Jiajun Ruan

 jiajunruan | jruan@umn.edu | +1 447-446-0374

Current Address: 1010 W University Ave 241C, Urbana, IL 61801

Permanent Address: 171 Qingzhiwu 4-306, Hangzhou, Zhejiang 310013, China

EDUCATION

2024.9 - 2025.8	Master of Engineering at **Department of Electrical and Computer Engineering University of Illinois Urbana-Champaign, USA** (GPA: 3.92/4.0)
2021.9 - 2025.6	Bachelor of Engineering in **College of Information Science and Electronic Engineering, Zhejiang University, China** (GPA: 3.88/4.0, Rank: 20/128)

WORK EXPERIENCE

Research Assistant at University of Minnesota   July 2025 - present

- Conducted unlearning evaluation, assessing the performance of unlearning methods using the pass@k metric. Trained nine state-of-the-art unlearning models on the TOFU benchmark and evaluated their worst-case performance by repeatedly querying each model with the same input. My work explains the lack of robustness in current unlearned large language models and will be published in a forthcoming paper

- Contributed to the AI for Land Management project, converting the land management problem into a multi-objective optimization problem. Developed and optimized the problem model and incorporated constraints for smooth transitions between management patterns, making the project more applicable to real-world scenarios.

Research Intern at Ant Group   June 2025 - present

- Developed a novel automated traffic alert handling system by researching LLM-based approaches for network security. This system integrated Chain-of-Thought (CoT), Retrieval-augmented generation (RAG) and Standard Operating Procedure (SOP) techniques to enhance LLM capabilities, achieving over 95% recall and automating the daily processing of 50,000 alerts, with the work leading to a forthcoming paper publication.

PROJECTS

NetPress: Dynamically Generated LLM Benchmarks for Network Applications   Link

- We develop this benchmark to evaluate whether LLMs can solve real-world computer networking problems, such as diagnosing and fixing misconfigurations to restore full connectivity among all hosts in a network.
- This benchmark can dynamically generate queries by leveraging the hierarchical structure of computer networks, randomly selecting nodes or links and injecting predefined errors into a healthy network. To increase difficulty, multiple errors can be injected sequentially. The large and diverse query space ensures uniqueness and broad coverage.
- We design a simulator-based environment where LLMs act as network engineers, using multi-turn commands to diagnose and fix injected issues like routing, capacity, and Kubernetes errors. We comprehensively evaluate LLMs by measuring accuracy, iteration count, and use of unsafe commands.
- This dynamic benchmark avoids data leakage and improves accuracy by increasing the number of queries, and it also evaluates a model's generalization ability—for example, by finetuning on a subset of queries and testing on the rest.

Leak@k: Unlearning Does Not Make LLMs Forget Under Probabilistic Decoding

- We propose an innovative meta-metric Leak@k to evaluate information leakage after unlearning. Our evaluation metric first introduces probabilistic decoding and multiple generations, enabling it to realistically assess the risk of information leakage.
- We implement our evaluation metric in three widely-used unlearning datasets and finish extensive experiments. All our results show that LLM does not truly forget under our evaluation metric, revealing a huge gap of current unlearning methods.

Compressive Sensing Kernel Optimization for Direction of Arrival Estimation Based on Information Theory and Neural Networks

- Traditional compressive uses random gaussian kernel which causes some information loss. In this research, we use mutual information of prior knowledge and compressed signal as objective function to optimize compressive sensing kernel, in order to save more information in the compressed signal.
- Update neural networks to estimate DoA of signals based on the compressed signal and the ground truth DoA, and the output from neural networks is the prior knowledge for optimization of compressive sensing kernel next turn
- Iteratively update parameters in neural networks and compressive sensing kernel until convergence, then test the accuracy of DoA estimation compared with traditional compressive sensing.

OBJECTIVES

I want to advance **post-training optimization** algorithms for large language models, focusing on **unlearning, alignment, and safety**. In the short term, I aim to advance unlearning methods for large language models, developing **safer** and more **reliable** forgetting mechanisms. In the long term, I hope to make LLMs safer and more trustworthy, while also developing efficient training and optimization methods that **enhance their capabilities**, enabling them to **better support humanity and generate enduring societal benefits**.

ACADEMIC AWARDS

- Second-Class Scholarship of Zhejiang University   Oct 2023
- Second-Class Scholarship of Zhejiang University   Oct 2022
- First Award in the National Math Competition   Oct 2022
- Second Award in Physics Competition in Zhejiang Province   April 2023

PUBLICATIONS

Zhou, Yajie, Jiajun Ruan, Eric S Wang, Sadjad Fouladi, Francis Y Yan, Kevin Hsieh, and Zaoxing Liu (2025). "NetPress: Dynamically Generated LLM Benchmarks for Network Applications". In: *arXiv preprint arXiv:2506.03231*.

SKILLS

Scripting Language	TOEFL: 102 (Writing: 27, Speaking: 25).
Programming skills	Python, MATLAB, C, C++.
Research Skills	Familiar with Github for Project Management