Huicheng Zhang

+86 13998723304 · 2112241@mail.nankai.edu.cn

EDUCATION

The Hong Kong University of Science and Technology (Guangzhou)
Ph.D. in Artificial Intelligence — *2025 – Present*

Nankai University
B.Eng in Computer Science and Engineering — *2021 – 2025*

RESEARCH EXPERIENCE

- **National-level Undergraduate Student Innovation Project** — ***Co-leader***, *2023–2024*
  - Developed an intelligent art education teacher by constructing a text-based virtual teacher with LLMs.
  - Focused on ensuring safety and alignment of the virtual teacher using AI safety strategies and reinforcement learning (PPO, DPO).
- **Enhance the Safety in Reinforcement Learning by ADRC Lagrangian Methods** — ***Co-First Author***, *2024*
  - Authored and submitted a co-first author paper to **NeurIPS-2025**.
  - Proposed a novel ADRC Lagrangian framework to improve the safety and robustness of reinforcement learning algorithms against environmental disturbances.
- **The Learnability Dilemma in LLM Watermarking** — ***First Author***, *2024*
  - Authored a first-author paper for submission to **IEEE Transactions on Pattern Analysis and Machine Intelligence** (T-PAMI).
  - Investigated the trade-offs between watermark robustness, learnability, and anti-forgery capabilities in large language models.
- **Debate Framework for LLM-as-a-judge (Ongoing Project)** — ***Leader***, *2025–Present*
  - Proposed a novel framework to enhance the fairness and accuracy of the evaluator.
  - The core idea utilizes a multi-agent debate to extract nuanced features from the content under evaluation, effectively mitigating the inherent self-preference bias found in existing methods.

RELEVANT COURSEWORK PERFORMANCE

Course	Credits	Term	Score
Deep Learning	2.5	2024 Spring	97/100
Machine Learning and Application	2.5	2023 Fall	96/100
Introduction to Parallel Programming	2.5	2023 Spring	96/100
College Physics	4	2022 Spring	94/100
Project Training and Practice	2	2023 Fall	92/100
Natural Language Processing	2.5	2024 Fall	91/100
Introduction to Artificial Intelligence	2.5	2023 Spring	90/100
Bachelor Degree Thesis	6.0	2025 Spring	90/100

HIGHLIGHTS

- **Academic Excellence:** Maintained a weighted average score above 90 in core AI-related subjects.
- **Research Proficiency:** Rapidly produced high-quality research, authoring two papers for top-tier venues (NeurIPS and T-PAMI) within one year.

- **Self-Motivated Researcher:** Demonstrated strong passion and initiative by independently exploring cutting-edge topics and tackling complex research challenges.