

UXAgent: A System for Simulating Usability Testing of Web Design with LLM Agents

Yuxuan Lu^{1,2}, Bingsheng Yao², Hansu Gu¹, Jing Huang¹, Zheshen (Jessie) Wang¹, Yang Li¹, Jiri Gesi¹, Qi He¹, Toby Jia-Jun Li³, Dakuo Wang^{1,2}

¹Amazon.com, Inc., ²Northeastern University, ³ University of Notre Dame

Dec 1, 2025

What's the worst nightmare as a researcher?

- The myth of “Review 2”
- Paper being scooped
- Realizing that your experiment design is flawed one week before deadline

As someone from both NLP and HCI, I think . . .

Field	NLP	HCI
Experiment Subject	Models and Machines	Human Subjects
Experiment Design	Code and Data	Study Protocol
Experiment Cost	Money	Human Participants' Time
"Debugging" Method	Code Debugging	???

Human Participants' Time is Valuable and Limited

How can we better evaluate UX Research study design before running the study?

How can we better evaluate usability testing
study design before running the study

LLM Agent as a promising solution



- Generative Agents¹ – “Believable” human behavior
- SimUser² – simulate user and application for UX research
- AXNav³ – accessibility testing

¹Joon Sung Park et al. “Generative Agents: Interactive Simulacra of Human Behavior”. In: *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. UIST '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 1–22.

²Wei Xiang et al. “SimUser: Generating Usability Feedback by Simulating Various Users Interacting with Mobile Applications”. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–17.

³Maryam Taeb et al. “AXNav: Replaying Accessibility Tests from Natural Language”. In: *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. CHI '24. New York, NY, USA: Association for Computing Machinery, May 2024, pp. 1–16.

LLM Web Agent as a promising solution for web designs

- WebGPT⁴ – Search Engine enhanced QA
- WebAgent⁵ – Planning for Web Automation
- WebVoyager⁶, LASER⁷, ...

⁴Reiichiro Nakano et al. *WebGPT: Browser-assisted Question-Answering with Human Feedback*. June 2022. arXiv: 2112.09332 [cs].

⁵Izzeddin 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.

⁶Hongliang He et al. *WebVoyager: Building an End-to-End Web Agent with Large Multimodal Models*. June 2024. arXiv: 2401.13919 [cs].

⁷Kaixin Ma et al. *LASER: LLM Agent with State-Space Exploration for Web Navigation*. Feb. 2024. arXiv: 2309.08172 [cs].

Challenges

- Existing LLM Agent systems mostly works in **sandboxed environments**
- Existing LLM Web Agents focus on **task completion rate**, not simulating complex and dynamic human behavior
- Existing reasoning architecture of LLM Agents/LLM Web Agents either fail to simulate human reasoning process (too simple) or introduces additional latency (too complex) for real time simulation.

UXAgent

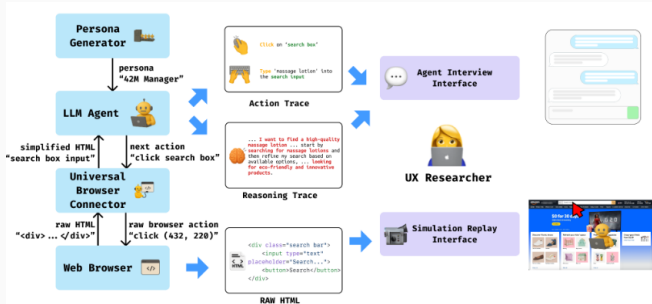


Figure 1: System Architecture of UXAgent

<https://broadcast.amazon.com/videos/1691301>

Persona Generator

1 Participant Recruitment

2 Survey

3 Review

Configure participant demographics

Provide a questionnaire

Confirm & Run

Recruitment Target Setting

URL of website being tested *

https://www.example.com

Number of Participants *

20

Participant Task *

Buy a jacket

Example Persona *

Persona: Clara
Background:
Clara is a PhD student in Computer Science at a

Demographics

Field Name

Age

Value	Weight	Actions
18-55	1	+ Add Choice Remove Value Remove Field

Field Name

Gender

Value	Weight	Actions
Male	1	+ Add Choice Remove Value Remove Field
Female	1	+ Add Choice Remove Value Remove Field
Non-binary	1	+ Add Choice Remove Value Remove Field

+ Add Field

Reset Form

Next

(a) Participant Task Config

(b) Demographic Distribution Config

Figure 2: Study Configure Interface

Agent Architecture

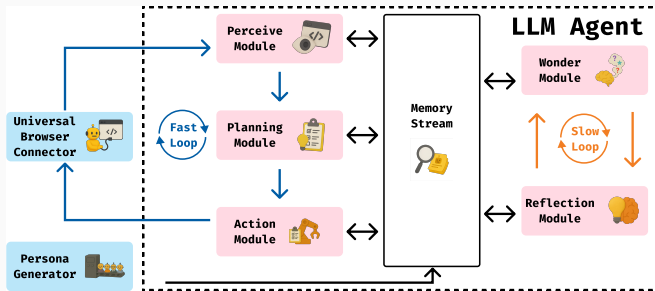


Figure 3: Agent Architecture

Agent Architecture: Two-Loop + Memory

- To allow both in-depth reasoning and real-time simulation:
 - Fast Loop: rapid response to the outside environment
 - Slow Loop: in-depth reasoning and thinking
- Asynchronous; interact through Memory Stream
- Each module retrieves and generate memories

Result Viewer Interface

Agent List

Agent ID: 1
He has a bachelor's degree and lives in Portland, Oregon, a...

Agent ID: 2
They have a bachelor's degree and live in Portland, Oregon,...

Agent ID: 3
She has a bachelor's degree and lives in Portland, Oregon, ...

Agent ID: 4
They have a bachelor's degree and live in Portland, Oregon, ...

Agent ID: 5
They have a bachelor's degree and live in Portland, Oregon, ...

Agent ID: 6
She has a bachelor's degree and lives in Portland, Oregon, ...

Agent ID: 7
She has a bachelor's degree and lives in Portland, Oregon, ...

Agent ID: 8
He has a bachelor's degree and lives in Portland, Oregon, i...

Action Trace

Action 1: [click](#)
Target: grocery_gourmet_food
Clicking on 'Grocery & Gourmet Food' category from the main navigation menu

Action 2: [click](#)
Target: meat_substitutes_79_
Clicking on the 'Meat Substitutes' subcategory from the left sidebar to navigate to meat substitute products

Action 3: [click](#)
Target: 100_00_199_99_4_item
Clicking on a product to view its details and rating

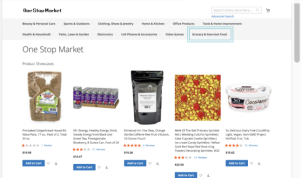
Action 4: [click](#)
Target: beyond_meat_beef_bee
Clicking on Beyond Meat Beefy Crumble product to view its details and rating

Action 5: [back](#)
Navigating back to meat substitutes category page

Action 6: [click](#)
Target: assi_luncheon_loaf_a
Clicking on ASSI Luncheon Loaf product to view its details and check its rating

Observation Snapshot

Screenshot



Interview

Agent Memory Trace

Kind	Importance	Content
Timestamp: 1 Interview Agent at This Moment		
> observation	0.8	I can see the One Stop Market e--...
> thought	0.8	I need to find

Figure 4: Result Viewer Interface

10 / 12

Universal Browser Connector

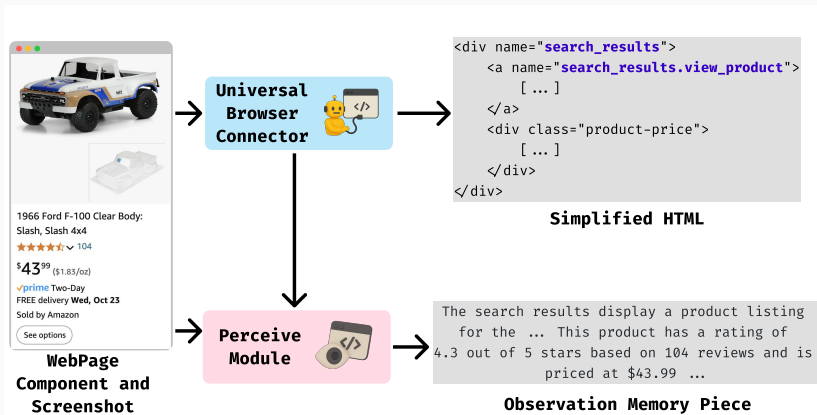


Figure 5: Universal Browser Connector

Generalizability

- The agent is designed to be generalizable to different websites and tasks
- Rufus: <https://broadcast.amazon.com/videos/1685504>
- Customer Service Agent:
<https://broadcast.amazon.com/videos/1685507>
- AC3: <https://broadcast.amazon.com/videos/1769741>

Questions?