Wenjie Du

duwenjie24@gmail.com • Homepage • Google Scholar • Singapore

Education

Nanyang Technological University, Singapore — Jan. 2026 – Present

M.Sci. in Artifact Intelligence at College of Computing and Data Science

Sichuan University, Chengdu, China — Sept. 2020 – June 2024

B.Eng. in Computer Science and Technology, **Average Scores:** 89.4/100, **GPA:** 3.74/4.00, **Rank:** 20/318

Publications

- Keda Tao, **Wenjie Du**, Bohan Yu, Weiqiang Wang, Huan Wang, "OmniAgent: Audio-Guided Active Perception Agent for Omnimodal Audio-Video Understanding" in **arXiv**, [Page], [PDF]
- **Wenjie Du**, Li Jiang, Keda Tao, Xue Liu, Huan Wang, "Which Heads Matter for Reasoning? RL-Guided KV Cache Compression" in **arXiv**, [Page], [PDF]
- Hao Wen, Shizuo Tian, Borislav Pavlov, **Wenjie Du**, Yixuan Li, Ge Chang, Shanhui Zhao, Jiacheng Liu, Yunxin Liu, Ya-Qin Zhang, Yuanchun Li, "AutoDroid-V2: Boosting SLM-based GUI Agents via Code Generation" in **MobiSys'25**, **Best Artifact Award**, [PDF]
- Shanhui Zhao, Hao Wen, **Wenjie Du**, Cheng Liang, Yunxin Liu, Xiaozhou Ye, Ye Ouyang, Yuanchun Li, "LLM-Explorer: Towards Efficient and Affordable LLM-based Exploration for Mobile Apps" in **MobiCom'25**, [PDF]

Experience

Nanyang Technological University — Singapore

MS Research Project at MMLab, supervised by Ziwei Liu — Jan. 2026 – Present

- Conducting research on MLLM reward model for reinforcement learning.

Westlake University — Hangzhou, China

Visiting Student at EncodeLab, supervised by Huan Wang — July 2025 – Jan. 2026

- Developed OmniAgent, an audio-guided active perception agent for fine-grained understanding.
- Proposed RLKV, a reasoning-critical head identification method for guiding KV cache compression tailored to reasoning models. RLKV leverages reinforcement learning to assess each head's contribution to reasoning behaviors under sparse KV-cache allocation, achieving 20-50% KV cache reduction with near-lossless performance.

Hong Kong University of Science and Technology — Hong Kong SAR

Research Assistant at MINSys Group, supervised by Xiaomin Ouyang — Sept. 2024 – Mar. 2025

- Applied text-semantic embeddings to guide the model's inference pipeline, leading to improved cross-domain generalization in Human Activity Recognition tasks.
- Explored fine-tuning LLMs to understand IMU time-series sensor data using LLaVA and LLaMA-AdapterV2 architectures, aiming to align sensor signals with textual labels.

Institute for AI Industry Research, Tsinghua University — Beijing, China

Research Intern at AIoT Group, supervised by Yuanchun Li — Dec. 2023 – Aug. 2024

- Designed and implemented AutoDroid-V2, a script-based GUI Agent, which pre-constructed App API documentation, reducing token consumption by 90% and latency by 93.1% compared to traditional methods.
- Designed a UI reassembling framework based on LLM to identify important UI components and provide better visualization solutions as low-code schemes.
- Developed a low-cost, high-efficiency App exploration Agent LLM-Explorer, using LLM to memorize and explore Apps and improving coverage by 12% compared to traditional solutions.

ByteDance Ltd. — Shanghai, China

Golang Software Engineer Intern — May 2023 – Nov. 2023

- Developed features for a multi-source alert auto-processing platform using the Go language.
- Refactored system using Domain-Driven Design and Chain of Responsibility to enhance the code quality.

Uranus Quant Fin-tech Co. — Chengdu, China

C++ Software Engineer Intern — June 2022 – Sept. 2022

- Built high-performance log collection program for linux with multi-threading and asynchronous IO.

Selected Awards

- National Scholarship (Top 3/318) 2023
- Comprehensive First Class Scholarship of Sichuan University (Top 4/318) 2023
- Comprehensive Third Class Scholarship of Sichuan University 2022
- Comprehensive Second Class Scholarship of Sichuan University 2021
- Outstanding Graduate Student of Sichuan Province 2024
- Merit Graduate Student of Sichuan University 2024
- Merit Student of Sichuan University 2023
- Excellent Student Cadre of Sichuan University 2021,2022

Skills

Python, C, C++, Go, JavaScript, Linux, Git, LaTeX, Torch, Triton.