

Jichu Li

Tel: +86-181-5508-2105 | Email: lijichu52@gmail.com
Google Scholar | Github | Homepage | Last update: September, 2026

EDUCATION

Renmin University of China

B.S. in Statistics (Honors Program), School of Statistics
B.Econ.(Minor) in Finance, Hillhouse Academy, Gaoling School of AI

Beijing, China

Sep 2023 - Jun 2027 (expected)

► **Overall GPA:** 3.97/4.00 **Ranked:** 2/168

► **Relevant Coursework:** Mathematical Analysis I(93) II(94) III(93), Advanced Algebra I(98) II(97), Probability Theory(98) Real Analysis(96), Mathematical Statistics(94), Data Structure and Algorithm(94), Machine Learning(98), Stochastic Process(97), Optimization(97), Nonparametric Statistics(96), Statistical Computing(95), Reinforcement Learning(94)

Toyota Technological Institute at Chicago

Research Visiting Student

Host Faculty: Prof. Zhiyuan Li

Chicago, IL, USA

Jun 2026 – Sep 2026

PUBLICATIONS (*equal contribution)

[C5] **Learning with Synthetic Data via SGD in High-Dimensional Linear Regression**

Jichu Li, Difan Zou.

Neural Information Processing Systems (NeurIPS) 2026.

ICML 2026 Workshop on High-dimensional Learning Dynamics (HiLD).

[C4] **The Implicit Bias of Steepest Descent with Mini-batch Stochastic Gradient**

Jichu Li, Xuan Tang, Difan Zou.

International Conference on Machine Learning (ICML) 2026.

[C3] **A Convergence Analysis of Adaptive Optimizers under Floating-point Quantization**

Xuan Tang*, Jichu Li*, Difan Zou.

International Conference on Learning Representations (ICLR) 2026.

[C2] **Long-range Modeling and Processing of Multimodal Event Sequences**

Jichu Li*, Yilun Zhong*, Zhiting Li, Feng Zhou, Quyu Kong.

International Conference on Learning Representations (ICLR) 2026.

[C1] **DanmakuTPPBench: A Multi-modal Benchmark for Temporal Point Process Modeling and Understanding**

Yue Jiang, Jichu Li, Yang Liu, Dingkang Yang, Feng Zhou, Quyu Kong.

Neural Information Processing Systems (NeurIPS) 2025.

RESEARCH INTEREST

My research interests lie at the intersection of machine learning theory and optimization, with a focus on theoretical understanding of modern optimization algorithms in deep learning, analyzing their convergence, implicit bias, and generalization [C3][C4][C5]. I am also interested in empirically understanding the training dynamics and learning behavior of LLMs/MLLMs, with prior experience in LLM-based approaches for modeling complex sequential and event-based data [C1][C2].

RESEARCH EXPERIENCE

Toyota Technological Institute at Chicago Research Visiting Student

Advisor: Prof. Zhiyuan Li and Prof. Tianhao Wang

Apr 2026 – Present

Chicago, USA

Project: Exploring the Benefits of Spectral Optimization Through Minimal Learning Models (On-going)

Investigated how optimizer geometry interacts with model architecture, data distributions, and prediction structure. Derived further theoretical results on spectral optimization in Gaussian associative memory, including multi-step convergence behavior and the role of adaptive learning rates. Developed a minimal next-token prediction model to systematically examine how token frequencies, language conditional dependencies, and model dimensions influence the relative performance of different optimizer geometries.

The University of Hong Kong Research Intern

Advisor: Prof. Difan Zou

May 2025 – Present

Hong Kong SAR, China

Project (i): Convergence analysis of optimizers in mixed-precision training [C3]

Developed a rigorous theoretical framework to analyze the convergence properties of the Quantized Adam and Muon algorithm, a crucial component in mixed-precision training for deep learning models. Proved the convergence of Quantized Adam and Muon,

demonstrating that the algorithm reliably converges under specific conditions. Analyzed the combined impact of weight and gradient quantization on the training process, providing a theoretical basis for understanding and optimizing mixed-precision training.

Project (ii): The Implicit bias of stochastic steepest descent [C4]

Analyzed the implicit bias of mini-batch stochastic steepest descent (e.g., SignSGD, Muon) in multi-class classification, rigorously characterizing how stochasticity and batch size affect the convergence direction under general norm geometries. Proved that momentum ensures convergence in small-batch regimes, while variance reduction removes batch-dependent margin gaps, enabling stochastic algorithms to recover the full-batch max-margin solution, with fully explicit, dimension-free convergence rates improving prior bounds. Investigated the vanilla batch-size-one steepest descent without momentum or variance reduction and revealed its convergence to a fundamentally different bias.

Project (iii): Learning with synthetic data via SGD [C5]

Analyzed how synthetic data affects the generalization of one-pass SGD in high-dimensional linear regression with model shift. Proved finite-sample risk bounds for mixed and two-stage training, separating standard bias and variance from mismatch-induced fluctuation and persistent drift under mixing and filtered initialization bias under two-stage, revealing that mixed training induces strong model collapse while two-stage training avoids the floor by confining synthetic data to the first stage. Derived random-sketch scaling laws for both protocols and an exact finite-sample necessary-and-sufficient condition for when two-stage training strictly outperforms real-only training under the same real-data budget and identical real-stage updates.

Renmin University of China Research Intern
Advisor: **Prof. Feng Zhou**

Jan. 2025 – Sep. 2025
Beijing, China

Project: LLM-based multimodal event sequence modeling [C1][C2]

Designed and constructed a novel multimodal dataset benchmark incorporating diverse data types (text, timestamps, figures) to establish a standard for model evaluation. Integrated Large Language Models (LLMs) with Temporal Point Processes (TPPs) to significantly enhance inference and predictive capabilities for complex event sequences.

HONORS & AWARDS

NeurIPS Travel Award. Neural Information Processing Systems Foundation	2026
Outstanding Student Award. Renmin University of China	2026
Academic Excellent Scholarship. School of Statistics, Renmin University of China. CNY 8000	2026
National Scholarship (twice). highest scholarship for Chinese university students. CNY 20000	2024, 2025
HKU Summer Research Stipend. The University of Hong Kong. HK\$ 19600	2025
Honorable Mention. Mathematical Contest in Modeling and Interdisciplinary Contest in Modeling.	2025
Provincial First Prize. Mathematics competition of Chinese College Students, CMC	2024
Provincial First Prize. Contemporary Undergraduate Mathematical Contest in Modeling.	2024

PROFESSIONAL SERVICE

Conference Reviewer ICML 2026, NeurIPS 2026, ICLR 2027.
Teaching Assistant Optimization Methods, Fall 2026.
Instructor - Statistical Investigation Association of Renmin University of China (2024-now):
Taught Python programming and wrote tutorial articles for student groups at Renmin University of China.

SKILLS

Programming: Python(Proficient), C/C++, R.
Typesetting: $\text{\LaTeX}$, Markdown.
Language: English (TOEFL: 110), Mandarin (Native)