JIARUI WANG

Email: wangjiarui1@sjtu.edu.cn | Homepage: https://jruiwang.github.io

Summary

My research focuses on parametric memory and next-generation LLM architectures. I have two accepted papers at **NeurIPS 2025** and **ICLR 2026**. My representative works, **Memory Decoder** and **MLP Memory**, introduce a novel parametric memory paradigm at the pre-training stage, enabling more reliable knowledge storage and decoupling long-tail knowledge from LLM's reasoning capabilities. I am further working on memory for enhancing specialized capabilities of **frontier LLMs** while avoiding catastrophic forgetting. I have hands-on experience in large-scale LLM pre-training, SFT, and RAG architectures.

Education

Shanghai Jiao Tong University, Shanghai, China — Sept. 2025 - Jun. 2030
Ph.D. in Computer Science and Technology
Advisor: Prof. **Zhouhan Lin** (LUMIA Group)

Shanghai Jiao Tong University, Shanghai, China — Sept. 2021 - Jun. 2025
B.S. in Computer Science and Technology

Publications

- **Memory Decoder: A Pretrained, Plug-and-Play Memory for Large Language Models**
  Jiaqi Cao*, **Jiarui Wang***, Rubin Wei, Qipeng Guo, Kai Chen, Bowen Zhou, Zhouhan Lin
  NeurIPS 2025.
- **MLP Memory: A Retriever-Pretrained Memory for Large Language Models**
  Rubin Wei*, Jiaqi Cao*, **Jiarui Wang**, Jiashi Kai, Qipeng Guo, Bowen Zhou, Zhouhan Lin
  ICLR 2026.

Research Experiences

Research Intern, Shanghai AI Lab — Aug. 2025 - Present
- Training a parametric memory module on scientific domains to enhance the capabilities of *Intern-S1-235B*, a state-of-the-art foundation model.
- Building supporting infrastructure, including vLLM-based inference engine adaptation and optimization, large-scale data processing and retrieval components for new LLM architectures.

Research Assistant, DDST Lab, Shanghai Jiao Tong University — May. 2023 - Nov. 2023
- Researched and developed DebugPilot, an interactive code debugging navigation tool.
- Engineered the system to automate code error diagnosis and generate debugging procedures by leveraging probabilistic models and data dependency analysis.
- Implemented key features using Java Instrumentation for dynamic program analysis.

Honors and Awards

Second Place, RoboCup China Open - Small Size League, 2024
Third Prize, Prototype System Competition, CCF Chinasoft, 2023
Provincial First Prize, Senior Group, National Olympiad in Informatics in Provinces (NOIP), 2018