

# Haoxuan Li

✉ lihaoxua@usc.edu | ☎ +1 (657) 243 - 7436 | 🌐 <https://haoxuanli.com/>

## EDUCATION

---

**University of Southern California (USC)**, Los Angeles, CA Aug. 2023 — May 2025  
Master of Science in Computer Science

**University of California, Irvine (UCI)**, Irvine, CA Sept. 2022 — Jun. 2023  
Electrical Engineering and Computer Science, Joint Program with HIT

**Harbin Institute of Technology (HIT)**, Harbin, China Aug. 2019 — Jun. 2023  
Bachelor of Engineering in Computer Science and Technology

## SELECTED PROJECTS

---

### LLM Multi-Agent Framework for Structured Data Bias Detection 🔗

- Proposed and implemented the first end-to-end multi-agent LLM framework for automated bias detection in structured datasets using collaborative planning and tool execution.
- Designed a multi-agent architecture with Primary and Advisor Agents to perform iterative task planning, preprocessing, bias analysis, visualization, and report generation.
- Built an extensible bias detection ecosystem containing 46 predefined analytical tools and 100+ retrievable bias detection methods covering categorical and numerical distribution/correlation bias.
- Developed a RAG-based method retrieval pipeline enabling dynamic selection, generation, and execution of statistical bias detection workflows from a structured method library.
- Constructed BIASBENCHMARK, a comprehensive benchmark with 100 structured-data bias detection tasks and automated evaluation pipelines for both end-result and intermediate-process assessment.
- Conducted extensive experiments across GPT-4o and Llama 3.3 70B frameworks, demonstrating substantial improvements over ReAct and self-reflection agent baselines in structured data bias detection.

### Context-Aware Multimodal Privacy Detection System 🔗

- Developed a context-aware multimodal privacy detection system that evaluates whether image sharing may cause privacy leakage under different real-world usage scenarios.
- Constructed SituationalPriv, a benchmark with 440 privacy-relevant images and paired sensitive/non-sensitive contextual scenarios to evaluate VLMs' contextual privacy understanding.
- Designed a training-free LLM-guided framework integrating GPT-4o, LLaVA, and RAM-Grounded-SAM to identify privacy-sensitive objects and generate context-aware visual queries.
- Conducted systematic evaluations across multiple LLM/VLM configurations, demonstrating improved recall and F1 score over LLM-only and VLM-only baselines for privacy leakage detection.

### Causal Reasoning Analysis and Enhancement for Vision-Language Models 🔗

- Built VQA-Causal and VCR-Causal, two causal-order reasoning benchmarks with 5,000+ image-text instances and 12 causal conjunction patterns to isolate VLMs' causal reasoning ability from object and activity recognition.
- Evaluated 12 widely used score-based and generative VLMs, including CLIP, BLIP, FLAVA, LLaVA, and Qwen-VL, showing that most models perform near random guessing on causal reasoning despite strong visual recognition performance.
- Conducted controlled object/activity understanding tests and data-level analyses on LAION-400M, MSCOCO, VQA, and VCR, revealing that VLM failures are strongly associated with the scarcity of explicit causal expressions in training and benchmark data.
- Developed CausalCLIP with hard-negative contrastive fine-tuning, improving causal reasoning performance on both in-domain VQA-Causal and zero-shot VCR-Causal benchmarks while maintaining downstream image-text retrieval performance.

## ACADEMIC EXPERIENCE

---

**Language, Intelligence, and Model Ethics (LIME) Lab**, USC Jul. 2024 — May. 2025  
*Research Assistant | Advisor: Prof. Jieyu Zhao*

- Conducted research on trustworthy AI systems, focusing on leveraging large language models and AI agents to improve fairness, reduce bias, protect privacy, and enhance multimodal reasoning capabilities.
- Led the design and implementation of an LLM-based multi-agent framework for bias detection, a multimodal privacy leakage detection system integrating LLMs and VLMs, and research projects on causal reasoning evaluation and enhancement for vision-language models.
- Independently completed the full research and development lifecycle of multiple AI projects, including problem formulation, framework design, model implementation, dataset and benchmark construction, experimental analysis, model optimization, and paper writing.
- Authored research works accepted at SoCalNLP Symposium 2024, NeurIPS 2025 Workshop on Efficient Reasoning, and EACL 2026.

**Computer Vision Lab, UCI**

Oct. 2022 — Jun. 2023

*Research Assistant | Advisor: Prof. Glenn Healey*

- Processed and analyzed hyperspectral image data for scientific research and visualization tasks.
- Developed data processing and visualization workflows for Mollweide projection analysis of baseball pitch distributions.

**Massive Data Computing Lab, HIT**


Mar. 2021 — Jun. 2023

*Research Assistant | Advisor: Prof. Hongzhi Wang, Dr. Xiaou Ding*


- Conducted research on data quality, data mining, automated error detection, and data cleaning for structured datasets.
- Developed and evaluated data cleaning algorithms and data quality-aware analysis frameworks through experiments with state-of-the-art baseline methods.
- Authored research works published in Journal of Software 2023 and Intelligent and Converged Networks 2022.


**PUBLICATIONS**


---

Haoruan Li, Mingyu Derek Ma, Jen-tse Huang, Zhaotian Weng, Wei Wang, Jieyu Zhao. **BIASINSPECTOR: Detecting Bias in Structured Data through LLM Agents**. SoCal NLP Symposium 2024. 

Zhaotian Weng\*, Haoruan Li\*, Jieyu Zhao. **SituationalPriv: A Context-Aware Framework for Privacy Detection and Protection in Vision-Language Models**. NeurIPS 2025 Workshop on Efficient Reasoning. 

Zhaotian Weng, Haoruan Li, Kuan-Hao Huang, Jieyu Zhao. **What's Missing in Vision-Language Models? Probing Their Struggles with Causal Order Reasoning**. EACL 2026. 

Xiaou Ding, Yingze Li, Chen Wang, Hongzhi Wang, Haoruan Li. **Time Series Data Quality Rules Discovery with Both Row and Column Dependencies**. Journal of Software 2023. 

Xiaou Ding, Hongzhi Wang, Genglong Li, Haoruan Li, Yingze Li, Yida Liu. **IoT data cleaning techniques: A survey**. Intelligent and Converged Networks 2022. 

**SKILLS**

---

- **Programming:** Python, SQL, Java, TypeScript, JavaScript, C/C++, R, MATLAB.
- **AI/ML Frameworks:** PyTorch, Hugging Face, LangChain, LangGraph, Scikit-learn, NumPy, Pandas.
- **Tools & Platforms:** Git, Docker, Linux, Jupyter Notebook, VS Code, MySQL, LaTeX.
- **APIs & Models:** GPT-series APIs, DeepSeek APIs, Llama, CLIP, BLIP, LLaVA, Qwen-VL.