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

Yuxuan Lu[1,2], Bingsheng Yao[2], Hansu Gu[1], Jing Huang[1], Zheshen (Jessie) Wang[1], Yang Li[1], Jiri Gesi[1], Qi He[1], Toby Jia-Jun Li[3], Dakuo Wang[1,2]

[1]Amazon.com, Inc., [2]Northeastern 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 Agents[1] – "Believable" human behavior
- SimUser[2] – simulate user and application for UX research
- AXNav[3] – accessibility testing

[1] Joon 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.

[2] Wei 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.

[3] Maryam 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

- WebGPT[4] – Search Engine enhanced QA
- WebAgent[5] – Planning for Web Automation
- WebVoyager[6], LASER[7], ...

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

[5] 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.

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

[7] Kaixin 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

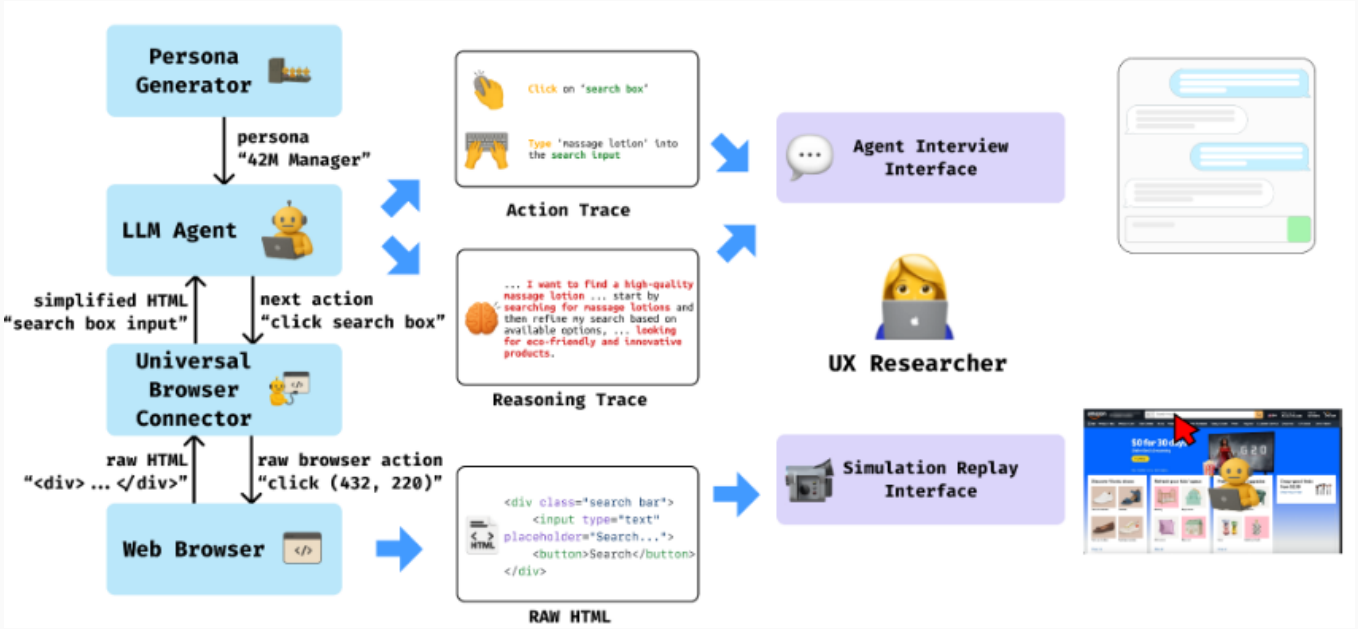
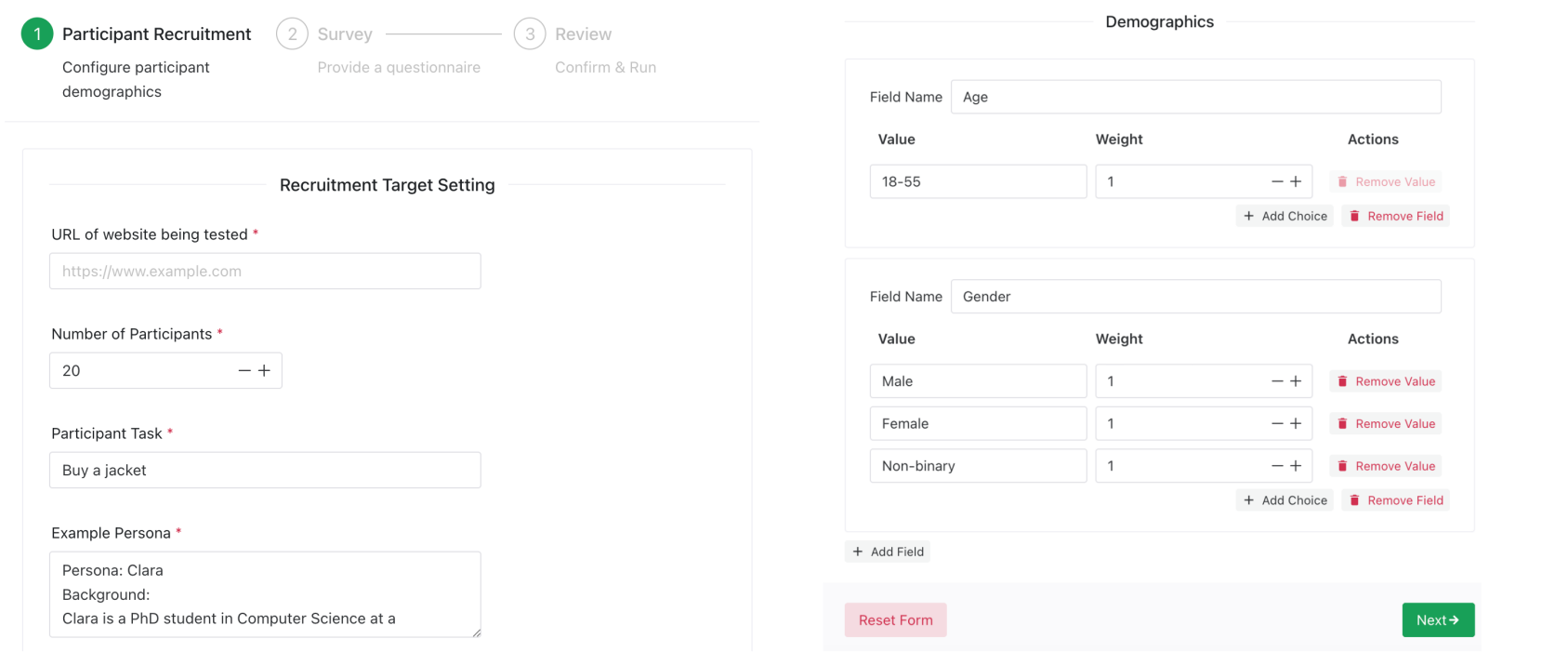


Figure 1: System Architecture of UXAgent

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

Persona Generator



(a) Participant Task Config **(b) Demographic Distribution Config**

Figure 2: Study Configure Interface

Agent Architecture

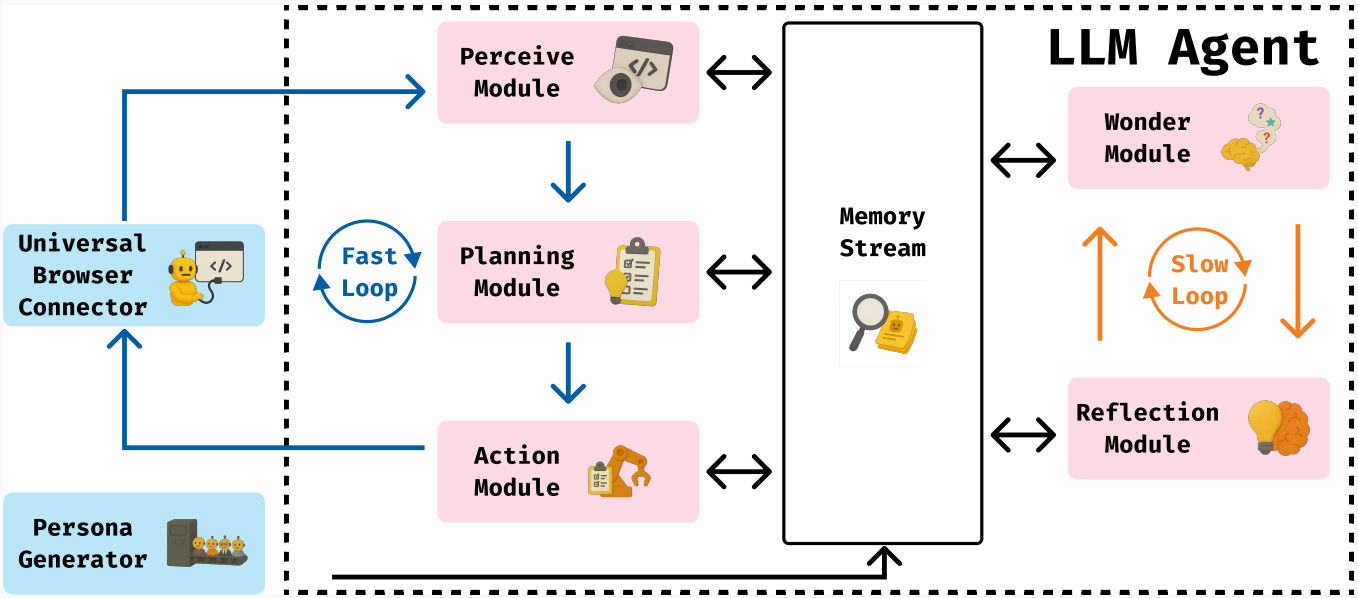


Figure 3: Agent Architecture

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

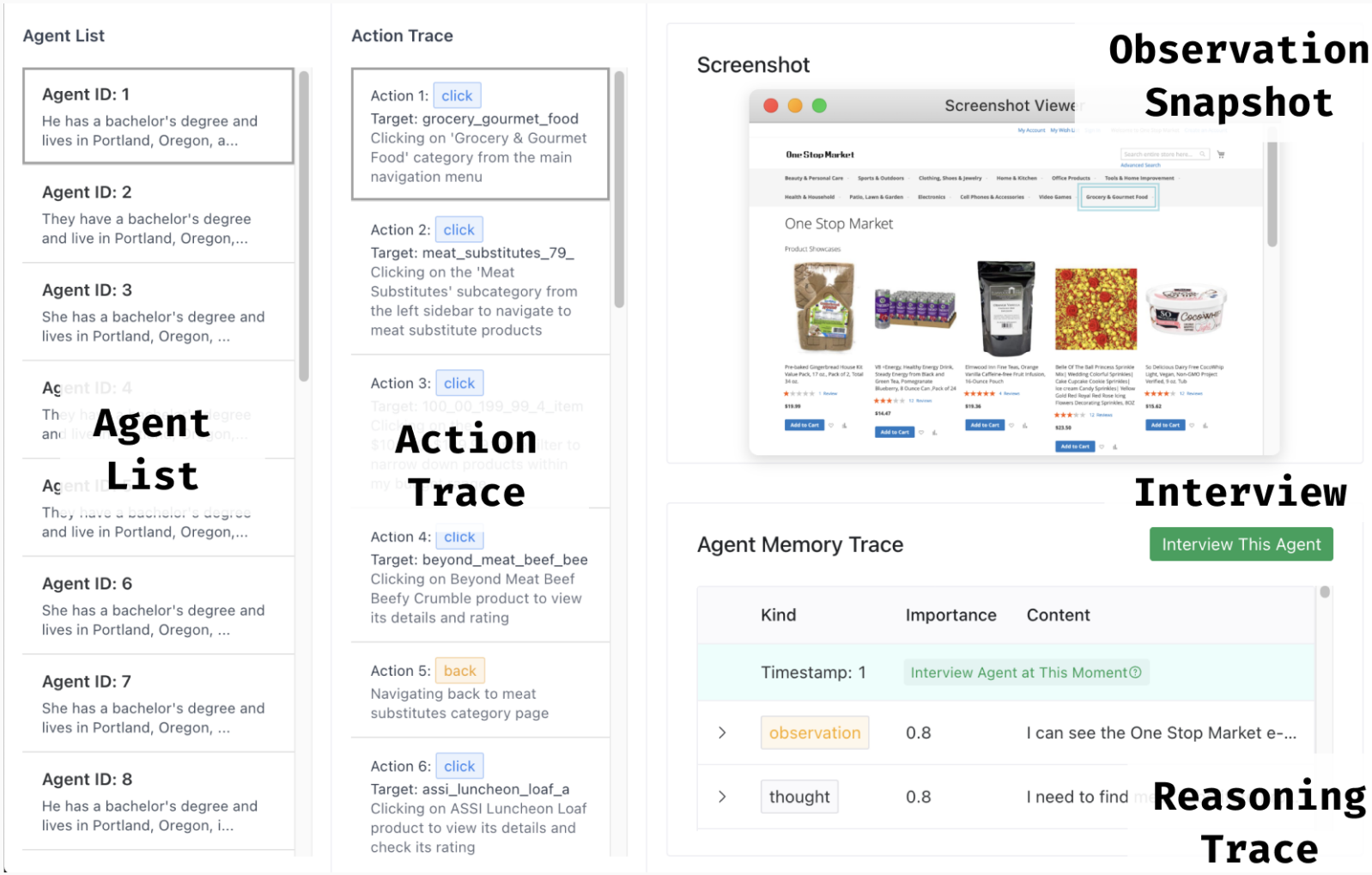


Figure 4: Result Viewer Interface

Universal Browser Connector

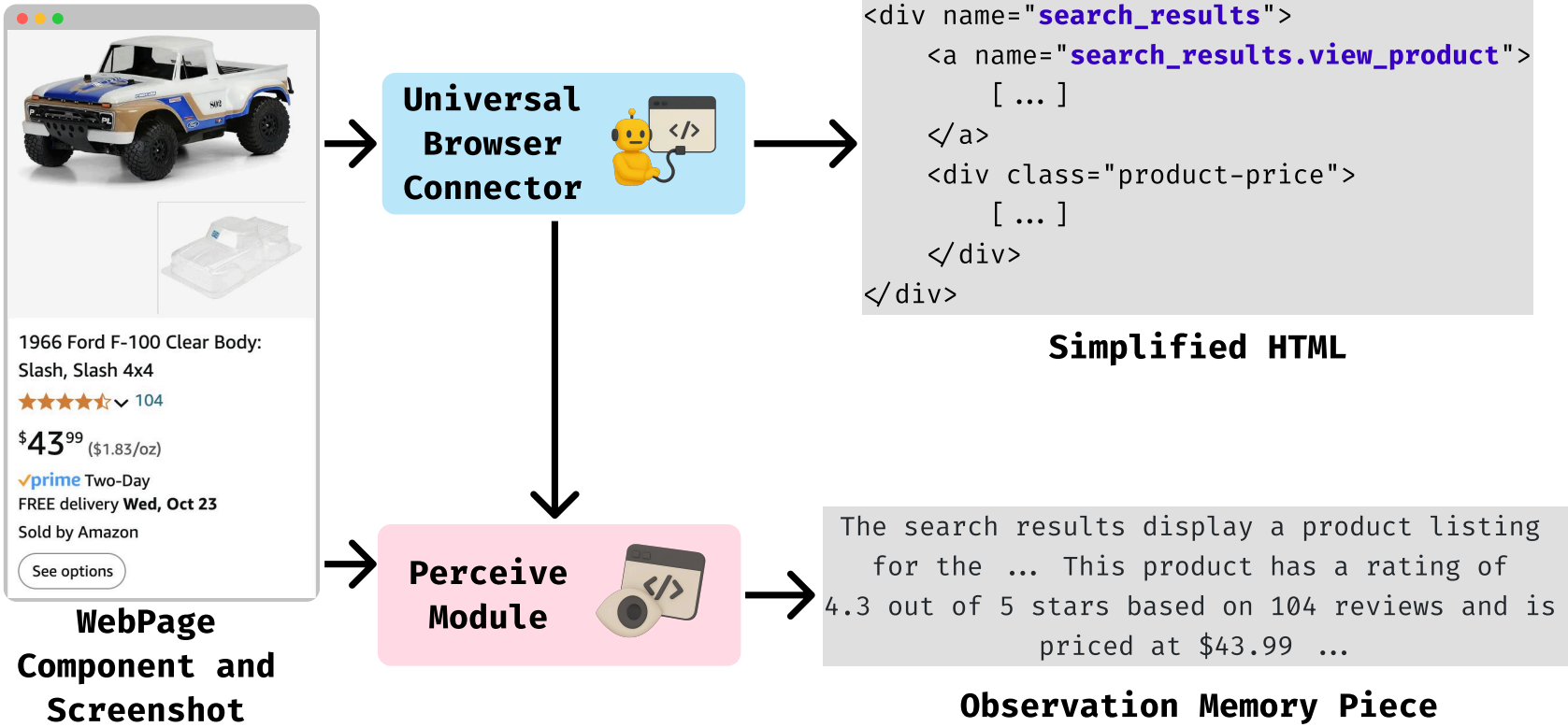


Figure 5: Universal Browser Connector

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?