



香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

CSC6052/5051/4100/DDA6307/ MDS5110 Natural Language Processing

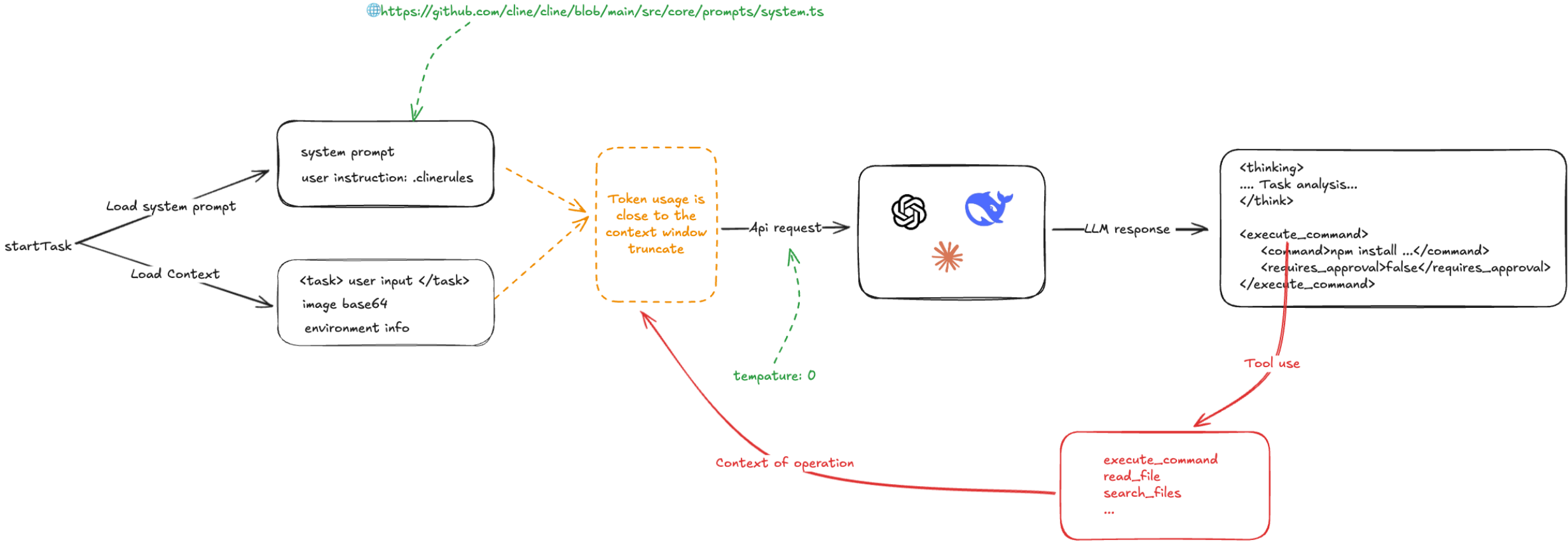
Lecture 7-2: LLM Agents Cont.

Spring 2025
Benyou Wang
School of Data Science

Agent Application - Cline

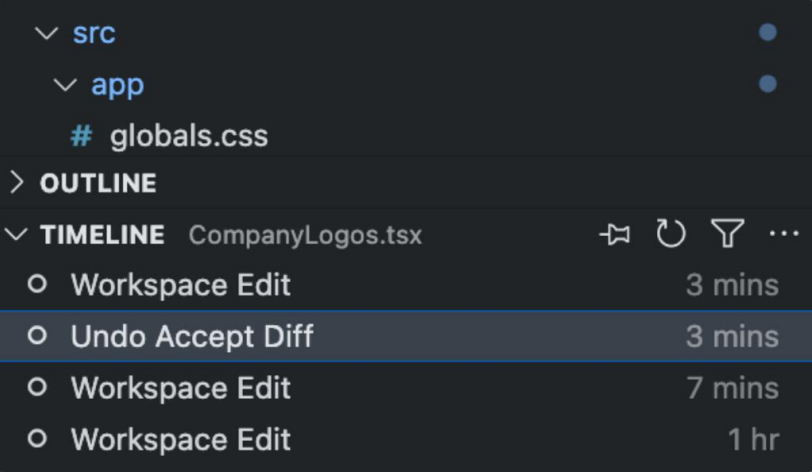
What is Cline?

Cline is an AI assistant built in VSCode that can use your CLI and Editor.

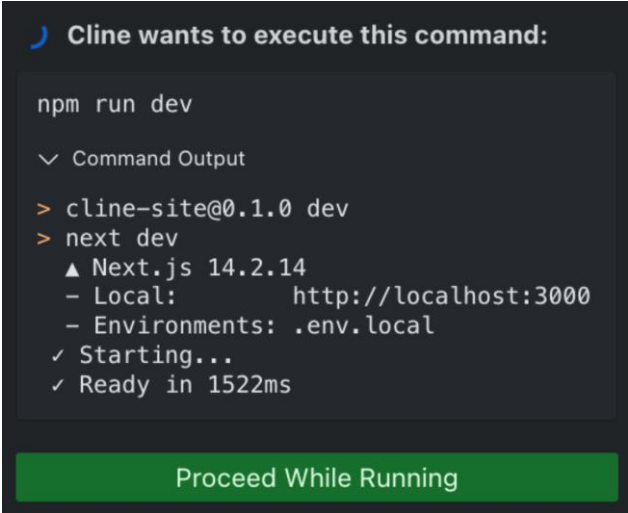


What can Cline do?

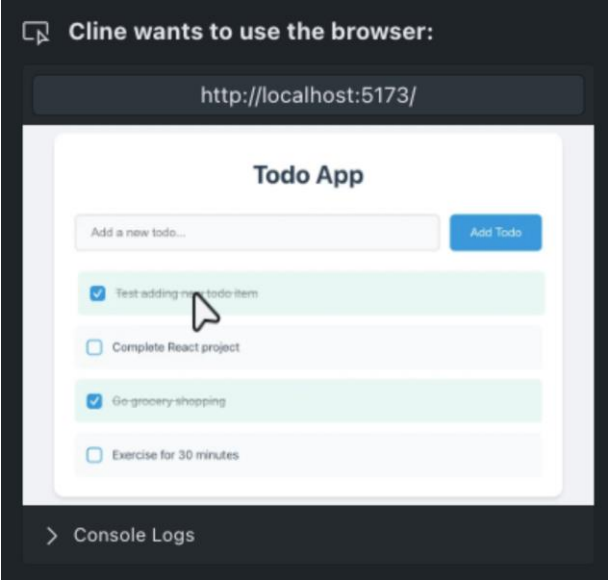
Create and Edit



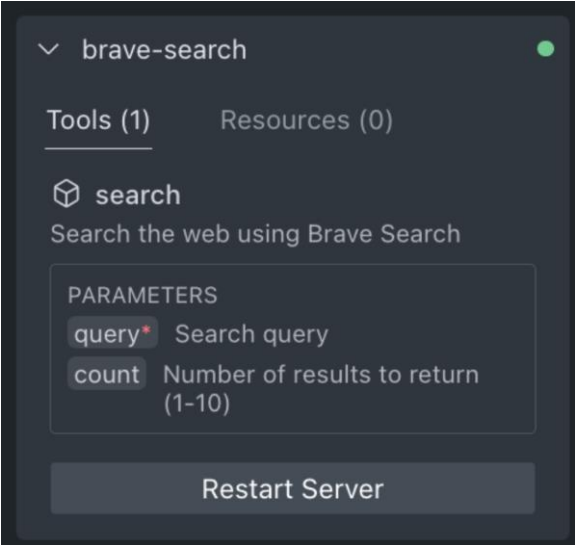
Run Commands in



Use the

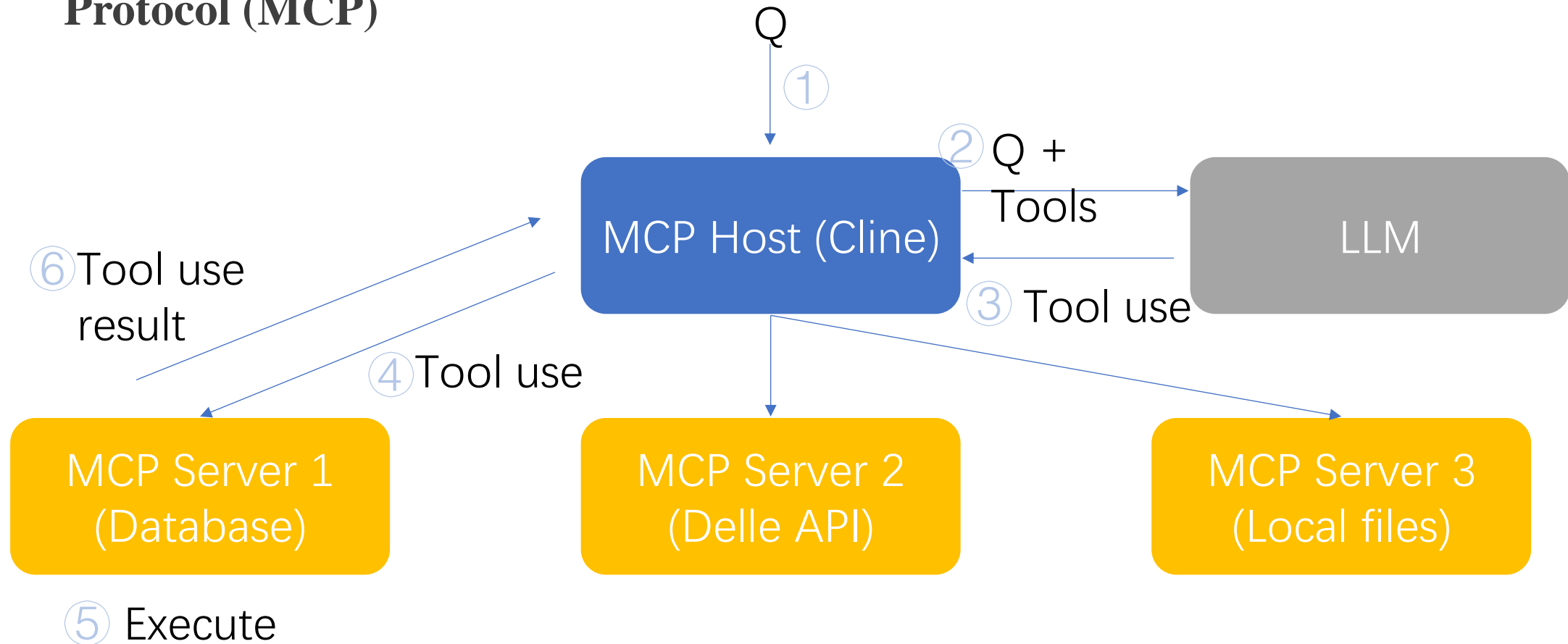


add a tool



How does Cline use tool?

Cline interact with tools according to **Model Context Protocol (MCP)**



However...

- Cline is very powerful with the support of LLMs, however, it still cannot operate without human intervention in
 - difficult problems and
 - complex contexts
- Therefore, the current stage of Cline is still somewhat distant from a General AI agent, and for now, it serves more as a programming copilot for humans.
- Try [Cline](#) out if you are interested. Alternatively, you can try [cursor](#), another AI programming copilot.

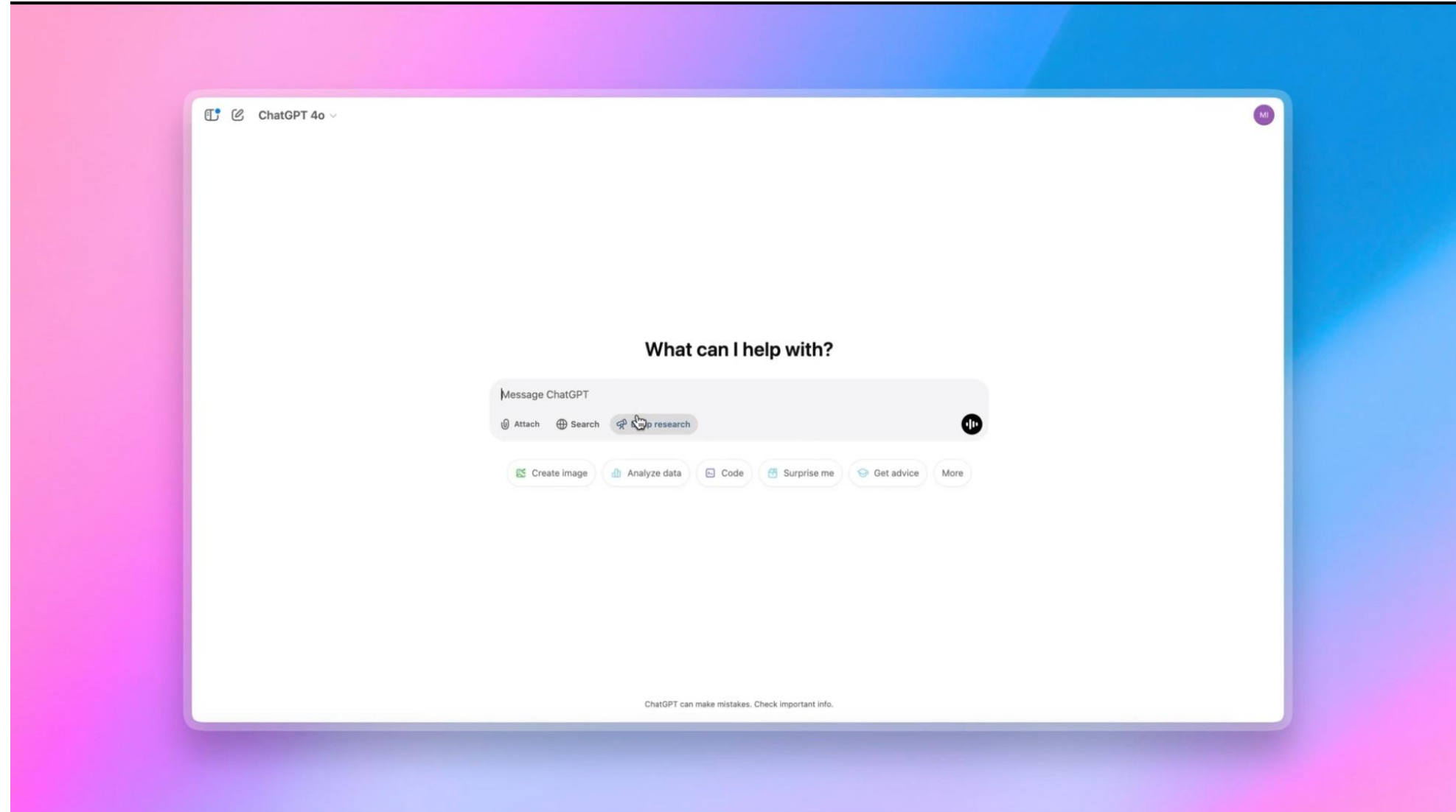
Agent Application

- Deep Research

What is Deep Research?

- **Deep Research** is an AI agent released by OpenAI, that can do autonomous tasks in an unsupervised way
- It was trained using end-to-end reinforcement learning on hard browsing and reasoning tasks across a range of domains
- It will find, analyze, and synthesize hundreds of online sources to create a comprehensive report at the level of a research analyst

What can Deep Research do?



What can Deep Research do?

Mar. 2023, Jason Wei

Rule of thumb: language models can do (with decent accuracy) most things that an average human can do in **1 minute**.

Feb. 2025, OpenAI

In an internal evaluation of expert-level tasks across a range of areas, deep research was rated by domain experts to have automated **multiple hours** of difficult, manual investigation.

Comparison between ChatGPT and Deep Research

- Question: Help me find iOS and android adoption rates, % who want to **learn another language**, and change in **mobile penetration**, over the **past 10 years**, for **top 10 developed** and **top 10 developing** countries by GDP. Lay this info out in a table and separate stats into columns, and include recommendations on markets to target for a new iOS translation app from ChatGPT, focusing on markets ChatGPT is currently active in.

Comparison between ChatGPT and Deep Research

GPT-4o
Top 10 Developed Countries by GDP:
1. United States
2. Japan
3. Germany
4. United Kingdom
5. France
6. Italy
7. Canada
8. South Korea
9. Australia
10. Spain
Top 10 Developing Countries by GDP:

Deep research				
Top 10 Developing/Emerging Economies (by GDP) – (ChatGPT availability considered)				
Country	iOS Market Share	Android Market Share	Interest in New Language (notable stats)	Mobile Penetration Change (2013→2023)
China <i>^(ChatGPT restricted)</i>	~24% iOS vs 75% Android	(see <i>left</i>)	Very High: Hundreds of millions learning English (but ChatGPT not officially accessible)	47%→68% (↑21 pp)

Comparison between ChatGPT and Deep Research

GPT-4o

Recommendations:

Considering the available data and focusing on markets where ChatGPT is currently active, the following developed countries are promising targets for a new iOS translation app:

- **United States:** High iOS market share and significant interest in language learning.
- **United Kingdom:** Substantial iOS user base and a multicultural population with diverse language needs.
- **Canada:** Notable iOS adoption and a bilingual population (English and French), increasing the demand for translation services.

Among developing countries, the following

Deep research

Recommendations: Based on the above, prime targets for a new **ChatGPT-powered iOS translation app** would be countries with **large iOS user bases and strong language-learning demand**, where ChatGPT is accessible. Notably, **Japan, the U.S., UK, Canada, and Australia** emerge as top choices. These markets pair high iOS adoption (50–60%+ share) with significant enthusiasm for language learning – for example, Japan’s iPhone dominance and widespread desire to learn English make it ideal. English-speaking countries like the US, UK, Canada, and Australia, despite being monolingual societies, show substantial interest in learning other languages (e.g. Spanish or French), and all have over 80% smartphone penetration and millions of iPhone users.

Among emerging economies, **Mexico and Brazil** stand out as secondary targets: although Android prevails, their sheer population sizes mean tens of millions of iPhone users exist,

Benchmark Results

Humanity's Last Exam

Model	Accuracy (%)
GPT-4o	3.3
Grok-2	3.8
Claude 3.5 Sonnet	4.3
Gemini Thinking	6.2
OpenAI o1	9.1
DeepSeek-R1*	9.4
OpenAI o3-mini (medium)*	10.5
OpenAI o3-mini (high)*	13.0
OpenAI deep research**	26.6

* Model is not multi-modal, evaluated on text-only subset.

**with browsing + python tools

GAIA

GAIA	Level 1	Level 2	Level 3	Avg.
<u>Previous SOTA</u> ↗	67.92	67.44	42.31	63.64
Deep Research (pass@1)	74.29	69.06	47.6	67.36
Deep Research (cons@64)	78.66	73.21	58.03	72.57

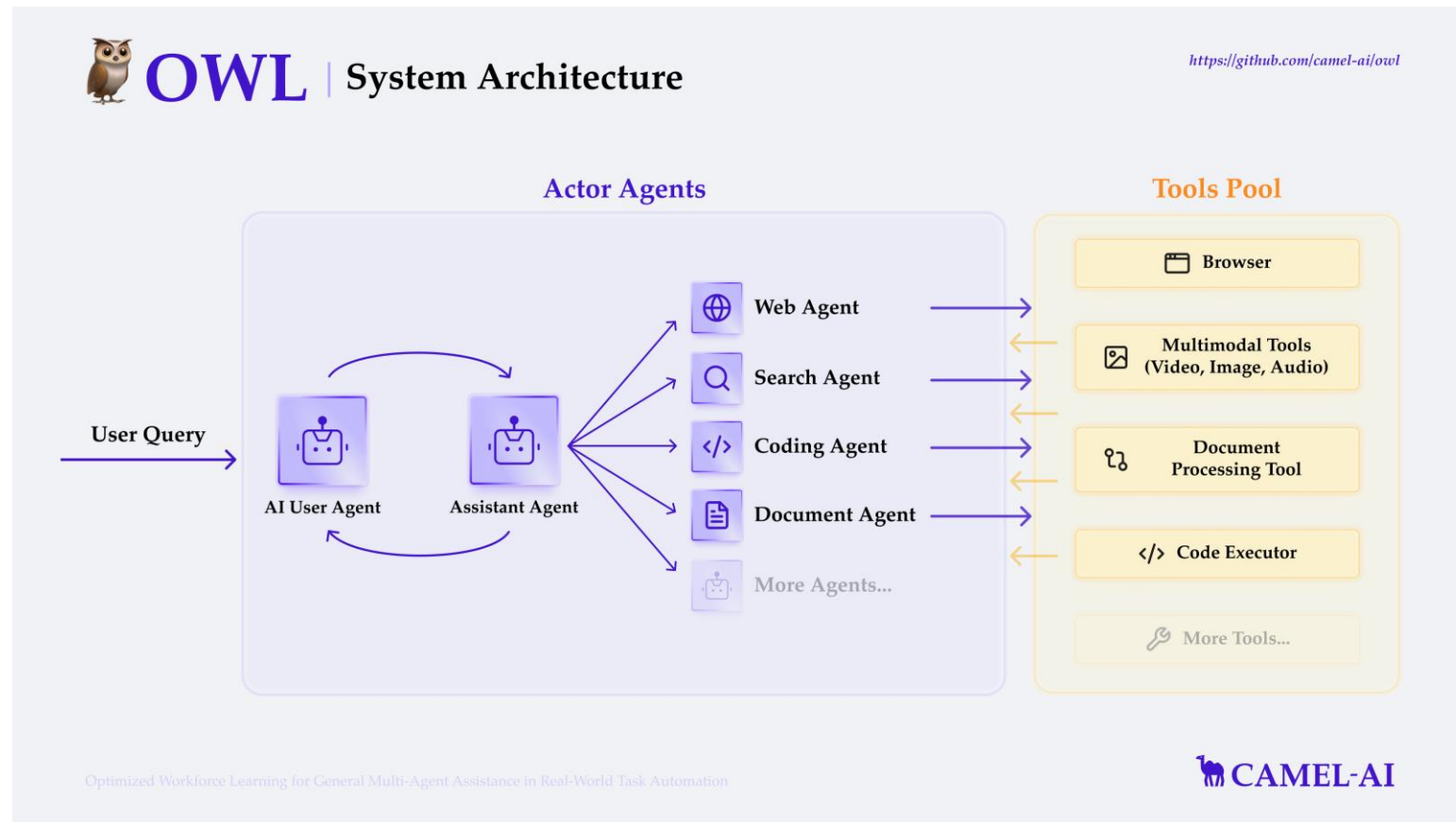
However...

- Deep Research seems to be really powerful at reasoning to synthesize large amounts of online information and complete multi-step research tasks
- However, the model and workflow are not open-sourced, only ChatGPT Pro users (\$200/month) can access it

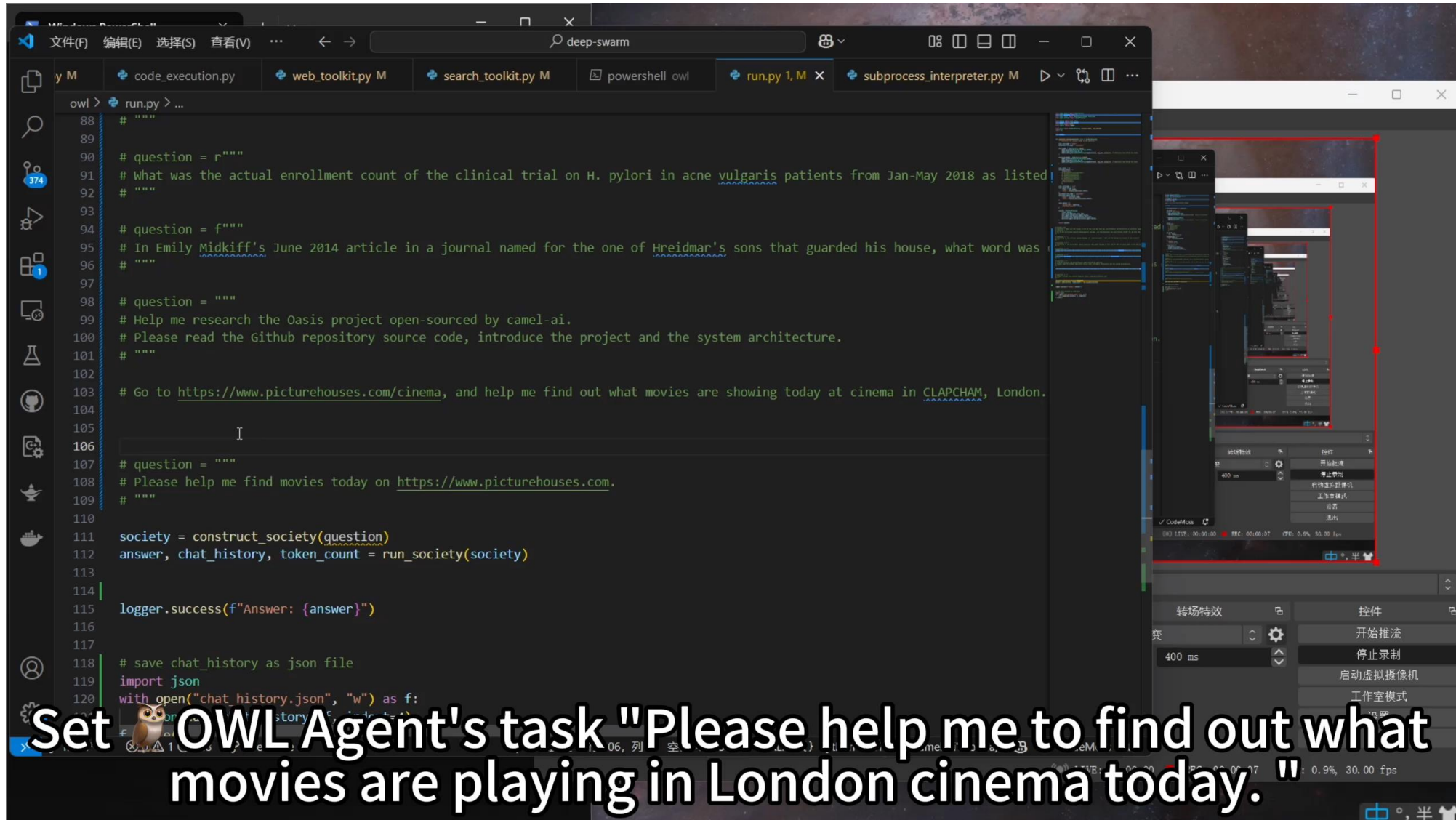
Agent Application - OWL

What is OWL?

- **OWL** is an open-sourced cutting-edge framework for multi-agent collaboration.



What can OWL do?



```
owl > run.py > ...
88 # """
89
90 # question = r"""
91 # What was the actual enrollment count of the clinical trial on H. pylori in acne vulgaris patients from Jan-May 2018 as listed
92 # """
93
94 # question = f"""
95 # In Emily Midkiff's June 2014 article in a journal named for the one of Hreidmar's sons that guarded his house, what word was
96 # """
97
98 # question = """
99 # Help me research the Oasis project open-sourced by camel-ai.
100 # Please read the Github repository source code, introduce the project and the system architecture.
101 # """
102
103 # Go to https://www.picturehouses.com/cinema, and help me find out what movies are showing today at cinema in CLAPCHAM, London.
104
105
106
107 # question = """
108 # Please help me find movies today on https://www.picturehouses.com.
109 # """
110
111 society = construct_society(question)
112 answer, chat_history, token_count = run_society(society)
113
114
115 logger.success(f"Answer: {answer}")
116
117
118 # save chat_history as json file
119 import json
120 with open("chat_history.json", "w") as f:
```

Set OWL Agent's task "Please help me to find out what movies are playing in London cinema today."

What can OWL do?

OWL supports various toolkits that can be customized by modifying the `tools` list in your script:

```
# Configure toolkits
tools = [
    *WebToolkit(headless=False).get_tools(), # Browser automation
    *VideoAnalysisToolkit(model=models["video"]).get_tools(),
    *AudioAnalysisToolkit().get_tools(), # Requires OpenAI Key
    *CodeExecutionToolkit(sandbox="subprocess").get_tools(),
    *ImageAnalysisToolkit(model=models["image"]).get_tools(),
    SearchToolkit().search_duckduckgo,
    SearchToolkit().search_google, # Comment out if unavailable
    SearchToolkit().search_wiki,
    *ExcelToolkit().get_tools(),
    *DocumentProcessingToolkit(model=models["document"]).get_tools(),
    *FileWriteToolkit(output_dir=".").get_tools(),
]
```



What can OWL do?

- OWL is still in its early stage: just a week since it has been developed
- More functionalities will be integrated in the short future
- Try [OWL](#) out if you are interested

Can we imitate Manus?

How to use Cursor and MCP to do a market research

- https://eegb6fzscd.feishu.cn/wiki/FHtRwrmPeikgDmk9tINcQS7gny0?from=from_copylink

Acknowledgements

- <https://github.com/Paitesanshi/LLM-Agent-Survey>
- <https://github.com/WooooDyy/LLM-Agent-Paper-List>
- [Generative Agents: Interactive Simulacra of Human Behavior.](#)
- <https://wenting-zhao.github.io/complex-reasoning-tutorial/>
- <https://acl2023-retrieval-lm.github.io/>
- <https://github.com/xlang-ai/llm-tool-use>