WULyucheng

Contact: +86-15824110303 | E-mail: 3210100887@zju.edu.cn

EDUCATION

Chu Kochen Honors College, Zhejiang University (ZJU), China *09/2021-06/2025*
College of Computer Science & Technology, ZJU, China *09/2021-06/2025*

- Major: **Computer Science and Technology**, Bachelor of Science
- Overall GPA: 3.78/4.0 (86.42/100) | Top 15%
- Core Courses: *C Programming, Computing Systems, Digital Logic Design, Machine Learning and Data Analysis, Database System, Advanced Data Structure & Algorithm Analysis, Computer Organization, Object-oriented Programming, Computer Networks, Compiler Principle, Software Engineering, Natural Language Processing*
- Math Courses: *Mathematical Analysis, Linear Algebra, Game Theory, Mathematical Modeling, Probability Theory & Mathematical Statistics, Discrete Mathematics*

RESEARCHINTERESTS

Artificial Intelligence, Natural Language Processing, Multimodal Large Language Model, Fine-tuning, Steering

PUBLICATIONS

Wu, L. (2025). Automating Steering for Safe Multimodal Large Language Models (Manuscript submitted for EMNLP).

RESEARCH&PROJECTEXPERIENCE

ZJU Graduation Thesis & Project: Automating Steering for Safe Multimodal Large Language Models
09/2024-present
Research Assistant, ZJU-Ant Group Joint Research Center for Knowledge Graphs | Supervisor: Prof. Ningyu Zhang

- Conducted research to optimize MLLMs with the "Steering" method, aiming to enhance their safety
- Implemented LM-Steers, a steering method of LLM output word embeddings, on MLLM; completed testing on VLSafe, achieved positive results
- Compared the performance of steering in single versus hybrid transformer architectures

NUS Summer Research Project: Research on the Chinese Large Language Model Database *06/2024-09/2024 Research Assistant, High Performance Computing for Artificial Intelligence (HPC-AI) Lab, National University of Singapore (NUS)* | Supervisor: Prof. Yang You

- Created the world's first Chinese Presupposition Dataset to evaluate LLMs' reasoning abilities with implicit assumptions in Chinese
- Conducted experiments and determined the prompts for the data generation pipeline; employed GPT-4 for automated generation of presupposition sentence pairs; generated and processed data for 1,500 simple and 1,400 complex cases

MindSpore Model Development Challenge, Huawei Technologies Co., Ltd. *07/2024-08/2024*
Team Leader

- Fine-tuned a Llama2-based model to enhance mathematical abilities, using LoRA and prompt-tuning techniques
- The model achieved an accuracy score of 51.38%, ranked 5th out of 104 in a competition, and won a gold prize along with a 10,000 RMB award (5,000 for each member)

EXTRACURRICULARACTIVITIES

BaiDangHai Community Service, Hangzhou *07/2022-09/2023*

- Engaged in community improvement work, including aspects like culture, environment, and safety

HONORSANDSCHOLARSHIPS

Gold Prize, MindSpore Model Development Challenge, Huawei Technologies *2024*
First-class Scholarship for Outstanding Students, ZJU *2024*
Five-star Volunteer (250+ hours), ZJU *2023*
Third-class Scholarship for Outstanding Students, ZJU *2022*
Model Student of Excellent Academic Performance, ZJU *2022-2024*

SKILLS

English: TOEFL: 107 | GRE: 321
Computer Languages: C, C++, Python, Verilog | LLMs: Transformer, Chameleon, BERT
Software Skills: Dev C++, Visual Studio, VS Code, Pycharm, Anaconda, Matlab, PyTorch, Yuque