WebDancer: Towards Autonomous Information Seeking Agency

Jialong Wu, Baixuan Li , Runnan Fang , Wenbiao Yin, Liwen Zhang, Zhenglin Wang, Zhengwei Tao, Dingchu Zhang, Zekun Xi, Xiangru Tang, Yong Jiang, Pengjun Xie, Fei Huang, Jingren Zhou
DeepResearch Team, Tongyi Lab, Alibaba Group
wujialongml@gmail.com

NEURAL INFORMATION PROCESSING SYSTEMS

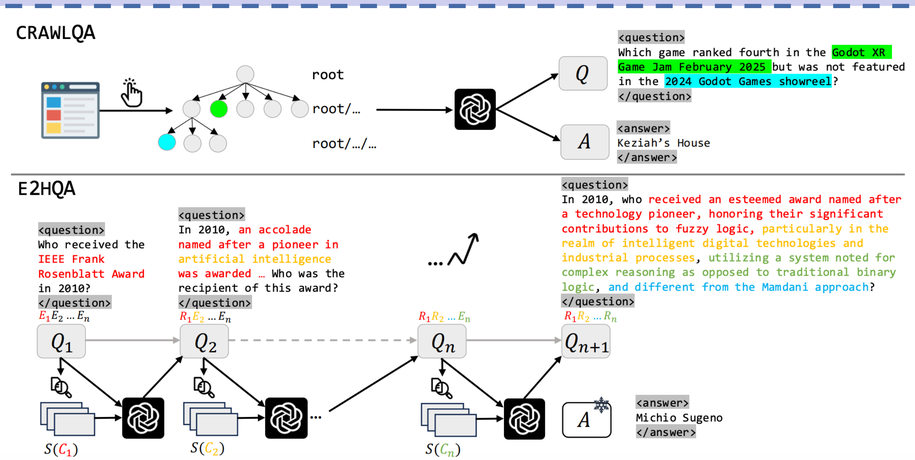
Github HuggingFace arXiv Paper

Introduction

Deep Research has demonstrated strong deep information seeking capabilities through end-to-end reinforcement learning (RL) training.
How to build a web agent like Deep Research from scratch?

(1) Acquiring high-quality synthetic **QA pairs**
(2) Constructing **reliable trajectorie**s that support long-horizon reasoning and task decomposition
(3) Agentic **training strategies**

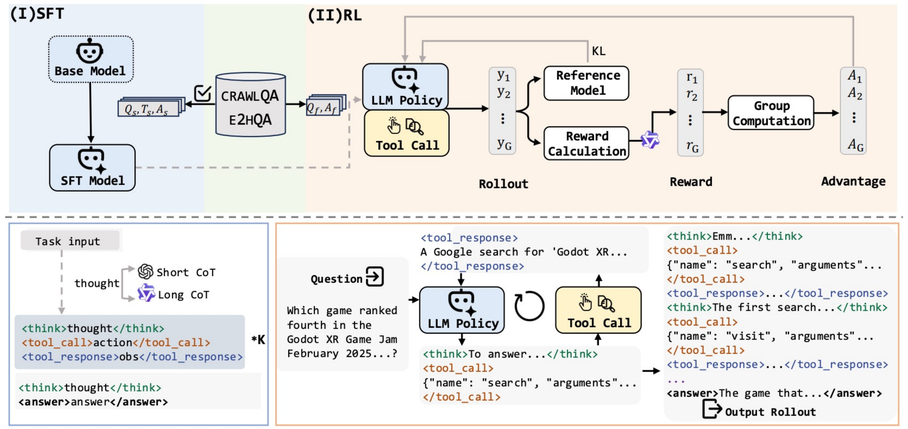
Data



CRAWLQA We begin by collecting the root URLs of official and knowledgeable websites spanning *arxiv*, *github*, *wiki*, etc. To emulate human browsing behavior, we recursively navigate subpages by following accessible hyperlinks from each root site.
E2HQA We begin from large QA pairs in SimpleQA style where each answer is a concise, fact-seeking entity. By continuously searching, we can gradually rephrase an initially simple question into a more complex multi-step one.

High-quality synthetic data is the foundation of agent RL.

Training



(I) The SFT stage for cold start utilizes the reformatted **ReAct** datasets, where the thought includes both short and long CoT, respectively.

(II) The RL stage performs rollouts with the tool calls on the QA pairs that are not utilized during the SFT stage, and optimizes the policy using the DAPO algorithm..

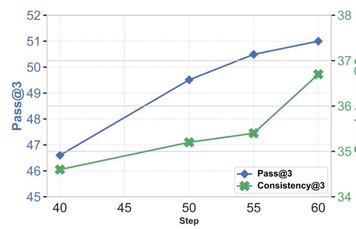
Experiments

Backbone	Framework	GAIA Level 1	GAIA Level 2	GAIA Level 3	GAIA Avg.	WebWalkerQA Easy	WebWalkerQA Medium	WebWalkerQA Hard	WebWalkerQA Avg.
No Agency									
Qwen-2.5-7B	Base	12.8	3.8	0.0	6.8	1.25	0.8	0.7	0.8
Qwen-2.5-32B	Base	20.5	9.6	8.3	13.6	3.8	2.5	3.3	3.1
	RAG	12.8	11.8	8.3	11.8	23.1	14.3	11.3	15.3
Qwen-2.5-72B	Base	20.5	13.5	0.0	14.6	9.4	7.1	3.3	6.3
GPT-4o	Base	23.1	15.4	8.3	17.5	6.7	6.0	4.2	5.5
QwQ-32B	Base	30.8	15.4	25.0	22.3	7.5	2.1	4.6	4.3
	RAG	33.3	36.5	8.3	32.0	36.9	26.1	33.5	31.2
DeepSeek-R1-671B	Base	43.6	26.9	8.3	31.1	5.0	11.8	11.3	10.0
Close-Sourced Agentic Frameworks									
	OpenAI DR	74.3	69.1	47.6	67.4	-	-	-	-
Open-sourced Agentic Frameworks									
Qwen-2.5-7B	Search-o1	23.1	17.3	0.0	17.5	-	-	-	-
	R1-Searcher	28.2	19.2	8.3	20.4	-	-	-	-
Qwen-2.5-32B	Search-o1	33.3	25.0	0.0	28.2	-	-	-	-
QwQ-32B	Search-o1	53.8	34.6	16.7	39.8	43.1	35.0	27.1	34.1
	WebThinker-Base	53.8	44.2	16.7	44.7	47.2	41.1	39.2	41.9
	WebThinker-RL	56.4	50.0	16.7	48.5	58.8	44.6	40.4	46.5
	Simple DS	-	-	-	50.5	-	-	-	-
ReAct Agentic Frameworks									
Qwen-2.5-7B	Vanilla ReAct	28.2	15.3	0.0	18.4	28.1	31.2	16.0	24.2
	WebDancer	41.0	30.7	0.0	31.0	40.6	44.1	28.2	36.0
Qwen-2.5-32B	Vanilla ReAct	46.1	26.9	0.0	31.0	35.6	38.7	22.5	31.9
	WebDancer	46.1	44.2	8.3	40.7	44.3	46.7	29.2	38.4
QwQ-32B	Vanilla ReAct	48.7	34.6	16.6	37.8	35.6	29.1	13.2	24.1
	WebDancer	**61.5**	**50.0**	**25.0**	**51.5**	**52.5**	**59.6**	**35.4**	**47.9**
GPT-4o	Vanilla ReAct	51.2	34.6	8.3	34.6	34.6	42.0	23.9	33.8

Our method significantly enhances agentic capabilities over the underlying base model, validating the strength and generality of our approach.

Framework	Browsing	*En.*	*Zh.*
GPT-4o	✘	0.6	6.2
	✔	1.9	-
QwQ-32B	✘	-	11.1
WebDancer	✔	3.8/**7.9**	18.0/**31.5**

Results on BrowseComp (En.) and BrowseComp-zh (Zh.).



Performance across training steps using the DAPO algorithm.

End-to-end agentic RL is difficult, yet highly potential-rich.

Tongyi DeepResearch

Tongyi DeepResearch, developed by Tongyi Lab, is specifically designed for **long-horizon, deep information-seeking** tasks. Tongyi DeepResearch also has an extensive deep researc h agent family.

If you like our project, feel free to give us a ⭐ on GitHub☺