

Haojin(Sean), WANG

Email: seanwang027@gmail.com | Tel: +86-15828330768
Address: 4800 Cao'an Highway, Shanghai 201804, P.R. China

EDUCATION

-
- Tongji University, China** 09/2021 - 06/2025
- Bachelor of Engineering in Computer Science and Technology (Expected) CGPA: 4.90/5.0
 - Core modules: C/C++ Programming, Advanced Mathematics, High-Level Language Programming, Linear Algebra, Discrete Mathematics, Introduction to Computer Science, Advanced Language Programming Practice, Data Structures, Probability and Mathematical Statistics, Web Technology, Formal Languages and Automata, Assembly Language Programming, Artificial Intelligence Principles and Technologies, Computer Architecture, Computer Organization, Algorithm Analysis and Design, Cocoa based Apple Software Development, Machine Learning, Brain and Cognitive Computing, Principles of Database Systems, Operating Systems, etc.
- National University of Singapore, Singapore** 01/2024 - 05/2024
- Exchange Student
 - Core modules: Introduction to Media Computing, Information Retrieval, Data Engineering

RESEARCH INTERESTS

-
- Natural Language Processing; Information Theory; Language Modeling; AI Safety.

RESEARCH EXPERIENCE

-
- **Research on Next Token Distribution Analysis** 05/2024 - Present
Advisor: Prof. Freda Shi, David R. Cheriton School of Computer Science, University of Waterloo
The project focuses on uncovering the mechanisms of natural language generation from the perspective of next token distribution.
 - Gained experience in organizing the code structure of a scientific project and deepened understanding of large language models (LLMs), particularly the transformer architecture;
 - Will be a Visiting Student at the University of Waterloo from September to December 2024, conducting research on natural language processing, including analyzing the probability distributions defined by language models and how these distributions vary with the changes in input sequences;
 - Currently planning to write and publish a research paper based on this work as the first author.
 - **Research on Estimation of Uncertainty on NLG tasks** 06/2024 - Present
Advisor: Prof. Hanjie Chen, Department of Computer Science, Rice University
 - As the primary researcher responsible for this project, conducted extensive literature reviews on uncertainty quantification, leading to the development of several well-formed ideas;
 - Gained foundational knowledge in uncertainty estimation, with a particular focus on adapting conformal prediction methods from classification tasks to natural language generation tasks.
 - **Knowledge Graph Construction for ChangYuan, ADMIS Lab, Tongji University** 07/2024 - Present
Advisor: Professor Wengen Li, College of Electronics and Information Engineering, Tongji University
 - In charge of LLM based work, utilizing LLM(Qwen model) to extract entity sets from unstructured data for the construction of a knowledge graph;
 - Currently refining the methodology to enhance accuracy based on the company's requirements.
 - **Research in Explainable and Controllable AI Lab, Stevens Institute of Technology** 04/2024 - 07/2024
Advisor: Professor Zining Zhu, Stevens Institute of Technology

-Acquired foundational knowledge in the concepts of Machine Unlearning;

-Developed an idea to apply machine unlearning to datasets in pretraining corpora to investigate the effects on model performance; however, a similar concept was published before our work could be completed.

- **Research in StatNLP Lab, Singapore University of Technology and Design** 01/2024 - 05/2024

Advisor: Dr. Lu Wei, Singapore University of Technology and Design

-Learned basic knowledge in NLP;

-Attended weekly meetings and discussed the fundamentals of LLM alignment, with a particular focus on RLHF area.

- **Research on Duplicate Text Detection, ADMIS Lab, Tongji University** 04/2023 - 09/2023

Advisor: Professor Wengen Li, College of Electronics and Information Engineering, Tongji University

-Gained basic knowledge in prompt engineering, taking a leading role in a project focused on detecting duplicate text in data associated with private company information;

-Constructed a Knowledge Graph for the company's information, and developed a search engine based on that knowledge graph.

EXTRACURRICULAR ACTIVITIES

- **Main Code Writer** in Project License Plate Detection and Recognition at Large Angles 01/2023 - 12/2023 (National Training Program of Innovation and Entrepreneurship for Undergraduates)

- **Project Leader** in Sarcasm Detection in Headlines based on BERT (HuaWei Mindspore Training Program) 08/2023 - 09/2023

-Led the team to migrate the training process from pytorch to mindspore, gained great knowledge in architecture of Transformer and fundamentals in LLMs;

-Obtained the second prize in Innovation.

- **Team Leader** in China Collegiate Computing Contest: 2023 Mobile Application Innovation Contest 03/2023 - 06/2023

-Served as the HCI designer and Idea Provider for the team work;

-Obtained the second prize in East China.

AWARDS & HONORS

- **Excellent Student Scholarship Second Prize** in December 2023.
- **Second Prize** in 2023 9th China Undergraduate Statistical Modeling Competition in August 2023.
- **Second Prize** in China Collegiate Computing Contest: 2023 Mobile Application Innovation Contest in July 2023.
- **National Scholarship** (as the top 5 student in the department) in July 2022.

SKILLS

- **Languages:** Mandarin (native), English(Proficient): TOEFL 109; GRE 328+3.5
- **Technical Skills:** Proficient in MS-Office Software: Word, Excel, PowerPoint; C, C++, Python, JavaScript; Fluent Latex skills.
- **Hobbies:** Fitness, Music, Movie