



香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

COMP 4901B

Large Language Models

Language Agents and Tools

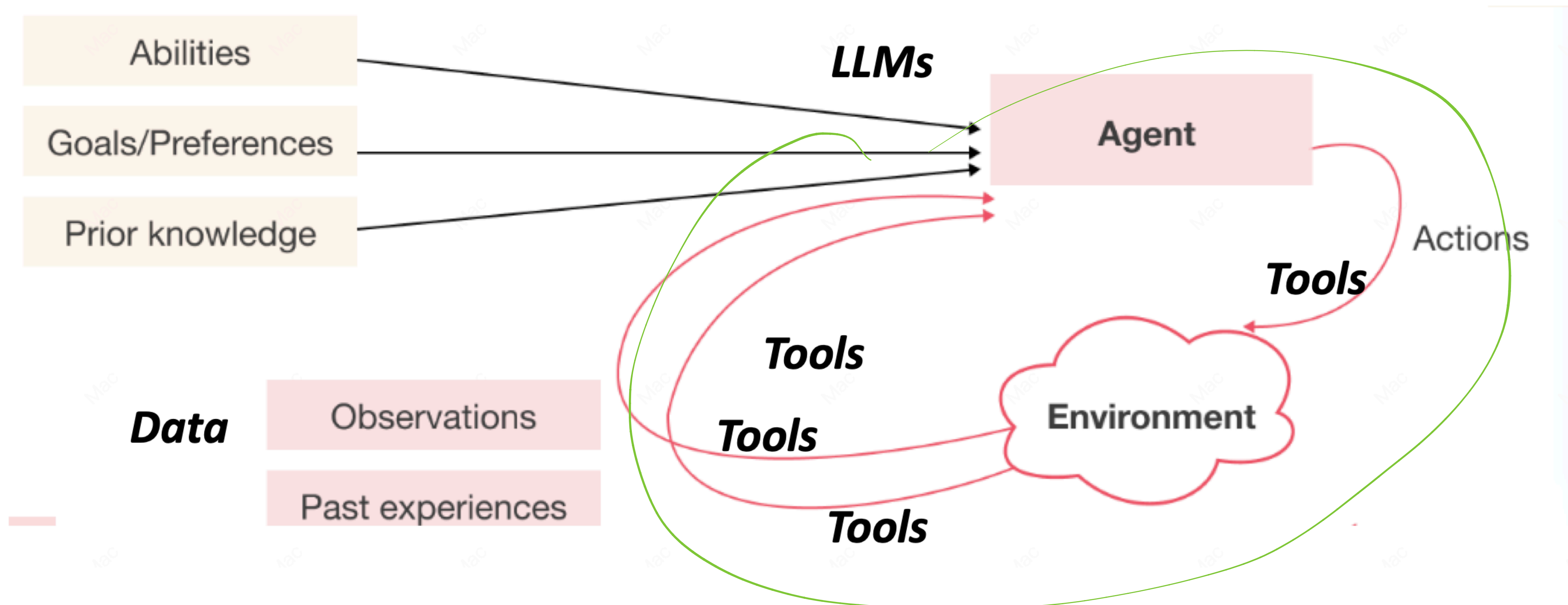
Junxian He

Nov 12, 2025

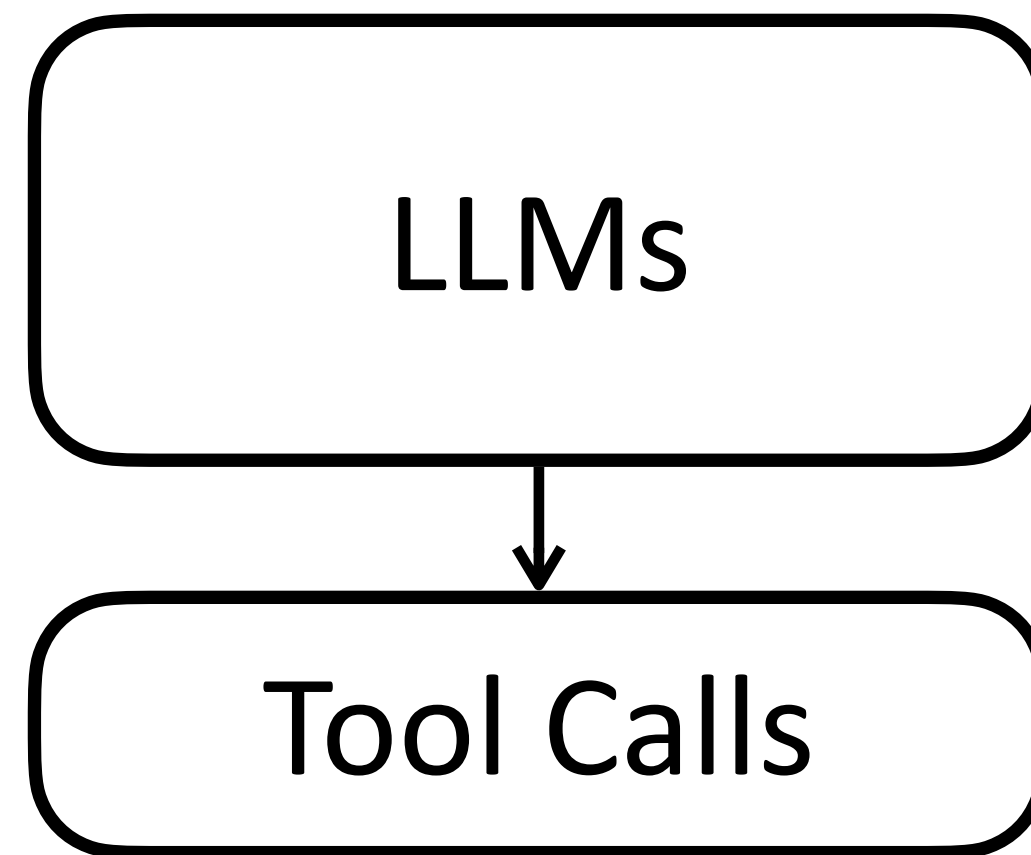
Part of slides are adapted from CMU 11711

Recap: What are Agents

Anything that can be viewed as **perceiving** its environment through sensors and **acting** upon that environment through actuators.



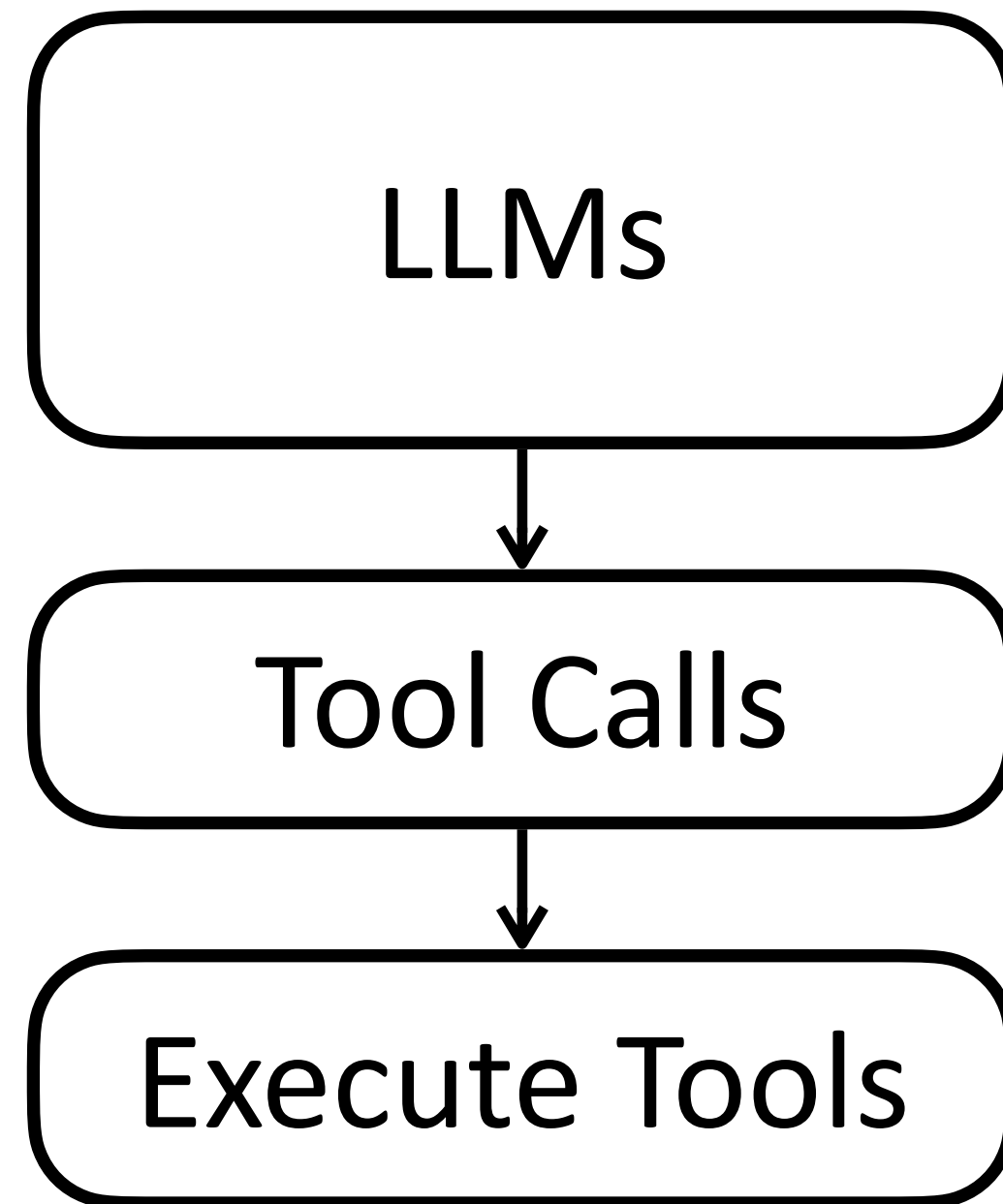
One-Step Tool Call -> Agents



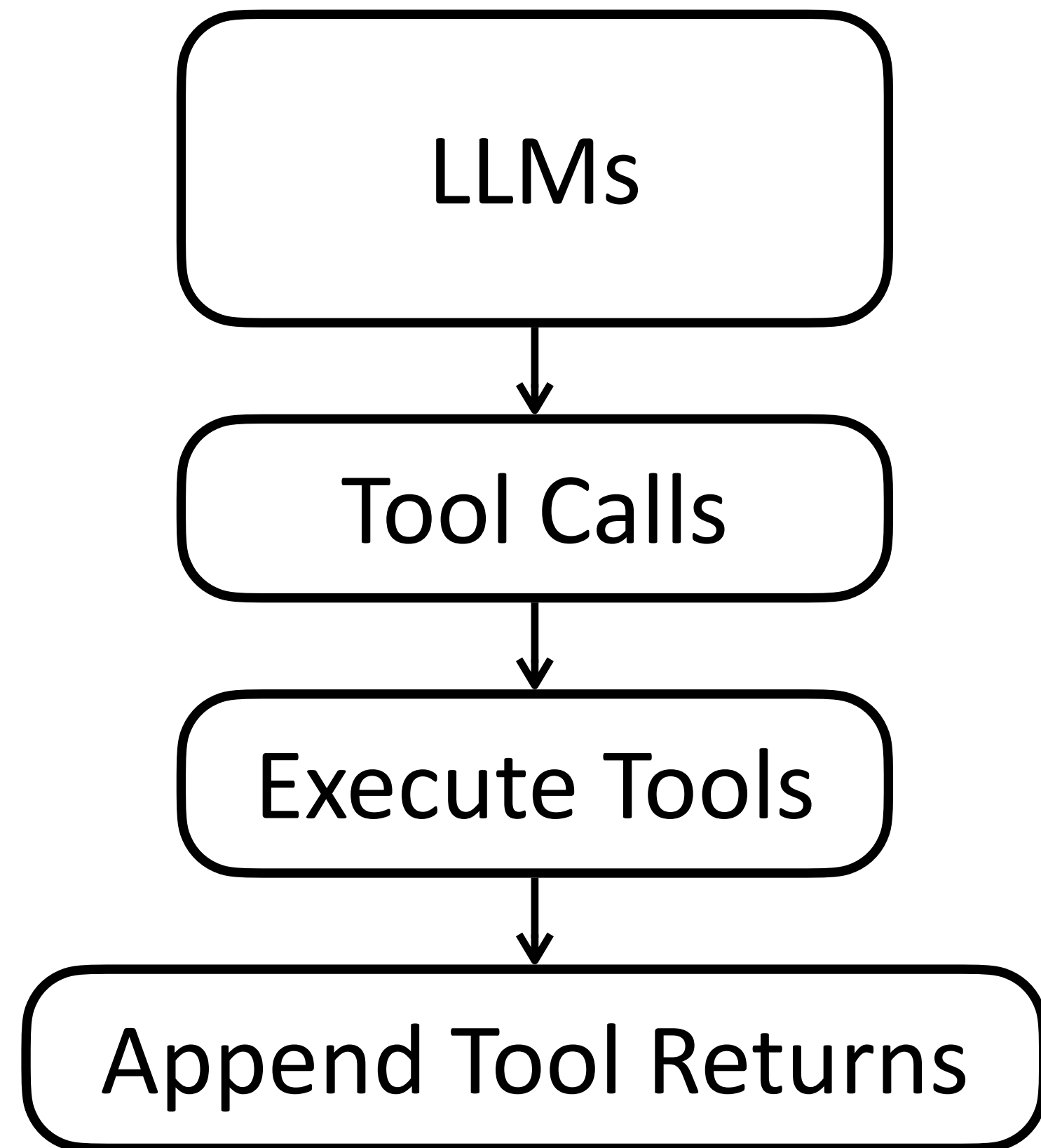
not exact

```
{
  "response": "Sure, I'll check the current weather for you.",
  "reasoning": "I need real-time conditions so the user's route recommendation is accurate.",
  "tool_calls": [
    {
      "name": "get_weather",
      "arguments": {
        "location": "San Jose, CA, US",
        "date": "2025-11-07"
      }
    }
  ],
}
```


One-Step Tool Call -> Agents



One-Step Tool Call -> Agents

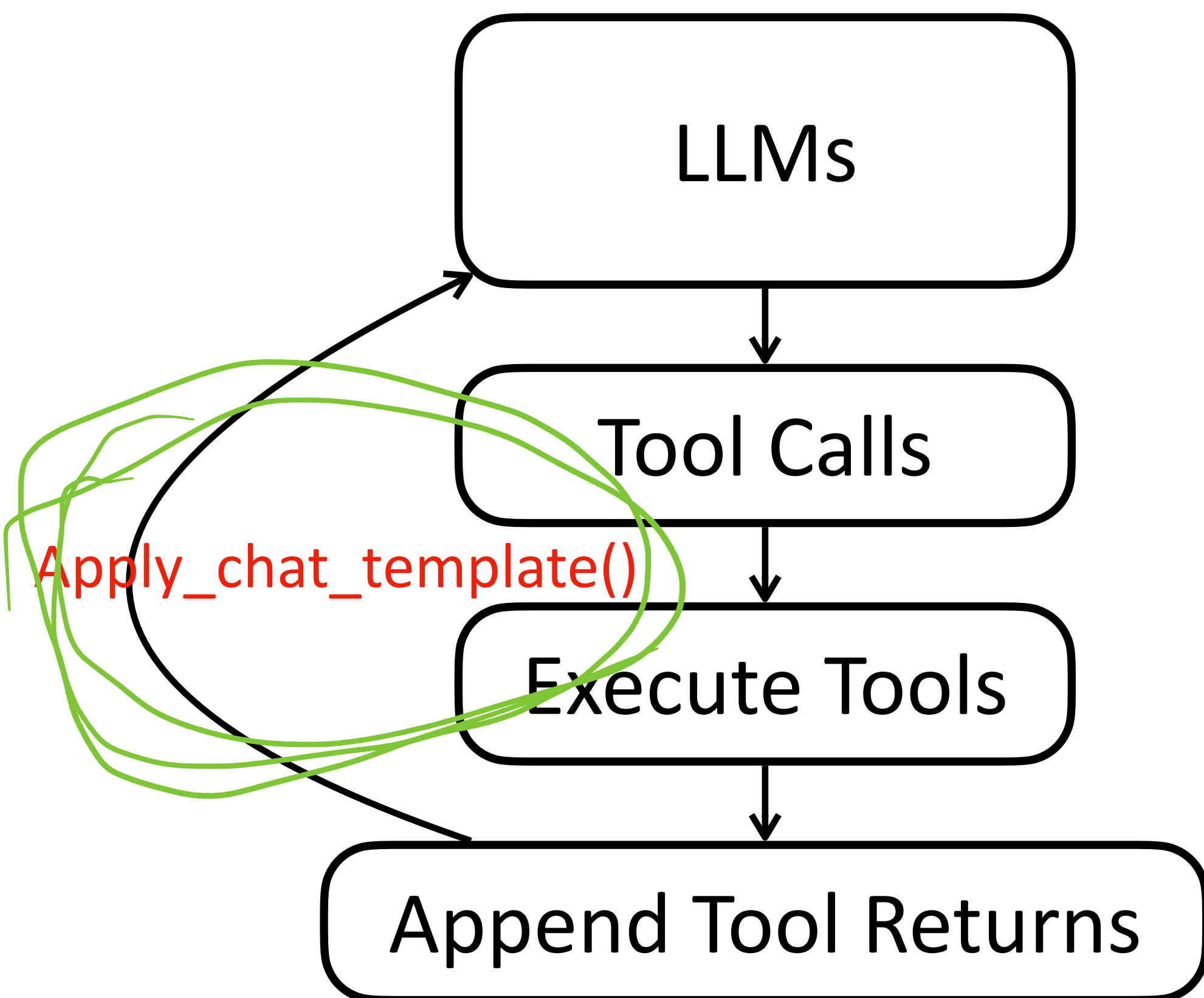


```
{
  "response": "Sure, I'll check the current weather for you.",
  "reasoning": "I need real-time conditions so the user's route recommendation is accurate.",
  "tool_calls": [
    {
      "name": "get_weather",
      "arguments": {
        "location": "San Jose, CA, US",
        "date": "2025-11-07"
      }
    }
  ],
  "tool_return": {
    "temperature": 21.5,
    "condition": "clear",
    "humidity": 60,
    "wind_speed": 10,
    "location": "San Jose, CA, US",
    "date": "2025-11-07"
  }
}
```

A green curved arrow originates from the 'Append Tool Returns' box and points to the 'tool_return' field in the JSON object, indicating a feedback loop where the tool's output is fed back into the LLM's reasoning.

openai'
url

One-Step Tool Call -> Agents



```
{  
  "response": "Sure, I'll check the current weather for you.",  
  "reasoning": "I need real-time conditions so the user's route recommendation is  
accurate.",  
  "tool_calls": [  
    {  
      "name": "get_weather",  
      "arguments": {  
        "location": "San Jose, CA, US",  
        "date": "2025-11-07"  
      }  
    }  
  ],  
  "tool_return": {  
    "temperature": 21.5,  
    "condition": "clear",  
    "humidity": 60,  
    "wind_speed": 10,  
    "location": "San Jose, CA, US",  
    "date": "2025-11-07"  
  }  
}
```

Sure, I'll check the current weather for you.

close-call

[thinking] I need real-time conditions so the user's route recommendation is accurate. [/thinking]

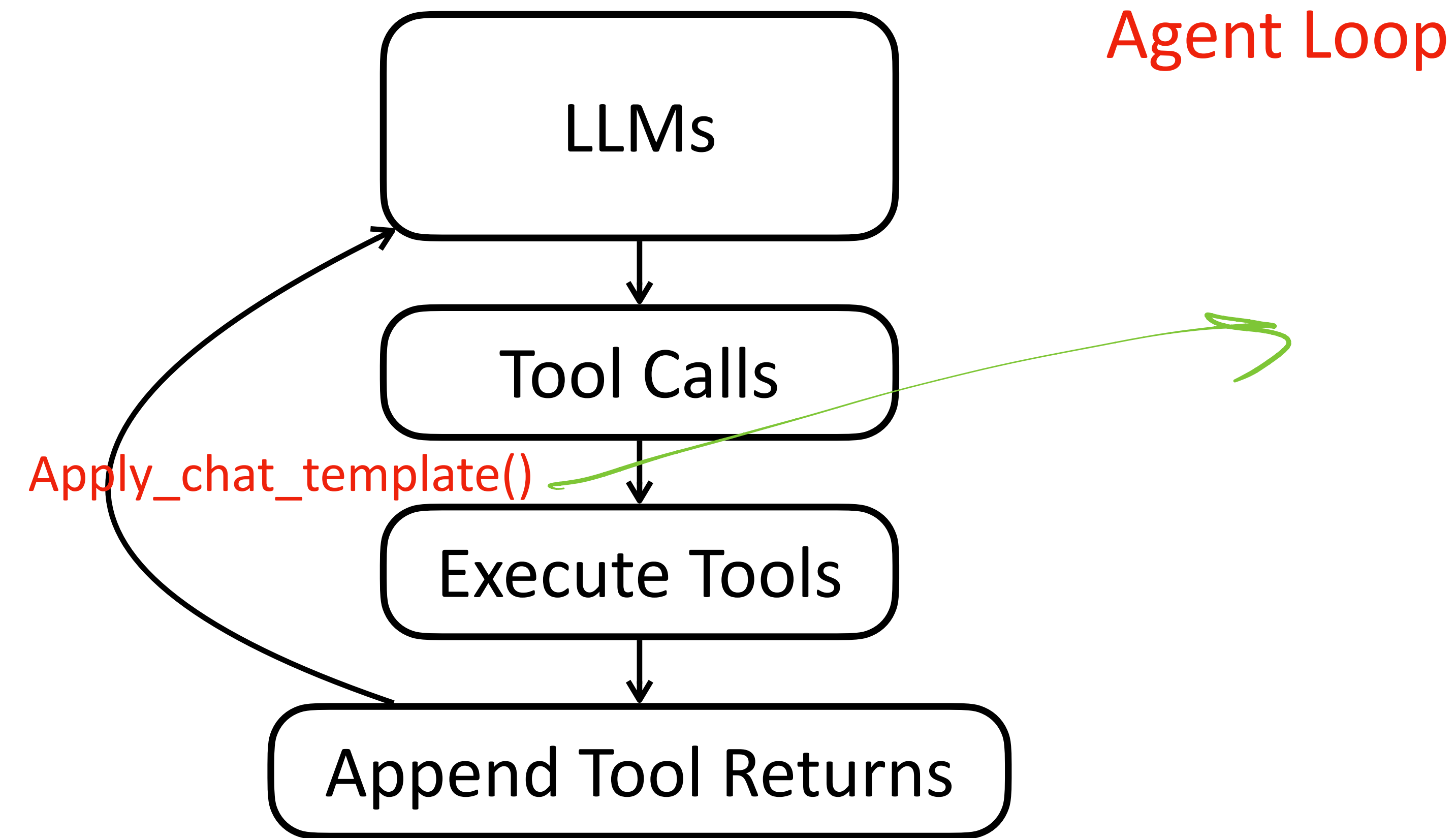
```
<tool_call>
{"name": "get_weather", "arguments": {"location": "San Jose, CA, US", "date": "2025-11-07"}}
</tool_call>
<tool_return>
{
  "temperature": 21.5,
  "condition": "clear",
  "humidity": 60,
  "wind_speed": 10,
  "location": "San Jose, CA, US",
  "date": "2025-11-07"
}
</tool_return>
```

This is the context fed back to the model to continue generation

tokenizer. apply - detect - template

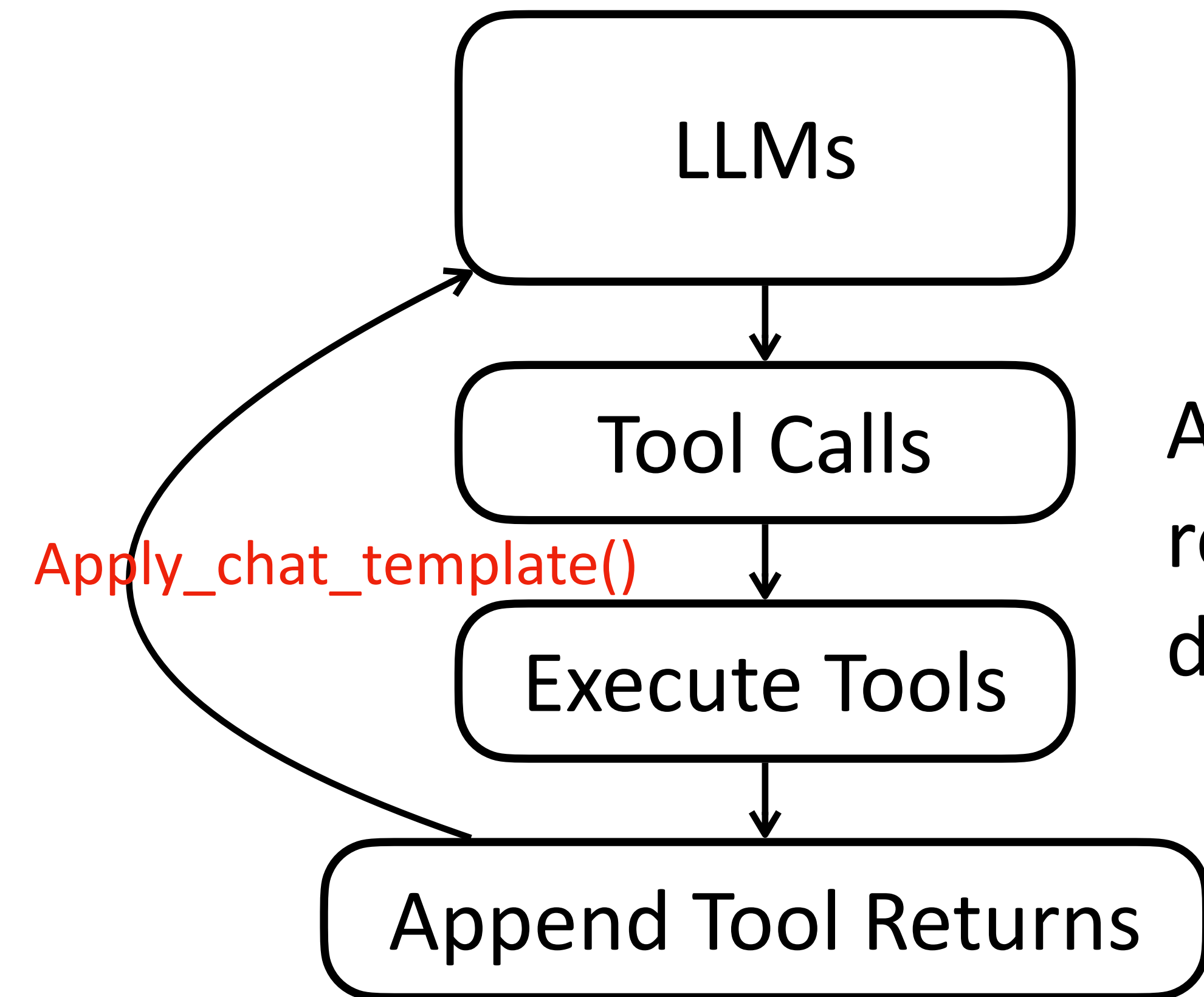
parser

One-Step Tool Call -> Agents



One-Step Tool Call -> Agents

Agent Loop



Agent loop ends when the model generates a text response without tool calling, or sometimes we define a “done” tool for the model to call

```

def send_messages(messages):
    response = client.chat.completions.create(
        model="deepseek-chat",
        messages=messages,
        tools=tools
    )
    return response.choices[0].message

client = OpenAI(
    api_key="<your api key>",
    base_url="https://api.deepseek.com",
)

tools = [
    {
        "type": "function",
        "function": {
            "name": "get_weather",
            "description": "Get weather of a location, the user should supply a location first.",
            "parameters": {
                "type": "object",
                "properties": {
                    "location": {
                        "type": "string",
                        "description": "The city and state, e.g. San Francisco, CA",
                    }
                },
                "required": ["location"]
            }
        }
    }
]

messages = [{"role": "user", "content": "How's the weather in Hangzhou, Zhejiang?"}]
message = send_messages(messages)
print(f"User>\t {messages[0]['content']}")

tool = message.tool_calls[0]
messages.append(message)

messages.append({"role": "tool", "tool_call_id": tool.id, "content": "24°C"})
message = send_messages(messages)
print(f"Model>\t {message.content}")

```

API key

OpenAI C

tool schema

list

One-step Example

tool creation

tool return

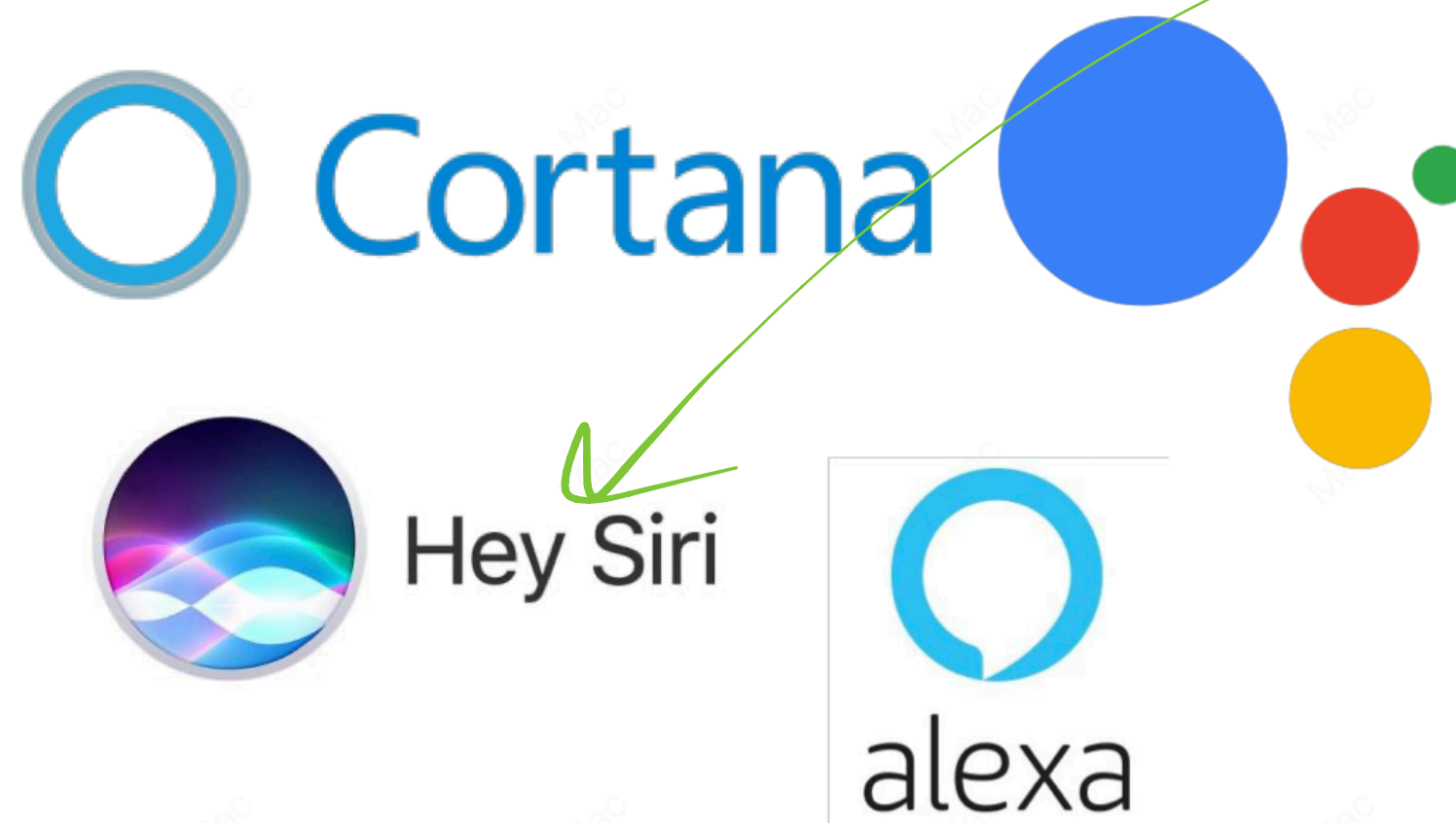
tool return

Why Do We Want Agents

cursor

Imagine if things get done by just talking...

coding agent

A screenshot of a Python IDE window titled 'Untitled-1'. The window contains a list of five lines of code:

```
1 my_list = [3, 5, 1]
2 sort in descending order →
3 sorted(my_list, reverse=True)
4
5
```

The third line is highlighted in green. The IDE interface includes a toolbar at the top with window control buttons and a status bar at the bottom showing 'master*' and 'Python 3.6.5 64-bit'.

Virtual Assistants

-  *Set an alarm at 7 AM*
-  *Remind me for the meeting at 5pm*
-  *Play Jay Chou's latest album*

Natural Language Programming

-  *Sort my_list in descending order*
-  *Copy my_file to home folder*
-  *Dump my_dict as a csv file output.csv*

Training-free Methods for Building Agents

How to Let LLM Become an Agent

How to Let LLM Become an Agent

Sure, I'll check the current weather for you.

[thinking] I need real-time conditions so the user's route recommendation is accurate. [/thinking]

<tool_call>

{"name": "get_weather", "arguments": {"location": "San Jose, CA, US", "date": "2025-11-07"}}

</tool_call>

How to Let LLM Become an Agent

Sure, I'll check the current weather for you.

[thinking] I need real-time conditions so the user's route recommendation is accurate. [/thinking]

<tool_call>

{"name": "get_weather", "arguments": {"location": "San Jose, CA, US", "date": "2025-11-07"}}

</tool_call>

We just need the LLMs to output certain formats of tool calls that we can parse

How to Let LLM Become an Agent

Sure, I'll check the current weather for you.

[thinking] I need real-time conditions so the user's route recommendati

<tool_call>

{"name": "get_weather", "arguments": {"location": "San Jose, CA, US", "dat

</tool_call>

parser

QA on step

The New England Journal of Medicine is a registered trademark of [QA("Who is the publisher of The New England Journal of Medicine?") → Massachusetts Medical Society] the MMS.

Out of 1400 participants, 400 (or [Calculator(400 / 1400) → 0.29] 29%) passed the test.

The name derives from "la tortuga", the Spanish word for [MT("tortuga") → turtle] turtle.

The Brown Act is California's law [WikiSearch("Brown Act") → The Ralph M. Brown Act is an act of the California State Legislature that guarantees the public's right to attend and participate in meetings of local legislative bodies.] that requires legislative bodies, like city councils, to hold their meetings open to the public.

easy

We just need the LLMs to output certain formats of tool calls that we can parse

Just Prompt

Supposing you have the following APIs, answer the given question.

Available APIs:

1. weather(location): for getting the weather information given a location
2. location(): for getting the location of the user.
3. bus_route(start, end): for finding the current bus route from the start position to the end position.
4. count_characters(s): for counting the number of characters in a string s.

Question: Is it okay to go hiking today?

To know the answer, we first need to know the current location of the user.

API call: location() = Seattle.

Next, we need to know the current weather in Seattle.

API call: weather(Seattle) = Cloudy with a chance of rain.

Based on this information, it is not recommended to go hiking today in Seattle.

Available APIs:

1. weather(location): for getting the weather information given a location
2. location(): for getting the location of the user.
3. bus_route(start, end): for finding the current bus route from the start position to the end position.
4. count_characters(s): for counting the number of characters in a string s.

Question: how long does it take to take a bus to Vancouver?

To answer this question, we first need to know the current location of the user.

API call: location() = Seattle.

Next, we need to find the bus route from Seattle to Vancouver.

API call: bus_route(Seattle, Vancouver) = 4 hours.

Therefore, it takes 4 hours to take a bus from Seattle to Vancouver.

Just Prompt

Supposing you have the following APIs, answer the given question.

Available APIs:

1. weather(location): for getting the weather information given a location
2. location(): for getting the location of the user.
3. bus_route(start, end): for finding the current bus route from the start position to the end position.
4. count_characters(s): for counting the number of characters in a string s.

Question: Is it okay to go hiking today?

To know the answer, we first need to know the current location of the user.

API call: location() = Seattle.

Next, we need to know the current weather in Seattle.

API call: weather(Seattle) = Cloudy with a chance of rain.

Based on this information, it is not recommended to go hiking today in Seattle.

Available APIs:

1. weather(location): for getting the weather information given a location
2. location(): for getting the location of the user.
3. bus_route(start, end): for finding the current bus route from the start position to the end position.
4. count_characters(s): for counting the number of characters in a string s.

Question: how long does it take to take a bus to Vancouver?

To answer this question, we first need to know the current location of the user.

API call: location() = Seattle.

Next, we need to find the bus route from Seattle to Vancouver.

API call: bus_route(Seattle, Vancouver) = 4 hours.

Therefore, it takes 4 hours to take a bus from Seattle to Vancouver.

→ system → prompt

We just need the LLMs to output certain formats of tool calls that we can parse

code agent

computer

Evaluating Language Agents

Environment

Evaluation of LLM Agents

- Simplified environments and basic tasks
- Performance is saturating.

1. Stateless, non interactive environment, e.g. *Mind2Web* (Deng et al. 2023) has only dumped pages.

2. Checking action sequence accuracy (step-wise, surface form only)

3. Simple interactive environment, short horizon, e.g. *WebShop* (Yao et al. 2023), *MiniWoB++* (Humphreys et al. 2022)

action space

Task

Follow one of the team leaders of one of the NHL teams from the Atlantic Division

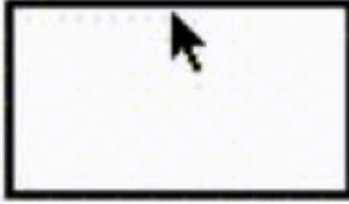
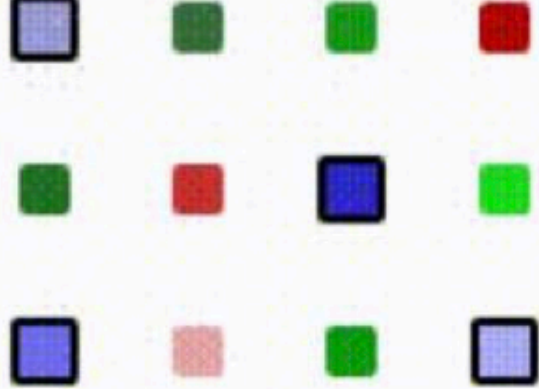
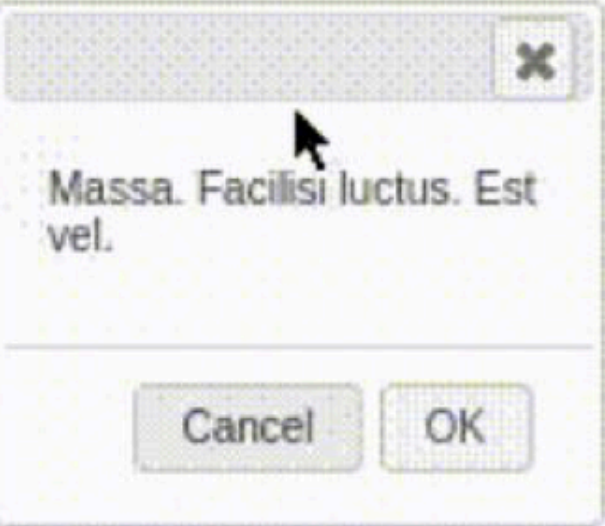
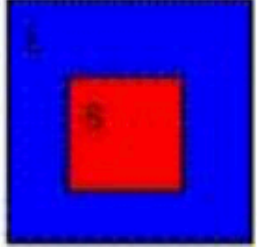
Actions

1. HOVER ➤ [link]
ESPN NHL Home Page
2. CLICK ➤ [link]
Teams
3. CLICK ➤ [heading]
Boston Bruins
4. CLICK ➤ [link]
Full Team Statistics
5. CLICK ➤ [heading]
David Pastrnak RW
6. CLICK ➤ [button]

one-action prediction

trajectory length

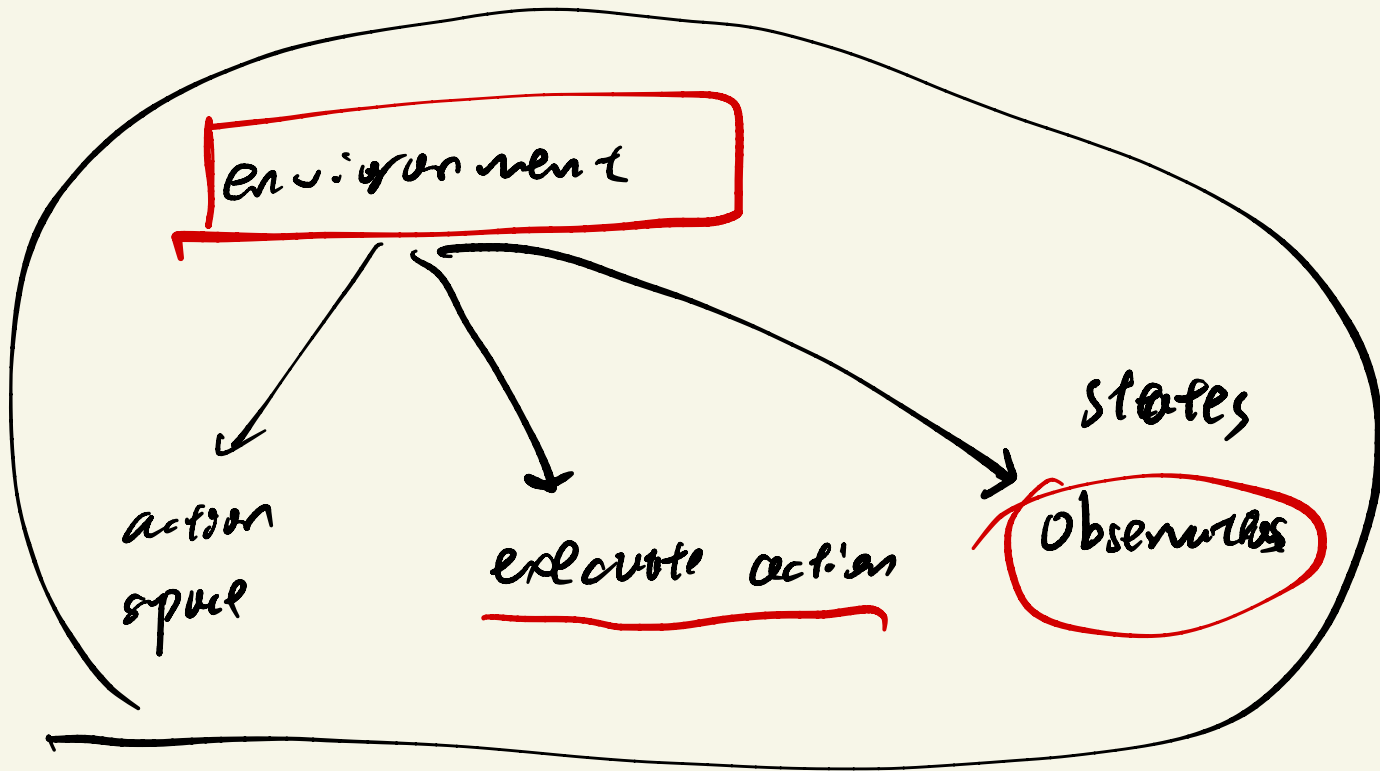
50

Move the cube around so that "5" is the active side facing the user.  Submit	Set the sliders to the combination [13,20,13] and submit. 14 19 17 Submit	Draw the number "2" in the checkboxes using the example on the right and press Submit when finished. ☒ ☒ ☒ ☐ ☐ ☒ ☐ ☐ ☐ ☐ ☒ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ Submit	Drag Ree to the 4th position. ↓ Jade ↓ Ree ↓ Karlen ↓ Millie ↓ Noelyn	Keep your mouse inside the circle as it moves around. 	Enter the value of Country into the text field and press Submit. <table border="1"> <tr><td>Gender</td><td>Male</td></tr> <tr><td>First name</td><td>AnneCorinne</td></tr> <tr><td>Country</td><td>Guam</td></tr> <tr><td>Year of Birth</td><td>1334</td></tr> <tr><td>Religion</td><td>Hinduism</td></tr> </table> Submit	Gender	Male	First name	AnneCorinne	Country	Guam	Year of Birth	1334	Religion	Hinduism
Gender	Male														
First name	AnneCorinne														
Country	Guam														
Year of Birth	1334														
Religion	Hinduism														
Drag all triangles into the black box.   Submit	Select 09/23/2016 as the date and hit submit. Date:  Submit	Sort the numbers in increasing order, starting with the lowest number at the top of the list. ↓ 9 ↓ -12 ↓ 49 ↓ -28 Submit	Copy the text from the 1st text area below and paste it into the text input. Blandit quisque. Amet ac odio aliquam. Ultrices ornare Submit	Select all the shades of blue and press Submit.  Submit	Find the 4th word in the paragraph, type that into the textbox and press "Submit". Non arcu ut ultricies est. Gravida gravida. Porta erat nulla eget condimentum posuere a Submit										
Click the button in the dialog box labeled "Cancel". 	Highlight the text in the paragraph below and click submit. Submit Ultricies. Sagittis in.	Highlight the text in the paragraph below and click submit. Tempor posuere nibh. Vel nisl, faucibus. Feugiat condimentum Submit	Find the 11th word in the paragraph, type that into the textbox and press "Submit". Ullamcorper aliquet amet ullamcorper. Elit. Mattis luctus diam. Lobortis nulla fermentum ornare faucibus Submit	Move the cube around so that "2" is the active side facing the user.  Submit	Drag the smaller box so that it is completely inside the larger box.  Submit										

environment

action

move coordinates/
click



Instruction: i am looking for x-large, red color women faux fur lined winter warm jacket coat, and price lower than 70.00 dollars

Current Query: women fur jacket coat

Results

Page 1 (1-10) of 50 total results

[Back to Search](#)

[Next >](#)



[B09KP78G37](#)

Women Faux Fur Lined Jacket Coat
Winter Warm Thick Fleece Outwear
Trench Zipper Plus Size Long
Sleeve Plush Overcoat



[B07ZXBGDXF](#)

Women's Coat, FORUU Winter Faux
Fur Fleece Outwear Warm Lapel
Biker Motor Aviator Jacket



[B098XT346Y](#)

Fjackets Real Lambskin Sherpa
Jacket - Mens Leather Jacket

★★★★★ 4.7

Current Action: click[Fjackets Real Lambskin...]

Key to Agent Benchmarks

Key to Agent Benchmarks

Environment:

Realistic

- Diverse functionality.
- Rich and realistic content.
- Interactive
- Easily Extendable
- Reproducible

self-driving

Key to Agent Benchmarks

Benchmark

lighthouse

Environment:

- Diverse functionality.
- Rich and realistic content.
- Interactive
- Easily Extendable
- Reproducible

Tasks:

- Long horizon tasks
- Enough difficulty
- Involves multiple websites

facilitate

model development

Key to Agent Benchmarks

short answer

Environment:

- Diverse functionality.
- Rich and realistic content.
- Interactive
- Easily Extendable
- Reproducible

Tasks:

- Long horizon tasks
- Enough difficulty
- Involves multiple websites

Evaluation:

- Reliable metrics
- Encourage final goal rather than partial satisfaction.

LLM as judge

WebArena

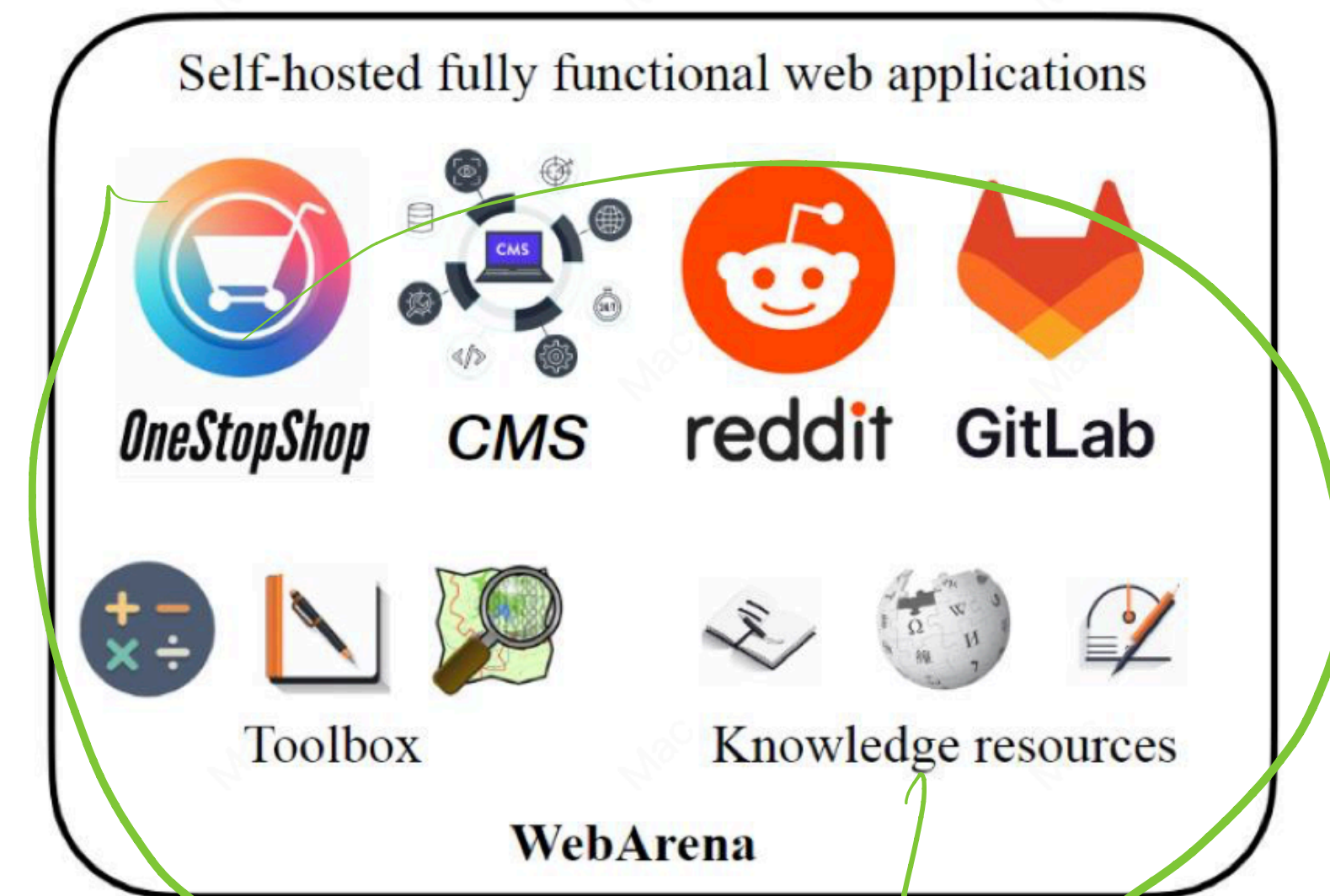
Environment:

- Diverse functionality.
- Rich and realistic content.
- Interactive
- Easily Extendable
- Reproducible

A sandbox Internet:

- Open source, production-ready implementation of the websites
- Data populated from real-world websites
- Easily distributable – Dockers, AWS images, etc.

simulate



Example Tasks in WebArena

Sequence of actions



“Create a plan to visit Pittsburgh's art museums with minimal driving distance starting from Schenley Park. Log the order in my “awesome-northeast-us-travel” repository”

webarena.wikipedia.com

Wikipedia Pittsburgh museums

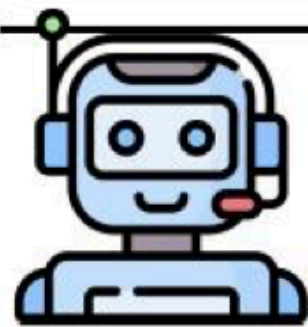
List of museums in Pittsburgh

This list of museums in **Pittsburgh, Pennsylvania** encompasses museums defined for this context as institutions (including nonprofit organizations, government entities, and private businesses) that collect and care for objects of cultural, artistic, scientific, or historical interest and make their collections or related exhibits available for public viewing. Also included are university and non-profit art galleries. Museums that exist only in cyberspace (i.e., virtual museums) are not included.

Wikimedia Commons has media related to [Museums in Pittsburgh](#).

See also: [List of museums in Pennsylvania](#)

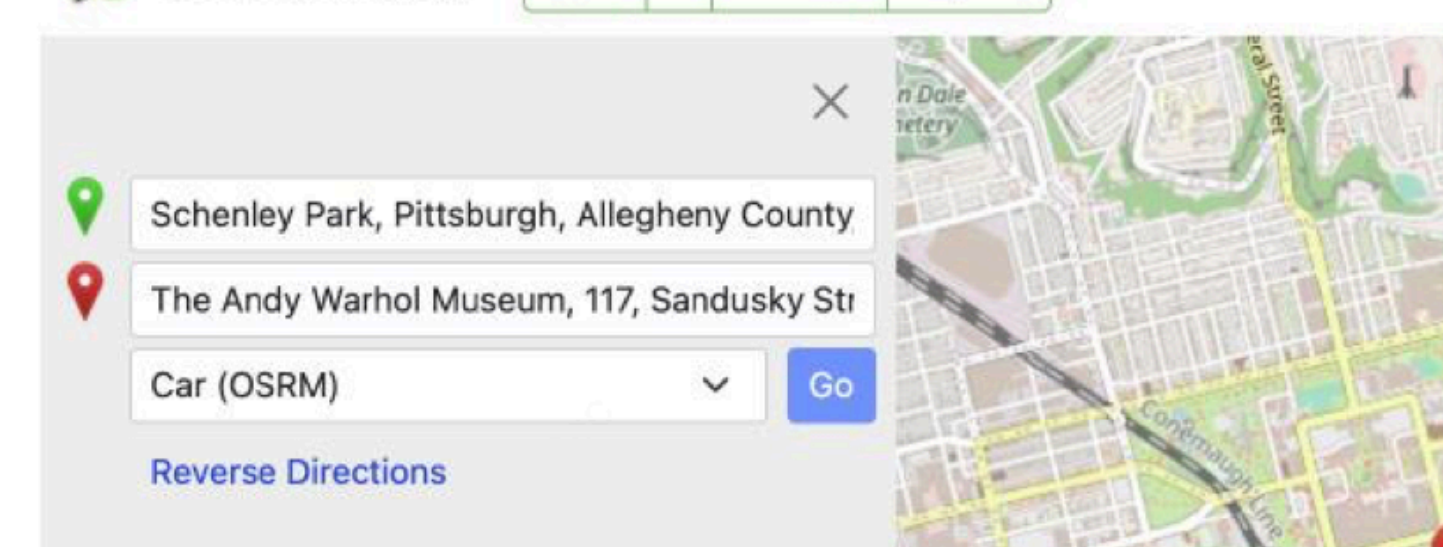
▼ Museums



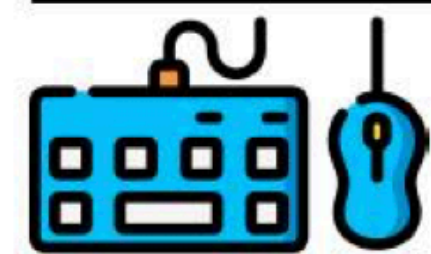
Search for museums in Pittsburgh

webarena.openstreetmap.com

OpenStreetMap Edit History Export



Directions
Distance: 7.1km. Time: 0:10.
1. Start on **Panther Hollow Road** 300m
2. Slight right onto unnamed road 160m



Search for each art museum on the Map

webarena.gitlab.com

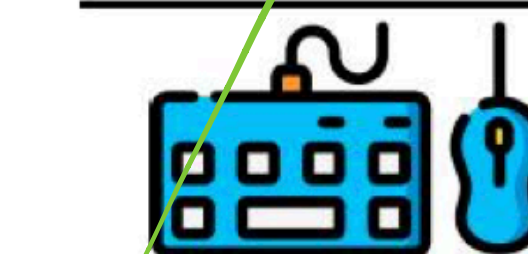
Update README.md

README.md 158 B Edit Replace

Travel in Northeast US

Pittsburgh

- + Miller Gallery at Carnegie Mellon University
- + American Jewish Museum
- + Carnegie Museum of Art



Record the optimized results to the repo

Outcome/Execution-based Evaluation

realistic

final

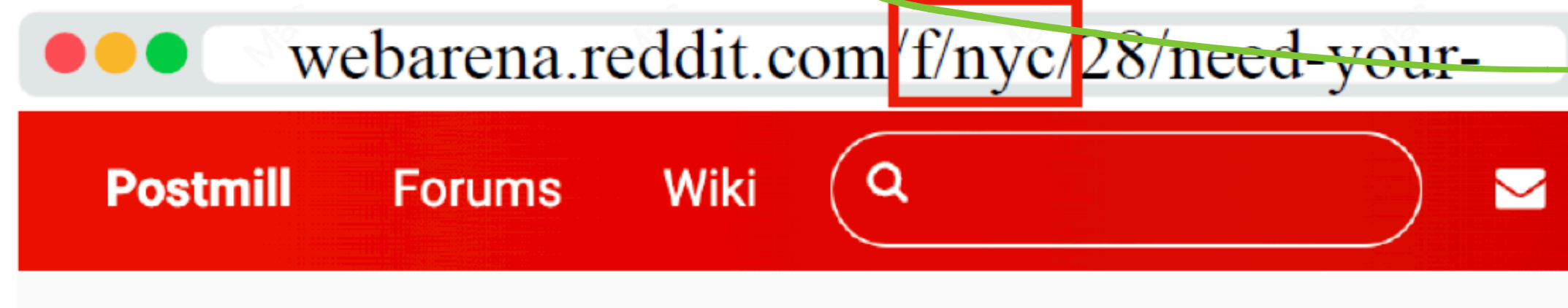
Goal: directly validate the correctness of the execution

- “When was the last time I bought shampoo?”
- **Directly compare with the annotated answer:** Answer is “Dec 15th, 2022”

Shore answers

Outcome/Execution-based Evaluation

Post my question, "is car necessary in NYC", in a subreddit where I'm likely to get an answer



← /f/nyc

↑
1
↓
Need your answer
Submitted by **convexeggstarxxx** 0 seconds ago in **nyc**
is car necessary in NYC?
No comments Edit Delete

```
▼<div class="submission__row"> flex
  ▼<div class="submission__inner"> == $0
    ▶<header class="submission_header"> ... </header>
    ▼<div class="submission_content flow-slim">
      ▼<div class="submission_body break-text text-flow">
        <p lang="en" dir="ltr">is car necessary in NYC?</p>
      </div>
    ...
```

"f/nyc" in page.url

"Is car necessary in NYC?" in
document.querySelector(".submission__inner").outText

Observation & Action Space

webarena.onestopshop.com

Patio, Lawn & Garden

Shop By

Shopping Options

Category

Gardening & Lawn
Care(168)

Patio Furniture &
Accessories(92)

Price

\$0.00 - \$999.99(311)

\$1,000.00 - \$1,999.99(8)

\$3,000.00 and above(1)

Compare Products

You have no items to
compare.

My Wish List

You have no items in
your wish list.

Items 1-12 of 320

Sort By Position 1



Outdoor Patio Folding Side Table
Square Metal End Table, Portable
Small Bistro Coffee Table, Green

★★★★★ 12 Reviews

\$49.99

Add to Cart



Shop Succulents | Assorted
Collection of Live Air Plants, Hand
Selected Variety Pack of Air
Succulents | Collection of 6

\$21.96

Add to Cart



ENEVOTX
Front Door Side Window Covering
Alligator and Cactus
Decor for Front Door Durable Fabric
Decor for Door Multi Size Door
Protector for Bedroom Home
Kitchen Party Decoration

\$38.00

Screenshot

webarena.onestopshop.com

```
<li>
  <div>
    <a href="..."></a>
    <div class>
      <a href="...">Outdoor Patio ...
    </a>
    <div>
      <span>Rating:</span>
      <div>
        <span>82%</span>
      </div>
      <a href="...#reviews">12
    </a>
    <span>Reviews</span></a>
  </div>
</li>
```

Text

webarena.onestopshop.com

RootWebArea 'Patio, Lawn ..'
link 'Image'
img 'Image'
link 'Outdoor Patio..'
LayoutTable ''
StaticText 'Rating:'
generic '82%'
link '12 Reviews'
StaticText '\$49.99'
button 'Add to Cart' focusable: True
button 'Wish List' focusable: ...
button 'Compare' focusable: ...

Accessibility tree

Keyboard: type

Mouse: click, hover, scroll

Browser: New tab, go back

action space

flexible
↑

Observation & Action Space

webarena.onestopshop.com

Patio, Lawn & Garden

Shop By

Shopping Options

Category

Gardening & Lawn Care(168)

Patio Furniture & Accessories(92)

Price

\$0.00 - \$999.99(311)

\$1,000.00 - \$1,999.99(8)

\$3,000.00 and above(1)

Compare Products

You have no items to compare.

My Wish List

You have no items in your wish list.

Outdoor Patio Folding Side Table
Square Metal End Table, Portable Small Bistro Coffee Table, Green
★★★★★ 12 Reviews
\$49.99
Add to Cart

Shop Succulents | Assorted Collection of Live Air Plants, Hand Selected Variety Pack of Air Succulents | Collection of 6
\$21.96
Add to Cart

ENEVOTX Front Door Side Window Covering Alligator and Cactus Decor for Front Door Durable Fabric Decor for Door Multi Size Door Protector for Bedroom Home Kitchen Party Decoration
\$38.00
Add to Cart

webarena.onestopshop.com

```
<li>
  <div>
    <a href="..."></a>
    <div class>
      <a href="...">Outdoor Patio ...
    </a>
    <div>
      <span>Rating:</span>
      <div>
        <span>82%</span>
      </div>
      <a href="...#reviews">12
    </a>
    <span>Reviews</span></a>
  </div>
</li>
```

webarena.onestopshop.com

RootWebArea 'Patio, Lawn ..'

link 'Image'

img 'Image'

link 'Outdoor Patio..'

LayoutTable ''

StaticText 'Rating:'

generic '82%'

link '12 Reviews'

StaticText '\$49.99'

button 'Add to Cart' focusable: True

button 'Wish List' focusable: ...

button 'Compare' focusable: ...

Screenshot

Text

Accessibility tree

Keyboard: type

Mouse: click, hover, scroll

Browser: New tab, go back

Another type of web agents, GUI agents, directly takes image as input observations

GUI

google_search('Hkust')

which one is better for
model?

API-based

Hkust → action

move mouse
to input

google search

move mouse → browser → type google

browse Wiki Hkust page

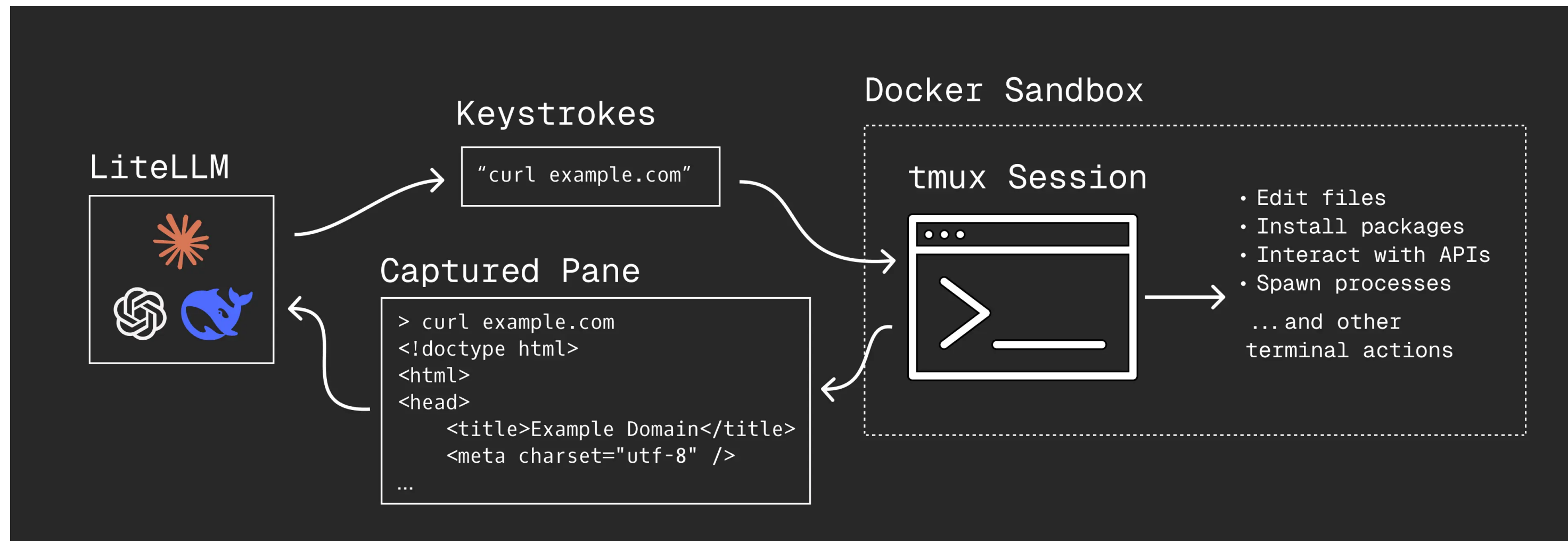
strength → scroll down → screenshot

no dls:

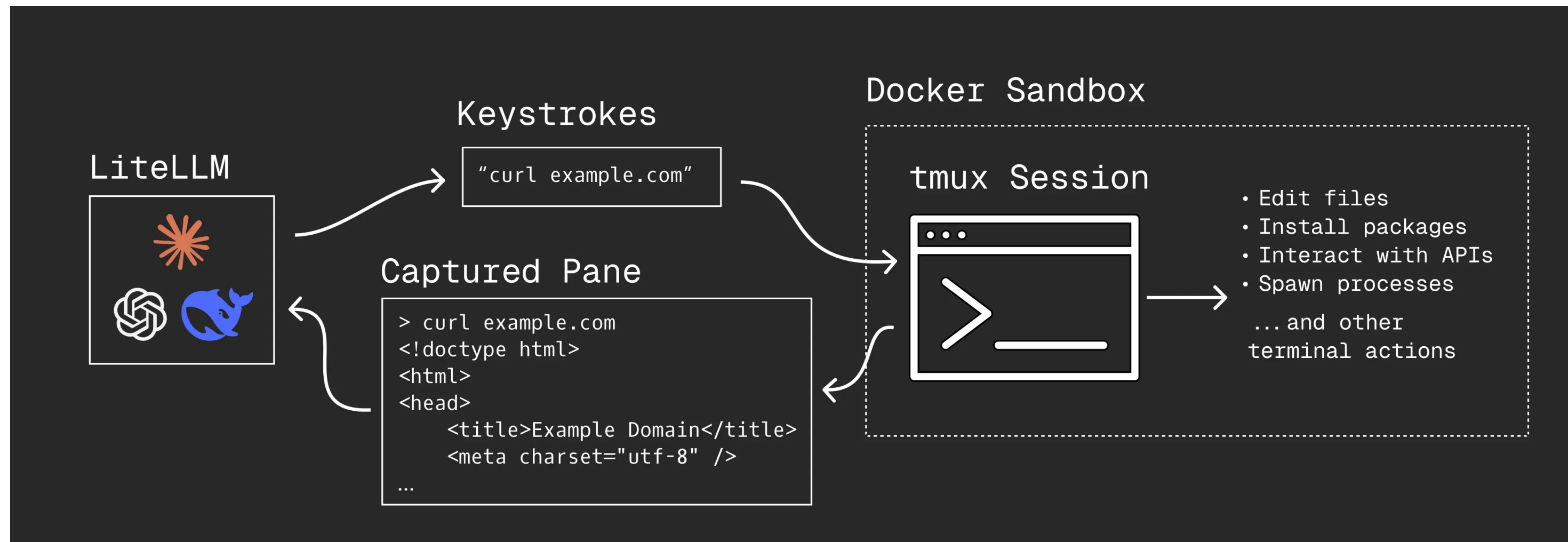
google - search (Hkust) → wiki url

browse (wiki url) → fill doc

TerminalBench



TerminalBench



Terminal-Bench: Possible Agent Actions

- Run shell commands (`ls`, `cd`, `make`, `python`, etc.)
- Manage tmux sessions/panes (`new-session`, `split-window`, `select-pane`)
- Edit files (`vim`, `nano`, `echo > file`, etc.)
- Install/build software (`apt install`, `gcc`, `make`)
- Read & analyze outputs/logs (`cat`, `less`, `grep`)
- Navigate directories and view help (`cd`, `ls`, `--help`)
- Verify or fix results (re-run scripts, check outputs)

Training Methods for Improving Agents

Learning of LLM Agents

- In-Context Learning – Learning from few-shot exemplars
- Supervised Finetuning – Learning From *Experts*
- Reinforcement Learning – Learning from *Environment*

Supervised Finetuning

- Collect large amount of expert trajectories (e.g. from human annotation)

Instruction
task_intent, [(obs_1, action_1), ..., (obs_N, action_N)] *traj.*

<think> action obs <think>

Story

acc

integrate the by

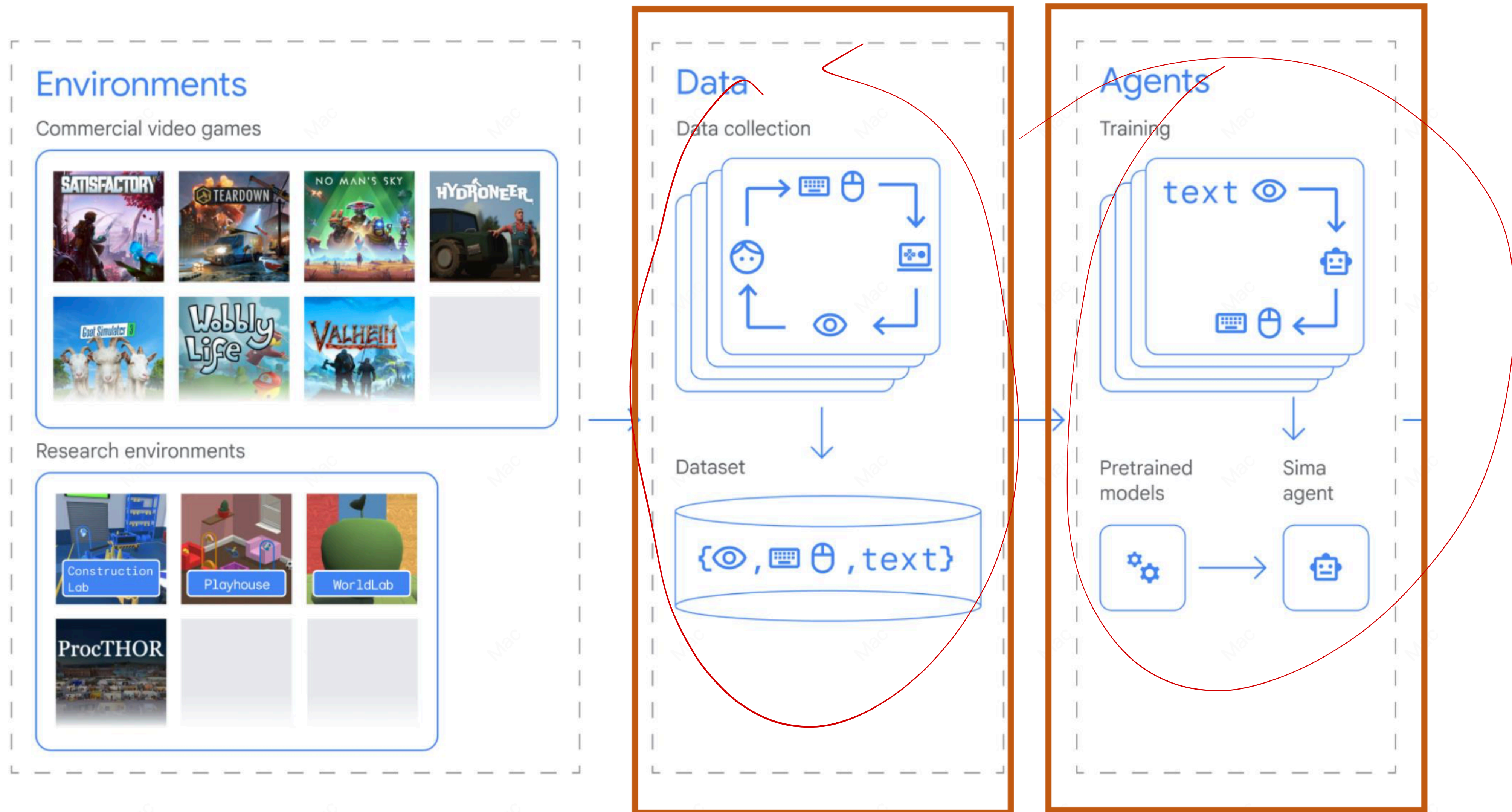
Supervised Finetuning

- Collect large amount of expert trajectories (e.g. from human annotation)

task_intent, [(obs_1, action_1), ..., (obs_N, action_N)]

- Finetune the LLM with standard cross-entropy loss.

Supervised Finetuning



Supervised Finetuning

- Data hungry

- Cannot learn much from failed trajectories

- $a_1, a_2, a_3, \dots, a_{10}$ - Success
- $a_1, a_2, a_3, \dots, a_{10}$ - Fail (Wasted)

- Need human trajectory?

- Data augmentation techniques

Remand

Create More Training Data

Open-ended Environments

Generalist Agent

Internet-scale Knowledge Base

Craft Glass Bridge



Build Oak House



Make Ice Igloo



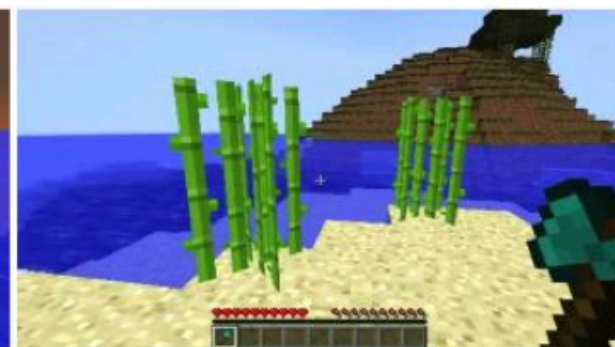
Combat Zombie



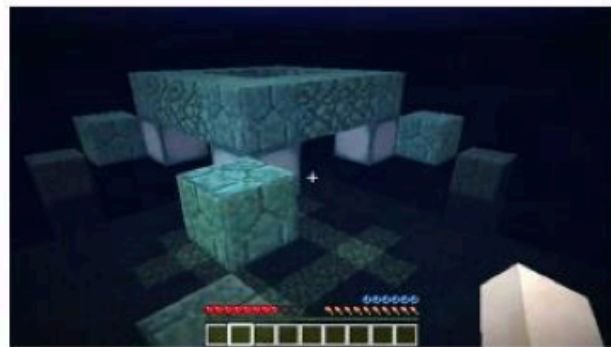
Fish Squid



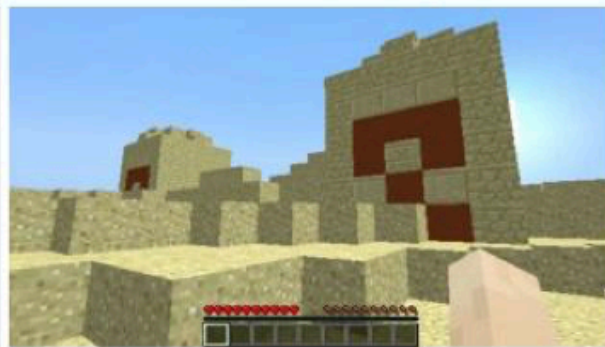
Farm Sugar Cane



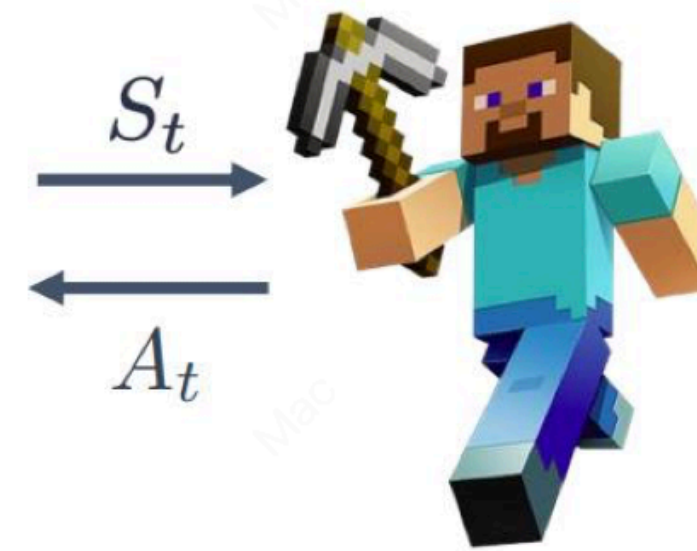
Find Ocean Monument



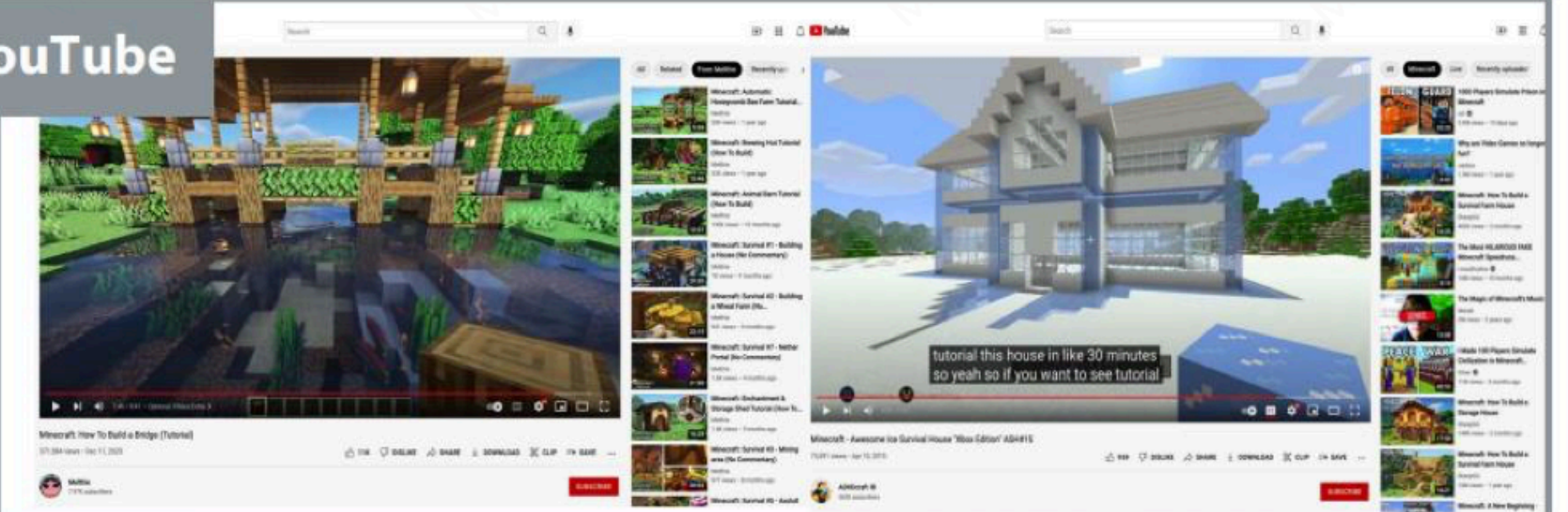
Explore Desert Temple



Treasure Hunt in End City



YouTube



Wiki

Features	Description	Screenshot	[hide]
 River	Water, Sand, Clay, Sugar Cane, Seagrass, Salmon, Squid, Drowned Temperature: 0.5. Rainfall: 0.5. A biome that consists of water blocks in an elongated, curving shape similar to a real river. Rivers are a reliable source of clay. They are good for fishing, but drowned can spawn at night.	 River	

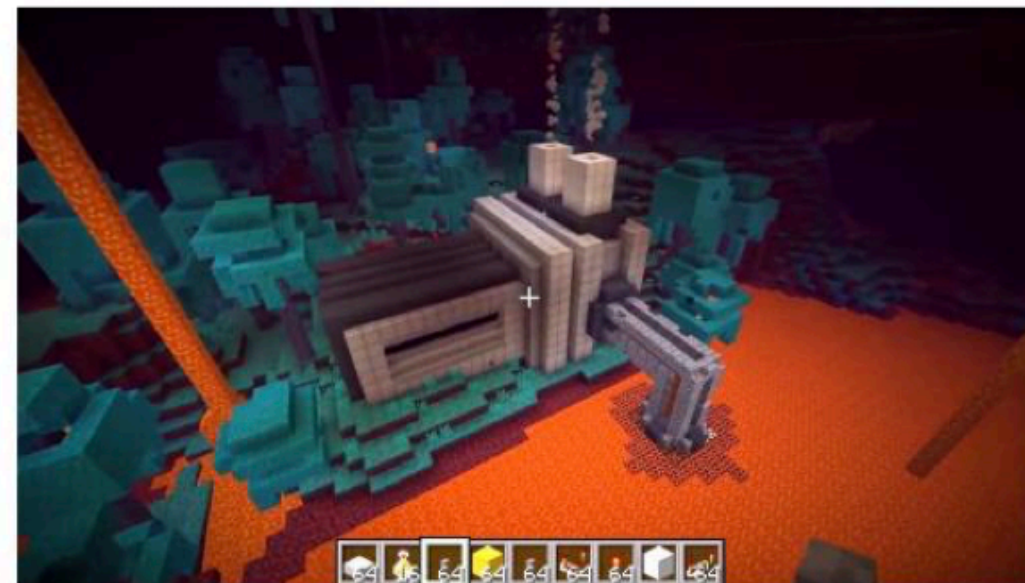
Reddit



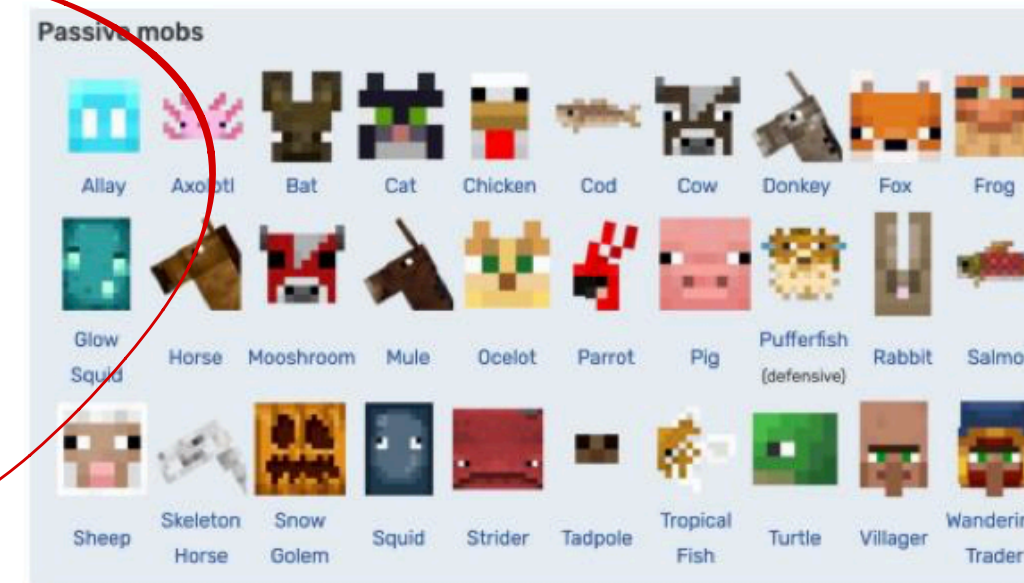
Data Augmentation

- Continue pre-train on large amount of data automatically mined
- Even noisy, not clear trajectories, provide domain adaptation.

YouTube



Wiki



Name	Ingredients	Crafting recipe
Cake	Milk Bucket + Sugar + Egg + Wheat	
Golden Apple	Gold Ingot + Apple	

Reddit



MineDojo, Fan et al. 22'

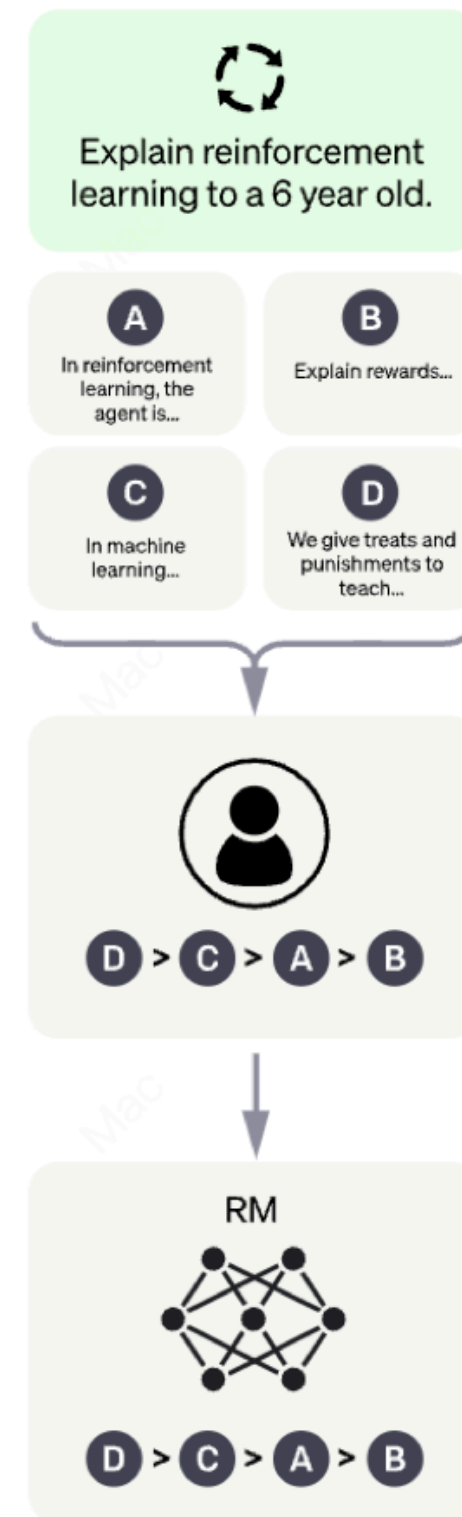
Don't Stop Pretraining, Gururangan et al., 20'

Reinforcement Learning

Lots of on-going research in this area!

Recall RLHF: Reinforcement Learning from Human Feedback:

A prompt and several model outputs are sampled.



A labeler ranks the outputs from best to worst.

This data is used to train our reward model.

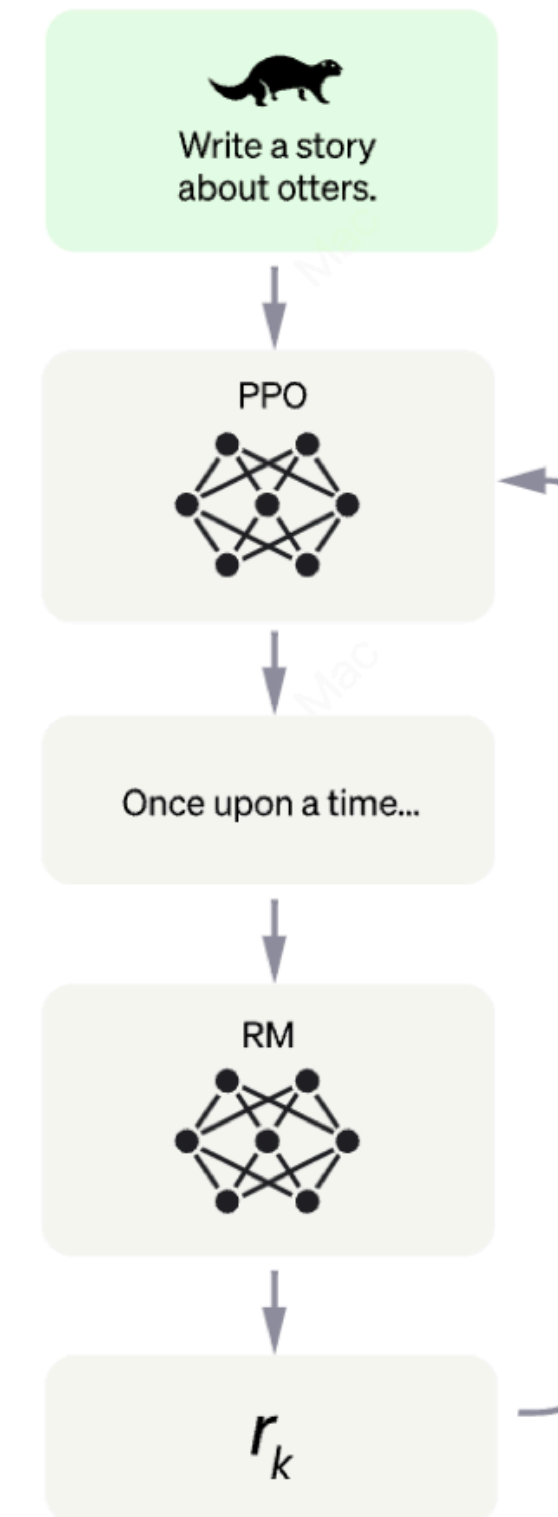
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

The policy generates an output.

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.



RLVR

RLHF, Ouyang, et al. 22'

Reinforcement Learning

Compared to RLHF:

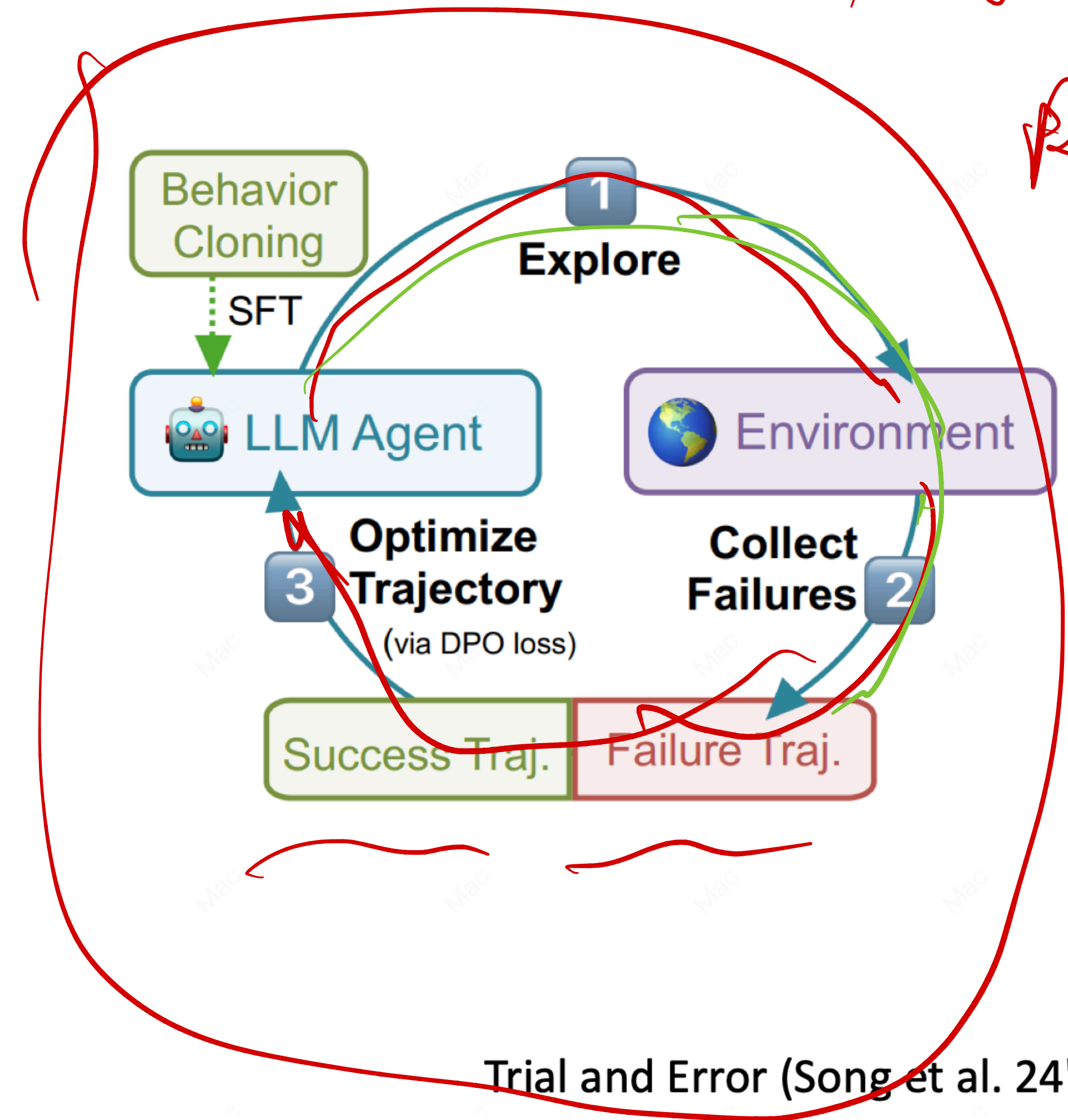
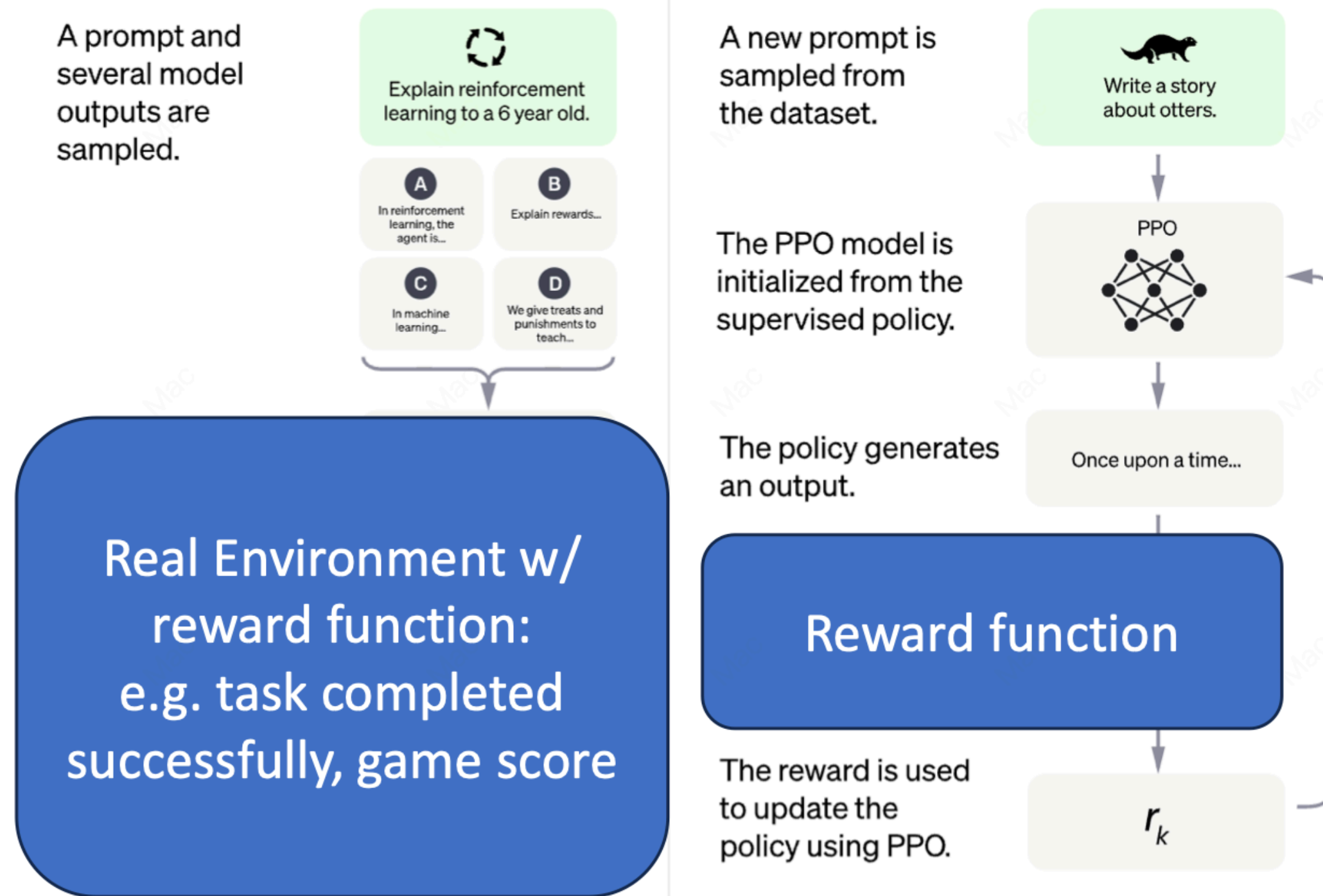
Given environment, *reward function*
(*trajectory, reward*) pairs without human

RLVR

Policy gradient

Reinforce

PPO



Trial and Error (Song et al. 24')

Reinforcement Learning

- Closed loop, interactive environment

Reinforcement Learning

- Closed loop, interactive environment
- Need good reward functions
 - What if the task success/fail is not easy to automatically assess?

Reinforcement Learning

- Closed loop, interactive environment
- Need good reward functions
 - What if the task success/fail is not easy to automatically assess?
- Need good initial models
 - Has decent basic knowledge ability, sparse rewards

Reinforcement Learning

- Closed loop, interactive environment

infra may be the muse

- Need good reward functions

important in RL

- What if the task success/fail is not easy to automatically assess?

- Need good initial models

1 trajectory → 50 steps

- Has decent basic knowledge ability, sparse rewards

100 steps → 500 seconds

- Scalability

0.01 second

1 min

builds good

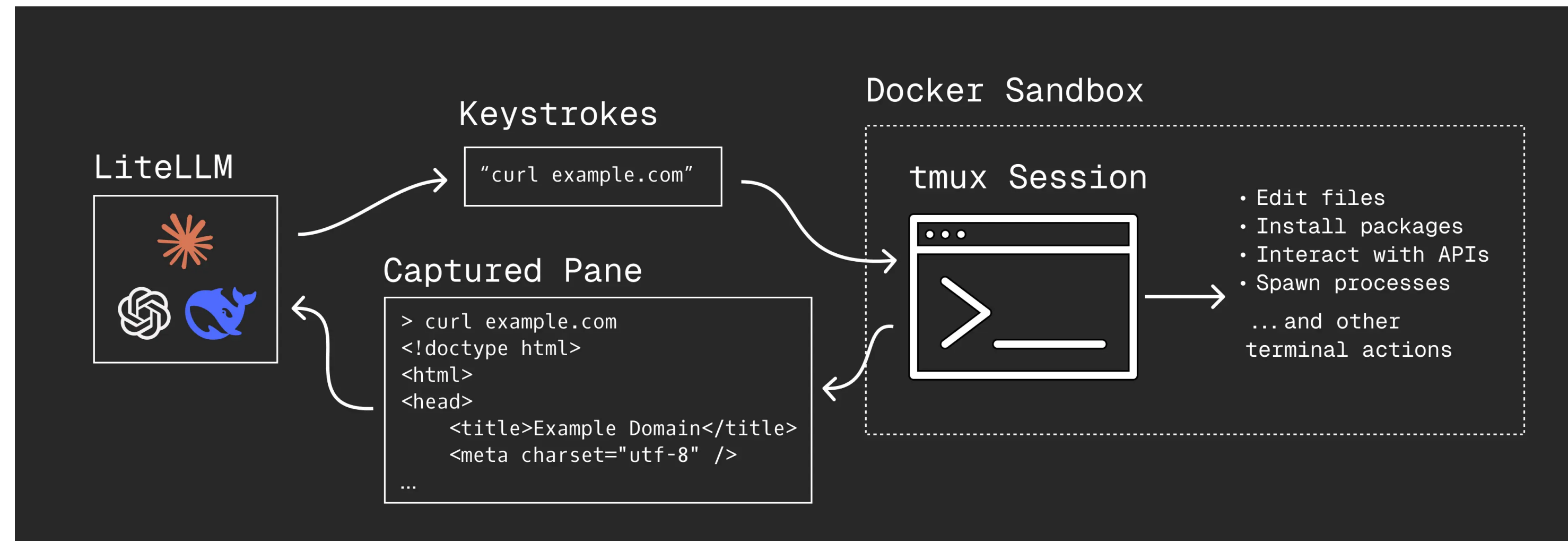
500 x 10

- The environment takes 10 seconds to `env.step()`

- The reward function takes 100 seconds to get a scalar reward

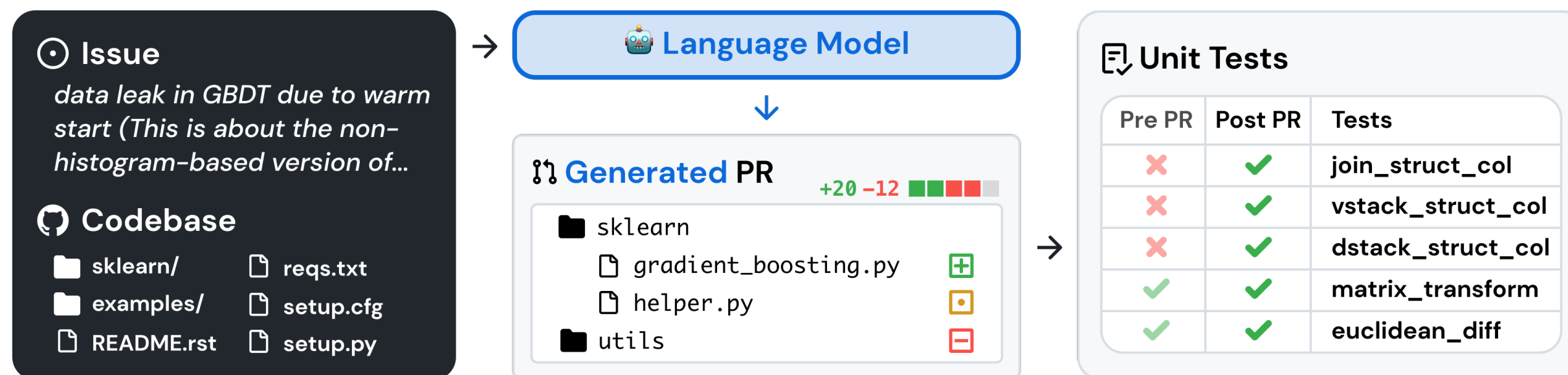
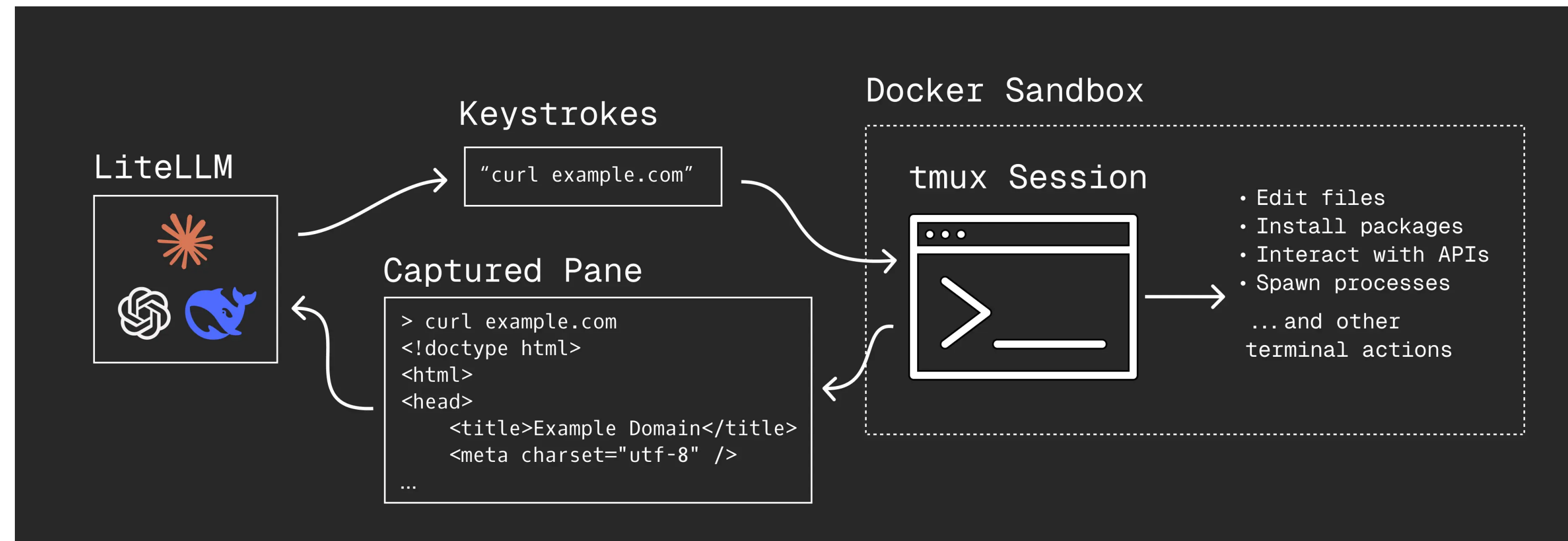
RL Environments

Environments and benchmarks typically come together



RL Environments

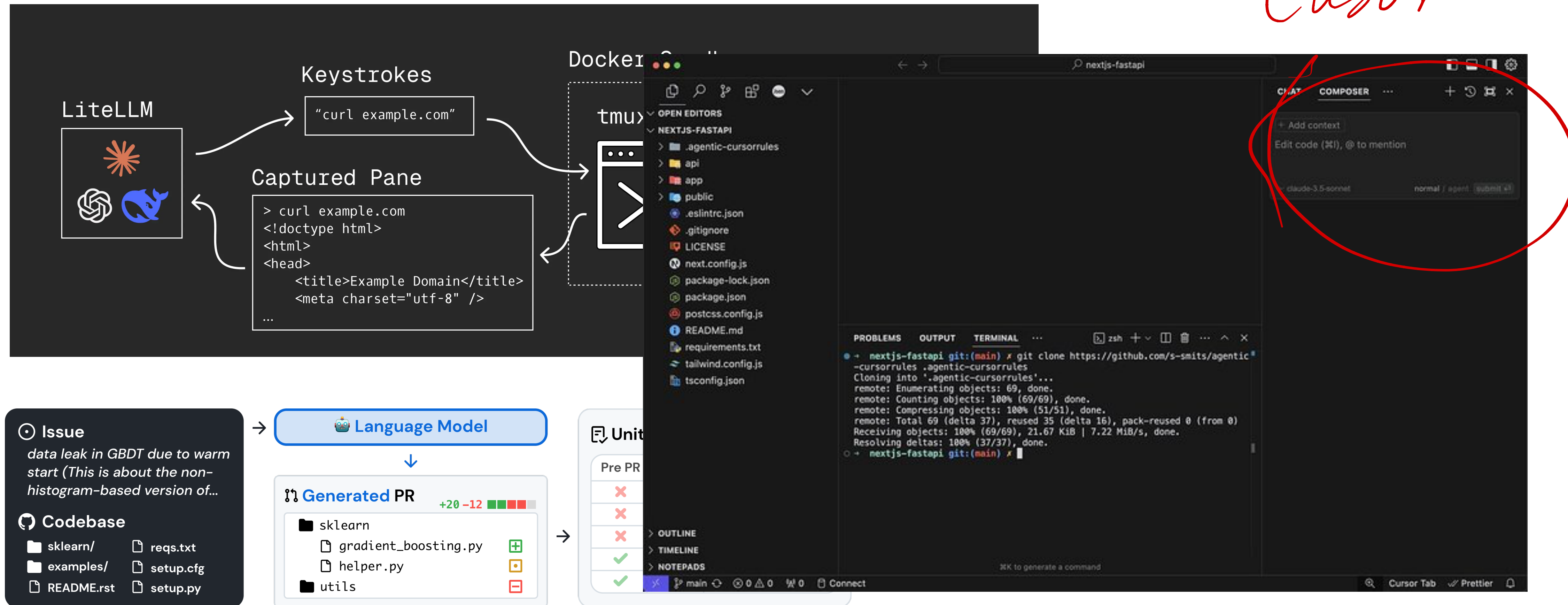
Environments and benchmarks typically come together



RL Environments

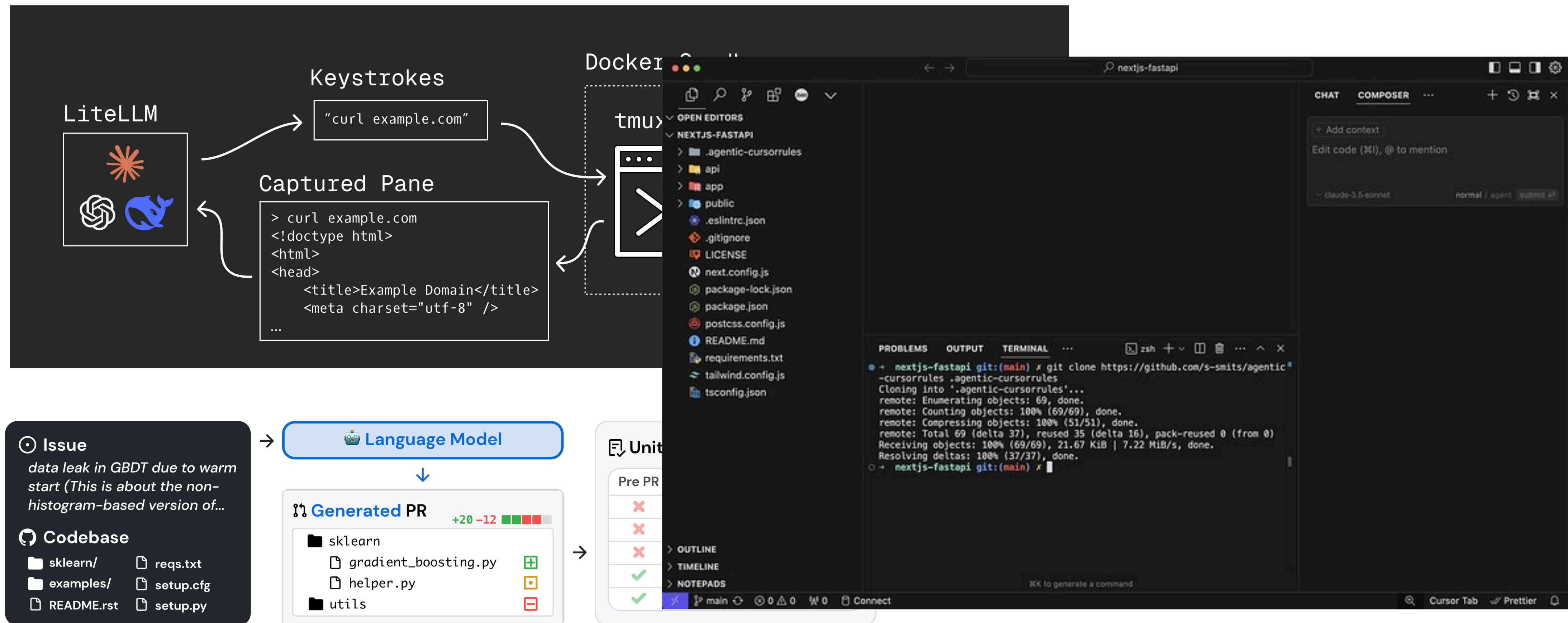
Environments and benchmarks typically come together

Cursor



RL Environments

Environments and benchmarks typically come together



Research and Products are really close nowadays, and we can directly RL in real, product-level environments

Thank You!