


RL for LLM Agent : OpenAI's Deep Research

Tianbing Xu, March, 2025


Deep Research (OpenAI 2025): An Example Task

I am a researcher interested in bringing mixed-gas sorption capabilities to my lab. Please discuss the differences between pure- and mixed-gas sorption for glassy polymers, how the dual-mode sorption model can be used to predict mixed-gas sorption behavior in glassy polymers (include equations where applicable), and what challenges there are in accurately predicting pure- and mixed-gas sorption using the dual-mode sorption model. Only include open-source information as of October 1, 2024.

 **Pure vs. mixed-gas sorption in glassy polymers:** In a glassy polymer (below T_g), a single gas typically exhibits a concave sorption isotherm: uptake rises sublinearly with pressure, unlike the linear Henry's-law behavior above T_g . This is attributed to "dual-mode" sorption: some penetrant dissolves in the dense polymer matrix (obeying Henry's law) while additional penetrant saturates specific microvoid "sites" (a Langmuir-type adsorption in the polymer's nonequilibrium free volume). By contrast, when two or

Understanding sorption models

I'm gathering info on pure and mixed-gas sorption in glassy polymers, focusing on the dual-mode sorption model and addressing challenges in modeling predictions.

 [search action]

Assessing open access

OK, let me see. The result from MDPI on mixed-gas sorption in glassy polymers is probably open access.

 [browse action]

 [browse action]

Deep Research

[General AI Assistant Benchmark](#) (Feb. 2025)

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

Deep Research - Trained with RL

Deep research independently discovers, **reasons** about, and consolidates insights from across the web. To accomplish this, it was **trained on real-world tasks** requiring **browser** and Python **tool** use, using the same **reinforcement learning** methods behind OpenAI o1, our first reasoning model.

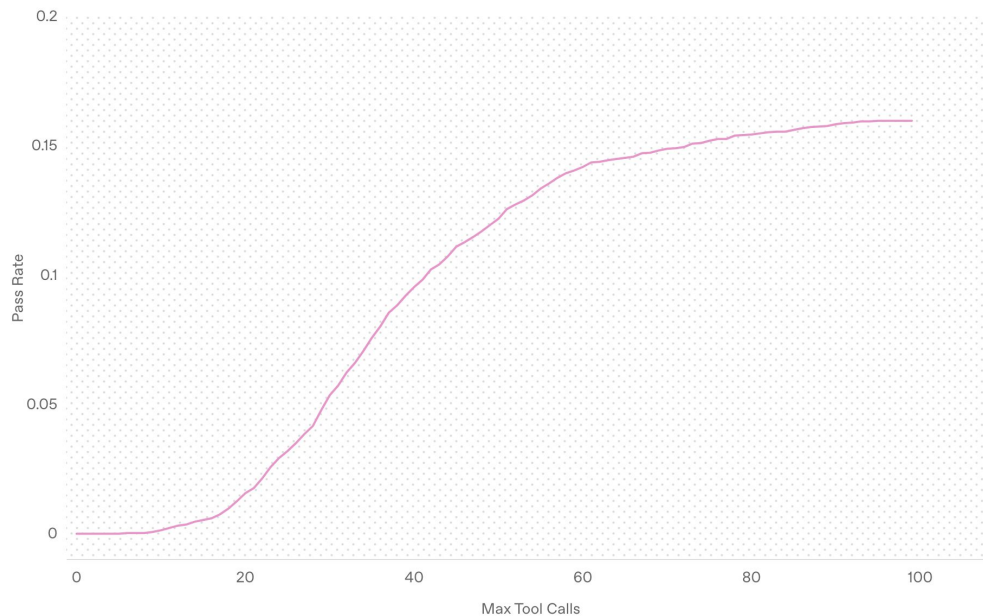
Learning to Plan and Execution for long multi-step tasks

Deep research was trained using **end-to-end reinforcement learning** on hard **browsing** and **reasoning** tasks across a range of domains. Through that training, it **learned to plan and execute a multi-step trajectory** to find the data it needs, backtracking and reacting to real-time information where necessary. The model is also able to browse over user uploaded files, plot and iterate on graphs using the python tool, embed both generated graphs and images from websites in its responses, and cite specific sentences or passages from its sources.

Deep Research

Expert Level Task

Pass Rate vs Max Tool Calls



The more the model browses and thinks about what its browsing, the better it does, which is why giving it time to think is important.

Agentic Reasoning: Reasoning LLMs with Tools for the Deep Research (2025)

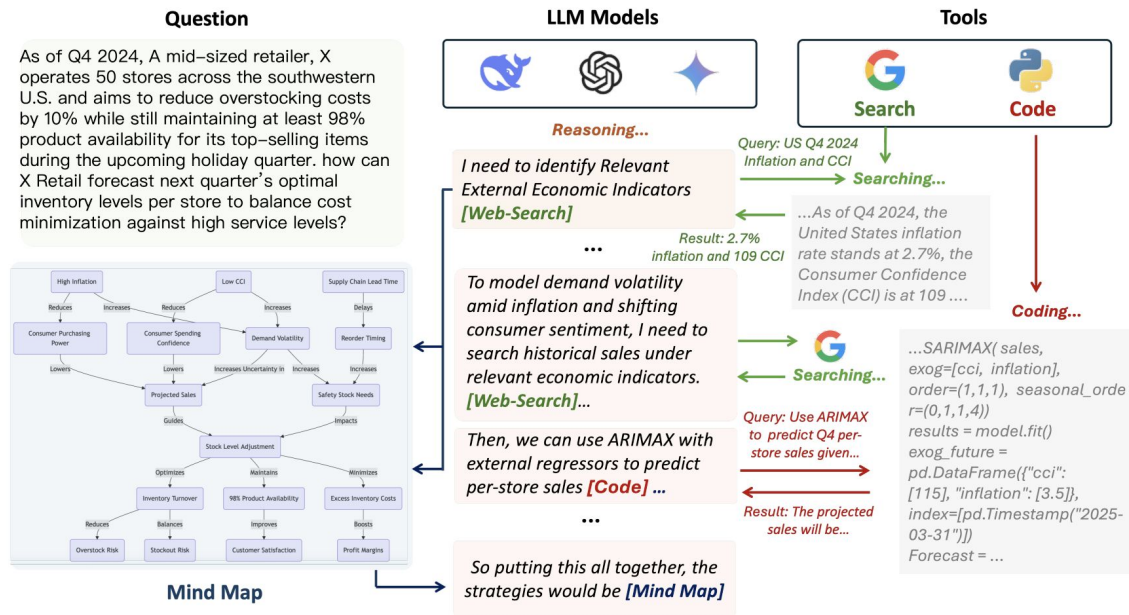


Figure 1: The overall workflow of Agentic Reasoning.

A Novel Reinforcement Learning Framework of LLM Agents (2024)

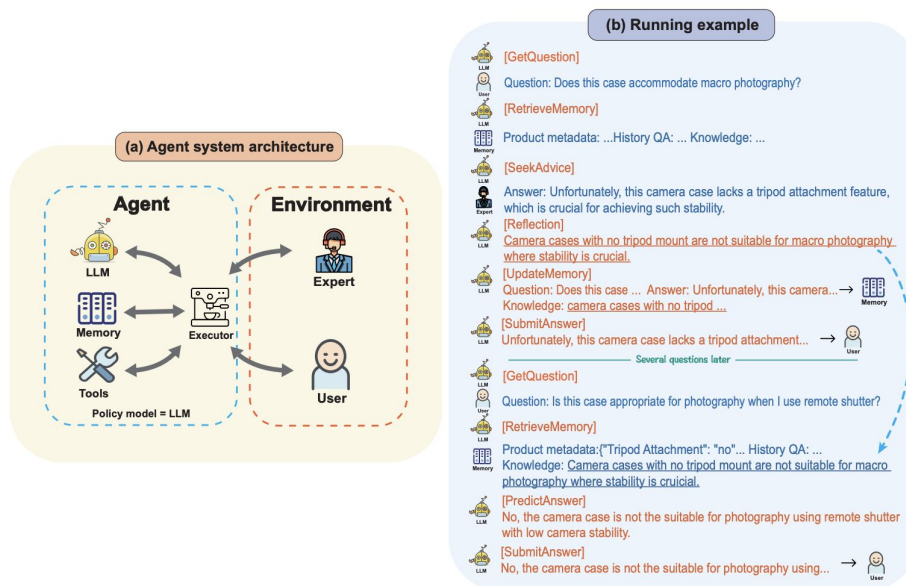


Figure 1: (a) Architecture of our agent system, including LLM, memory, tools, and executor. (b) A running example of AGILE in a customer service QA environment. The tokens (actions) generated by the LLM are in orange color and the tokens appended by the executor are in blue color.

Reference

1. [OpenAI, Deep Research](#)
2. [Agentic Reasoning: Reasoning LLMs with Tools for the Deep Research](#)
3. [A Novel Reinforcement Learning Framework of LLM Agents](#)