Shiyun Xiong

📞 +86 15571600452   ✉ christine_shiyunx@outlook.com

EDUCATION

Harbin Institute of Technology   **Weihai, China**

Bachelor of Software Engineering   *Sep. 2021 – Jun. 2025(Expected)*

- Weighted Average: 90.0/100
- Main course: Linear Algebra(92), Calculus(92), Algorithm Analysis and Design(97), Probability theory and mathematical statistics(90), Introduction to Artificial Intelligence(92), Data Structure(94), Database System(97), Operating Systems(94)

RESEARCH EXPERIENCES

MLR-Agent: Autonomous Agents for Machine Learning Research   **Work with Zhiyuan Hu**

National University of Singapore, SoC/IDS   *Dec. 2024 – Present*

- This project aims to develop a agents framework capable of autonomously generating research ideas, conducting experiments, and writing research papers, with the long-term goal of having a paper accepted at a reputable venue.
- My contributions: The development of a framework for agents to autonomously conduct experiments is currently underway. Including an iterative process where agents write and run code, evaluate the results, and adjust the research plan based on feedback.

A Graph-Enhanced Multi-Modal Approach to Long Video Question-Answering   **Advisor: Prof. Yue Feng**

University of Birmingham   *Oct. 2024 – Present*

- This project aims to enhance the understanding of long videos and support question-answering tasks by utilizing heterogeneous graph algorithms combined with multi-modal large language models (MLLMs).
- My contributions: Designed a framework and developed algorithms to convert long video data into event-based graphs. Integrated heterogeneous graph algorithms with MLLMs to facilitate better semantic understanding and reasoning over the video content. Future work involves conducting phased fine-tuning and training on various datasets, followed by testing the framework's effectiveness in video-based question-answering tasks.

Self-Evolving Intelligent Agents in Visual Web Environments   **Work with Zhiyuan Hu**

National University of Singapore, SoC/IDS   *Apr. 2024 – Oct. 2024*

- This project aims to enhance agent interaction within various visual digital platforms using a proprietary reward model. The goal is to enable agents to adapt based on user interactions and environmental feedback.
- My contributions: Designed an agent framework where agents interact with the different environments using various prompting methods and constructed training data for different reward models based on existing data features. Conducted experiments in both static and dynamic environments to evaluate the effectiveness of the framework and its performance in diverse scenarios.

Multi-Agent Framework for Web Services with Fuzzy Requirements   **Advisor: Prof. Zhiying Tu**

Research Center of Intelligent Computing for Enterprises & Services (ICES, HIT)   *Jan 2024 – Apr.2024*

- This project implemented H-Tower, the first LLM-based multi-agent framework simulating multi-turn interactions between humans and websites. It achieves emotion-aware elicitation of complex requirements and mass-produces training data for this task.
- My contributions: Addressed challenges posed by fuzzy user requirements in conversational settings, developed techniques utilizing multiple agents to efficiently elicit comprehensive user needs. Additionally, validated H-Tower's effectiveness through experiments showing its accuracy in reasoning tasks and its zero-shot generalization ability on out-of-distribution websites.

A Benchmark for Universal Instruction Following in Realistic Web Services Navigation   **Advisor: Prof. Dianhui Chu**

Research Center of Intelligent Computing for Enterprises & Services (ICES, HIT)   *Oct. 2023 – Mar. 2024*

- This project constructed the first Chinese multi-modal benchmark and dataset to evaluate web service navigation tasks and devised an efficient multi-modal framework to evaluate performance on the dataset.
- My contributions: Developed a multi-modal framework integrating slot tree maintenance, instruction parsing, and web execution, achieving a 68.61% success rate in automatic web navigation. Proposed using object detection to filter noise from web pages, converting visual information into language data and fine-tuned a large language model for better slot tree utilization.

PUBLICATIONS

Bolin Zhang, **Shiyun Xiong**, Dianbo Sui, Yunzhe Xu, Zhiying Tu, and Dianhui Chu. 2024. **"RealWeb: A Benchmark for Universal Instruction Following in Realistic Web Services Navigation."** In *2024 IEEE International Conference on Web Services (ICWS)* (CCF B)

Zhiyuan Hu, **Shiyun Xiong**, Yifan Zhang, See-Kiong Ng, Anh Tuan Luu, Bo An, Shuicheng YAN, Bryan Hooi **"Guiding VLM Agents with Process Rewards at Inference Time for GUI Navigation."** *Preprint*

COMPETITIONS

China Undergraduate Mathematical Contest in Model (CUMCM) (<u>First Prize</u> in Shandong Division)   **Sep. 2023**
- **Title**: Optimization Model of Heliostat Field Based on Adaptive Gravitational Search Algorithm
- Developed an efficient ray field model for analyzing optical parameters and utilized the Adaptive Gravitational Search Algorithm (AGSA) to optimize heliostat configurations, resulting in substantial annual average thermal power output.

SKILLS

Technical Skills and Tools: Python, C++, C, Java, Matlab, Java, Pytorch, SQL, Selenium, Android Studio, etc.

Languages: Chinese Mandarin (naive), English (CET 4: 613, CET 6: 564)