WebServ: A Full-Stack and RL-Ready Web Environment for Training Web Agents at Scale

Yuxuan Lu[1,2], Ziyi Wang[1,2], Jing Huang[1], Hui Liu[1], Jiri Gesi[1], Yan Han[1], Shihan Fu[1,2], Tianqi Zheng[1], Xianfeng Tang[1], Chen Luo[1], Yisi Sang[1], Jin Lai[1], Dakuo Wang[1,2]

[1]Amazon.com, Inc., [2]Northeastern University

Jan 14 2026

Motivation

- Large Language Model (LLM) Web Agents have been developed and widely used
- WebGPT[1], Claude Computer Use, ChatGPT Atlas, Step[2], LASER[3], WebAgent[4], . . .

[1] Reiichiro Nakano et al. *WebGPT: Browser-assisted Question-Answering with Human Feedback*. June 2022. arXiv: `2112.09332 [cs]`.

[2] Paloma Sodhi et al. *SteP: Stacked LLM Policies for Web Actions*. Apr. 2024. arXiv: `2310.03720 [cs]`.

[3] Kaixin Ma et al. *LASER: LLM Agent with State-Space Exploration for Web Navigation*. Feb. 2024. arXiv: `2309.08172 [cs]`.

[4] Izzeddin Gur et al. “A Real-World WebAgent with Planning, Long Context Understanding, and Program Synthesis”. In: *The Twelfth International Conference on Learning Representations*. Oct. 2023.

Motivation

- These agents have become a primising approach for web automation tasks[5] and simulating user behaviors[6].
- However, existing Web Agents are mainly built with **prompting-based** and **behavior cloning-based** methods.
- RLHF and RLVR have been proven to work in other fields and tasks[7]
- RL Web Agents is limited by the absense of a full-stack, efficient and effective environment.

[5] Shuyan Zhou et al. *WebArena: A Realistic Web Environment for Building Autonomous Agents*. Apr. 2024. arXiv: 2307.13854 [cs].

[6] Yuxuan Lu et al. *Prompting Is Not All You Need! Evaluating LLM Agent Simulation Methodologies with Real-World Online Customer Behavior Data*. June 2025. arXiv: 2503.20749 [cs].

[7] Long Ouyang et al. *Training Language Models to Follow Instructions with Human Feedback*. Mar. 2022. arXiv: 2203.02155 [cs], DeepSeek-AI et al. *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*. Jan. 2025. arXiv: 2501.12948 [cs].

Web Browsers are NOT RL-Ready

- Web Browsers are NOT RL-Ready:
  - Requires per-site configuration or rely on site features
  - Noisy context and action
  - Lack of visual hint
  - Non-robust action execution

Web Servers are NOT RL-Ready

- Web Servers are usally packed in docker images
- Not possible to run hundsreds *isolated* web server containers to support parallel RL rollouts

Existing Work

- Researchers have argued that existing web environments are not RL-ready[8]
- Existing works are:
  - …off-policy[9]
  - …not isolated[10]
  - …uses synthetic environments[11]
- The community needs a scalable and efficient training environment

[8] Zhaorun Chen et al. *Scaling Agent Learning via Experience Synthesis*. Nov. 2025. arXiv: `2511.03773 [cs]`.

[9] Zehan Qi et al. *WebRL: Training LLM Web Agents via Self-Evolving Online Curriculum Reinforcement Learning*. Jan. 2025. arXiv: `2411.02337 [cs]`.

[10] Zhepei Wei et al. *WebAgent-R1: Training Web Agents via End-to-End Multi-Turn Reinforcement Learning*. May 2025. arXiv: `2505.16421 [cs]`.

[11] Zhaorun Chen et al. *Scaling Agent Learning via Experience Synthesis*. Nov. 2025. arXiv: `2511.03773 [cs]`.

WebServ Architecture Design

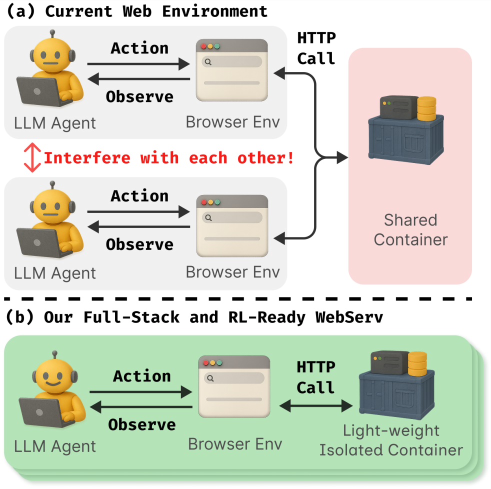


Figure 1: System Architecture of WebServ compared to existing environments

WebServ: Web Browser Interface Design

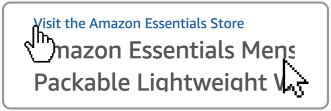
- To tackle challenges in web browser interface, we design:
  - Fully automatic parser
  - Uses a simplified HTML as context
    - LLM already recognizes HTML structure
  - Network-Idle based waiting
  - VLM Support
  -

WebServ: Web Browser Interface Design

- LLM Agent often click on non-interactive elements, but human can easily tell text and link apart

WebServ: Web Browser Interface Design

- LLM Agent often click on non-interactive elements, but human can easily tell text and link apart



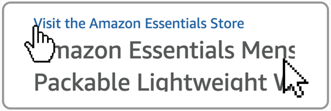
Visit the Amazon Essentials Store

mazon Essentials Mens
Packable Lightweight V

Figure 2: Visual Hint Example

WebServ: Web Browser Interface Design

- LLM Agent often click on non-interactive elements, but human can easily tell text and link apart



Visit the Amazon Essentials Store

mazon Essentials Mens

Packable Lightweight W

Figure 2: Visual Hint Example

- We made these visual hints available for agent as well

WebServ: Web Server Design

- Efficient Incus-based server container management
  - Run a server in seconds instead of minutes
- Reduce resource need by 240x
- Make it possible to run 200 parallel instances

Evaluation and Experiments

Evaluation

- Web browser I/O interface: Task successful rate on WebArena[12]
- Server manager effiency: Launch speed and resource need

[12] Shuyan Zhou et al. *WebArena: A Realistic Web Environment for Building Autonomous Agents*. Apr. 2024. arXiv: `2307.13854 [cs]`.

Evaluation Result

Model and Method	Shopping	CMS	GitLab
Vanilla WebArena			
GPT-4o	11.1	20	10.0
OpenAI-o3	33.3	45.7	46.7
Llama-3.1-8B	8.9	5.7	10.0
WebAgent-R1	**44.4**	**57.1**	**56.7**
WebServ			
GPT-4o	20.0	28.6	43.3
GPT-5	35.6	57.1	**53.3**
Claude 3.5 Sonnet	26.7	31.4	36.7
Claude 3.7 Sonnet	31.1	37.1	50.0
Claude 4 Sonnet	**42.2**	48.6	50.0
Claude 4.5 Sonnet	40.0	**62.9**	50.0

Table 1: Comparison of model accuracy (in %) across Shopping, CMS and GitLab tasks in WebArena-Lite.

Evaluation Result

- With WebServ, model can achieve better performancde than with vanilla WebArena
- Achieves tate-of-the-art single-prompt agent performance
- Prompting-based baseline models can achieve and even beat RL based methods

Ablation Study: Visual Cue

- How important are visual cues?
- Ablation study: remove visual cues and "clickable" element hint in the observation space.

Ablation: Visual Cue

Model	Δ Shopping (%)	Δ CMS (%)	Δ GitLab (%)
GPT-4o	-33.3%	-70.0%	-46.2%
GPT-5	-18.8%	-35.0%	-25.0%
Claude 3.5 Sonnet	-58.3%	-45.5%	-81.8%
Claude 3.7 Sonnet	-50.0%	-53.8%	-66.7%
Claude 4 Sonnet	-26.3%	-29.4%	-33.3%
Claude 4.5 Sonnet	-5.6%	-45.5%	6.7%

Table 2: Performance difference (%) after removing visual cues.

- Except for one setting, all models exhibit performance drops
- Weaker models suffer larger performance drops
- Proves that Visual Cues is important especially for weaker models.

Ablation: Vision Language Model

- How important is other vision signals?

Model	Shopping	CMS	GitLab
Claude 4 Sonnet	42.2	**48.6**	**50.0**
+ VLM	**44.4**	**48.6**	43.3
Claude 4.5 Sonnet	**40.0**	62.9	50.0
+ VLM	37.8	**65.7**	**56.7**

Table 3: Ablation results comparing text-only and VLM settings for Claude 4 Sonnet and Claude 4.5 Sonnet across WebArena-Lite tasks.

Ablation: Vision Language Model

- How important is other vision signals?

Model	Shopping	CMS	GitLab
Claude 4 Sonnet	42.2	**48.6**	**50.0**
+ VLM	**44.4**	**48.6**	43.3
Claude 4.5 Sonnet	**40.0**	62.9	50.0
+ VLM	37.8	**65.7**	**56.7**

Table 3: Ablation results comparing text-only and VLM settings for Claude 4 Sonnet and Claude 4.5 Sonnet across WebArena-Lite tasks.

Visual signals are not universally beneficial for general web agent tasks

Evaluation: Launch Speed and Resource Need

Metric	WebServ (Incus)	Naïve Docker
Launch speed	1.781 s	8.963 s
Storage	28.01 MiB	6.78 GiB
Memory	1.74 GiB	1.63 GiB

Table 4: Comparison of system efficiency between WebServ and Docker.

With Incus, WebServ can reduce launch latench and improved storage footprint with comparable memory usage, making large-scale RL rollouts possible.

Conclusion

WebServ enables:

- State-of-the-art single-prompt agent performance on WebArena-Lite
- No per-site configuration, able to operate on real complex website (Amazon, Google Flight, etc.)
- Efficient RL training with VeRL integration