

Language Models as Agents

Frank Xu

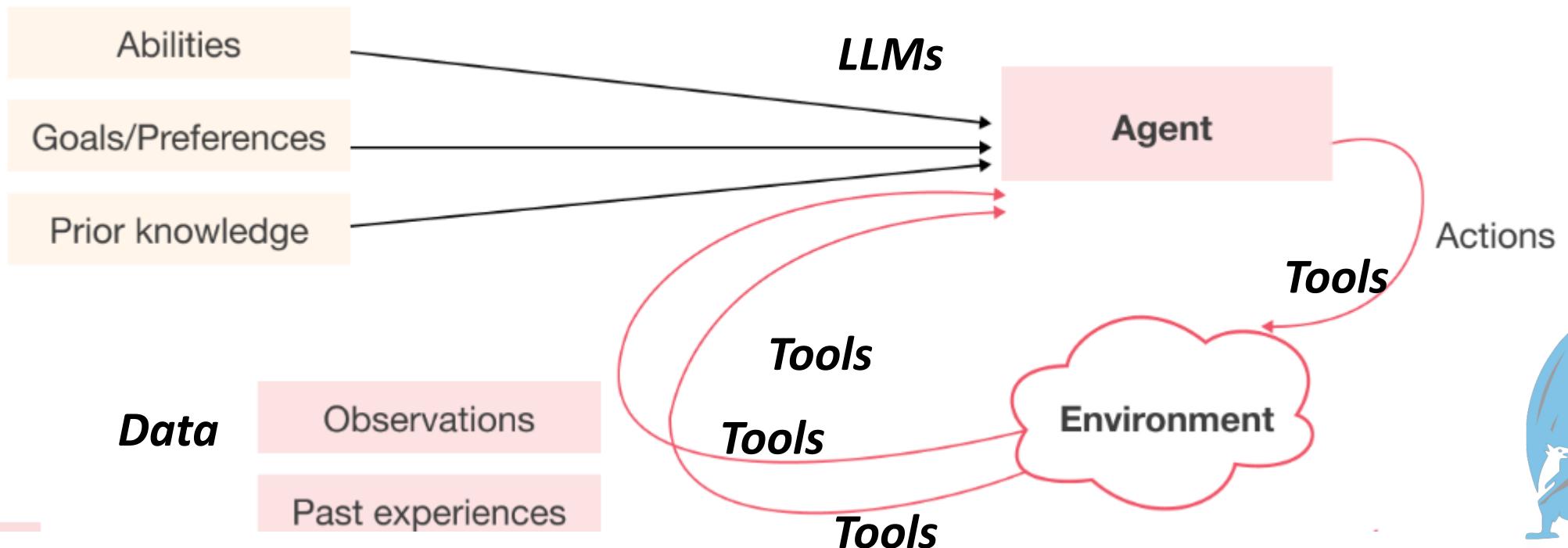


Carnegie Mellon University
Language Technologies Institute RG



What are agents?

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



How to get started in LLM Agents

- Tasks and Applications
- Training-free Methods for Building Agents
- Evaluation Environment and Benchmark
- Training Methods for Improving Agents

Tasks and Applications For LLM Agents

Why do we want agents?

Imagine if things get done by just talking...



How People Interact with Computers?

- Traditionally ...

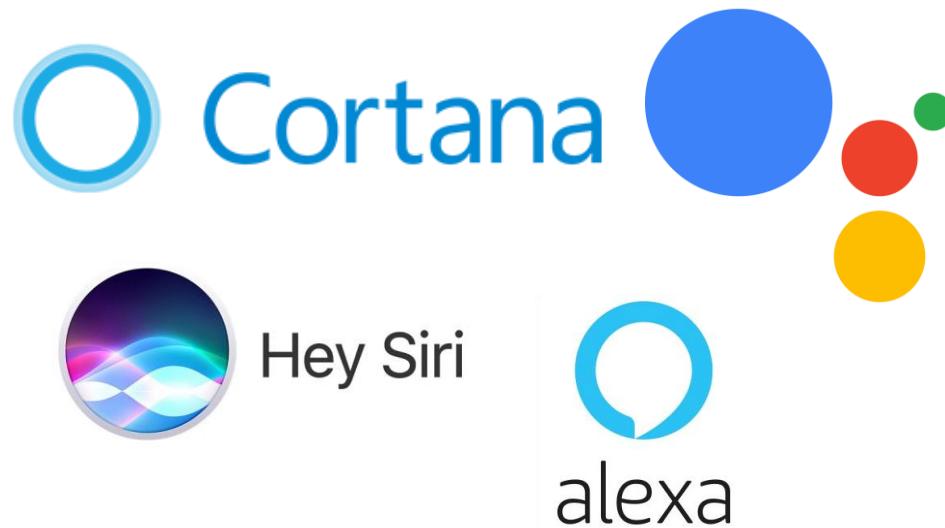
- Traditionally ...



- What about using natural language?
 - Save time, natural, accessible, no need to browse, no programming learning curve, etc.



Natural Language Interfaces to Computers



A screenshot of a Python code editor window titled 'Untitled-1'. The code is as follows:

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

The third line, 'sorted(my_list, reverse=True)', is highlighted in green. The status bar at the bottom shows '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



Tool Integrations into Chatbots

ChatGPT plugins

We've implemented initial support for plugins in ChatGPT. Plugins language models help ChatGPT do computations, o

ChatGPT plugins

 Expedia Bring your trip plans to life—get there, stay there, find things to see and do.	 FiscalNote Provides and enables access to select market-leading, real-time data sets for legal, political, and regulatory data and information.	 Instacart Order from your favorite local grocery stores.	 KAYAK Search for flights, stays and rental cars. Get recommendations for all the places you can go within your budget.
 Klarna Shopping Search and compare prices from thousands of online shops.	 Milo Family AI Giving parents superpowers to turn the manic to magic, 20 minutes each day. Ask: Hey Milo, what's magic today?	 OpenTable Provides restaurant recommendations, with a direct link to book.	 Shop Search for millions of products from the world's greatest brands.
 Speak Learn how to say anything in another language with Speak, your AI-powered language tutor.	 Wolfram Access computation, math, curated knowledge & real-time data through Wolfram Alpha and Wolfram Language.	 Zapier Interact with over 5,000+ apps like Google Sheets, Trello, Gmail, HubSpot, Salesforce, and more.	 ABN ASIA.ORG

Robots

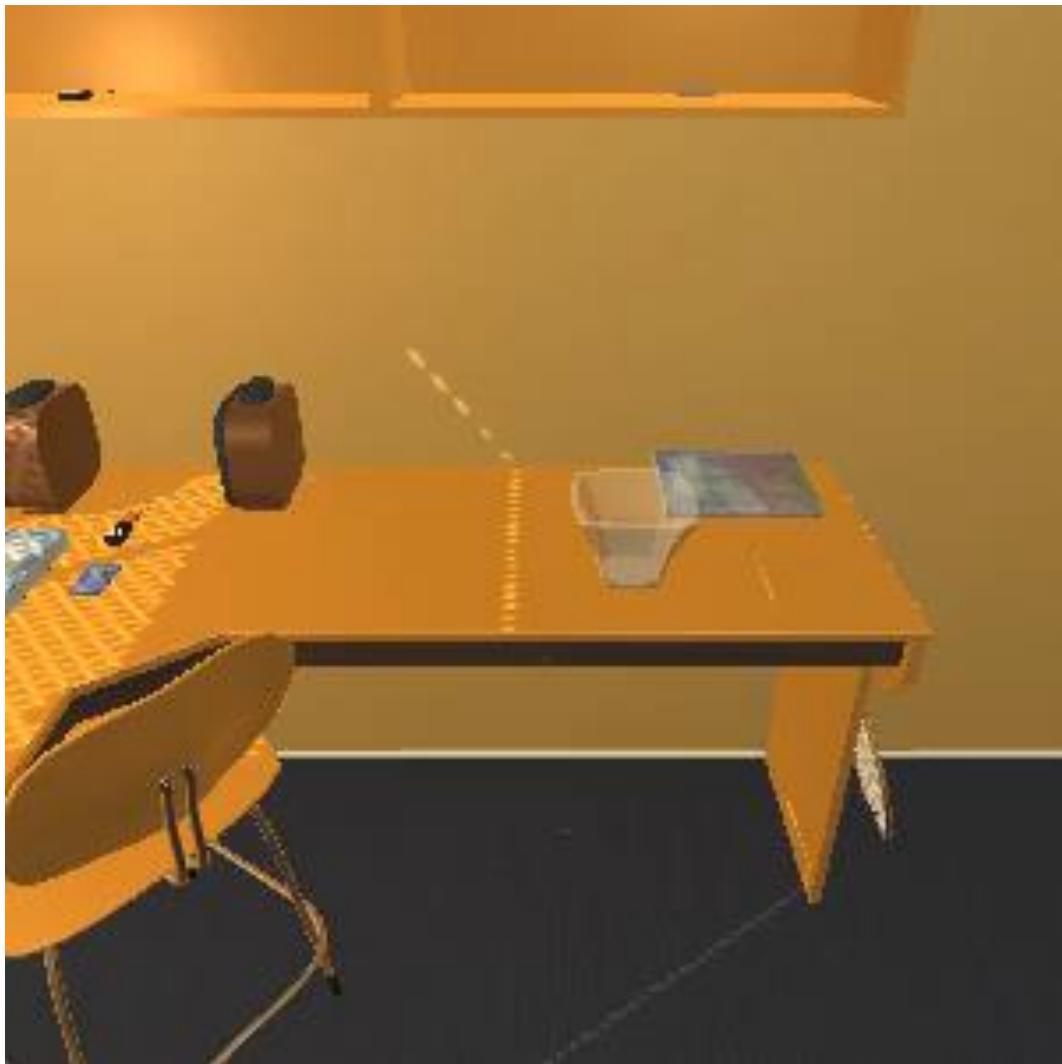


Turn and go with the flow of traffic. At the first traffic light turn left. Go past the next two traffic light, As you come to the third traffic light you will see a white building on your left with many American flags on it. Touchdown is sitting in the stars of the first flag.

⋮



Robots



You are in the middle of a room. Looking quickly around you, you see a safe 1, a shelf 4, a drawer 2, a bed 1, a drawer 1, a shelf 5, a shelf 2, a sidetable 2, a shelf 3, a drawer 3, a shelf 1, a sidetable 1, a desk 1, and a garbagecan 1.

Your task is to: examine an alarmclock with the desklamp.

> go to desk 1

You arrive at loc 8. On the desk 1, you see a pen 1, a bowl 1, a alarmclock 2, a pencil 2, a pencil 3, a creditcard 3, a book 1, a alarmclock 3, a keychain 3, and a book 2.

> take alarmclock 2 from desk 1

You pick up the alarmclock 2 from the desk 1.



Games

Open-ended Environments

Craft Glass Bridge



Build Oak House



Make Ice Igloo



Combat Zombie



Fish Squid



Farm Sugar Cane



Find Ocean Monument



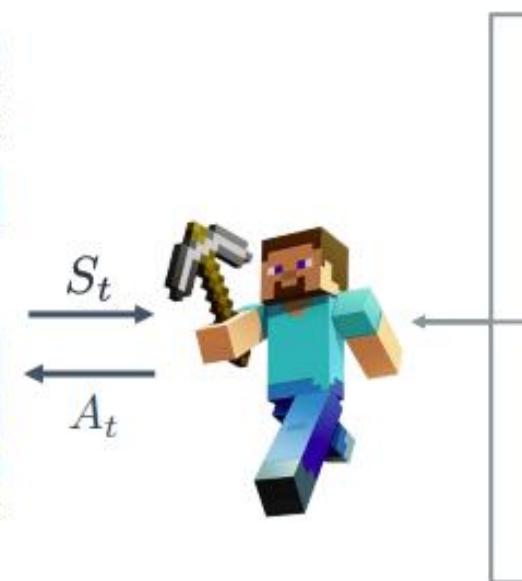
Explore Desert Temple



Treasure Hunt in End City



Generalist Agent



Games



No Man's Sky: Go to spaceship

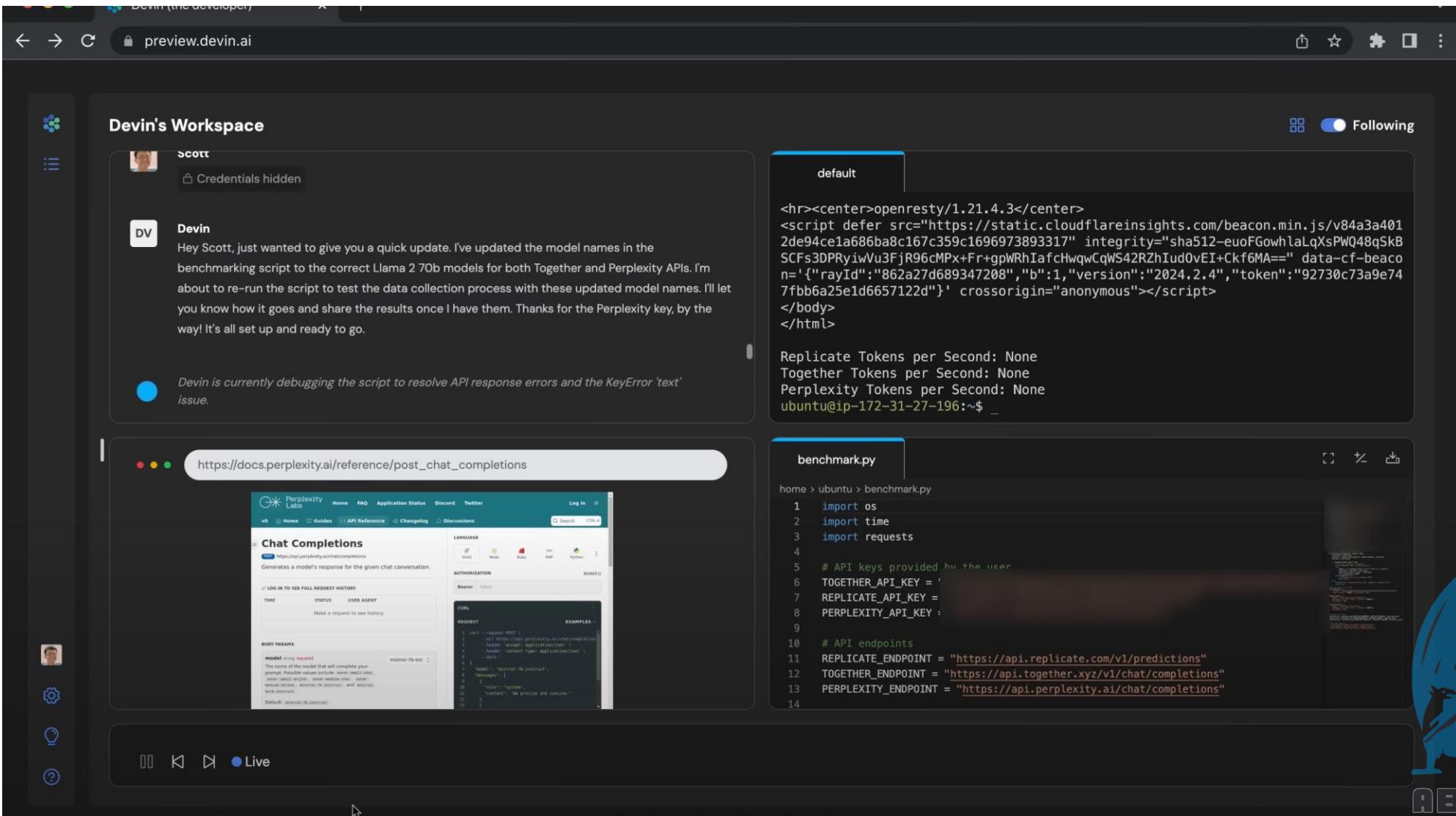


No Man's Sky: Shoot Asteroid



Software Development

Devin (https://www.cognition-labs.com/introducing-devin)



The screenshot shows a web browser window with two tabs. The top tab is titled "Devin (the developer)" and the URL is "preview.devin.ai". The content of this tab is "Devin's Workspace", showing a message from Devin to Scott. Devin's message says: "Hey Scott, just wanted to give you a quick update. I've updated the model names in the benchmarking script to the correct Llama 2 70b models for both Together and Perplexity APIs. I'm about to re-run the script to test the data collection process with these updated model names. I'll let you know how it goes and share the results once I have them. Thanks for the Perplexity key, by the way! It's all set up and ready to go." Below the message, a note says: "Devin is currently debugging the script to resolve API response errors and the KeyError 'text' issue." The bottom tab is titled "https://docs.perplexity.ai/reference/post_chat_completions" and the URL is "https://docs.perplexity.ai/reference/post_chat_completions". The content of this tab shows a screenshot of the Perplexity API documentation for "Chat Completions". The right side of the browser window displays a code editor with a Python script named "benchmark.py". The script is used for benchmarking API endpoints. It imports os, time, and requests, and defines API keys for TOGETHER_API_KEY, REPLICATE_API_KEY, and PERPLEXITY_API_KEY. It then defines API endpoints for REPLICATE_ENDPOINT, TOGETHER_ENDPOINT, and PERPLEXITY_ENDPOINT. The script includes examples of curl commands for making requests to these endpoints.

```
<hr><center>openresty/1.21.4.3</center>
<script defer src="https://static.cloudflareinsights.com/beacon.min.js/v84a3a4012de94ce1a686ba8c167c359c1696973893317" integrity="sha512-euoFGoWlalQxsPwQ48qSkBSCFs3DPRyiwVu3FjR96cMPx+Fr+gpWRhIafcHwqwCqWS42RZhIud0vEI+Ckf6MA==" data-cf-beacon={"rayId":"862a27d689347208","b":1,"version":"2024.2.4","token":"92730c73a9e747fb6a25e1d6657122d"}' crossorigin="anonymous"></script>
</body>
</html>

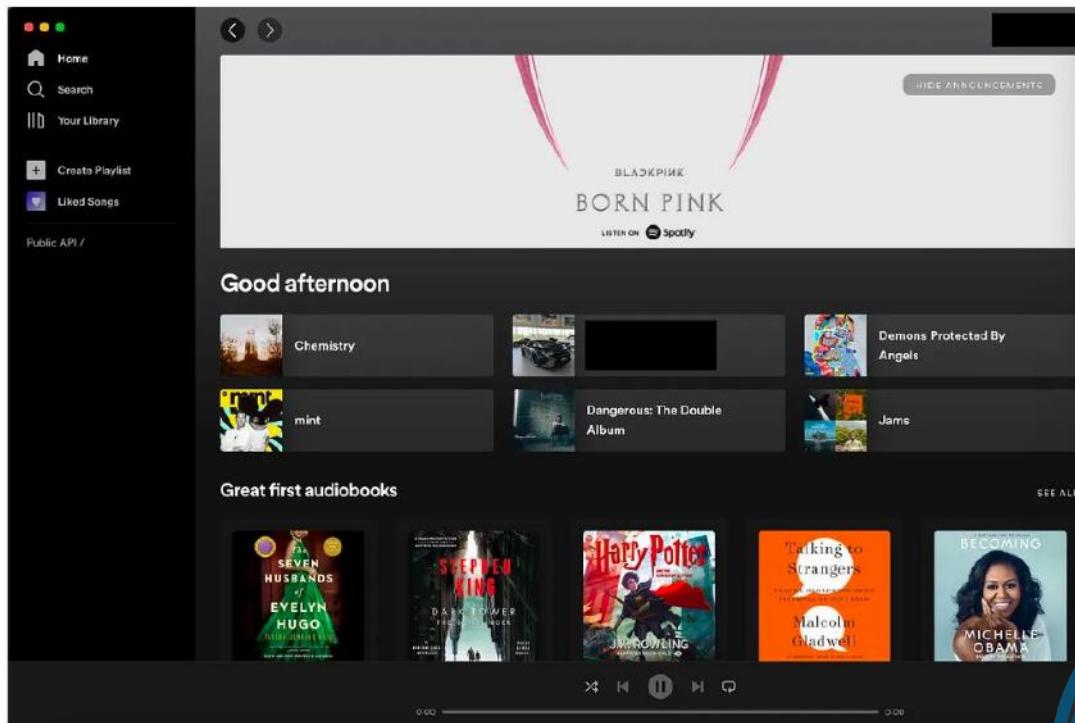
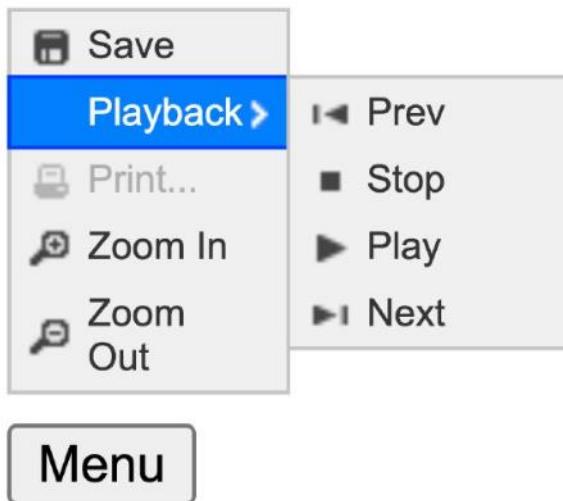
Replicate Tokens per Second: None
Together Tokens per Second: None
Perplexity Tokens per Second: None
ubuntu@ip-172-31-27-196:~$
```

```
1 import os
2 import time
3 import requests
4
5 # API keys provided by the user
6 TOGETHER_API_KEY =
7 REPLICATE_API_KEY =
8 PERPLEXITY_API_KEY =
9
10 # API endpoints
11 REPLICATE_ENDPOINT = "https://api.replicate.com/v1/predictions"
12 TOGETHER_ENDPOINT = "https://api.together.xyz/v1/chat/completions"
13 PERPLEXITY_ENDPOINT = "https://api.perplexity.ai/chat/completions"
14
```



UI Automation

Click the "Menu" button, and then find and click on the item with the ►| icon.

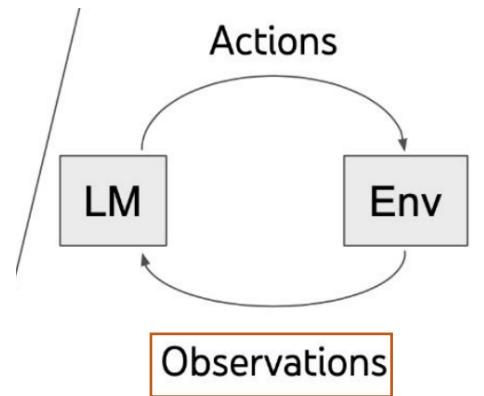


"Play some synthwave songs"

Training-free Methods for Building Agents



How to let LLM become an agent



- Observation

- Text input
- Visual Input
- Audio Input
- Structured Input

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.



- Need for Multimodal LLMs

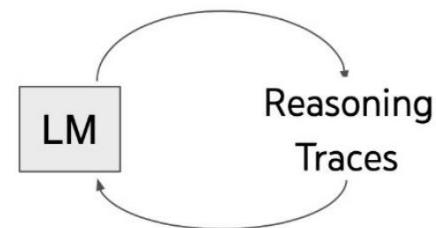


How to let LLM become an agent

- Planning and reasoning ability

Chain-of-thoughts (CoT)

"Let's think step by step"



CoT (Wei et al. 22')

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.

Ask LLM:

What should I do next? Let's think step by step:

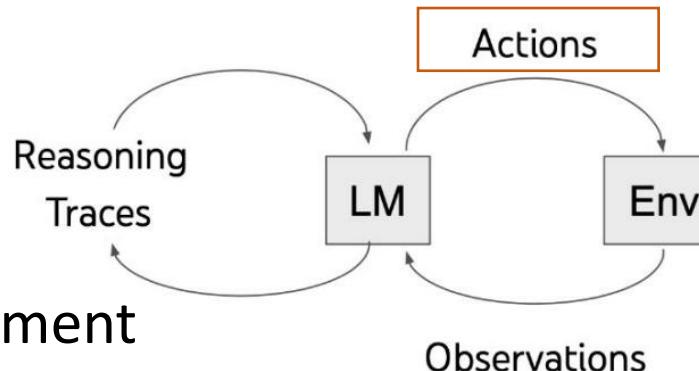
First I need to find a pepper shaker ... more likely to appear in cabinets (1-6), countertops (1-3) ...

After I find pepper shaker 1, next I need to put it on drawer 1



How to let LLM become an agent

- Tool-use ability
 - Generate action calls
 - Execute the actions in environment
 - Put new observation back in prompt



Toolformer (Schick et al. 23')
ReAct (Yao et al. 23')

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.

Ask LLM:

What should I do next? Let's think step by step:

First I need to find a pepper shaker ... more likely to appear in cabinets (1-6), countertops (1-3) ...

Action: GOTO Cabinet 1

Observation: On cabinet 1, there is a vase 2

...

How to let LLM become intelligent

- Planning and reasoning ability
 - Chain-of-thoughts (CoT)
 - "Let's think step by step"
- Tool-use ability
 - Generate API calls with arguments
 - Execute the API calls
 - Get API call results as new observation
- What if lots of APIs?

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.



How to let LLM become an AI developer

- Generate code to perform the task
Reasoning + planning + action unified
PAL (Gao et al. 23')

YI

Assuming you can use Python, you have installed all the common APIs and you have completed all the authentication steps, answer the following task.

Task 1: set up a meeting with Noah tomorrow at 10am.



I'm sorry, but as an AI language model, I don't have the capability to directly set up a meeting or send invitations. However, I can guide you on how to use Python to interact with calendar APIs to set up a meeting.

Here's a sample code to set up a meeting with Noah using the Google Calendar API in Python:

```
python

import datetime
import pytz
from google.oauth2.credentials import Credentials
from googleapiclient.discovery import build

# Set up the credentials
creds = Credentials.from_authorized_user_file('token.json', SCOPES)

# Set up the calendar service
service = build('calendar', 'v3', credentials=creds)

# Set up the meeting details
event = {
    'summary': 'Meeting with Noah',
    'location': 'Virtual',
    'description': 'Discussing project updates',
    'start': {
        'dateTime': datetime.datetime(2023, 2, 22, 10, 0, 0, tzinfo=pytz.UTC),
        'timeZone': 'UTC',
    },
    'end': {
        'dateTime': datetime.datetime(2023, 2, 22, 11, 0, 0, tzinfo=pytz.UTC),
        'timeZone': 'UTC',
    },
    'reminders': {
        'useDefault': True,
    },
}

# Create the event
event = service.events().insert(calendarId='primary', body=event).execute()

print(f'Event created: {event.get("htmlLink")}')
```



ABN ASIA.ORG

Evaluation Environment and Benchmark

Evaluation of LLM Agents

- Simplified environments and basic tasks
- Performance is saturating.

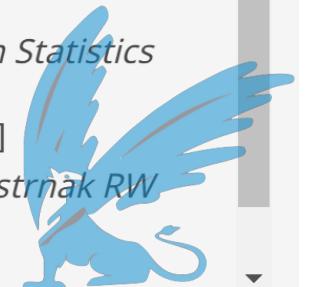
1. Stateless, non interactive environment, e.g. *Mind2Web (Deng et al. 2023)* has only dumped pages.
2. Checking action sequence accuracy (step-wise, surface form only)
3. Simple interactive environment, short horizon, e.g. *WebShop (Yao et al. 2023)*, *MiniWoB++ (Humphreys et al. 2022)*

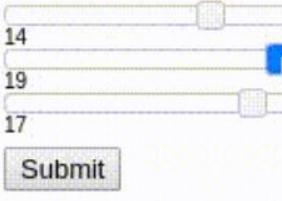
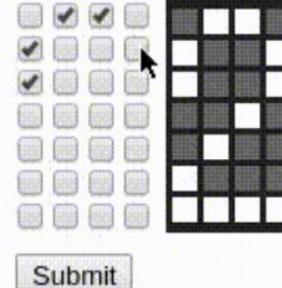
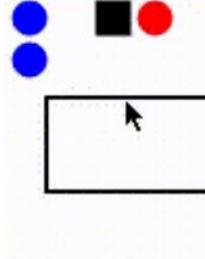
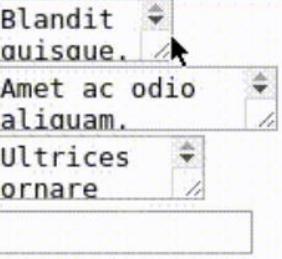
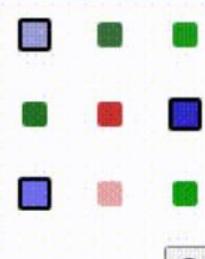
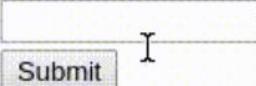
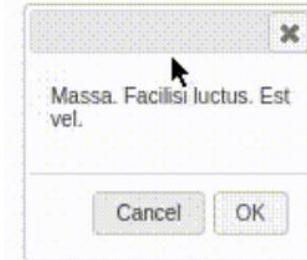
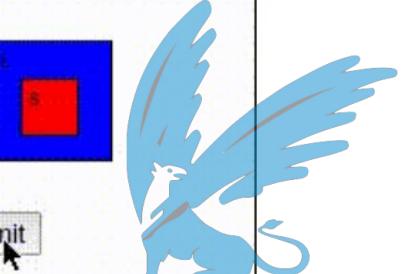
Task

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 Pastrňák RW
6. CLICK ➤ [button]



<p>Move the cube around so that "5" is the active side facing the user.</p>  <p>Submit</p>	<p>Set the sliders to the combination [13,20,13] and submit.</p> 	<p>Draw the number "2" in the checkboxes using the example on the right and press Submit when finished.</p>  <p>Submit</p>	<p>Drag Ree to the 4th position.</p>  <p>Submit</p>	<p>Keep your mouse inside the circle as it moves around.</p>  <p>Submit</p>	<p>Enter the value of Country into the text field and press Submit.</p> <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>1934</td></tr> <tr><td>Religion</td><td>Hinduism</td></tr> <tr><td colspan="2"> </td></tr> <tr><td colspan="2">Submit</td></tr> </table>	Gender	Male	First name	Annecorinne	Country	Guam	Year of Birth	1934	Religion	Hinduism			Submit	
Gender	Male																		
First name	Annecorinne																		
Country	Guam																		
Year of Birth	1934																		
Religion	Hinduism																		
Submit																			
<p>Drag all triangles into the black box.</p>  <p>Submit</p>	<p>Select 09/23/2016 as the date and hit submit.</p>  <p>Submit</p>	<p>Sort the numbers in increasing order, starting with the lowest number at the top of the list.</p>  <p>Submit</p>	<p>Copy the text from the 1st text area below and paste it into the text input.</p>  <p>Submit</p>	<p>Select all the shades of blue and press Submit.</p>  <p>Submit</p>	<p>Find the 4th word in the paragraph, type that into the textbox and press "Submit".</p> <p>Non arcu ut ultricies est. Gravida gravida. Porta erat nulla eget condimentum posuere a...</p>  <p>Submit</p>														
<p>Click the button in the dialog box labeled "Cancel".</p> 	<p>Highlight the text in the paragraph below and click submit.</p> <p>Ultricies. Sagittis in.</p> <p>Submit</p>	<p>Highlight the text in the paragraph below and click submit.</p> <p>Tempor posuere nibh. Vel nisl, faucibus. Feugiat condimentum</p> <p>Submit</p>	<p>Find the 11th word in the paragraph, type that into the textbox and press "Submit".</p> <p>Ullamcorper aliquet amet ullamcorper. Elit. Mattis luctus diam. Lobortis nulla fermentum ornare faucibus</p> <p>Submit</p>	<p>Move the cube around so that "2" is the active side facing the user.</p>  <p>Submit</p>	<p>Drag the smaller box so that it is completely inside the larger box.</p>  <p>Submit</p>														



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

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.

WebArena Environment Design

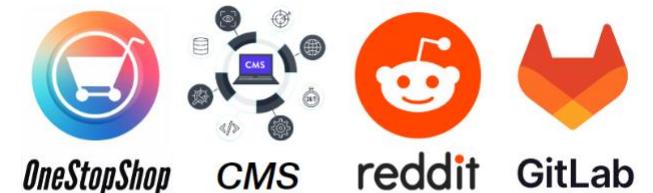
Environment:

- 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.

Self-hosted fully functional web applications



OneStopShop

CMS

reddit

GitLab



Toolbox

Knowledge resources

WebArena



WebArena (Zhou et al. 23')

ABMASIA.ORG

Collecting Realistic Intents

We checked our own web browsing history!

- Information seeking
 - *“When was the last time I bought shampoo?”*
- Site navigation
 - *“Checkout merge requests assigned to me.”*
- Content & configuration operation
 - *“Post my question, “is a car necessary in NYC”, in a subreddit where I’m likely to get an answer.”*

Example Tasks in WebArena



“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

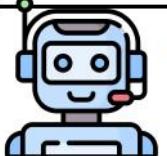
List of museums in Pittsburgh

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](#)

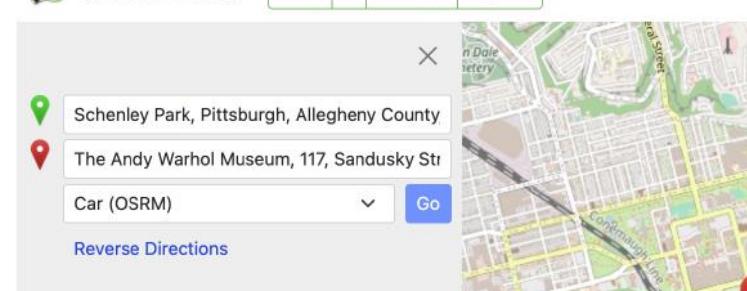
Museums



Search for museums in Pittsburgh

webarena.openstreetmap.com

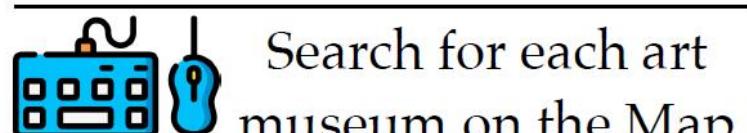
OpenStreetMap Edit History Export



Directions

Distance: 7.1km. Time: 0:10.

1. Start on Panther Hollow Road
2. Slight right onto unnamed road



Search for each art museum on the Map

webarena.gitlab.com

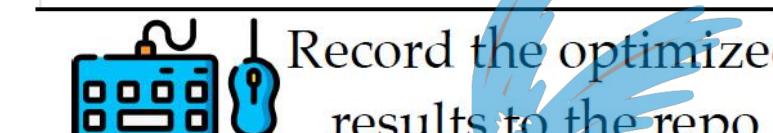
Update README.md

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



ABMASIA.ORG

Outcome/Execution-based Evaluation

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”

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

Need your answer

1

Submitted by convexeggtarxxx 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>
      ...
    ...
  ...
</div>
```

“f/nyc” in page.url

“Is car necessary in NYC?” in
document.querySelector(“.su
bmission__inner”).outText



ABN ASIA.ORG

Observation & Action Space



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.

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

★★★★★ 12 Reviews

\$49.99

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

Screenshot

Keyboard: type

Text

```
<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
<span>Reviews</span></a>
```

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

Mouse: click, hover, scroll

Browser: New tab, go back



Prompting LLM as Agent

- ***Few-shot in-context learning:*** General guideline + two examples

You are an autonomous intelligent agent tasked with navigating a web browser. You will be given web-based tasks. These tasks will be accomplished through the use of specific actions you can issue.

You can observe the following information:

...

You can do the following actions:

...

...

Prompting LLM as Agent

- ***Few-shot in-context learning:*** General guideline + two examples

Example input:

OBSERVATION:

[1744] link 'HP CB782A#ABA 640 Inkjet Fax Machine (Renewed)'

[1757] button 'Add to Cart'

URL: <http://onestopmarket.com/office-products/office-electronics.html>

OBJECTIVE: What is the price of HP Inkjet Fax Machine

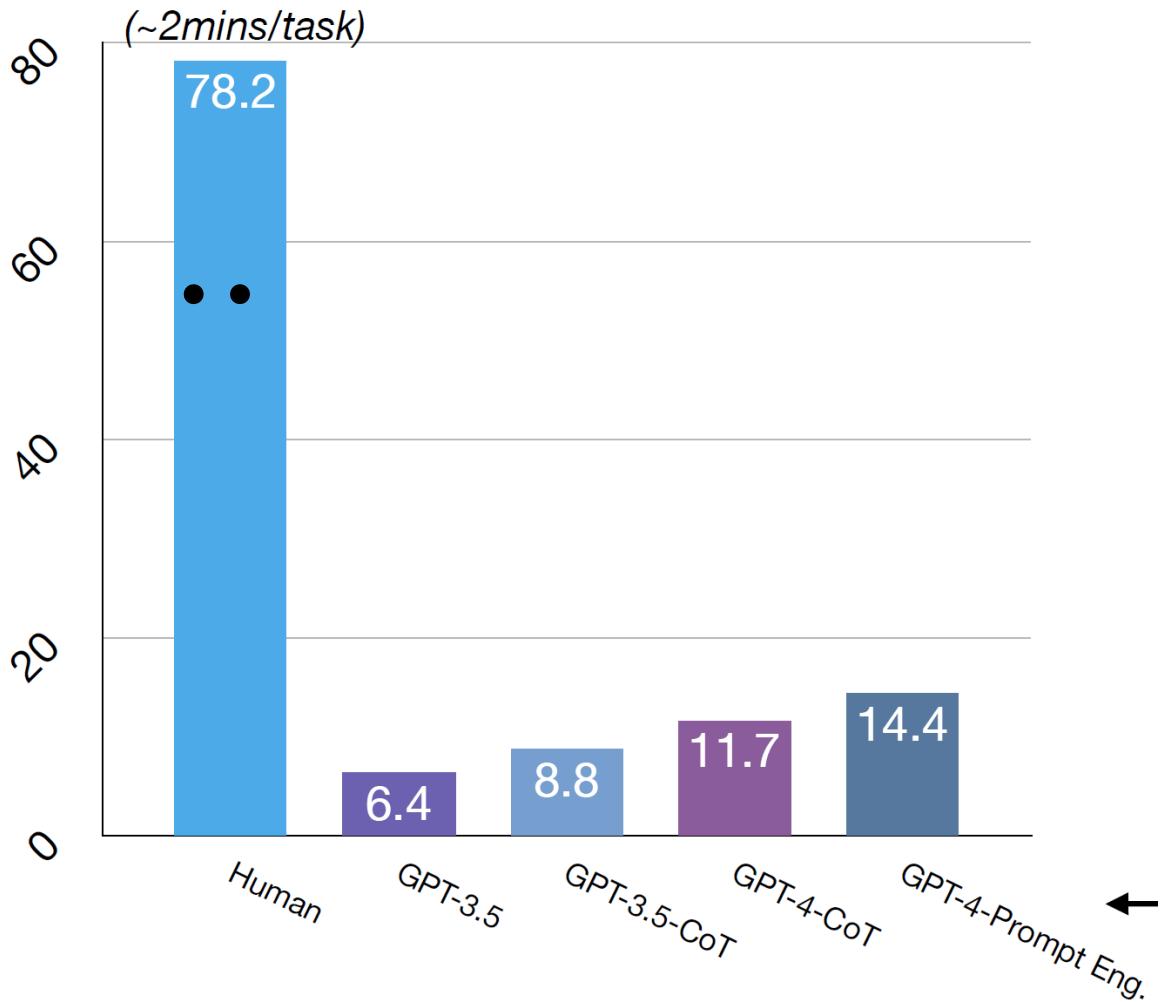
PREVIOUS ACTION: None

Example output:

Let's think step-by-step. This page lists ... the next action I will perform is **click [1744]**
(Optional chain-of-thought reasoning)



WebArena is Challenging



- Chain-of-thought prompting provides *limited* benefits.
- GPT-4 remains *significantly behind* human performance.
- Prompt engineering emphasizes LLMs' *sensitivity* to subtle instruction changes that typically don't affect humans.



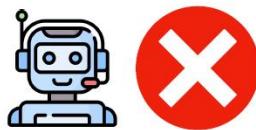
Failures: Not Knowing How



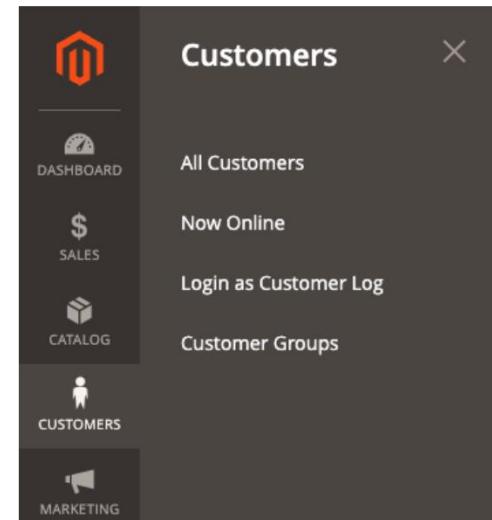
Show me the customers who have expressed dissatisfaction with Olivia zip jacket



Either going to the **catalog (product)** section or the **marketing (review)** section



Decided to go to **customers** section which is not easy to select and filter reviews



Failures: Not being Accurate

“... and set the due date to 2023/12/23”

Due date

2023/12/23



“... and set the due date to 2023-12-13”

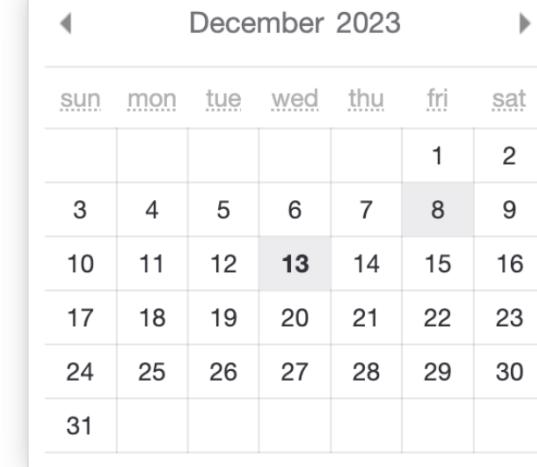
Due date

2023-12-13



Due date

2023-12-13



Failures: Trivial Errors



Search

Search query

DMV area

Search

50 results for **DMV area**:

[2430] **searchbox 'Search query'**
[5172] **StaticText 'DMV area'**

- GPT-4 : 21% examples failed due to repeated typing.
- May be related to hallucination effect, generates repeated actions
- Irrelevant content in a webpage hurts!

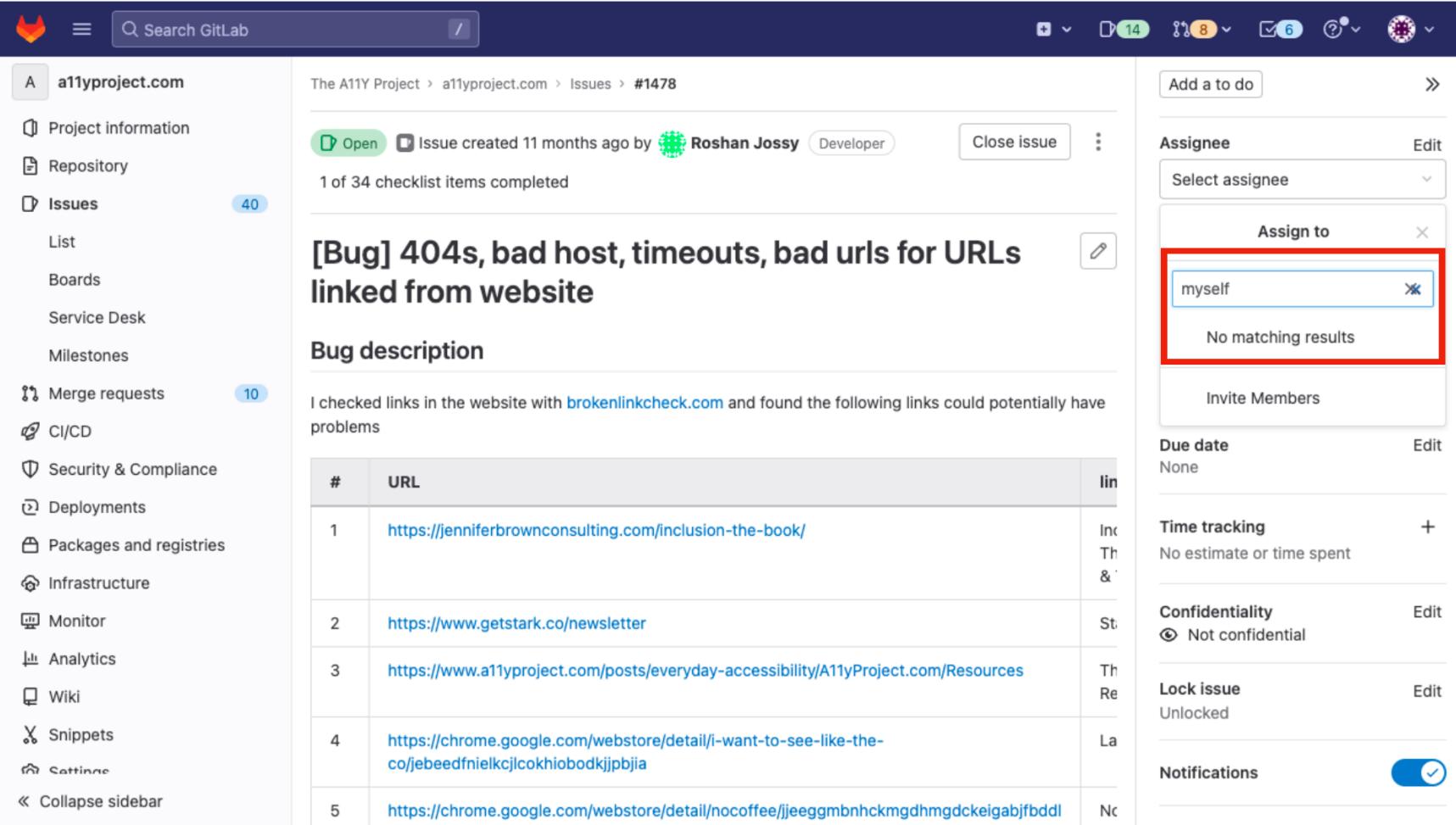
Search query

DMV areaDMV areaDMV areaDMV area

Search

Failures: (Not So) Trivial Errors

- “Assign this issue to myself.”



The screenshot shows a GitLab issue page for a11yproject.com. The sidebar on the left lists various project sections: Project information, Repository, Issues (40), List, Boards, Service Desk, Milestones, Merge requests (10), CI/CD, Security & Compliance, Deployments, Packages and registries, Infrastructure, Monitor, Analytics, Wiki, Snippets, and Settings. The main content area shows an issue titled "[Bug] 404s, bad host, timeouts, bad urls for URLs linked from website". The issue was created 11 months ago by Roshan Jossy (Developer). The description states: "I checked links in the website with brokenlinkcheck.com and found the following links could potentially have problems". A table lists five URLs with their respective line numbers and descriptions. The right sidebar contains fields for "Add a to do", "Assignee" (set to "Select assignee"), "Assign to" (with "myself" selected, highlighted with a red box), "Due date" (None), "Time tracking" (No estimate or time spent), "Confidentiality" (Not confidential), "Lock issue" (Unlocked), and "Notifications" (switched on). A "No matching results" message is displayed in the assignee dropdown.

#	URL	lin
1	https://jenniferbrownconsulting.com/inclusion-the-book/	Inc Th &
2	https://www.getstark.co/newsletter	St
3	https://www.a11yproject.com/posts/everyday-accessibility/A11yProject.com/Resources	Th Re
4	https://chrome.google.com/webstore/detail/i-want-to-see-like-the-co/jebeedfnielkjcjlcohobiobdkjjpbjia	La
5	https://chrome.google.com/webstore/detail/nocoffee/jieggbmnbhckmgdmgdckiegabjfbddl	No



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*

In-context Learning

- LM performs a task just by ***conditioning on input-output examples***, without optimizing any parameters.

Circulation revenue has increased by 5% in Finland.

\n Positive

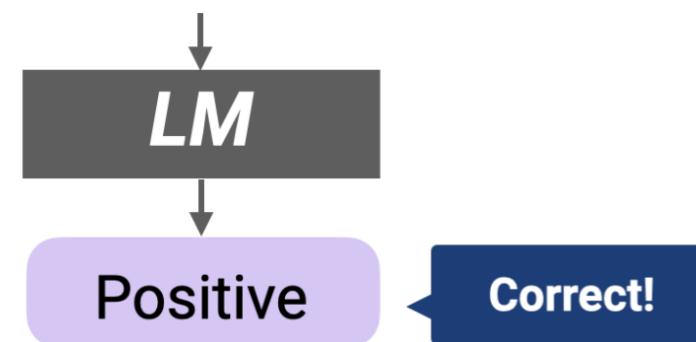
Panostaja did not disclose the purchase price.

\n Neutral

Paying off the national debt will be extremely painful.

\n Negative

The company anticipated its operating profit to improve. \n _____



In-context Learning

example_user

OBSERVATION:

[1744] link 'HP CB782A#ABA 640 Inkjet Fax Machine (Renewed)'

[1749] StaticText '\$279.49'

[1757] button 'Add to Cart'

[1760] button 'Add to Wish List'

[1761] button 'Add to Compare'

URL: <http://onestopmarket.com/office-products/office-electronics.html>

OBJECTIVE: What is the price of HP Inkjet Fax Machine

PREVIOUS ACTION: None

example_assistant

Let's think step-by-step. This page lists the information of HP Inkjet Fax Machine, which is the product identified in the objective. Its price is \$279.49. I think I have achieved the objective. I will issue the stop action with the answer. In summary, the next action I will perform is stop
[\$279.49]` ``

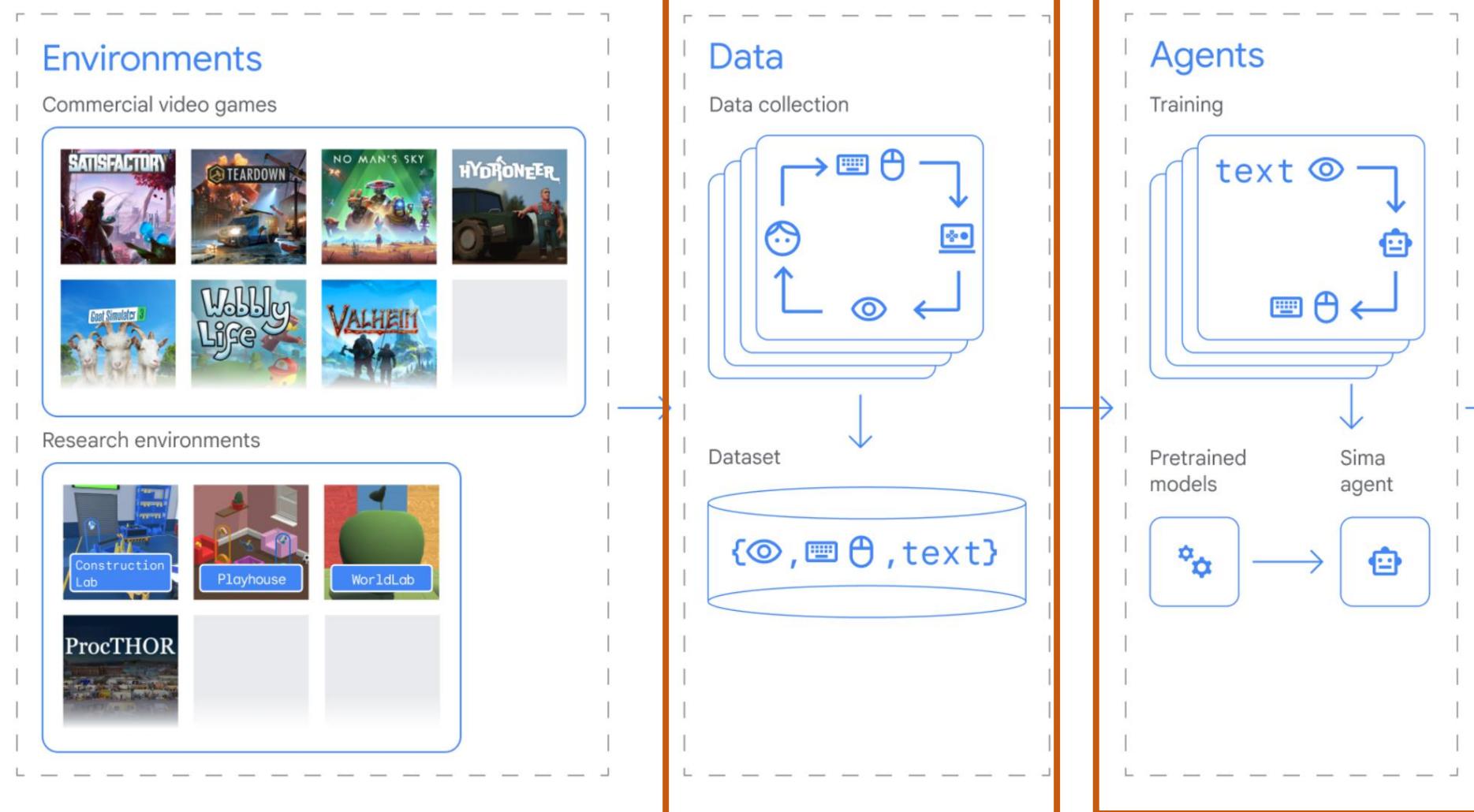
Supervised Finetuning

- 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

Create More Training Data

Open-ended Environments

Craft Glass Bridge



Build Oak House



Make Ice Igloo



Combat Zombie



Fish Squid



Farm Sugar Cane



Find Ocean Monument



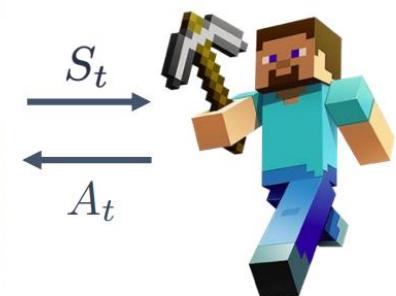
Explore Desert Temple



Treasure Hunt in End City



Generalist Agent

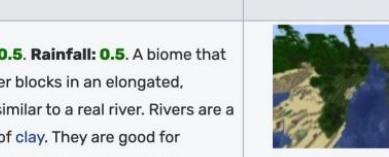


Internet-scale Knowledge Base

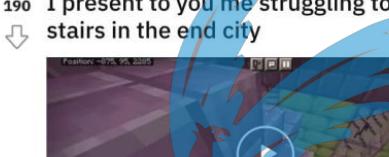
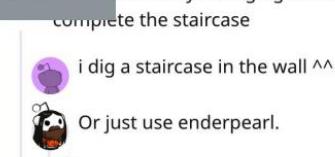
YouTube



Wiki

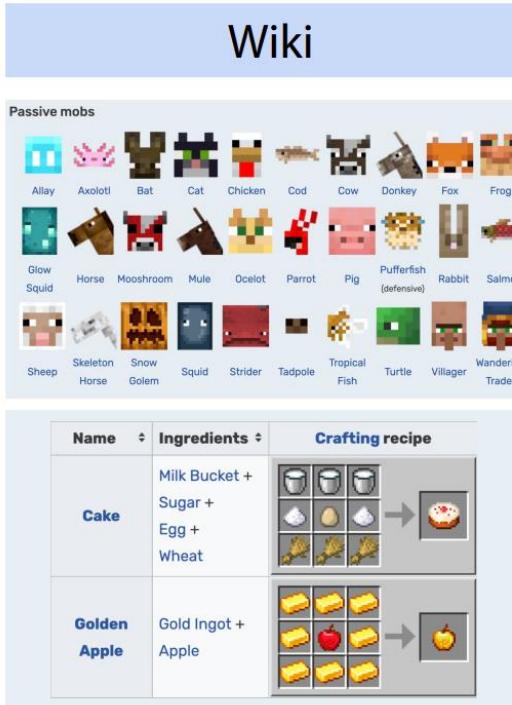
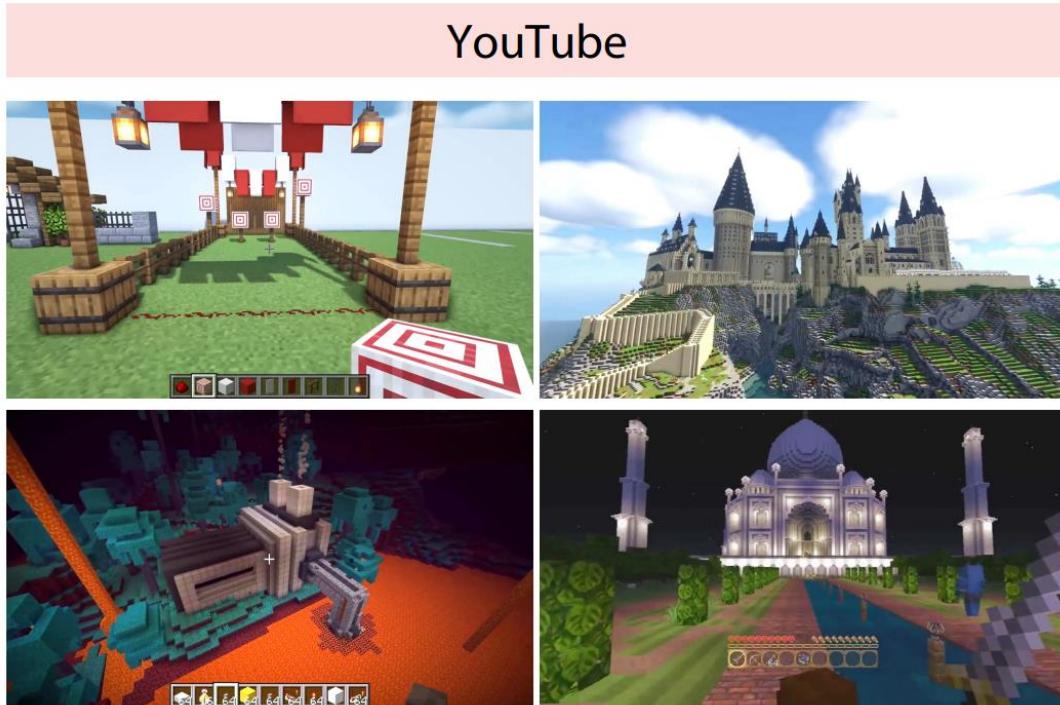


Reddit



Data Augmentation

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



MineDojo, Fan et al. 22'
Don't Stop Pretraining, 6

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.

Explain reinforcement learning to a 6 year old.

A
In reinforcement learning, the agent is...
B
Explain rewards...
C
In machine learning...
D
We give treats and punishments to teach...

A labeler ranks the outputs from best to worst.



D > C > A > B

This data is used to train our reward model.



D > C > A > B

A new prompt is sampled from the dataset.



Write a story about otters.



PPO

The PPO model is initialized from the supervised policy.

Once upon a time...



RM

The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

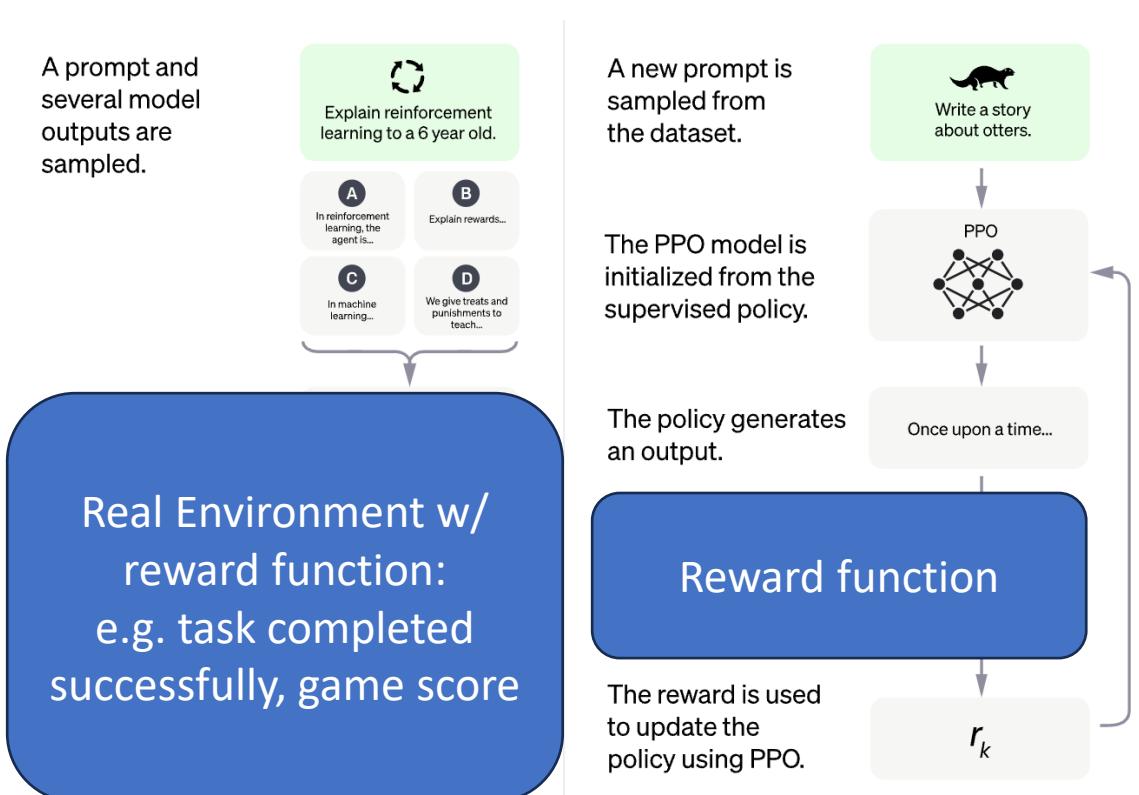
r_k



Reinforcement Learning

Compared to RLHF:

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



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
- Scalability
 - The environment takes 10 seconds to env.step()
 - The reward function takes 100 seconds to get a scalar reward

Reinforcement Learning

- No access to policy model parameters? (e.g. GPT-4)
- We can train with environment reward a second LLM (smaller open source models), that generates additional “reflection” prompts.

