

Yucheng Pan

+86 193 7248 5548 ◇ yuchengpan707@gmail.com ◇ Personal Homepage

EDUCATION

B.S. Mathematics and Physics, Tsinghua University

GPA: 3.87/4.00

Sep. 2021 - Jul. 2025 (expected)

Beijing, China

RESEARCH INTERESTS

Primary

Large Language Model Agents (LLM Agents), Reinforcement Learning

Supplementary

AI Safety, Multi-Agent Systems, Lifelong Learning, Bio-Inspired AI

RESEARCH EXPERIENCE

Department of Electrical and Computer Engineering, Princeton University

Remote Research Intern

May. 2024 - Present

Mentor: Prof. Mengdi Wang (Princeton ECE), Prof. Huazheng Wang (Oregon State EECS)

· ***LLM Agents Can Deceive and Be Constrained by Social Norms and Payoff Allocations***

Investigate how LLM agents exhibit deceptive behaviors in multi-agent games and design mechanisms to enforce compliance with social norms and fair payoff distributions.

- Literature review and idea development.
- Designed and implemented the overall framework and constraint mechanisms; conducted extensive experiments with comprehensive analysis.
- Submitted to *Nature Computational Science*.

· ***Data Augmentation for Generalizable LLM Agents***

Enhance the generalization capabilities of LLM agents through innovative data augmentation techniques.

- Literature review, idea development and preliminary experimentation.
- Proposed novel methods to redesign existing agent environments for fine-tuning pre-trained LLMs.

Department of Statistics and Operations Research, University of North Carolina at Chapel Hill

Remote Research Intern

Feb. 2024 - Oct. 2024

Mentor: Prof. Yao Li (UNC Chapel Hill STOR), Prof. Minhao Cheng (Penn State IST)

· ***Traceable Neural Audio Watermarking with Multi-Embedding***

Develop a novel audio watermarking model that enables embedding multiple watermarks from different sources into a single audio file to improve robustness and traceability.

- Literature review and idea development.
- Redesigned the architecture and fine-tuning pipeline of a state-of-the-art audio watermarking neural network for multi-embedding support.

Department of Statistics and Data Science, Tsinghua University

Undergraduate Research Assistant

Sep. 2023 - Jul. 2024

Mentor: Prof. Ke Deng

· ***Enhancing Precision in Isotope Nuclear Radius Estimation through Statistical Analysis***

Improve the precision of isotope nuclear radius estimation using advanced statistical analysis of laser spectroscopy data.

- Literature review, data collection, and data pre-processing.
- Developed statistical models; implemented statistical computing algorithms to enhance the precision and validated the results.

· *Nucleotide Sequences Prediction and Protein Design Based on Transformers*

Utilize a BERT model to predict nucleotide sequences and explore their applications in artificial protein design.

- Led the implementation and experimentation of the model.

Department of Automation, Tsinghua University

Undergraduate Research Assistant

Sep. 2024 - Present

Mentor: Prof. Wenhui Fan

· *Experimental Teaching Platform for Group Cooperative Control of Intelligent Unmanned Systems*

- Designed and implemented the interface of interaction between humans, LLMs, and robotic arms.
- Completed the preliminary platform.

SCIENTIFIC TALKS

Enhancing Isotope Charge Radius Measurement Precision with Statistical Analysis Jul. 2024

Topics on Frontiers of Cross-Sciences, Beijing

- Oral presentation, on the advancements and results of my research project.

PLMs as Meta-function: Learning In-context Learning for Named Entity Recognition Dec. 2023

Tsinghua Text Analysis Symposium, Beijing

- Oral presentation, a literature review on this paper.

SCHOLARSHIPS AND AWARDS

Academic Excellence Scholarship Tsinghua University, 2024

Progress in Academic Performance Scholarship Tsinghua University, 2022

Outstanding Innovation Scholarship Philip K.H. Wong Foundation, 2022

SELECTED COURSE PROJECTS

Deep Learning Sep. 2024 - Present

- Investigating the analogs of brain sleep mechanisms in neural networks and their role in improving continual learning capabilities.

Large Language Models and Alignment Sep. 2024 - Present

- Pre-training, instruction fine-tuning, and RLHF on LLMs, with a focus on CUDA/DPU programming.

Deep Reinforcement Learning Mar. 2024 - Jun. 2024

- Conducted literature reviews, designed experiments, and developed algorithms to improve offline RL performance under limited data scenarios.
- Delivered a project paper and oral presentation.

Machine Learning and Big-data Nov. 2023 - Dec. 2023

- Designed and implemented neural networks (ANN, RNN, CNN) to predict autonomous underwater vehicle health.

TECHNICAL SKILLS

Programming Languages Python, R, L^AT_EX, C/C++, Mathematica, HTML, CSS, JavaScript