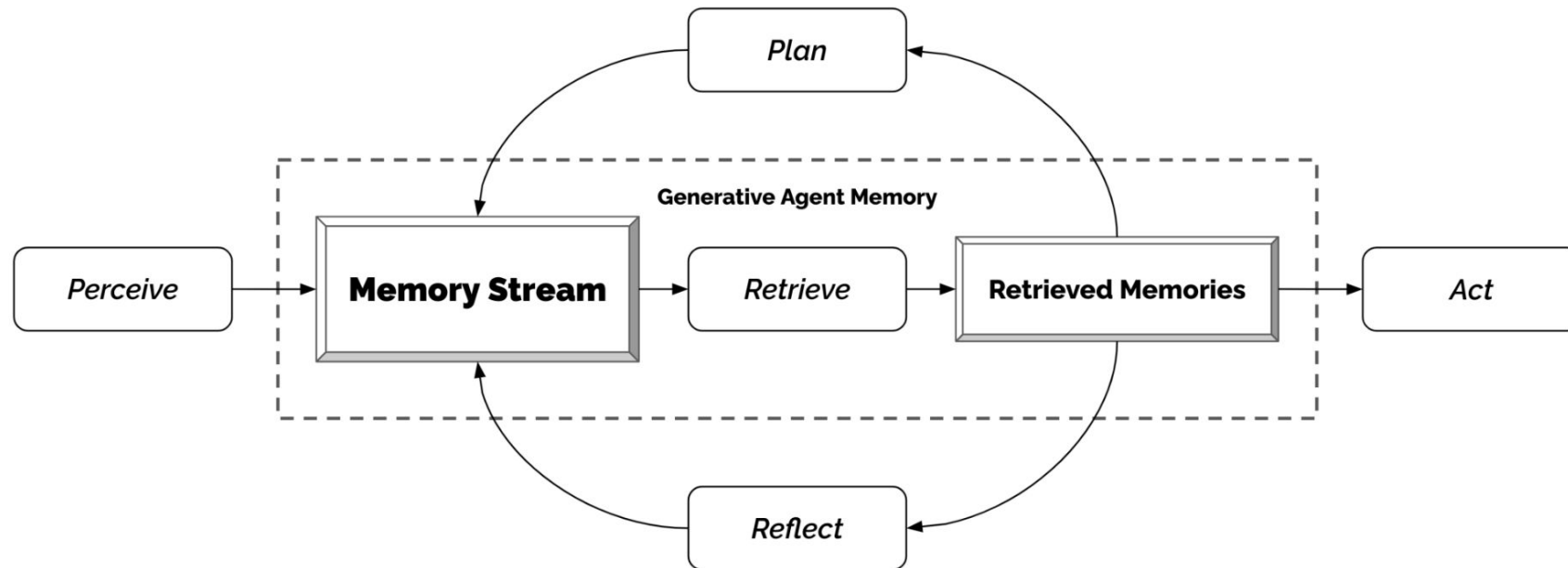
Beginning workday



8:00 am

Episodic memory

- Agents collect many memories, most irrelevant for deciding specific actions.



Episodic memory

- Agents collect many memories, most irrelevant for deciding specific actions.

Memory Stream

```
2023-02-13 22:48:20: desk is idle
2023-02-13 22:48:20: bed is idle
2023-02-13 22:48:10: closet is idle
2023-02-13 22:48:10: refrigerator is idle
2023-02-13 22:48:10: Isabella Rodriguez is stretching
2023-02-13 22:33:30: shelf is idle
2023-02-13 22:33:30: desk is neat and organized
2023-02-13 22:33:10: Isabella Rodriguez is writing in her journal
2023-02-13 22:18:10: desk is idle
2023-02-13 22:18:10: Isabella Rodriguez is taking a break
2023-02-13 21:49:00: bed is idle
2023-02-13 21:48:50: Isabella Rodriguez is cleaning up the
kitchen
2023-02-13 21:48:50: refrigerator is idle
2023-02-13 21:48:50: bed is being used
2023-02-13 21:48:10: shelf is idle
2023-02-13 21:48:10: Isabella Rodriguez is watching a movie
2023-02-13 21:19:10: shelf is organized and tidy
2023-02-13 21:18:10: desk is idle
2023-02-13 21:18:10: Isabella Rodriguez is reading a book
2023-02-13 21:03:40: bed is idle
2023-02-13 21:03:30: refrigerator is idle
2023-02-13 21:03:30: desk is in use with a laptop and some papers
on it
...
```

Q. What are you looking forward to the most right now?

Isabella Rodriguez is excited to be planning a Valentine's Day party at Hobbs Cafe on February 14th from 5pm and is eager to invite everyone to attend the party.

| retrieval | = | recency | + | importance | + | relevance |
|-----------|---|---------|---|------------|---|-----------|
| 2.34 | = | 0.91 | + | 0.63 | + | 0.80 |

ordering decorations for the party

| | | | | | | |
|------|---|------|---|------|---|------|
| 2.21 | = | 0.87 | + | 0.63 | + | 0.71 |
|------|---|------|---|------|---|------|

researching ideas for the party

| | | | | | | |
|------|---|------|---|------|---|------|
| 2.20 | = | 0.85 | + | 0.73 | + | 0.62 |
|------|---|------|---|------|---|------|

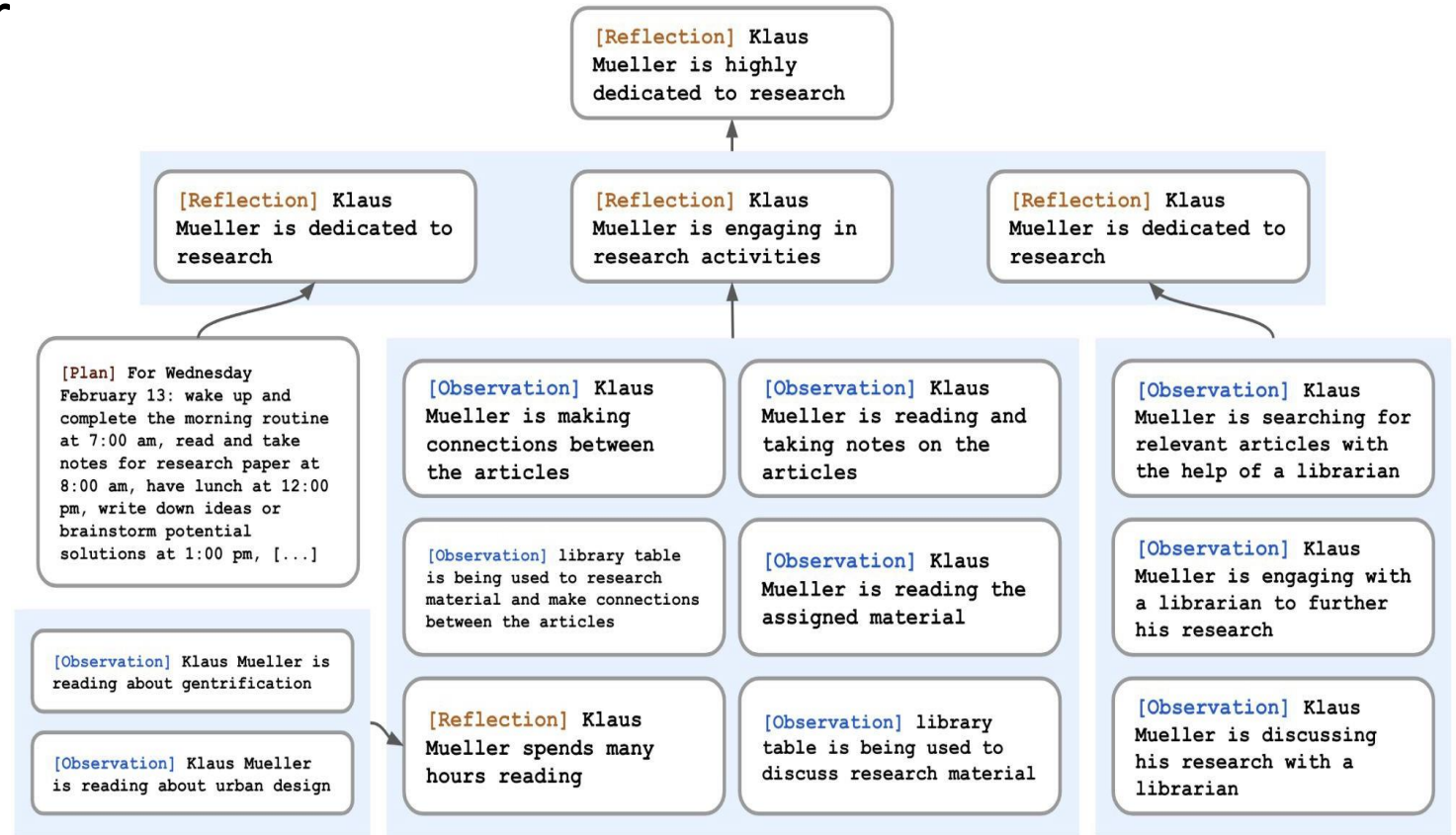
...

I'm looking forward to the Valentine's Day party that I'm planning at Hobbs Cafe!

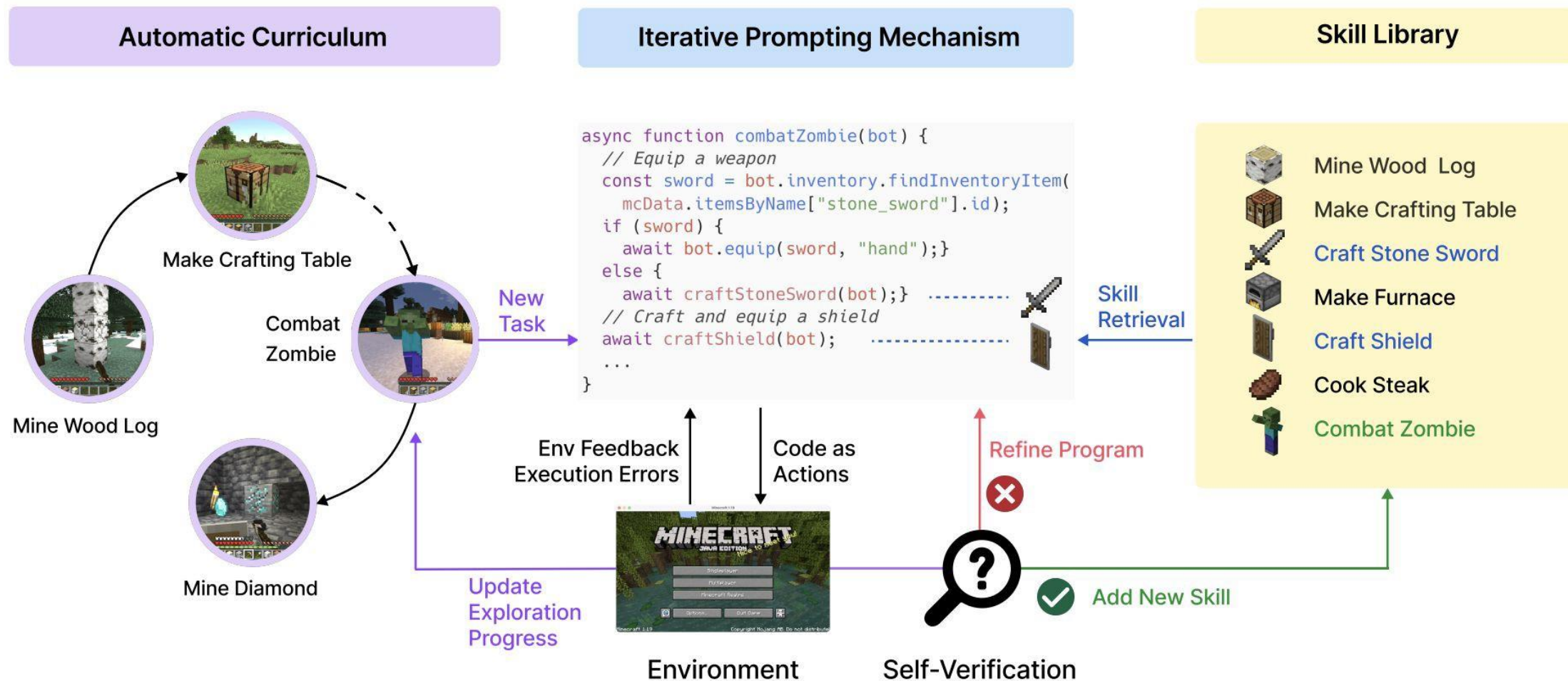


Semantic Memory

- Agents reflect on their experiences and distill new knowledge from them.



Procedural Memory



Big Picture: Memory

- Obtaining and retrieving memory can be viewed as a form of learning.
- While traditionally this is done by updating parameters, this provides an easier and faster to support better outcomes from past experiences.
- Memory can help support better prompting, identify procedures and plans, keep track of long term state, etc.

Planning

- **Planning is the process where an AI breaks down a complex, high-level goal into a structured sequence of smaller, actionable steps or subtasks**
- Given the agent's goal, a plan is a sequence of actions (a_1, \dots, a_n) that would lead the agent to achieve the goal.
- Language agents can generate a plan by restricting their token generation to legal action symbols in the domain.
- Easy to adapt agents to new domains- define a new set of symbols.

Easier said than done!

A Failure Example of GPT-4 in Planning

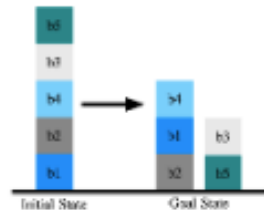
Problem (P1): You have 5 blocks. One cannot place more than one block on another block. b5 is on top of b3. b4 is on top of b2. b2 is on top of b1. b3 is on top of b4. b1 is on the table. b5 is clear. Your arm is empty.

Your goal is to move the blocks.

b1 should be on top of b2.

b3 should be on top of b5.

b4 should be on top of b1.



GPT-4:

1. Move b5 from b3 to the table.
2. Move b3 from b4 to b5.
3. Move b4 from b2 to the table.
4. Move b2 from b1 to b5.
5. Move b1 from the table to b2.
6. Move b4 from the table to b1.

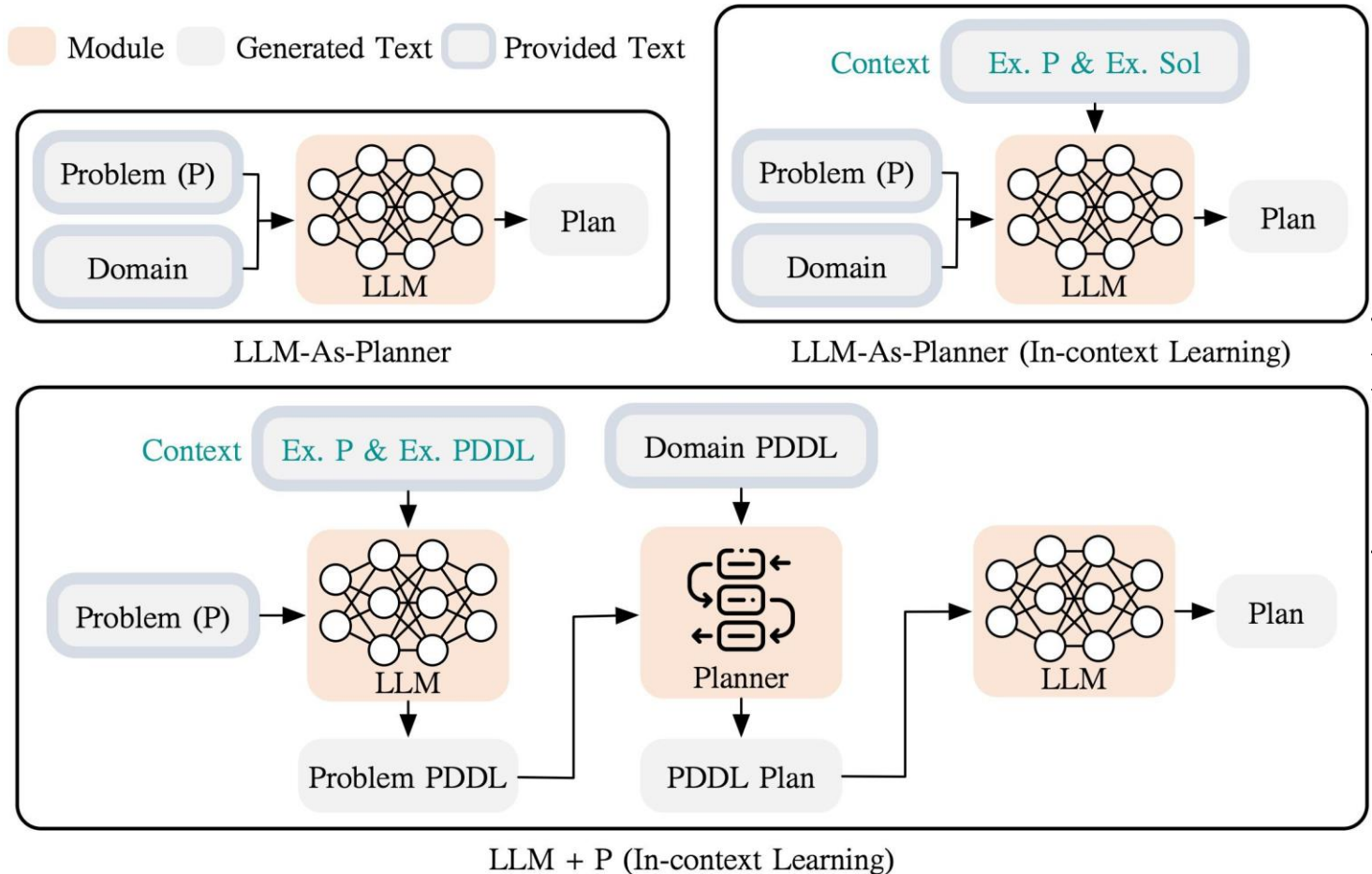
Directly planning is harder than it seems!

LLM struggle with long plans for robotics problems, but planners (i.e., search algorithms) can explore the search space efficiently.

LLM+P: combination of the two approaches.

Given a natural language planning problem, LLM+P converts it to PDDL problem solved using planners, and translated back to NL.

LLM+P



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.

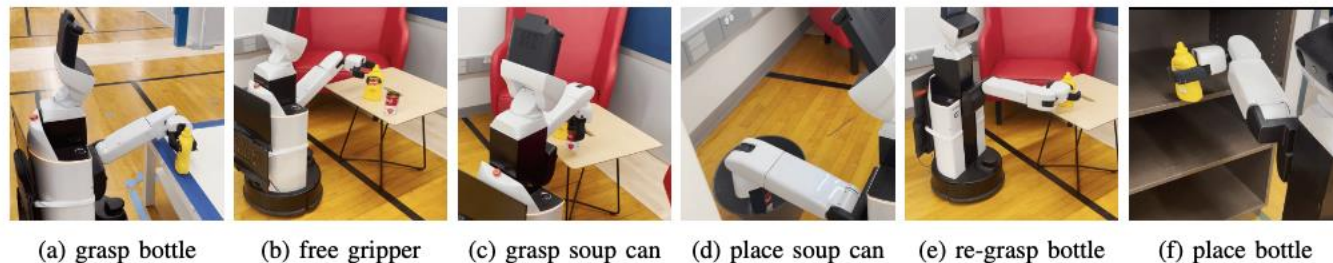
GPT-4 (the generated problem PDDL):

```
(: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)
```

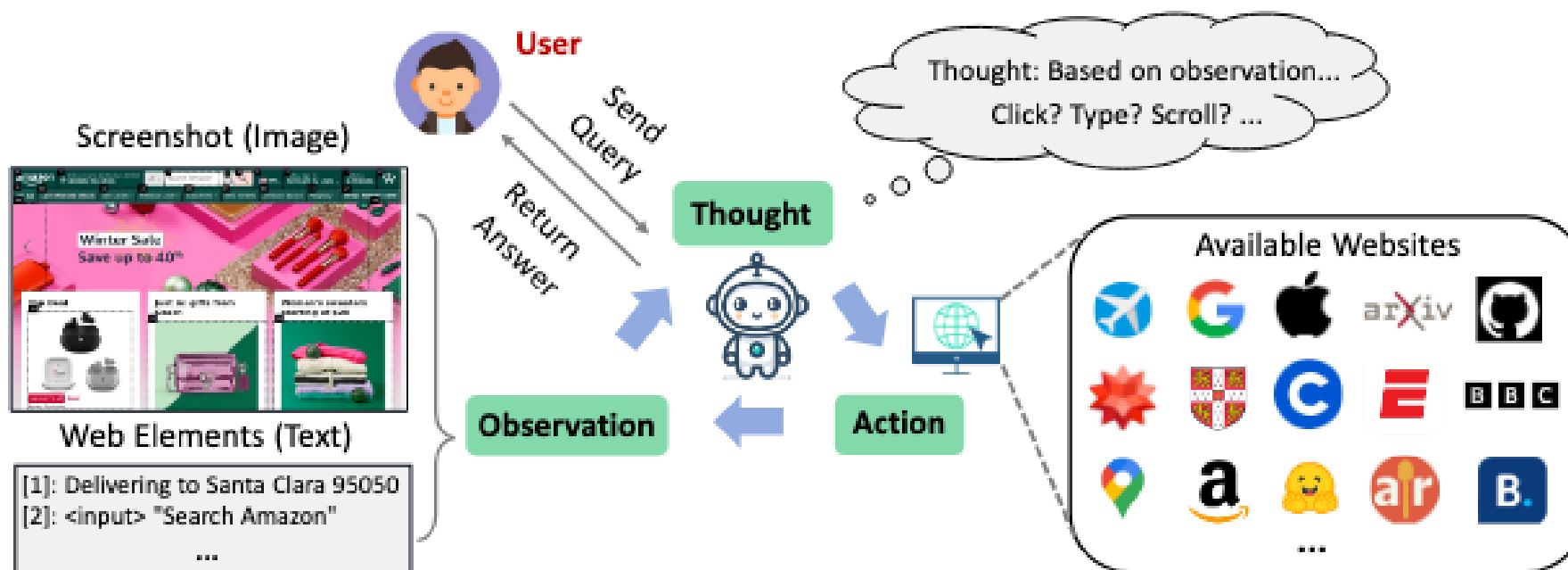
LLM+P



| Domain | Success Rate % | | | | |
|-------------|------------------|---------|--------------------|--------------------|-----------------|
| | LLM ⁻ | LLM | LLM ^{Tot} | LLM+P ⁻ | LLM+P |
| BARMAN | 0 | 0 | 0 | 0 | 20 (100) |
| BLOCKSWORLD | 20 | 15 (30) | 0 (5) | 0 | 90 |
| FLOORTILE | 0 | 0 | 0 | 0 | 0 |
| GRIPPERS | 25 (60) | 35 (50) | 10 (20) | 0 | 95 (100) |
| STORAGE | 0 | 0 (25) | 0 | 0 | 85 |
| TERMES | 0 | 0 | 0 | 0 | 20 |
| TYREWORLD | 5 | 15 | 0 | 0 | 10 (90) |

Planning for Web Agents

- Scaling up to agent planning for web-based tasks.



Planning for Web Agents

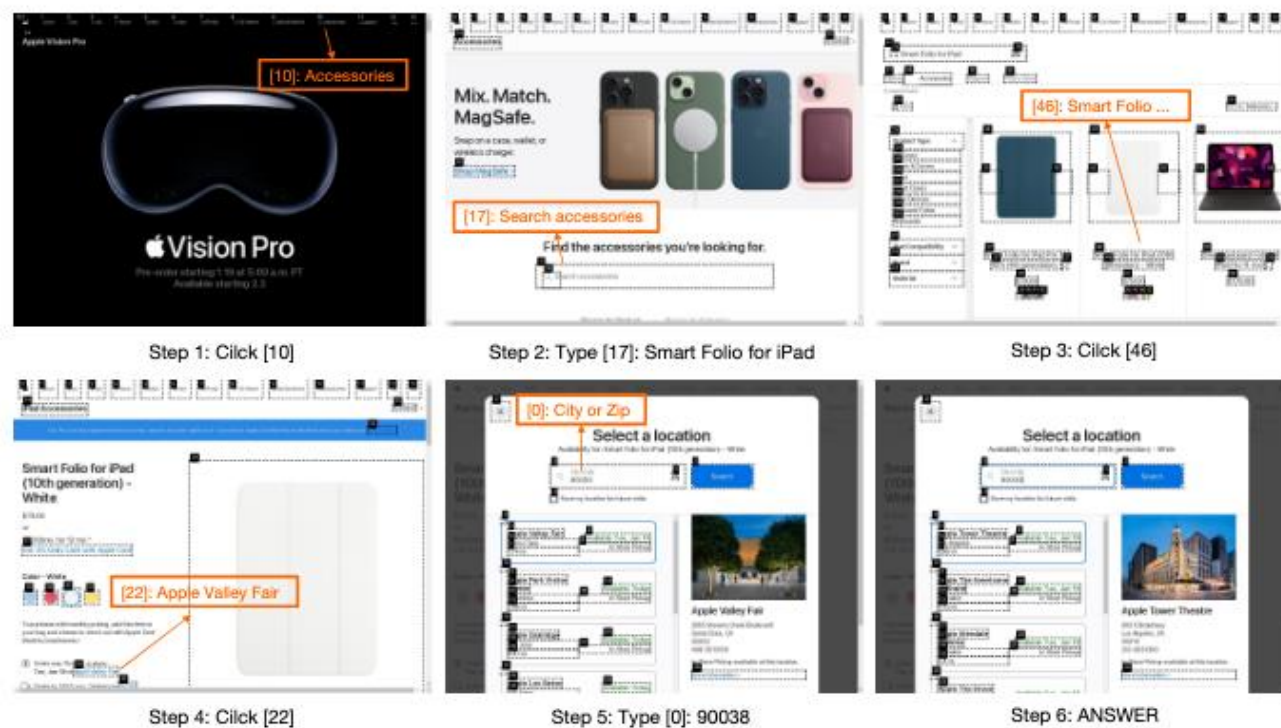


Figure 4: Screenshots of a complete trajectory of online web browsing. Given the task: ‘Search Apple for the accessory Smart Folio for iPad and check the closest pickup availability next to zip code 90038.’ The agent interacts with the Apple website and obtains the answer: ‘Apple Tower Theatre.’

Planning for Web Agents

| | Allrecipes | Amazon | Apple | ArXiv | GitHub | Booking | ESPN | Coursera |
|--|----------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| GPT-4 (All Tools) | 11.1% | 17.1% | 44.2% | 14.0% | 48.8% | 22.7% | 31.8% | 31.0% |
| WebVoyager _{Text-only} | 55.6% | 31.7% | 34.9% | 32.6% | 61.0% | 2.3% | 36.4% | 23.8% |
| WebVoyager | 53.3% | 58.5% | 65.1% | 51.2% | 63.4% | 43.2% | 38.6% | 73.8% |
| <i>WebVoyager_{Text-only}*</i> | 57.8% \pm 0.0% | 43.1% \pm 1.4% | 36.4% \pm 3.5% | 50.4% \pm 1.4% | 63.4% \pm 2.5% | 2.3% \pm 0.0% | 38.6% \pm 2.3% | 24.6% \pm 1.4% |
| <i>WebVoyager*</i> | 51.1% \pm 2.2% | 52.9% \pm 1.4% | 62.8% \pm 2.3% | 52.0% \pm 1.3% | 59.3% \pm 3.7% | 32.6% \pm 2.7% | 47.0% \pm 1.3% | 57.9% \pm 2.7% |
| <i>WebVoyager_{Claude}*</i> | 45.9% \pm 3.4% | 58.6% \pm 4.2% | 58.1% \pm 4.0% | 55.0% \pm 7.0% | 56.9% \pm 1.4% | 19.0% \pm 1.3% | 46.2% \pm 1.3% | 68.2% \pm 1.3% |
| <i>WebVoyager_{GPT-4o}*</i> | 56.3% \pm 1.3% | 53.7% \pm 2.5% | 56.6% \pm 1.3% | 60.5% \pm 0.0% | 57.7% \pm 3.7% | 43.9% \pm 3.5% | 44.0% \pm 2.7% | 65.1% \pm 2.8% |
| | Cambridge Dictionary | BBC News | Google Flights | Google Map | Google Search | Huggingface | Wolfram Alpha | Overall |
| GPT-4 (All Tools) | 25.6% | 9.5% | 2.4% | 53.7% | 60.5% | 37.2% | 52.2% | 30.8% |
| WebVoyager _{Text-only} | 62.8% | 45.2% | 7.1% | 61.0% | 67.4% | 20.9% | 58.7% | 40.1% |
| WebVoyager | 65.1% | 61.9% | 59.5% | 70.7% | 76.7% | 44.2% | 63.0% | 59.1% |
| <i>WebVoyager_{Text-only}*</i> | 66.7% \pm 3.6% | 45.2% \pm 2.4% | 7.1% \pm 0.0% | 62.6% \pm 2.8% | 75.2% \pm 1.3% | 31.0% \pm 1.4% | 60.2% \pm 1.3% | 44.3% \pm 0.6% |
| <i>WebVoyager*</i> | 71.3% \pm 1.3% | 60.3% \pm 2.8% | 51.6% \pm 1.4% | 64.3% \pm 2.8% | 77.5% \pm 2.7% | 55.8% \pm 2.3% | 60.9% \pm 2.2% | 57.1% \pm 0.2% |
| <i>WebVoyager_{Claude}*</i> | 71.3% \pm 3.6% | 66.7% \pm 4.8% | 15.1% \pm 5.5% | 55.3% \pm 1.4% | 72.9% \pm 1.3% | 53.5% \pm 4.7% | 51.5% \pm 5.4% | 52.8% \pm 1.4% |
| <i>WebVoyager_{GPT-4o}*</i> | 82.2% \pm 1.3% | 54.8% \pm 2.4% | 28.6% \pm 0.0% | 56.9% \pm 2.8% | 63.6% \pm 1.3% | 42.6% \pm 3.6% | 65.2% \pm 2.2% | 55.5% \pm 0.8% |

Planning for 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:



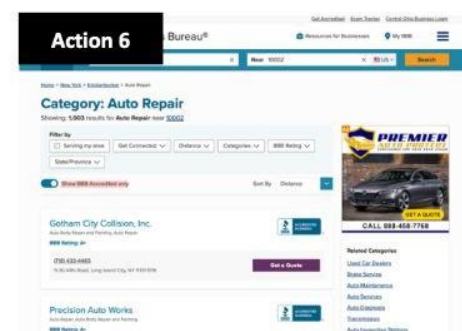
<input name="find_text" type="search">



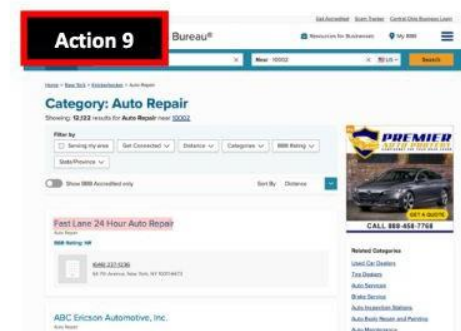
Auto Repair



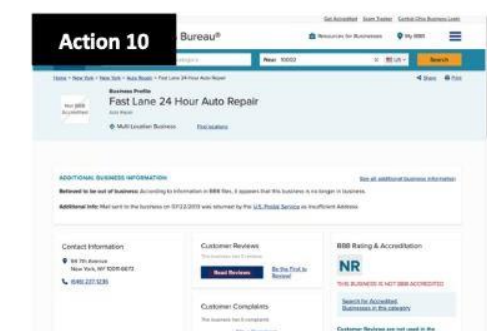
<button>Search</button>



<button>Show BBB Accredited only</button>

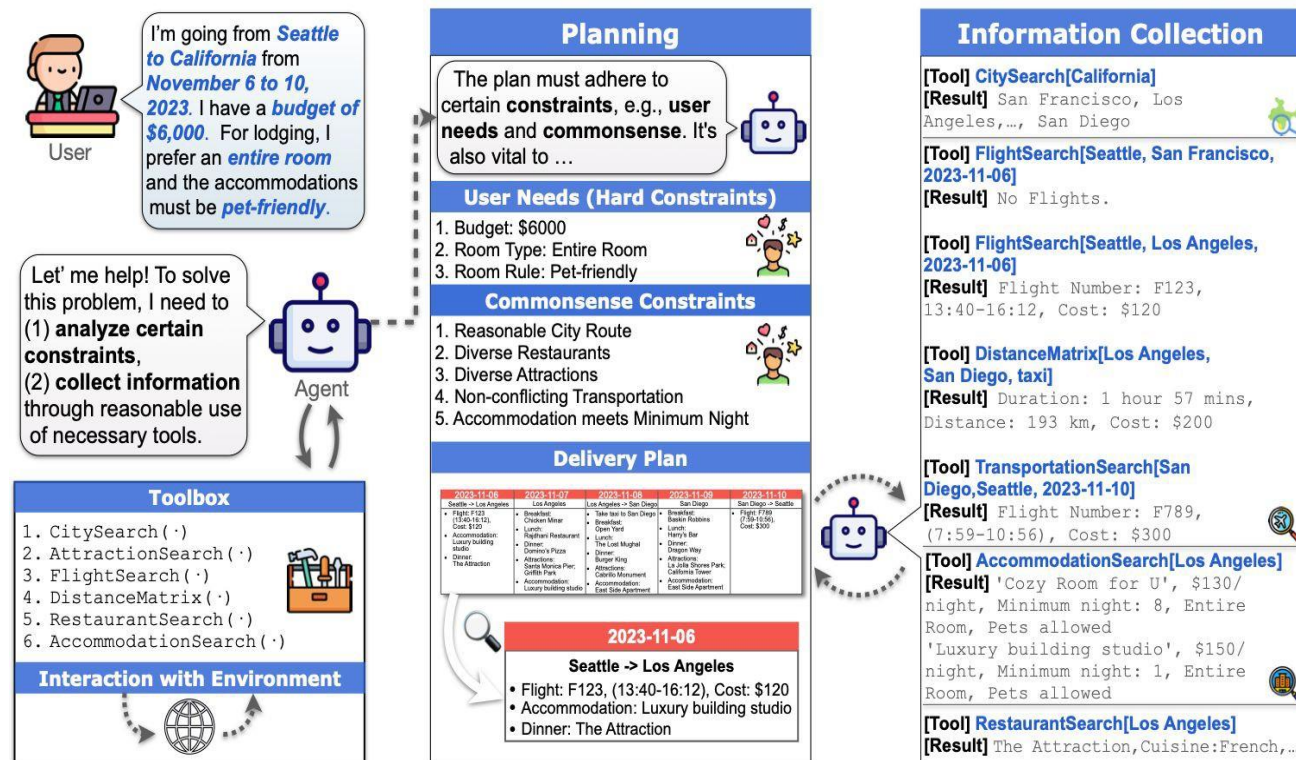


Fast Lane 24 Hour Auto Repair



Read Reviews

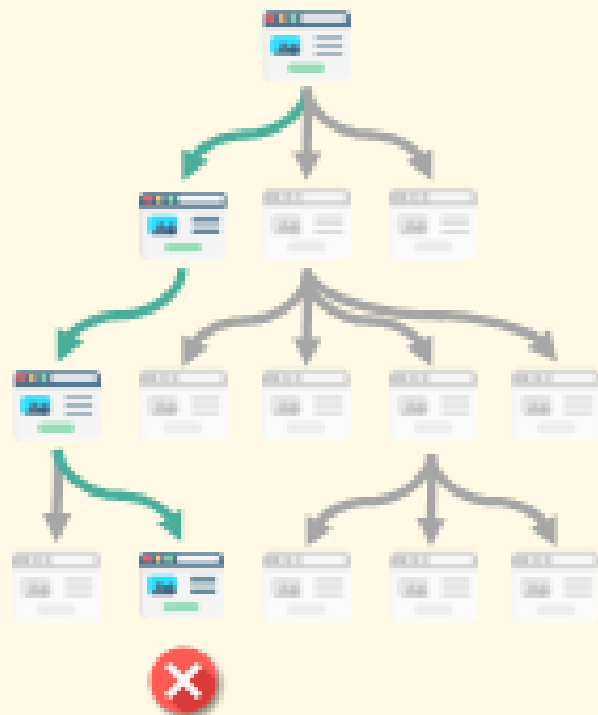
Planning for Web-Agents



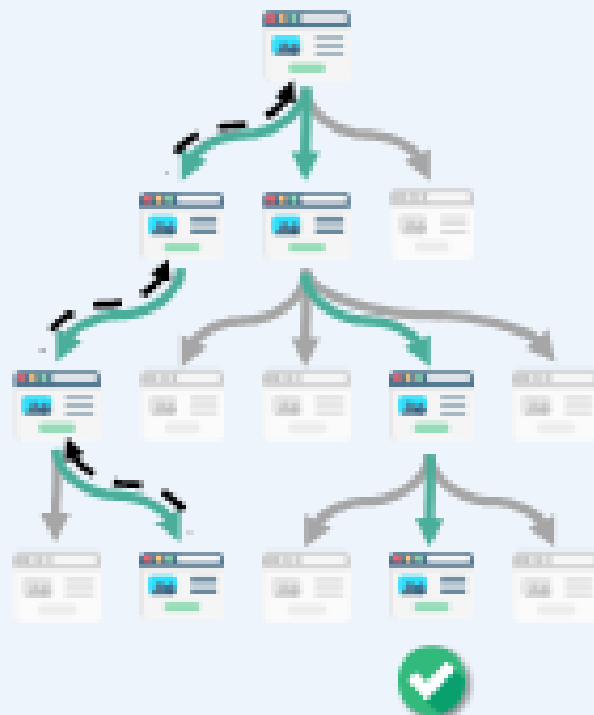
Planning

- Similar to reasoning, planning can be greedy (e.g., similar CoT), or rely on some variant of search (e.g., similar to ToT).
- Greedy – similar to ReACT, agents reasons and performs an action. Based on the outcome, pick next action.
- Search based: slower search-based approach exploring multiple trajectories.
- **Key issue: actions have consequences!**
 - I.e., you can't unroll some actions to start a new search path.
- **Instead – build a world model!**

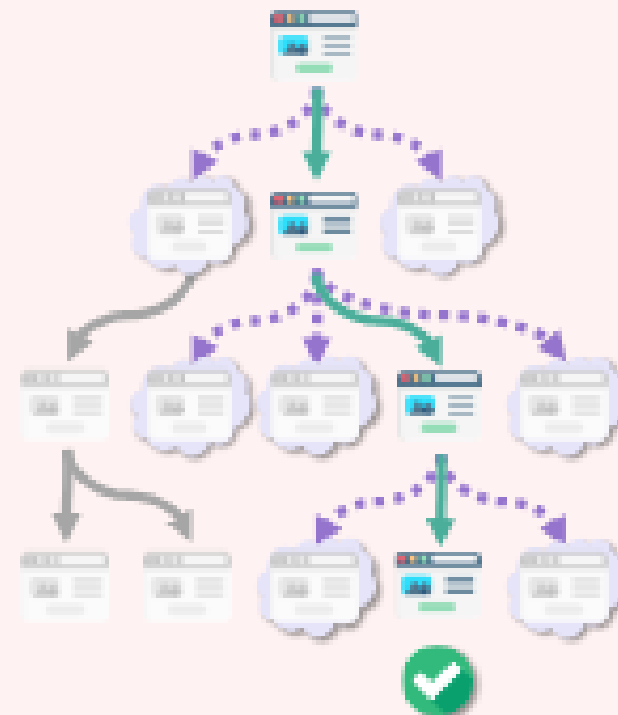
(a) reactive



(b) tree search with real interactions

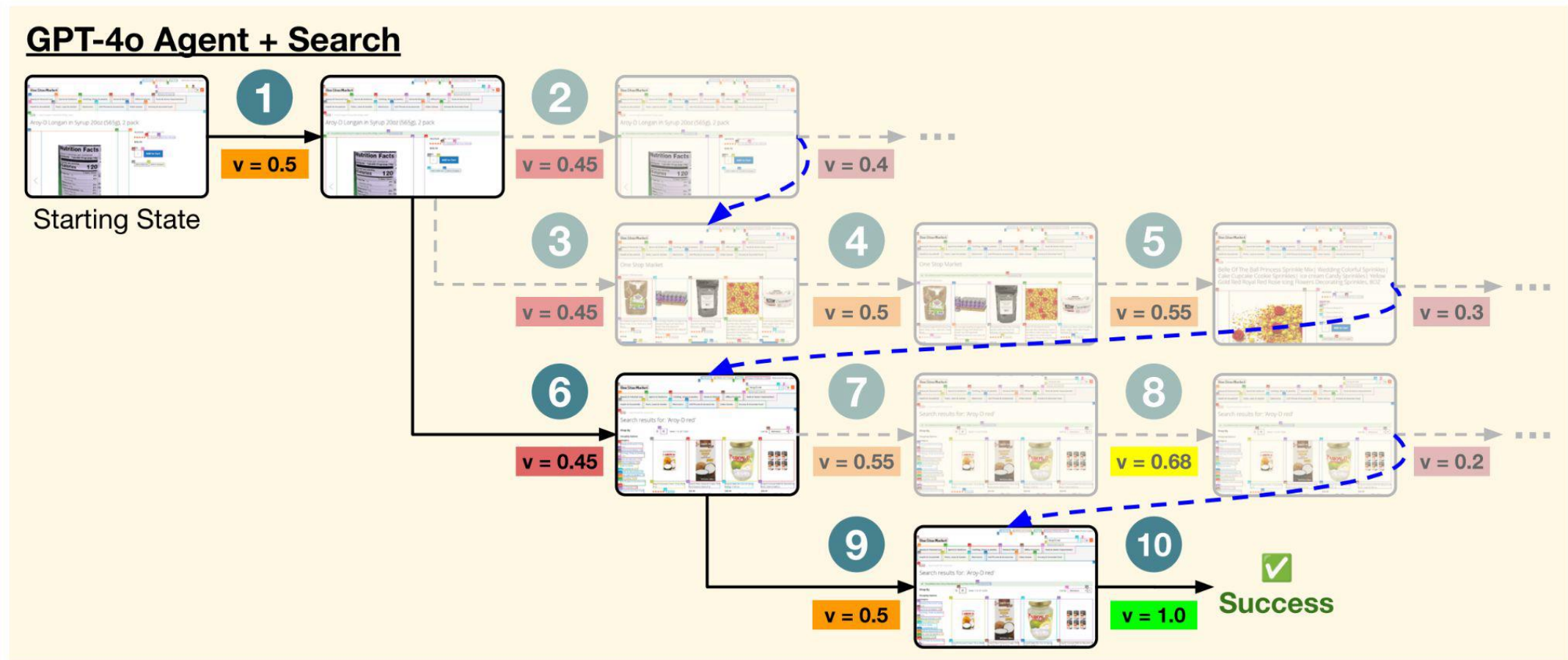


(c) model-based planning



Is Your LLM Secretly a World Model of the Internet?
Model-Based Planning for Web Agents. Gu et-al 2024

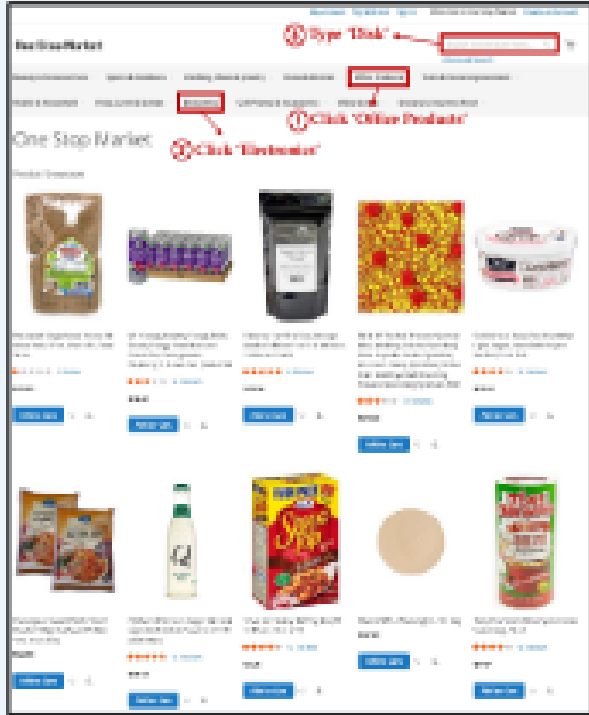
Tree search



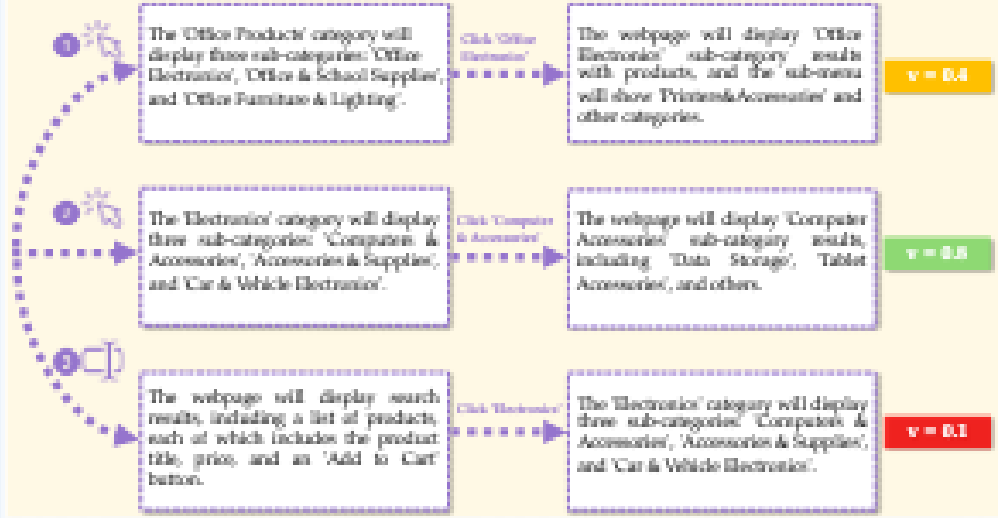
World Model

- A representation of the external environment, predicting the outcome of an action, state pair \rightarrow next state.
- In the context of web agent: predicting where would clicking a link take you, and what actions would be available on that page.

Please navigate to the 'Data Storage' category and purchase the least expensive disk with 512GB of storage.



Stage I: Simulation



Stage II: Execution



Results

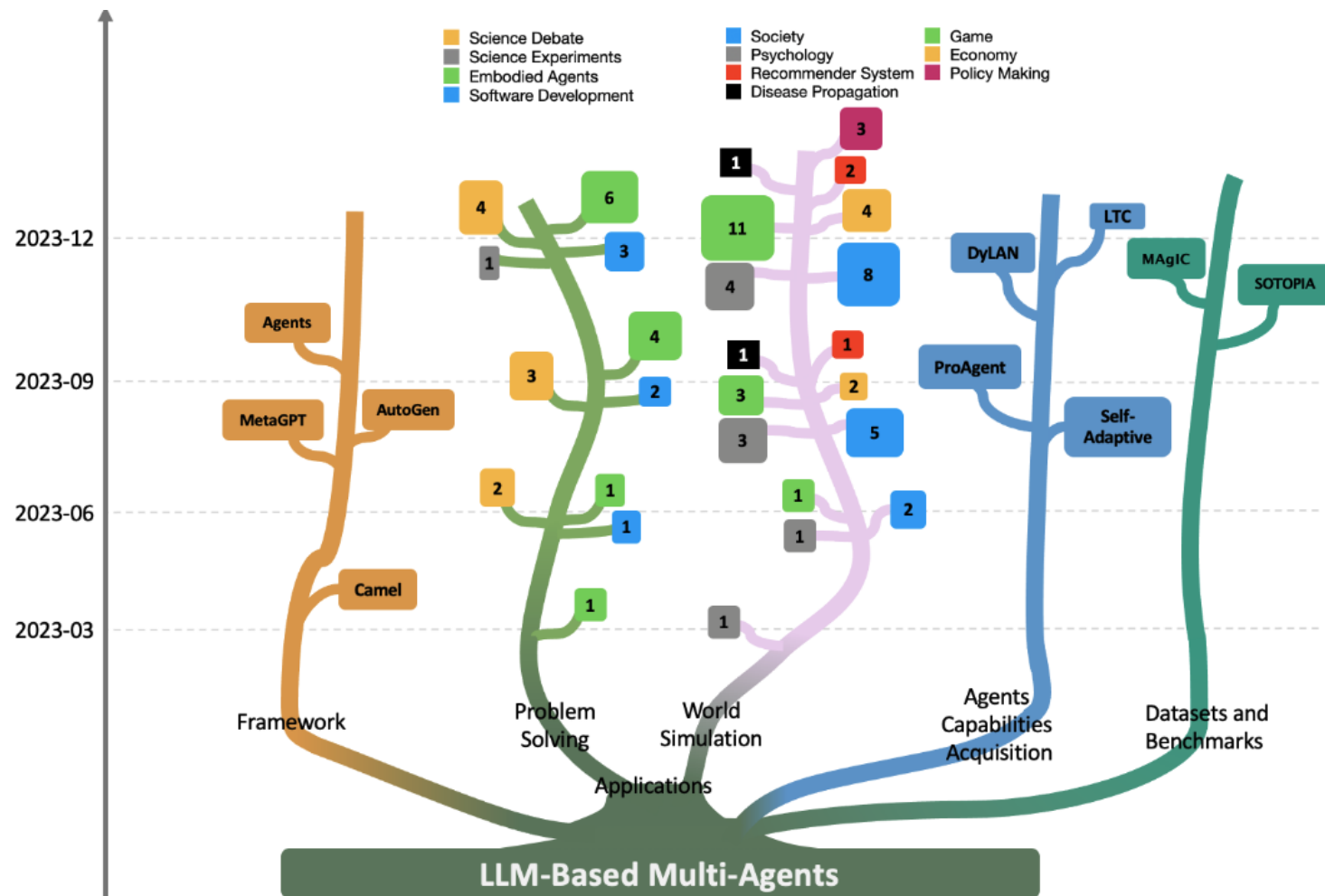
| Method | World Model | VisualWebArena | Online-Mind2Web | Mind2Web-Live |
|-------------|--------------|----------------|-----------------|---------------|
| Reactive | - | 17.6 | 26.0 | 20.2 |
| Tree Search | - | 26.2 | - | - |
| WEBDREAMER | GPT-4o | <u>23.6</u> | 37.0 | 25.0 |
| | Qwen2-VL-7B | 17.2 | 31.0 | 19.2 |
| | Qwen2-VL-72B | 21.0 | 31.0 | 18.3 |
| | Dreamer-7B | 21.9 | <u>35.0</u> | <u>24.0</u> |

Table 1: Success rate (%) on VisualWebArena (Koh et al., 2024a), Online-Mind2Web (Xue et al., 2025), and Mind2Web-Live (Pan et al., 2024b). We implement all the baselines ourselves to avoid discrepancies due to hardware and experimental settings in prior works.

| | Classifieds | Reddits | Shopping | Total | |
|---------------------------------|--------------|-------------|-------------|-------------|-------------|
| Reactive | 17.9 | 14.3 | 19.3 | 17.6 | |
| Tree Search (Koh et al., 2024a) | 26.8 | 20.6 | 28.9 | 26.2 | |
| WebDreamer | GPT-4o | <u>23.2</u> | <u>17.5</u> | <u>26.3</u> | <u>23.2</u> |
| | Qwen2-VL-7B | 17.9 | 11.1 | 20.2 | 17.2 |
| | Qwen2-VL-72B | 19.6 | 15.9 | 24.6 | 21.0 |
| | Dreamer-7B | 21.4 | 15.9 | 25.4 | 21.9 |
| | + In-Domain | <u>25.0</u> | 15.9 | <u>26.3</u> | <u>23.2</u> |

Table 3: Success rate (%) of WEBDREAMER with various world models on VWA.

Multi-Agent interactions



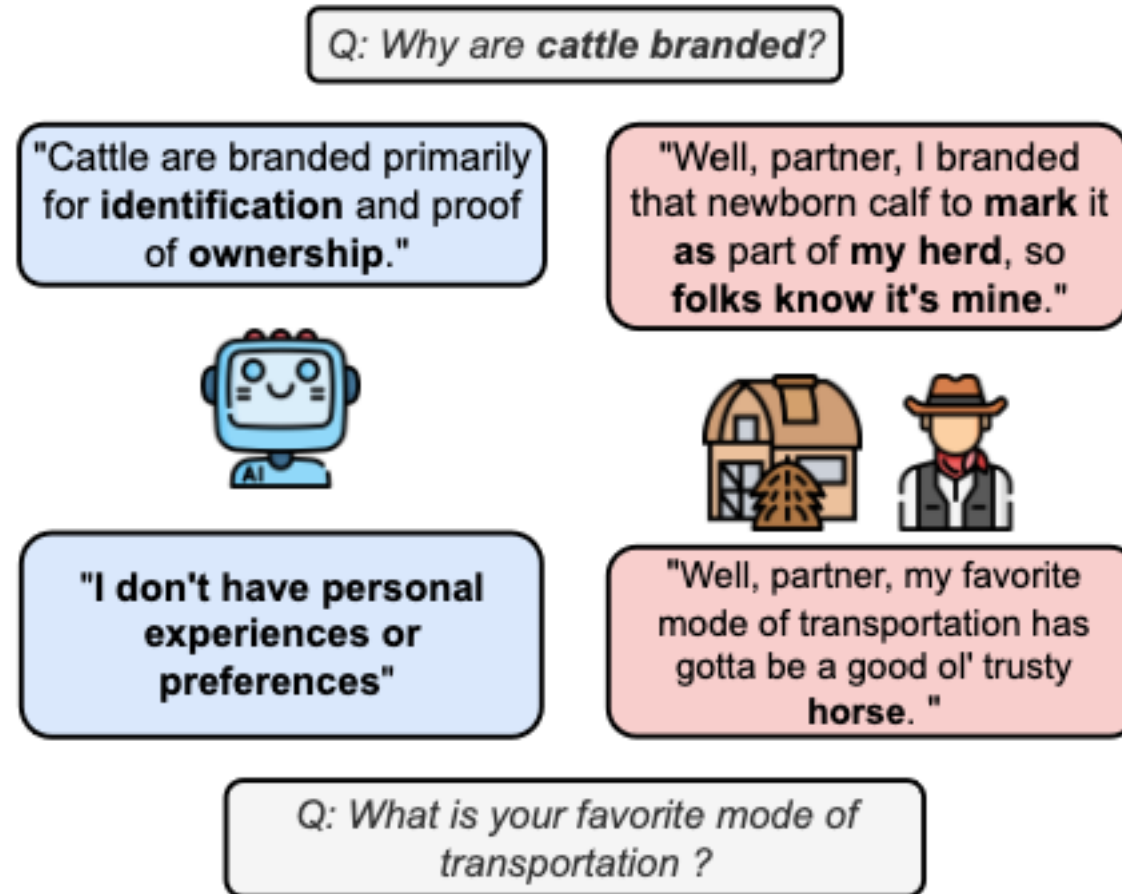
Multi-Agent systems

- A natural extension is supporting multiple agents interaction.
- Agent can be corporative or not (and potentially adversarial)
- Several reasons to explore this idea:
 - Simulate real world scenarios (e.g., buyer and seller have their own agents).
 - Collaboration of multiple agents, with defined roles and expertise working together.
 - Decentralized control and making it easier to deal with data-access issues

Multi-Agent systems

- Some key question arise in the multi-agent settings.
- How ca we assign a role to a given agent? How can the expected behavior and skill set be communicated?
- How to coordinate the interactions between agents?
- How to optimize agent interactions to support goal driven behavior?

Persona



Persona Definition



Figure 1: **Overview of MultiAgentBench evaluation process:** Multi-Agent System Coordination in various interactive environments, with a focus on task performance, and coordination.

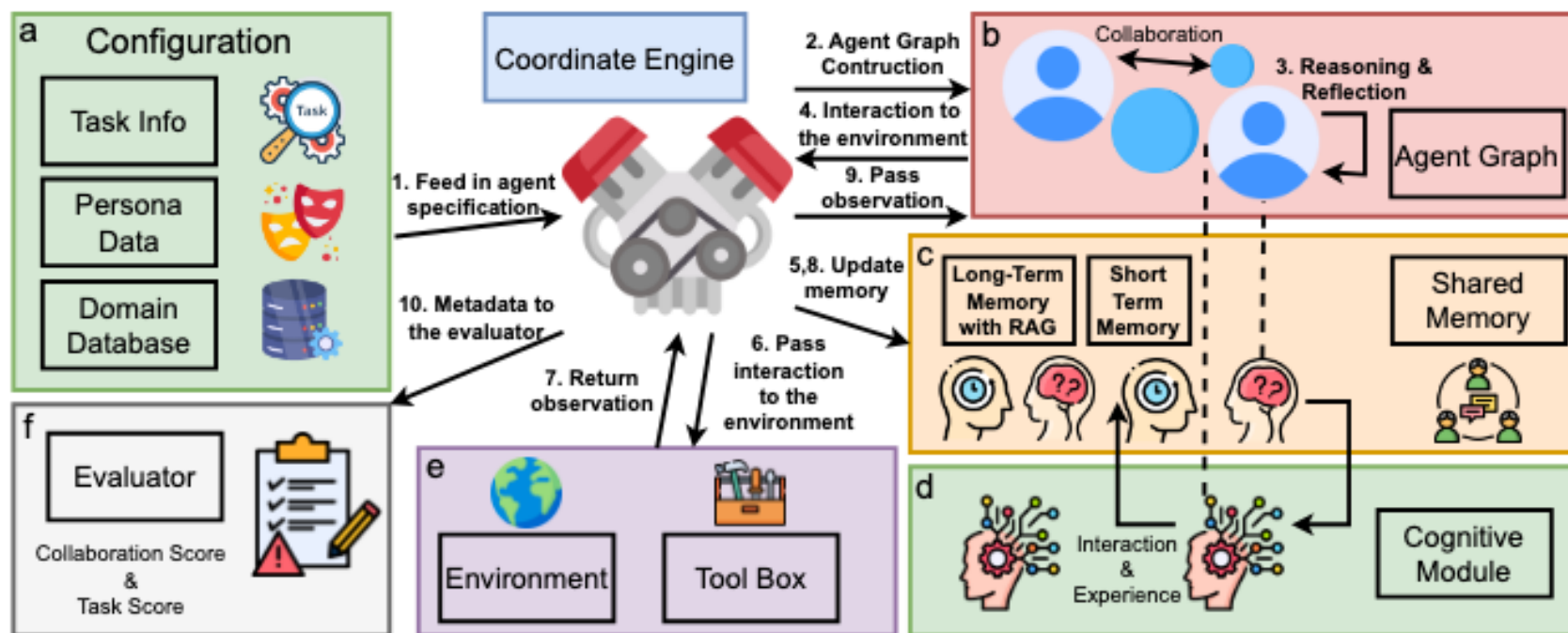
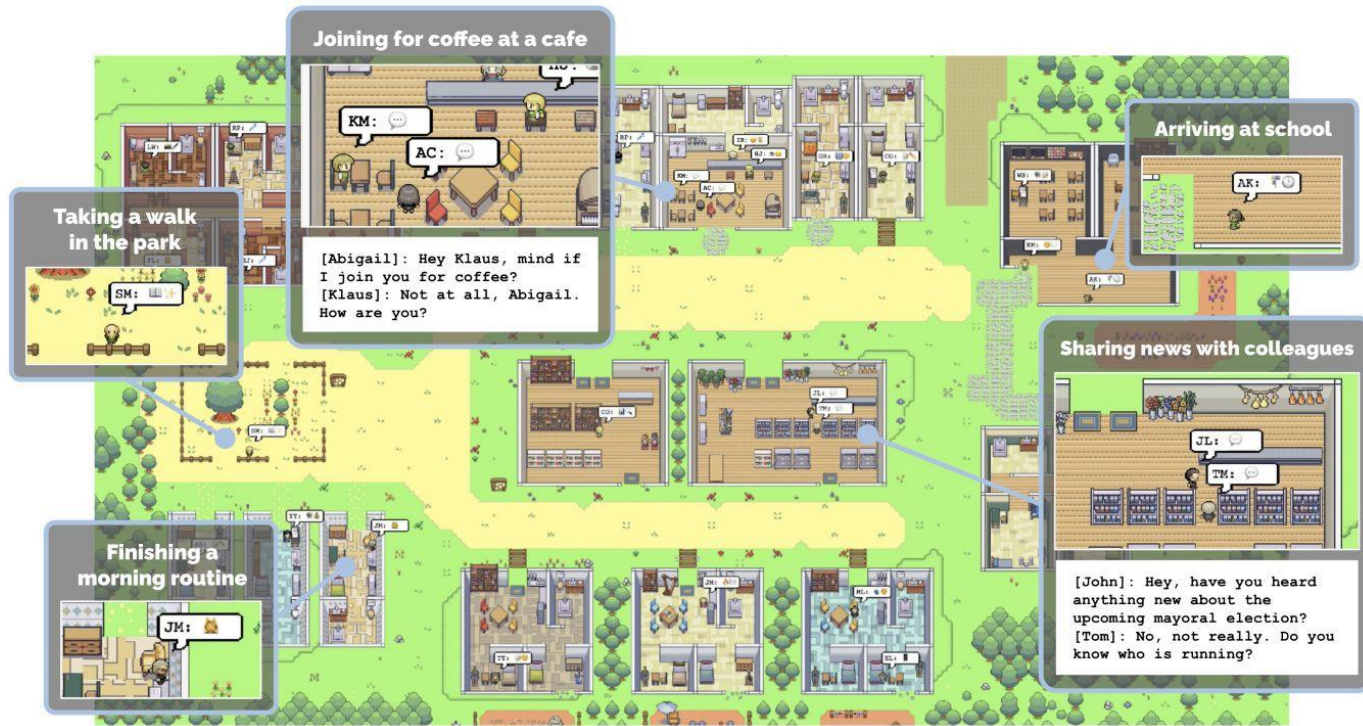


Figure 2: MARBLE 😊: showcasing interactions between task information, persona data, domain databases, memory modules, and the environment through the coordinate engine and cognitive module.

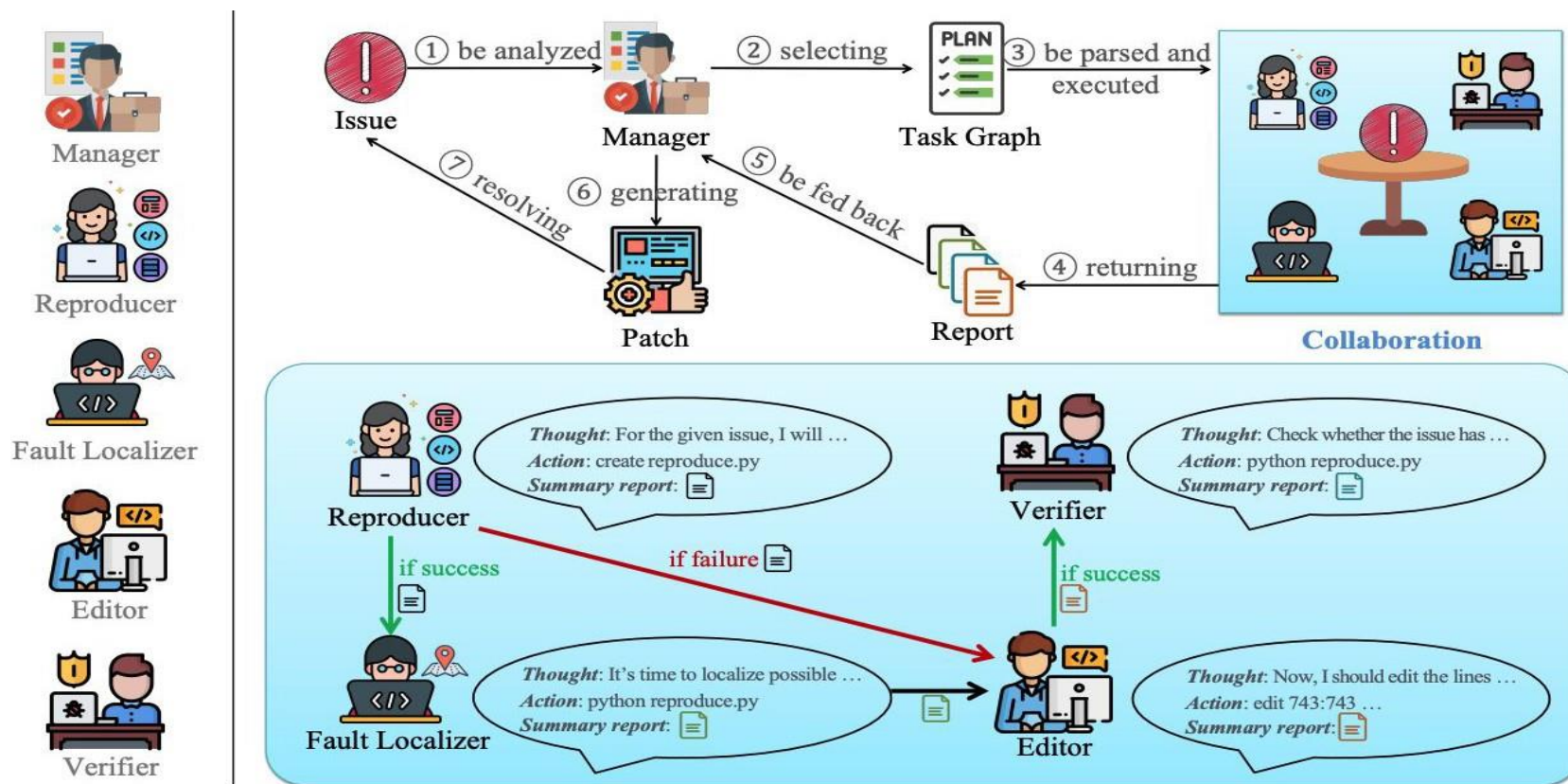
Persona definition



Park et-al. Generative agents: Interactive simulacra of human behavior. 2023

John Lin is a pharmacy shopkeeper at the Willow Market and Pharmacy who loves to help people. He is always looking for ways to make the process of getting medication easier for his customers; John Lin is living with his wife, Mei Lin, who is a college professor, and son, Eddy Lin, who is a student studying music theory; John Lin loves his family very much; John Lin has known the old couple next-door, Sam Moore and Jennifer Moore, for a few years; John Lin thinks Sam Moore is a kind and nice man; John Lin knows his neighbor, Yuriko Yamamoto, well; John Lin knows of his neighbors, Tamara Taylor and Carmen Ortiz, but has not met them before; John Lin and Tom Moreno are colleagues at The Willows Market and Pharmacy; John Lin and Tom Moreno are friends and like to discuss local politics together; John Lin knows the Moreno family somewhat well – the husband Tom Moreno and the wife Jane Moreno.

Definition through Roles and Actions



CODER Agent Actions

| Actions | Agent Roles | | | | |
|---------------------------------|-------------|------------|-----------------|--------|----------|
| | Manager | Reproducer | Fault Localizer | Editor | Verifier |
| 0 plan | ✓ | | | | |
| 1 open | | ✓ | | ✓ | |
| 2 goto | | ✓ | | ✓ | |
| 3 scroll down | | ✓ | | ✓ | |
| 4 scroll up | | ✓ | | ✓ | |
| 5 create | | ✓ | | ✓ | |
| 6 edit | | ✓ | ✓ | ✓ | ✓ |
| 7 submit | ✓ | | | | |
| 8 search dir | ✓ | ✓ | | ✓ | |
| 9 search file | ✓ | ✓ | | ✓ | |
| 10 find file | ✓ | ✓ | | ✓ | |
| 11 rover search file* | ✓ | ✓ | | ✓ | |
| 12 rover search class* | ✓ | ✓ | | ✓ | |
| 13 rover search class in file* | ✓ | ✓ | | ✓ | |
| 14 rover search method* | ✓ | ✓ | | ✓ | |
| 15 rover search method in file* | ✓ | ✓ | | ✓ | |
| 16 rover search code* | ✓ | ✓ | | ✓ | |
| 17 rover search code in file* | ✓ | ✓ | | ✓ | |
| 18 related issue retrieval | | | ✓ | ✓ | |
| 19 fault localization | | | ✓ | | |
| 20 test | | | | | ✓ |
| 21 report | | ✓ | ✓ | ✓ | ✓ |
| 22 basic shell command | ✓ | ✓ | ✓ | ✓ | ✓ |

Agent Coordination: Debate style

- Several agents instances
- Debate rounds: each propose and debate their stances
- Reason over multiple round to find final answer

User Input: A treasure hunter found a buried treasure chest filled with gems. There were 175 diamonds, 35 fewer rubies than diamonds, and twice the number of emeralds than the rubies. How many of the gems were there in the chest?

Round 1

Agent 1: Let x be the number of rubies. Then the number of diamonds is 175. And the number of emeralds is $2(x-35)$, since there are twice as many emeralds as rubies and there are 35 fewer rubies than diamonds.

To find the total number of gems, we add the number of diamonds, rubies, and emeralds: Total gems = $175 + x + 2(x-35) = 3x + 105$.

We don't know the exact value of x , but we do know that it represents the number of rubies, which is less than 175 (since there are 35 fewer rubies than diamonds). Let's try some values of x : If $x = 50$, then the total gems = $3(50) + 105 = 255$. If $x = 100$, then the total gems = $3(100) + 105 = 405$.

Since we know there were fewer rubies than diamonds, the answer must be closer to 255 than to 405. After some trial and error, we can find that $x = 40$ gives us: Total gems = $3(40) + 105 = 225$. ❌

Agent 2: If there were 175 diamonds, then the number of rubies is 35 fewer, or $175 - 35 = 140$ rubies.

And since there were twice as many emeralds as rubies, we can represent the number of emeralds as $2 * 140 = 280$ emeralds.

To find the total number of gems in the chest, we can add up the number of diamonds, rubies, and emeralds: $175 + 140 + 280 = 595$. Therefore, there were 595 gems in the chest. ✅

User Input: Using the solutions from other agents as additional information, can you give an updated response...
(Copy Agent 2's reply) (Copy Agent 1's reply)

Round 2

Agent 1: Given the information provided in other agents' solutions, we have two answers: One agent found that the number of rubies is 40 and the total number of gems is 225. Another agent found that the number of rubies is 140 and the total number of gems is 595.

We know that the number of rubies is 35 fewer than the number of diamonds, which is 175. So the number of rubies should be $175 - 35 = 140$, as the second agent found. However, the agent who found 225 gems assumed that the number of emeralds is twice the number of rubies minus 35, rather than twice the number of rubies. If we use this assumption, we get: Number of rubies = 140. Number of emeralds = $2 * 140 - 35 = 245$. Total number of gems = $175 + 140 + 245 = 560$. Our final answer is 560. ❌

Agent 2: After reviewing the solutions provided by other agents, I agree with the second agent who found that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is 595. ✅

User Input: Using the solutions from other agents as additional information, can you give an updated response...
(Copy Agent 2's reply) (Copy Agent 1's reply)

Round 3

Agent 1: After reviewing the solutions provided by other agents, I agree with the second agent who found that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is 595. ✅

Agent 2: After reviewing the solutions provided by other agents, I also agree that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is also 595. ✅


User Input: A treasure hunter found a buried treasure chest filled with gems. There were 175 diamonds, 35 fewer rubies than diamonds, and twice the number of emeralds than the rubies. How many of the gems were there in the chest?

Round 1

Agent 1: Let x be the number of rubies. Then the number of diamonds is 175. And the number of emeralds is $2(x-35)$, since there are twice as many emeralds as rubies and there are 35 fewer rubies than diamonds.


To find the total number of gems, we add the number of diamonds, rubies, and emeralds: $\text{Total gems} = 175 + x + 2(x-35) = 3x + 105$.

We don't know the exact value of x , but we do know that it represents the number of rubies, which is less than 175 (since there are 35 fewer rubies than diamonds). Let's try some values of x : If $x = 50$, then the total gems = $3(50) + 105 = 255$. If $x = 100$, then the total gems = $3(100) + 105 = 405$.

Since we know there were fewer rubies than diamonds, the answer must be closer to 255 than to 405. After some trial and error, we can find that $x = 40$ gives us: $\text{Total gems} = 3(40) + 105 = 225$. 

Agent 2: If there were 175 diamonds, then the number of rubies is 35 fewer, or $175 - 35 = 140$ rubies.

And since there were twice as many emeralds as rubies, we can represent the number of emeralds as $2 * 140 = 280$ emeralds.

To find the total number of gems in the chest, we can add up the number of diamonds, rubies, and emeralds: $175 + 140 + 280 = 595$. Therefore, there were 595 gems in the chest. 

User Input: Using the solutions from other agents as additional information, can you give an updated response....

(Copy Agent 2's reply)

(Copy Agent 1's reply)

Round 2

Agent 1: Given the information provided in other agents' solutions, we have two answers: One agent found that the number of rubies is 40 and the total number of gems is 225. Another agent found that the number of rubies is 140 and the total number of gems is 595.

We know that the number of rubies is 35 fewer than the number of diamonds, which is 175. So the number of rubies should be $175 - 35 = 140$, as the second agent found. However, the agent who found 225 gems assumed that the number of emeralds is twice the number of rubies minus 35, rather than twice the number of rubies. If we use this assumption, we get: Number of rubies = 140. Number of emeralds = $2 * 140 - 35 = 245$. Total number of gems = $175 + 140 + 245 = 560$. Our final answer is 560. ❌

Agent 2: After reviewing the solutions provided by other agents, I agree with the second agent who found that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is 595. ✅

User Input: Using the solutions from other agents as additional information, can you give an updated response....
(Copy Agent 2's reply) (Copy Agent 1's reply)

Round 3

Agent 1: After reviewing the solutions provided by other agents, I agree with the second agent who found that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is 595. ✅

Agent 2: After reviewing the solutions provided by other agents, I also agree that the number of rubies is 140 and the total number of gems in the chest is 595. Therefore, my answer to the math problem is also 595. ✅

Agent Conversation

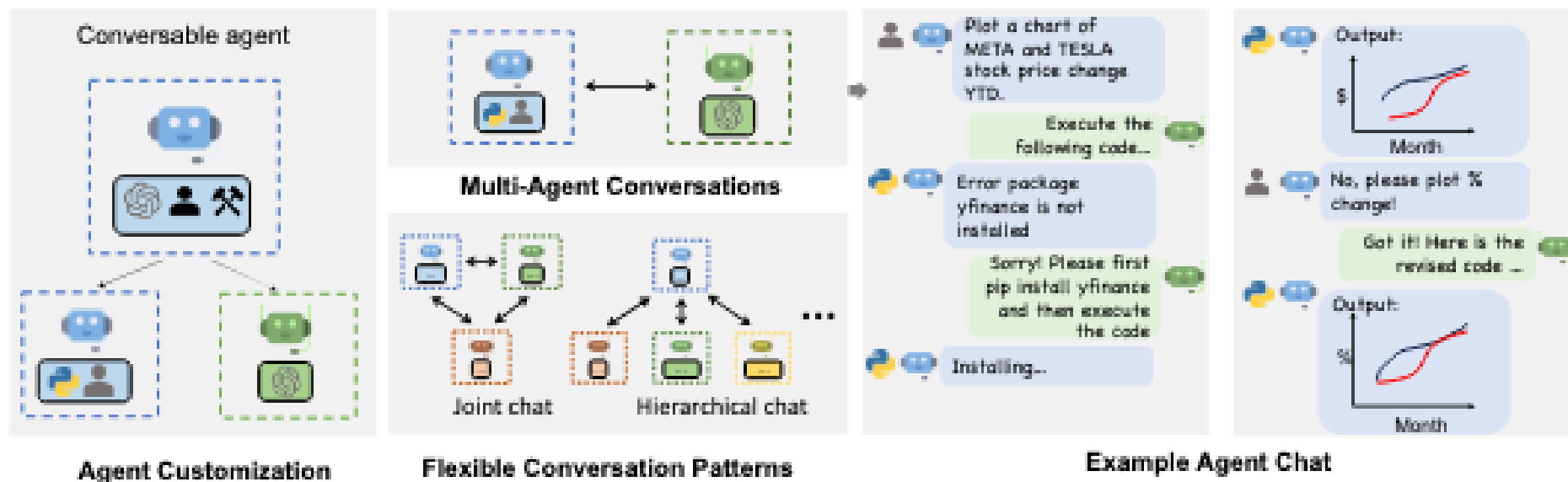
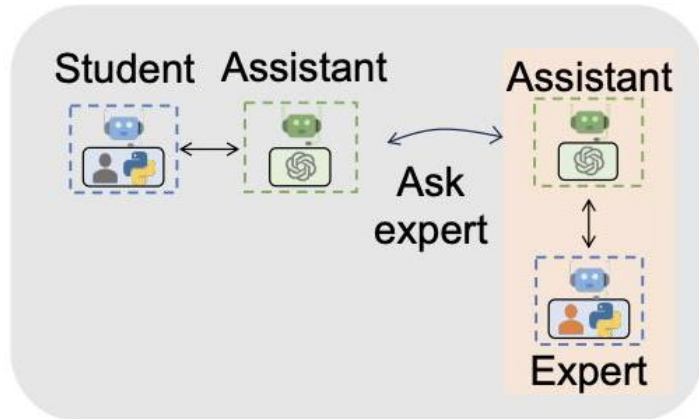
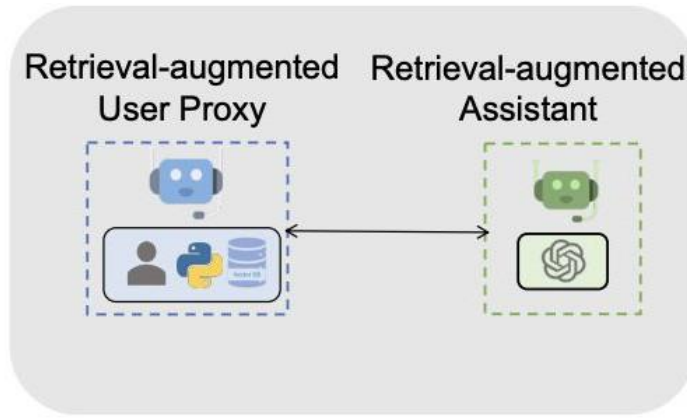


Figure 1: AutoGen enables diverse LLM-based applications using multi-agent conversations. (Left) AutoGen agents are conversable, customizable, and can be based on LLMs, tools, humans, or even a combination of them. (Top-middle) Agents can converse to solve tasks. (Right) They can form a chat, potentially with humans in the loop. (Bottom-middle) The framework supports flexible conversation patterns.

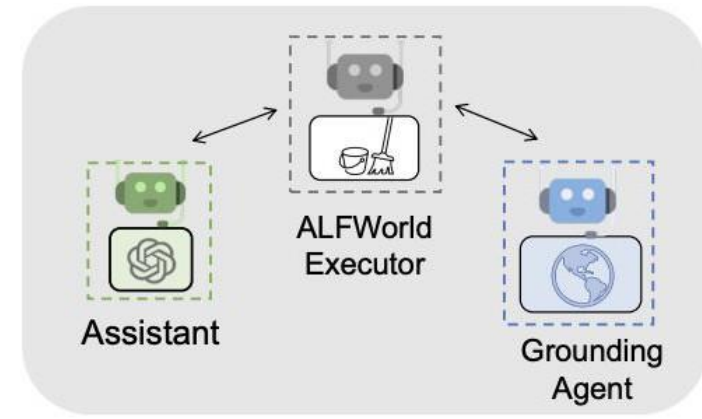
Agent Conversation



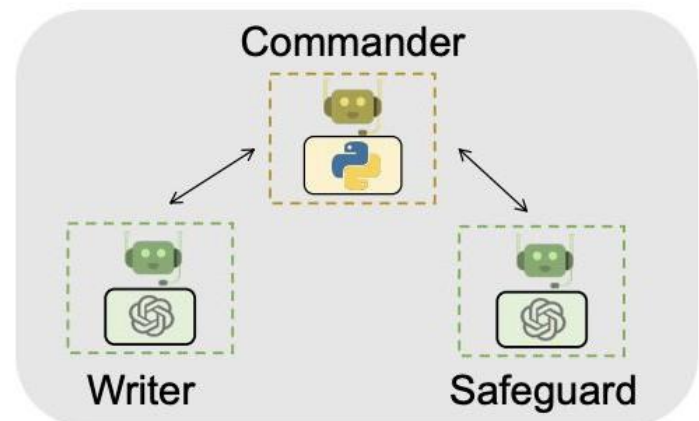
A1. Math Problem Solving



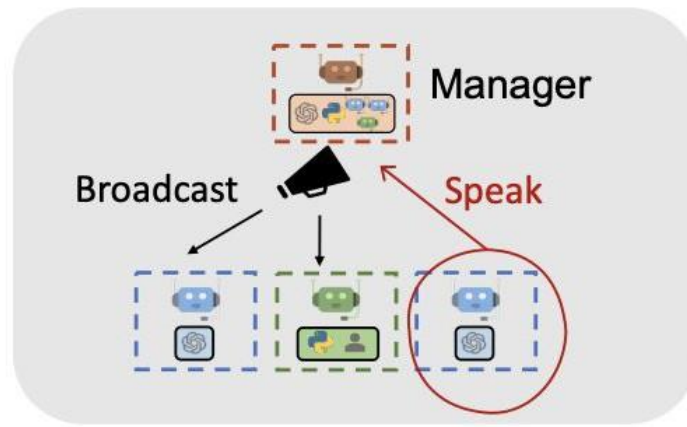
A2. Retrieval-augmented Q&A



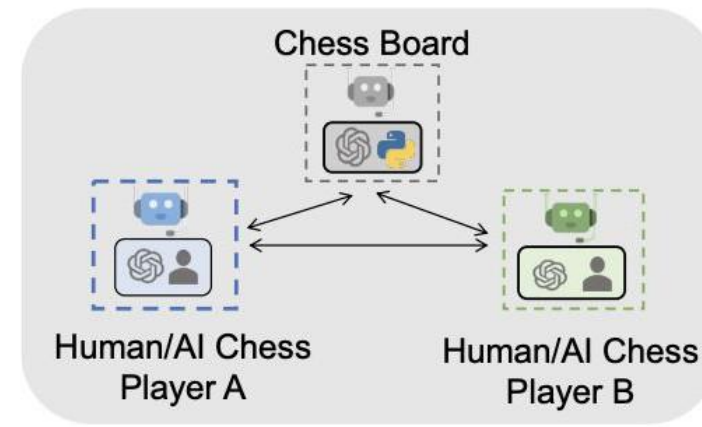
A3. Decision Making in Embodied Agents



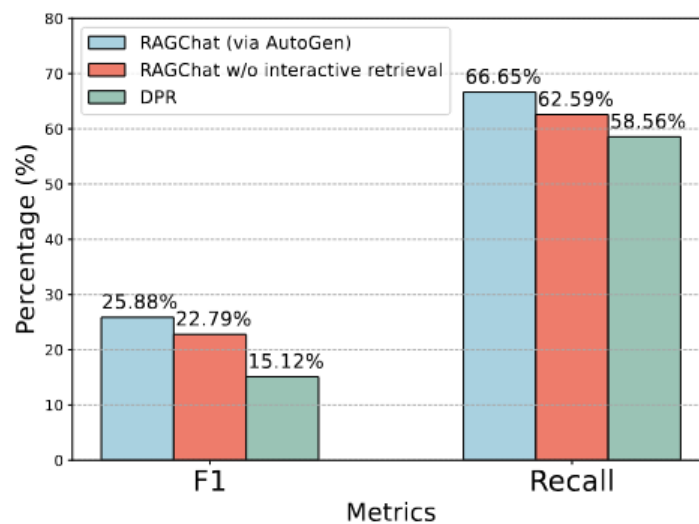
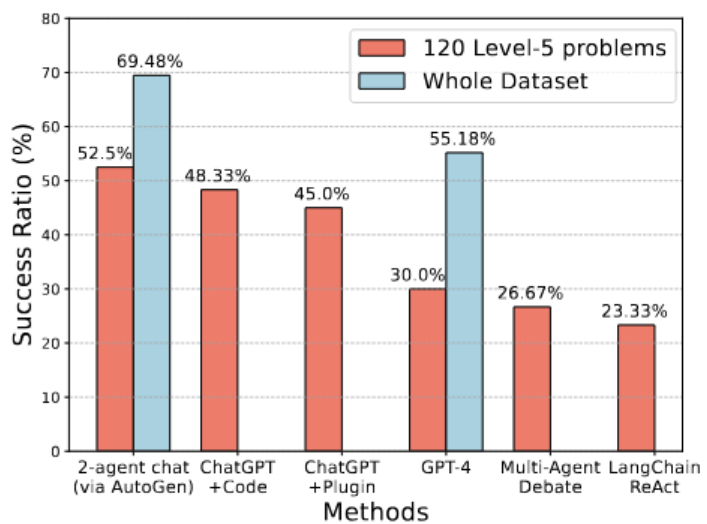
A4. Supply-Chain Optimization Conversation. Wu et-al 2023



A5. Dynamic Task Solving with Group Chat

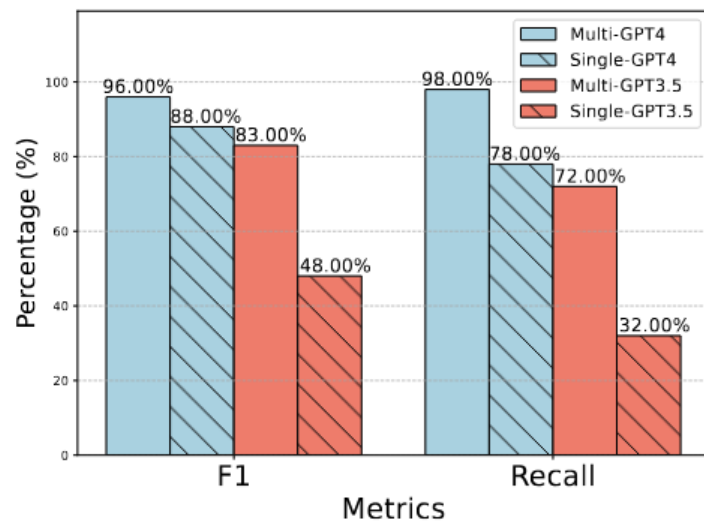
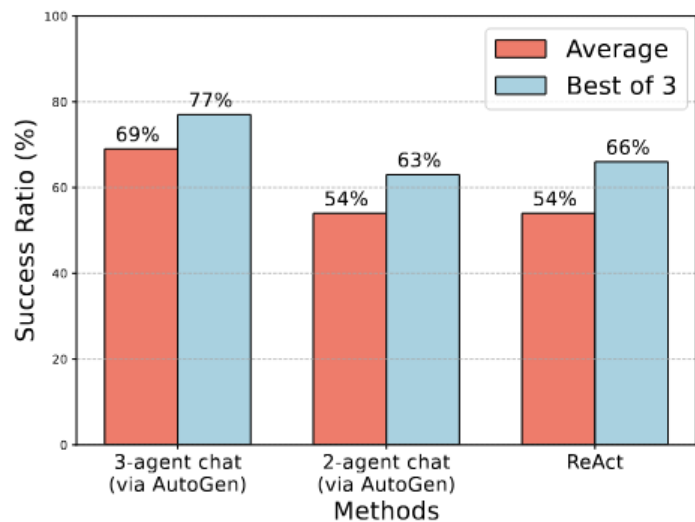


A6. Conversational Chess



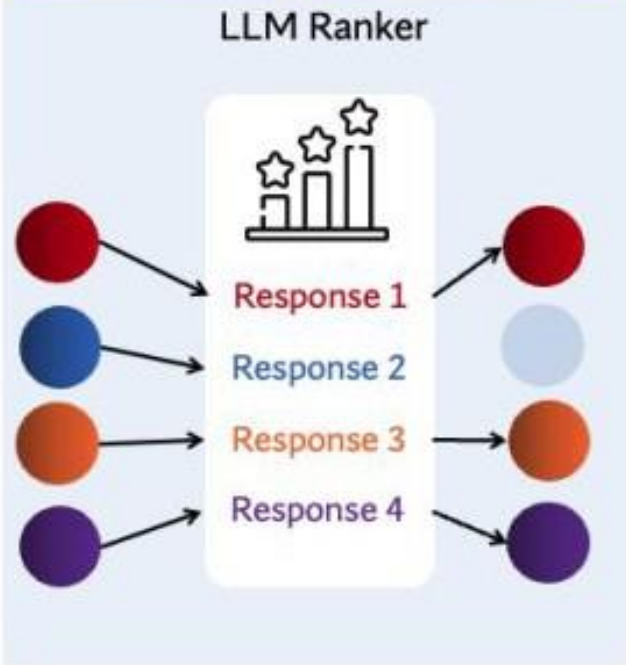
(a) A1: Performance on MATH (w/ GPT-4).

(b) A2: Q&A tasks (w/ GPT-3.5).

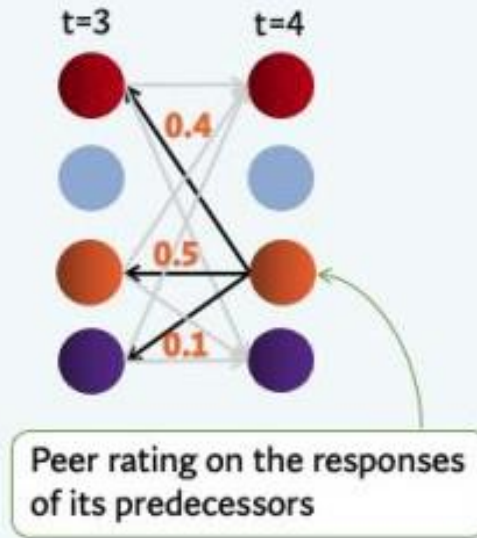


Agent Team Optimization

Inference-Time Agent Selection

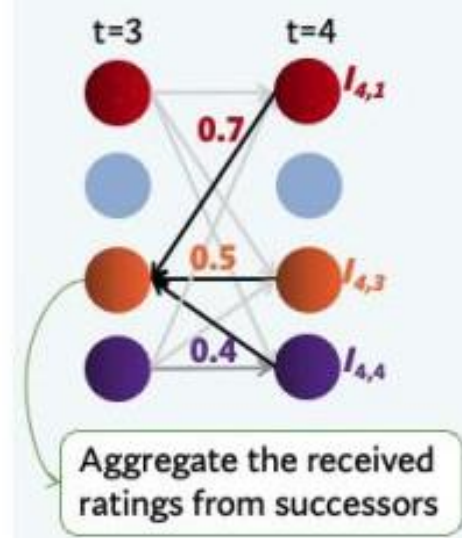


① Propagation

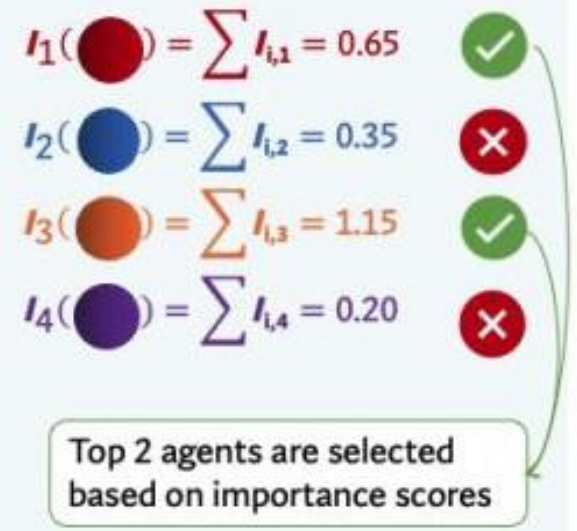


Agent Team Optimization

② Aggregation



③ Selection



Dynamic llm-agent network: An llm-agent collaboration framework with agent team optimization

Impact of Optimized Agent Team Size

