

Xuan Ouyang

✉ xouyang7@wisc.edu | ☎ 6083208044

Education

University of Wisconsin–Madison	2024/09 - 2026/06 (Expected)
Bachelor of Science in Computer Science	Wisconsin, USA
GPA: 3.871/4.000 (Good academic standing)	
Nanjing University of Science and Technology	2022/09 - 2024/06
Candidate for Bachelor of Science in Applied Statistics	Nanjing, Jiangsu, China
<ul style="list-style-type: none">• Science and Technology Activist Award, 2023. (1%)• School First-Class Academic Scholarship, 2023. (5%)	

Research Interests

I am broadly interested in **Multimodal Large Language Models** and **AI agent**. My previous work was on **dataset pruning and natural language processing application**. I recently collaborated with Dr. Yujian Gan from University College London to conduct some research on the evaluation and optimization of text-to-SQL tasks for large language models. Now my research is mainly about **Agent Tool use and Multimodal reasoning**.

Research Experience

Shanghai Jiao Tong University Epic Lab	2025/03 - now
Remote Research Intern (Collaborate with Shaobo Wang)	Madison, US
<ul style="list-style-type: none">• I designed and implemented a novel ways to filter existing Tool-LLM corpora, automatically identifying a small subset of high-quality tool-invocation examples based on diversity of tool calls, contextual complexity, data quality, and diversity. Trained agent models on the filtered dataset and demonstrated a relative improvement in tool-usage success rate over baseline agents trained on the unfiltered corpus.	
University College London	2024/08 - 2024/12
Research Assistant, advised by Dr. Yujian Gan	London, UK
<ul style="list-style-type: none">• My work is to build a database based on the spider database, which contains the questions and the two corresponding SQL queries for evaluating the robustness of text-to-SQL models. Try to address the mismatch between natural language descriptions and the corresponding SQL queries. My paper is under reviewed by EMNLP 2025.	
Nanjing University of Science and Technology X Nanjing Medical University	2023/09 - 2024/03
Research Assistant, advised by Prof. Luyin Gui	Nanjing, Jiangsu
<ul style="list-style-type: none">• My work is to use the deep learning MONAI framework to segment the airway medical image model provided by Nanjing Medical University, and segment the patient’s tracheal image into four parts for pathological research.	

Publications

Enhancing NLDBs: Advancing from Text-to-SQL to Text-to-Multi-SQL	EMNLP Reviewing, 2025
Xuan Ouyang*, Yuxi Lin*, Yujian Gan*, Zhilin Zhang, Jinxia Xie, Vassilis Routsis†	
MelTrim: Coarse-to-Fine Data Pruning for Speech Recognition	EMNLP Reviewing, 2025
Shaobo Wang*, Xintong Li*, Xuan Ouyang*, Tianle Niu, Zhengkun Ge, Yue Min, Xiaoqian Liu, Hankun Wang, Linfeng Zhang†	

Working Experience

Machine Learning Research Assistant of Marler Lab	2025/01 - now
University of Wisconsin–Madison	Madison, US
<ul style="list-style-type: none">• Collaborated with graduate students and faculty to optimize machine learning models for processing and synchronizing complex behavioral data, including semantic and syntactic analyses of mouse vocalizations.• Developed and maintained robust computational pipelines for analyzing rodent vocalizations and movements using cutting-edge tools such as DeepSqueak and Sleep.	
Machine Learning Tutor	2024/012 - 2025/03
University of Wisconsin–Madison	Madison, US

- Work every week to provide personalized support for lecture content, homework, and exam preparation, ensuring students gained a deep understanding of course material.
- Mentored students in implementing advanced machine learning algorithms—including PCA, clustering, CNNs, and Q-learning—facilitating hands-on learning and practical application of theoretical concepts.

Data Engineer, Intern
Tencent

2024/04 - 2024/08
Shenzhen, China

- Participated in the development and continuous optimization of the company’s internal big data infrastructure components, including performance optimization and secondary development of Hive, Spark, and other tools, which significantly improved data processing efficiency and stability.
- Responsible for the cleaning, conversion, integration and data warehouse model design of business data, built an efficient and scalable ETL process, and improved data processing efficiency by 30% by optimizing SQL query performance.

Selected Awards

I am the head of the Mathematical Modeling Association of Nanjing University of Science and Technology. I have participated in many mathematical modeling competitions in the past and have rich experience in machine learning. I am very happy to be selected as a **Dean’s List** student at UW-Madison for every semester.

Dean's List in UW-Madison

2024/09-2025/05

- High-achieving students in UW-Madison.

MCM/ICM (3-people Group)

2023/12 - 2024/02

- Honorable Mention, Second Prize in World Class.

May Day Mathematical Modeling Contest (3-people Group)

2023/05 - 2023/05

- First Prize in jiangsu Province.(0.1%)

Technical Skills

- **Languages:** Python, C/C++, java, Shell, LaTeX, Julia
- **Frameworks and Tools:** Pytorch, Docker, Sklearn, Numpy, Pandas
- **AI:** Natural language Processing (llama-2, ChatGLM-3, CPM-Bee) | Computer Vision (YOLO Series, OpenCV)