(984)259-5226 | Durham, NC
aihua.li@duke.edu

Aihua Li

linkedin.com/in/aihua-li
github.com/aihua-li12

PhD candidate developing generative models and LLM methodologies, seeking a Research Internship for December 2025 – Fall 2026, with a preference for a long full-time internship.

Education

Duke University — **Aug 2022 – May 2027 (Expected)**
Ph.D. in Statistical Science — Durham, NC

- Advisor: Dr. Li Ma (University of Chicago)

Stanford University — **Sep 2025 – May 2027 (Expected)**
Artificial Intelligence Graduate Certificate Program — Stanford, CA

- Core Courses: CS230 (Deep Learning), CS236 (Deep Generative Models)

Duke University — **Aug 2020 – May 2022**
M.S. in Statistics — Durham, NC

- GPA: 3.98/4.00 (**Rank 1/30**)

Shanghai University of Finance and Economics — **Sep 2016 – Jun 2020**
Bachelor's in Economic Statistics — Shanghai, China

- GPA: 3.75/4.00 (**Rank 1/90**)
- **Distinguished Undergraduate Award** and **National Scholarship** *(Top 0.2% of undergraduates nationwide)*

Research Experience

LLM | Language Generation by Learning the Probability Path of Multi-Logits — Jul 2025 – Present
Research Intern (Mentor: Dr. Yibing Song, BYD)

- Pioneered **FLoMA**, a novel non-autoregressive framework for parallel text generation using Flow Matching, achieving up to **3x faster inference** than standard autoregressive LLMs.
- Innovated a technique to **bridge discrete token spaces with continuous flow models** by modeling the probability path of multi-token logits on *multi-simplex*, circumventing the computational challenges of discrete state-space optimization.
- Architected a novel **Tripper module** to augment a standard encoder-decoder backbone, featuring *dual cross-attention* mechanisms that condition the flow model on both source and target representations for richer learning of vector fields.
- Developed **inference-time controllable generation** (e.g., style transfer, summarization) by directly interpolating the model's learned vector field, avoiding the expensive model retraining required by autoregressive models.
- Preparing a **first-author manuscript** for submission to ICML 2026.

Generative AI | Direct Generative Modeling of Sparse Compositional Data [Project Page] — May 2025 – Oct 2025
Ph.D. Researcher (Advisor: Dr. Li Ma, University of Chicago)

- Proposed **Simplex-Zero**, a deep generative framework that directly generates sparse compositional data on the probability simplex, overcoming geometric distortions and the inability to model exact zeros of traditional transformation-based methods.
- Designed a novel activation function with **bounded gradients**, **Lipschitz continuity**, and **Universal Approximation** properties.
- Achieved an **81% improvement** in generation quality (measured by Jensen-Shannon Divergence) over baseline methods that rely on logarithmic transformations.
- Achieved **4x faster computation** than complex, manifold-specific models through a simple integration with standard generative architectures.
- Finalizing a **first-author manuscript** for submission to ICML 2026.

Regularization | A Bayesian Decision-Theoretic Approach to Sparse Estimation [Paper] — Aug 2023 – Jan 2025
Ph.D. Researcher (Advisor: Dr. Surya Tokdar, Duke University)

- Developed an **L1 regularization** framework that, for the first time, *simultaneously* achieves high prediction accuracy (via bias reduction) and computational efficiency (via convexity), resolving a key limitation of existing methods.

- Established a new performance frontier for the **prediction-sparsity trade-off** along the entire regularization path, an improvement that is mathematically *unattainable* by existing methods through hyperparameter tuning.
- Outperformed baselines by **49% in L2 prediction error** and **86% reduction in false discovery rate** across extensive simulations.
- **First-author manuscript** under review at **Biometrika**, a top-tier statistics journal.
- Received the **PhD BEST Award** from Duke University and the **Best Poster Award** at the 2024 ISBA World Meeting.

Work Experience

AI Research Intern, BYD Jul 2025 – Present

Shanghai, China (Remote)

- Independently conceived and led the development of **FLoMA**, a large language model framework aimed at **fast and controllable text generation** for industrial applications.
- Engineered the complete training and inference pipeline in **PyTorch**, implementing **Distributed Data Parallel** training across multiple GPUs to improve efficiency.
- Designed and executed structured **benchmarking experiments** to evaluate generation speed, fluency, and controllability, achieving **3x faster inference** with competitive output quality.
- Delivered technical reports and presentations to the research group, summarizing progress, experimental results, and deployment considerations.

Projects

LLM + Computer Vision | Self-Supervised Learning of Causal Hierarchy for End-to-End Autonomous Driving

- Developed a **self-supervised fine-tuning framework** to mitigate causal confusion in end-to-end imitation learning by explicitly inferring **causal hierarchies** in a scene without human supervision.
- Augmented standard driving LLMs with an explicit **causal attention module**, trained using an **iterative distillation** scheme combined with **occlusion analysis**.

Publications

1. **Li, Aihua**, and Ma, L. (2025). *"Simplex-Zero: Direct Generative Modeling of Sparse Compositional Data."* To appear on arXiv. [Project Page]
2. **Li, Aihua**, Tokdar, S. T., and Xu, J. (2025). *"A Bayesian Decision-Theoretic Approach to Sparse Estimation."* Under review at *Biometrika*. [Paper]
3. **Li, Aihua** (2021). *"Conditional Estimates of Diffusion Processes for Evaluating the Positive Feedback Trading."* [Paper]

Professional Activities

Reviewer:	NeurIPS Position Paper Track	2025
Oral Presentation:	Joint Statistical Meetings (JSM)	2025
Poster Presentation:	Frontiers in Statistical Machine Learning (FSML)	2025
	ISBA World Meeting (with **Best Poster Award**)	2024

Selected Honors, Awards, and Fellowships

PhD BEST Award — Duke University 2025
Dean's Research Award — Duke University 2021
Distinguished Undergraduate Award — Shanghai University of Finance and Economics 2020
National Scholarship — Ministry of Education of China 2019
(The highest academic award in China, granted to the top 0.2% of undergraduates nationwide.)

Skills

Programming Languages:	Python, R
Deep Learning Framework:	PyTorch