

# Yucheng Pan

+86 19372485548 ◇ yuchengpan707@gmail.com ◇ Personal Homepage

## EDUCATION

---

**B.S. Physics, Tsinghua University**

*GPA: 3.874/4.000*

Sep. 2021 - Jun. 2025 (expected)

Beijing, China

## RESEARCH INTERESTS

---

LLM Agents, LLM × RL, Reinforcement Learning, Multi-Agent System, AI Safety, Machine Learning

## RESEARCH EXPERIENCE

---

**Research Internship, Princeton University (Remote)**

May. 2024 - Present

*Department of Electrical and Computer Engineering*

*Mentor: Prof. Mengdi Wang (Princeton), Prof. Huazheng Wang (Ohio State)*

· ***Deception and Defense of LLM Agents***

- Co-led the cross-disciplinary project with Princeton's Department of Psychology.
- Designed and implemented a Public Goods Game framework for LLM agents with communication features and defectors, as well as the prompts for LLM agents.
- Conducted experiments analyzing deceptive behaviors in LLM agents.
- Developed and evaluated defense strategies based on the results and insights.
- Authored the first draft of a research paper for submission.

· ***LLM Agent Data Augmentation***

- Conducted initial experiments and identified key research questions, focusing on the generalization capabilities of LLM agents.
- Redesigned environments/benchmarks to enhance and evaluate LLM agent performance across different generalization dimensions.

**Research Internship, University of Carolina at Chapel Hill (Remote)**

Feb. 2024 - Present

*Department of Statistics and Operations Research*

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

· ***Orthogonal Audio Watermarking***

- Led the project.
- Conducted literature reviews, determined research directions and questions.
- Redesigned and implemented AudioSeal, an open-source watermarking neural network, for multi-source watermark embedding without interference.
- Modified the architecture and training processes to enhance watermark robustness.

**Undergraduate Research Assistant, Tsinghua University**

Sep. 2023 - Present

*Center for Statistical Science*

*Mentor: Prof. Ke Deng*

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

- Led the project.
- Constructed and processed datasets of nuclear radii and isotope shifts across multiple spectra.
- Developed and implemented statistical computing algorithms to reduce estimation errors with theoretical measurement error models.

- Validated our methods using Bootstrap with simulated and real-world data.
  - Authored the first draft of a research paper for submission.
  - ***Nucleotide Sequences Based on Language Models***
    - Developing a BERT model to predict nucleotide sequences and explore artificial protein design by leveraging patterns in amino acid and nucleotide datasets.
  - ***Diffusion Model combined with Monte Carlo Markov Chain***
    - Conducting preliminary research on integrating diffusion models with Monte Carlo Markov Chain methods.
- Track the Light Research Program for Undergraduates, Tsinghua University** Jul. 2022 - Sep. 2023  
*Department of Astronomy*  
*Mentor: Prof. Zheng Cai*
- ***JWST-Based Study of High-Redshift Universe Through Stellar Mass Function***
    - Gathered and processed high-redshift galaxy data from James Webb Space Telescope (JWST).
    - Analyzed galaxy stellar mass functions and verified the  $\Lambda$ CDM model.

## SCIENTIFIC TALKS

---

**Oral Presentation** Jul. 2024

*Topics on Frontiers of Cross-Sciences, Beijing*

- Enhancing Isotope Charge Radius Measurement Precision with Statistical Analysis

**Oral Presentation** Dec. 2023

*Tsinghua Text Analysis Symposium, Beijing*

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

## SELECTED COURSE PROJECTS

---

**Large Language Models and Alignment** Sep. 2024 - Present

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

**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; achieved an A.

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

- Designed and implemented deep learning models (ANN, RNN, CNN) to predict Autonomous Underwater Vehicle health.

**C++ Programming for Linux** Jul. 2023

- Developed a simplified remote system administration tool for web servers.

**Observational Astronomy** Mar. 2023 - Jun. 2023

- Designed a spectral fitting pipeline for observational data; earned an A+ for the project.

## SCHOLARSHIPS AND AWARDS

---

Scholarship for Outstanding Technological Innovation	2022
Friends of Tsinghua - Qianheng Huang Scholarship	2022
Scholarship for Academic Progress	2022
First Prize in Public Welfare and Social Innovation Track of Creative Competition for Freshmen of Tsinghua	2021

## EXTRA-CURRICULAR

---

- Implementing Object Detection Applications Using Ascend Elastic Cloud Servers

Aug. 2023