# Guangyu Li

liny556.github.io +86 15569443425 flipped@mail.ustc.edu.cn, li5690@purdue.edu

## Education

**University of Science and Technology of China(USTC)** Sep. 2021 - Jun. 2025

*Bachelor of Science in Statistics*

GPA (overall): 3.5/4.3 (85.65/100); GPA (last two semesters): 3.8/4.3 (89.34/100)

Core Courses: Time Series Analysis (4.3), Regression Analysis (4.0), Multivariate Analysis (4.3), Nonparametric Statistics (4.3), Operations Research (4.0)

Awards: Outstanding Student Scholarship, USTC (2021-2024); USTC Fellowship, USTC (2024)

## Publications

**Accelerating Discrete Langevin Samplers via Continuous Intermediates**

***Guangyu Li**, Ruqi Zhang, Submitted to Transactions on Machine Learning Research (TMLR)*

openreview.net/forum?id=fNI2fPyAfQ

## Research Experiences

**Protein Agent** | *Python, TypeScript, React JSX* Oct. 2025 - Present

*Research Intern. Advisor: Prof. Tianlong Chen (Department of Computer Science, UNC)*

- Developed a unified web interface integrating multiple REST-based protein tools with an intelligent agent system.
- Built a tldraw-based front-end for interactive tool-chain generation and chatbot-driven workflow execution.
- Enabled visualization and modular invocation of protein tools.

**Continual In-Context Learning** | *Python* Oct. 2025 - Present

*Research Intern. Advisor: Prof. Lijie Hu (Department of Machine Learning, MBZUAI)*

- Studying the generalization and continual learning capabilities of in-context learning (ICL).
- Analyzing whether ICL mitigates catastrophic forgetting and exploring theoretical connections between them.
- Developing analytical tools to explain observed generalization behaviors beyond existing theoretical frameworks.

**In-Context Learning with Lasso** | *Python* Jun. 2025 - Present

*Research Assistant. Advisor: Prof. Canhong Wen (Department of Statistics and Finance, USTC)*

- Exploring the theoretical connection between in-context learning (ICL) and Lasso.
- Designed plug-and-play neural layers and modified Transformer architectures to induce sparsity in model outputs.
- Observed stable convergence of estimated parameters and potential directions for interpretable, sparse in-context learners.

**AI Healthcare** | *Python, Horos* May 2025 - Sep. 2025

*Research Intern. Advisor: Prof. Yanran Wang (Department of Biomedical Engineering, University of Michigan)*

- Explored computational pathology and applied foundation models for medical report question-answering and diagnostic reasoning.
- Performed large-scale anonymization and preprocessing of clinical imaging datasets from multiple hospital centers.
- Evaluated vision-language models (Qwen, MedGemma) for medical report generation and disease classification across MRI and CCTA data.

**MCMC Sampling** | *Python* Jul. 2024 - Mar. 2025

*Research Intern. Advisor: Prof. Ruqi Zhang (Department of Computer Science, Purdue University)*

- Proposed *Continuous-exploratory Discrete Langevin Sampler (cDLS)* to accelerate gradient-based MCMC for discrete distributions.
- Introduced continuous intermediates to bridge discrete and continuous domains, enabling geometry-aware large moves and faster mixing.
- Demonstrated substantially improved convergence efficiency over existing discrete Langevin methods across multiple high-dimensional tasks.

**Long Read Clustering** | *Python, C* May 2024 - Nov. 2024

*Research Intern. Advisor: Prof. Zhigang Yao (Department of Statistics, National University of Singapore)*

- Developed manifold clustering methods for RNA-Seq long-read data, integrating statistical and geometric perspectives.
- Implemented advanced sequence representations (Kmer2vec, Natural Vector) to capture contextual correlations in transcriptome data.
- Achieved high clustering accuracy on benchmark datasets (H-score, C-score, ARI≈0.95) and prepared methods for large-scale deployment.

## Activities

**Student Union of USTC** | *Member* Sep. 2021 - Sep. 2022

Responsible for relevant documents writing and publicity work

**Official Media Center of USTC** | *Leader* Oct. 2023 - Jun. 2025

Responsible for the management in office

## Skills

**Languages**: C, Python, R, LaTeX, Matlab

**Tools**: Git/GitHub, VS Code, Rstudio, LLMs