

EDUCATION

University of Minnesota

September 2026 - Present

- M.S., Computer Science (candidate).
- **Graduate Teaching Assistant:** Data Structures and Algorithms (Fall 2026).

University of California, Los Angeles

September 2020 - June 2024

- B.S., Applied Mathematics and B.S., Economics (Double Major).
- **GPA:** 3.68 (Math GPA: 3.82).
- **Honors:** William Lowell Putnam Mathematical Competition 2023 (Ranked in top 15% in the country).

PROFESSIONAL EXPERIENCE

Generative AI Engineer | Reality AI

November 2024 - September 2025

- Scaled system to support 4,000+ concurrent workflows by migrating containerized microservices from AWS EC2 to serverless Google Cloud Run, reducing workflow latency by 55% (from >1 minute to <30 seconds).
- Automated CI/CD pipelines via GitHub Actions and integrated MLflow for real-time observability, tracking generation latency and throughput.
- Built a semantic retrieval layer using Redis and OpenAI embeddings to execute real-time vector similarity search for the generation pipeline.
- Designed state-based control flows using LangGraph, modeling user interactions as deterministic DAG transitions.

ML Engineering Intern | UCLA Office of Advanced Research Computing

March 2024 - June 2024

- Engineered an offline recommendation pipeline in PyTorch, training a collaborative filtering model and text embedding extractors to learn complex user preferences and activity representations.
- Architected a hybrid candidate retrieval layer in PostgreSQL, leveraging PostGIS for geospatial indexing and pgvector for ANN search across semantic data.
- Improved NDCG@10 by 15% over baseline by co-developing a Flask REST API that aggressively re-ranks spatial and collaborative candidate pools in real time.
- **Award:** "Most Commercially Viable Application" at Innovate@UCLA.

RESEARCH PROJECTS

Metric Property Testing | Graduate Research - Advisor: Dr. Zihan Tan

February 2026 - Present

- Formulated a structural characterization of one-sided $O(1)$ -query testability for metric families, analyzing the combinatorial properties of symmetric distance matrices.
- Proposed sublinear $o(N)$ validation methods for massive LLM embedding spaces, enabling lightweight, continuous monitoring of structural degradation and catastrophic forgetting during fine-tuning.
- Targeting submission to ACM Symposium on Theory of Computing (STOC) 2027.

Identifying Auction Fraud with Ensemble Methods | Capstone - Advisor: Dr. Randall Rojas

March 2023 - June 2023

- Detected shill bidding in eBay auctions using ensemble methods, mitigating severe class imbalance by comparing synthetic minority over-sampling (SMOTE) with algorithmic class weighting.
- Outperformed existing literature ($F1 > 0.99$) in detecting eBay auction fraud by training a Weighted XGBoost model with rigorous stratified 5-fold cross-validation.
- Identified "successive outbidding" as the primary fraud predictor through variable importance analysis, extracting interpretable decision logic from the ensemble trees.

SKILLS

- **Languages:** Python, C/C++, JavaScript/TypeScript, SQL.
- **Frameworks & Libraries:** LangGraph, LlamaIndex, FastAPI, Flask, NumPy, Pandas, PyTorch.
- **Tools & Platforms:** Docker, Google Cloud Platform (GCP), AWS, Git, Redis, PostgreSQL.