

Ziyi Zhang

ziyizhan@andrew.cmu.edu

EDUCATION

Carnegie Mellon University

PhD in Electrical and Computer Engineering (GPA: 3.8 /4.0)

Duke University

MS in Computer Science (GPA: 3.9 /4.0)

University of Michigan

BS in Mathematics & Computer Science (GPA: 3.9/4.0)

Pittsburgh, PA

Expected Jan 2027

Durham, NC

Aug 2021 – Sep 2022

Ann Arbor, MI

Aug 2017 – Apr 2021

CURRICULUM & TECHNICAL SKILLS

- **Programming languages:** C++ (3 years), Java (3 years), Python (7 years)
- **Database:** Oracle (1 year), SQL (1 year), MongoDB (1 year)
- **Math Coursework:** Stochastic Calculus, Real Analysis, Functional Analysis, Partial Differential Equations, Graph Theory
- **Data Science:** Statistical Analysis, Regression
- **Computer Science:** Numerical Methods, Reinforcement Learning, Deep Learning

WORK EXPERIENCE

Google LLC.

Software Engineer Intern (Gmail Ads)

Sunnyvale, CA

May 2026 – Aug 2026

- Build an LLM-based agentic system to adjust bidding parameters produced by the hidden distribution model for impressions.
- By modeling bidding as a Markov Decision Process, the system aims to use the LLM's reasoning ability to adjust the range of maximal bidding prices on a daily basis.
- Used knowledge agent, critic agent, and Thompson sampling to select the most valuable knowledge on a weekly basis.
- For the selected adgroups, the desired-spend-over-spend (DSoS) moved closer to 1 by about 0.03, with reduced variance in daily DSoS fluctuations.

Aira Technologies, Inc.

AI System Intern

Saratoga, CA

May 2025 – Aug 2025

- Designed and implemented an agentic LLM network for root cause analysis (RCA) in AT&T network
- Implemented an agentic system using LangGraph with a team of researcher LLMs each focusing on different elements of KPI, alarms, and tickets. The team is led by a supervisor LLM, which produces the final report
- Used GraphDB to build the knowledge graphs for the researcher LLMs to identify the root cause more accurately
- Improved the accuracy of RCA from 30% in the prototype version to 85% in the production version

SELECTED PUBLICATIONS

[1] Predictive Control and Regret Analysis of Non-Stationary MDP with Look-ahead Information

Ziyi Zhang, Yorie Nakahira, Guannan Qu; TMLR <https://openreview.net/forum?id=uObsIYwXjQ>

- Designed and implemented an algorithm using exogenous prediction in reinforcement learning
- Demonstrated with mathematical proof and experiments that the proposed algorithm outperforms existing benchmarks in time-varying RL with even a few steps of predictions

[2] Stabilizing LTI Systems under Partial Observability: Sample Complexity and Fundamental Limits

Ziyi Zhang, Yorie Nakahira, Guannan Qu; NeurIPS 2025 <https://openreview.net/pdf?id=KwHsZJatB8>

- Designed and implemented an algorithm framework for stabilizing partially observable LTI systems
- Using Singular Value Decomposition, we isolate and stabilize only the unstable component of the system
- For systems with small unstable dimensions, the proposed algorithm has a drastic reduction of data required for system stabilization, compared with traditional adaptive control algorithms

[3] Polynomial Convergence of Riemannian Diffusion Models

Xinyu Xu, Ziyi Zhang, Yorie Nakahira, Guannan Qu, Yuejie Chi; ICLR 2026 <https://openreview.net/forum?id=IL0FR3UPhZ>

- Restricted high dimensional diffusion model to low dimensional manifold structure
- Implemented a score training method and tested the proposed algorithm with different score functions
- Demonstrated by both mathematical proof and simulation that the proposed algorithm achieves polynomial convergence

[4] HilbertGen-3D: Hilbert--Multifractal Conditioning for Topology-Aware 3D Generation <http://arxiv.org/abs/2610.22720>

Ziyi Zhang, Shixuan Li, Paul Bogdan, Yorie Nakahira, Guannan Qu; Neurips 2026

- Used diffusion to generate 3D point clouds for three different categories of surfaces: Chair, Airplane, and Car.
- Designed a Hilbert sampling strategy, which learns the density of points in voxels in 3D space and translates them into Hilbert signal. Then, a variational autoencoder learns from the signal and generates the signals at inference time.
- Improved 1-Nearest Neighbor Accuracy (1-NNA) and convergence by up to 8.9 and 5.8 percent, respectively.