JIAHAO HUO

+86 19822717806 | Homepage| Google Scholar | jiahaohuotj@gmail.com

EDUCATION

Tongji University, Shanghai, China Sept 2021-July 2025
Bachelor in Data Science ***Senior*** | *GPA:**91.52**/100*

Technical University of Munich, Munich, Germany 2024/10/1-2025/4/1
School of Computation, Information and Technology(CIT) *Exchange Student*

Hong Kong University of Science and Technology(Guangzhou) Sept 2025-Present
AI Thrust *MPhil*

SKILLS

Programming Languages: Python•C++•Matlab•SQL•Markdown
Libraries,Frameworks and Tools: PyTorch•Transformers•vLLM•Diffusers•Plotly•Flask•Git•MySQL•Latex
Mathematics: Mathematical Analysis•Advanced Algebra•Stochastic Process•Game Theory
Spoken Languages: English(102/120 for TOEFL), Chinese, Japanese(N4 level), Germany(Beginner)

RESEARCH EXPERIENCE

The Hong Kong University of Science and Technology February 2024 – Present
Research Intern; Advisor:Xuming Hu and Mingxun Zhou *Computer Science and Engineering*

- LLM watermarking with **distortion-free and robustness** guarantee.
- Interpretation of **cross-modality** ability of multimodal models.
- Dynamics of **cross-lingual knowledge transfer** in LLMs

Pixel Lab June 2025 – September 2025
MLE Intern; Mentor:Chengfei Lv *Alibaba Group*

- Efficient Audio-langauge Model for **Edge-side Speech Synthesis**.
- Leveraging LLMs for large-scale codebase understanding, while overcoming **context window limitations**. (Under Exploration)

Institute for Advanced Algorithms Research April 2025 – June 2025
Core Dev.; Advisor:Zhiyu Li *MemTensor (Shanghai) Technology Co., Ltd.*

- Developing **open source agentic memory** management system
- Enhances **compatiblility with LLM frameworks** like Ollama, Transformers, vLLM and MCP.
- **Evaluating system performance** across various competitors like Mem0 and Zep-cloud.

Squirrel AI Learning September 2024 – April 2025
Research Intern; Advisor:Shen Wang *AI For Eduacation*

- Improving Multimodal Large Language Models for **Mathematics Problem Solving**.
- Multimodal Large Language Models as agents for **realworld error detection** in exams.

Tongji University June 2023 – March 2024
Research Intern; Advisor:Zhihua Wei *Computer Science and Technology*

- Constructed an random-forest addiction prediction model based on gastrointestinal microbiota abundance
- Utilized **text-to-image diffusion models** for the generation of pixel-wise construction datasets, **with an associated paper currently in progress**

RECENT AWARDS

National Second Prize in the Contemporary Undergraduate Mathematical Contest in Modeling. September 2023
First Prize in Shanghai Region of The Chinese Mathematics Competitions. October 2022
Undergraduate Scholarship of Tongji University November 2022

PUBLICATION

- **PMark: Towards Robust and Distortion-free Semantic-level Watermarking with Channel Constraints (Under Review)** First Author 2025.09
- MemOS: A Memory OS for AI System **(Github 2.4k Stars)** Core-developer 2025.07
- Pierce the Mists, Greet the Sky: Decipher Knowledge Overshadowing via Knowledge Circuit Analysis **(EMNLP'25 Main)** Co-author 2025.05
- MathAgent: Leveraging a Mixture-of-Math-Agent Framework for Real-World Multimodal Mathematical Error Detection **(ACL'25 Industry Oral)** Co-author 2025.03
- **MMUnlearner: Reformulating Multimodal Machine Unlearning in the Era of Multimodal Large Language Models (ACL'25 Findings)** **First Author** 2025.02
- EssayJudge: A Multi-Granular Benchmark for Assessing Automated Essay Scoring Capabilities of Multimodal Large Language Models **(ACL'25 Findings)** Co-author 2025.02
- **Position: Multimodal Large Language Models Can Significantly Advance Scientific Reasoning** (Under Review) **Co-author** 2025.02
- **Explainable and Interpretable Multimodal Large Language Models: A Comprehensive Survey** (TPAMI Submitted) First Co-author 2024.12
- ErrorRadar: Benchmarking Complex Mathematical Reasoning of Multimodal Large Language Models Via Error Detection **(ICLR'25 Workshop)** Co-author 2024.10
- Miner: Mining the underlying pattern of modality-specific neurons in multimodal large language models (Under Review) Co-author 2024.10
- **MMNeuron: Discovering Neuron-Level Domain-Specific Interpretation in Multimodal Large Language Model (EMNLP 2024 Main)** **First Author** 2024.06

PERSONAL PROJECTS

Open Source Operation System for LLM-based Agent Memory Management (MemoryOS)
- Enhance AI assistants and agents with an intelligent memory layer, enabling personalized AI interactions.
- Compatible with popular LLM framework and Agent Protocol (such as Ollama, OpenAI, vLLM, MCP, etc.)
- Everything is a memory: memorize multimodal context including images, videos and audios.

Redefining Multimodal Machine Unlearning for Selective Visual Knowledge Erasure in MLLMs (ACL 2025 Findings)
- Reformulated multimodal machine unlearning to **erase entity-specific visual patterns** while preserving textual knowledge in language model backbones.
- Developed **MMUnlearner**, a geometry-constrained gradient ascent method that optimizes MLLM weights via concept-aware saliency maps to protect non-target knowledge during unlearning.
- Demonstrated state-of-the-art performance over GA and NPO baselines across metrics for MLLM Unlearning.

Multimodal Error Detection via Agent-Driven Toolkit in Mathematical Problem-Solving (ACL2025 Industry Oral)
- Designed a **Mixture-of-Math-Agent framework** with three specialized agents for multimodal error detection, enabling API-driven dynamic routing of image/text analysis and error reasoning workflows.
- Developed **MathToolKits** integrating OCR, symbolic computation, and diagram parsing libraries, optimizing API call efficiency by 40% through automated tool debugging and latency-aware scheduling.
- Deployed the framework in production with 98% API success rate, achieving 89.3% student satisfaction across 1M+ K-12 users and reducing manual grading costs by $2.1M annually.

Exploring Neurons Specific to Different Vision Domains in Large Vision Language Models (EMNLP'24, Main)
- Discovering neurons that are sensitive to specific image domains.
- Analyzing the distribution of domain-specific neurons(DSN), as well as their influence on model performance.
- Exploring post-projection visual embeddings through logit lens.

Construction Sites Dataset Generation Based on Diffusion Models (ICIC'24, Oral)
- Applying current **text-to-image diffusion models** for generating images with corresponding pixel-wise annotations
- Employing **prompt engineering** techniques to synthesize a comprehensive dataset of construction scenarios
- Conducting **quantitative experiments** to validate the effectiveness of the employed methods
- Authoring a research paper based on the findings and insights gained from this project