Yuantao Zhang

Tel: 17850565385 (CHN), 87310540 (SIN)
Email: yuantaozhang@u.nus.edu
yuantaozhang@link.cuhk.edu.cn
Address: Blk 312C, Clementi Ave 4, #25-193, Singapore, 123312
Wechat Number: zytttdwxh

Education

National University of Singapore 2024.8 – Present
Master of Computing (Artificial Intelligence Specialisation) *GPA: 4.83/5*

Nanyang Technological University, Singapore 2023.1 – 2023.5
Undergraduate Exchange Program

The Chinese University of Hong Kong (Shenzhen) 2020.9 – 2024.7
Degree - Bachelor of Engineering with Honours, First Class *CGPA: 3.706 / 4 (top 10%)*
Major - Computer Science and Engineering *MGPA: 3.798 / 4 (top 10%)*

Publications

M. Li*, **Y. Zhang***, W. Wang, W. Shi, Z. Liu, F. Feng, T. Chua.
Self-Improvement Towards Pareto Optimality: Mitigating Preference Conflicts in Multi-Objective Alignment.
Submitted to ACL Rolling Review, February 2025 (*Equal contribution). [Arxiv, GitHub]

Y. Zhang, Z. Yang.
A Perplexity and Menger Curvature-Based Approach for Similarity Evaluation of Large Language Models.
First submission to TACL received review scores of 5, 3, and 2 out of 5. [Arxiv, GitHub]

Patents

B. Yang, Z. Yang, Y. Zhang. (2024). *An Approach for Generating Diverse Images Based on Latent Diffusion Models.* CN Patent No. 202411135613.6.

Research Interests

Natural Language Processing LLM Alignment Reinforcement Learning

Research Experience

Research Intern in LLM Alignment at NExT++ 2024.8 – Present
Mentor: Dr. Moxin Li, Dr. Wenjie Wang (Postdoc), Prof. Tat-Seng Chua

- **Conducted preliminary experiments** to analyze the impact of conflicting preference pairs on the Pareto Front in Multi-Objective Alignment, **iteratively refined strategies** for generating Pareto-optimal responses, **performed ablation studies**, and **contributed to paper writing**.
- (Present) Investigating **Multi-Objective Alignment** in LLMs, with a particular focus on the **3H principles**: **harmlessness**, **helpfulness**, and **honesty**.

Research Intern at National Supercomputing Center in Shenzhen 2024.3 – 2024.7
Mentor: Dr. Zhankui Yang (Postdoc)

- Conceptualized a metric utilizing **Perplexity Curves**, which plot the perplexity of sub-sequences in a sentence against word indices, and **Menger Curvature Difference** of these curves to assess the similarity between LLMs. This metric can aid in **detecting unethical practices in LLM usage**, such as marginally altering an existing LLM to falsely claim a new development or covertly distilling domain-specific knowledge from one LLM to another.
- Assisted in developing the idea of a patent.

Large Language Model Research 2023.9 – 2023.12

Mentor: Dr. Feng Jiang (Postdoc), Prof. Haizhou Li

- Conducted literature review on **popular open-source LLMs**, including GLM and Baichuan; systematically organized findings and contributed to a public resource. [GitHub]
- Trained a **560M Phoenix LLM** and applied it to various language processing tasks.

Data Distillation Using the Stable Diffusion Model 2023.5 – 2023.8

Mentor: Dr. Haonan Wang, Prof. Kenji Kawaguchi

- Reviewed key literature on **Generative Models** and **Data Distillation** techniques.
- Developed a pipeline to map images to token embeddings using **Stable Diffusion**, employing **Gradient Matching Loss** to optimize these embeddings for generating distilled images. [GitHub]

Convergence Rate Analysis of FedDualAvg 2022.9 – 2023.5

Mentor: Prof. Ming Yan

- Reviewed literature on **Federated Optimization**, with a focus on the integration of the **Bregman Proximal Gradient Method** and **FedAvg**, particularly in algorithms like FedDualAvg.
- Analyzed the convergence rate of **FedDualAvg** in different settings (different constraints of data heterogeneity, non-convex setting) and conducted experiments to verify theoretical results.

Internships

Shenzhen Tibo Ruike Network Technology Co., LTD 2021.8 – 2022.1

Back-end Developer

- Developed and tested back-end APIs for web applications aimed at enhancing students' campus life experience, using Postman and unit testing frameworks.
- Contributed to the back-end development of "Yoummy," a food ordering mini-program in collaboration with CUPL, ensuring accurate processing of user requests. [GitHub]

Awards & Honors

AY2024-2025 Semester 1 Dean's List
National University of Singapore, School of Computing *2025.1*

AY2020-21, AY2021-22, AY2022-23, AY2023-24 SDS Dean's List
The Chinese University of Hong Kong (Shenzhen), School of Data Science *2021.9-2024.9*

AY2022-23 SDS TA/USTF Award
The Chinese University of Hong Kong (Shenzhen), School of Data Science *2023.11*

AY2021-22 Academic Performance Scholarship: Class C 20000 RMB
The Chinese University of Hong Kong (Shenzhen), School of Data Science *2022.12*

Undergraduate Research Award
The Chinese University of Hong Kong (Shenzhen), the URA Selection Committee *2022.8*

Bowen Scholarship 35000 RMB
The Chinese University of Hong Kong (Shenzhen) *2020.9*

Languages & Skills

Languages – English (**GRE**: 328 + 3.5, **TOEFL**: 102), Mandarin (native)
Skills – Pytorch, Tensorflow, SQL, C++, Java, Linux, Latex