

ANDY (ZEHAO) ZHAO

609-937-3646 | zhaozeha@usc.edu | [linkedin.com/in/zehao-zhao](https://www.linkedin.com/in/zehao-zhao) | zehao.org | No Sponsorship Needed (PR)

EDUCATION

UNIVERSITY OF SOUTHERN CALIFORNIA

Los Angeles, CA

PhD Candidate in Electrical Engineering & Computer Science, Center of Autonomy & Artificial Intelligence, GPA: 4/4 2023 –

- Awards: Ming Hsieh Fellowship, Viterbi Graduate Fellowship, Annenberg School Award, Early Career Innovator Award

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge, MA

Master of Science in Business Analytics & Operations Research, GPA: 5.0/5.0

2022 – 2023

- Awards: 2nd place in MIT Analytics Lab Contest out of 40 teams, Top 10 in MIT Hackathon out of 1000+ attendees

UNIVERSITY OF CALIFORNIA, BERKELEY (4 Majors with High Honors, First in UC Berkeley 150 years)

Berkeley, CA

Bachelor of Science in Computer Science, Statistics, Math, and Data Science, GPA: 3.9/4.0, Dean's List 8 Semesters 2018 – 2022

- Awards: 1st place in California actuarial league competition; Finalist for Mathematics Contest in Modeling
- Honors Thesis: Data Science Honors: Time series modelling on Finance; EECS Honors: PCA & Facial Recognition

TECHNICAL SKILLS

Python (TensorFlow, Keras, Pytorch, Pandas), AWS (Sagemaker, Bedrock, Kendra), LangChain/LangGraph, Git, RAG, RLHF, Prompt Engineering, MATLAB, R(ggplot, dplyr), Julia (gurobi), C, C++, Java, Rust, LaTeX, RISC-V, SQL, CUDA, ArcGIS

EXPERIENCE

SAMSUNG RESEARCH AMERICA

Mountain View, CA

Generative AI Research Intern, Advanced ML Lab (AWS Bedrock, Kendra, LangChain/LangGraph)

May 2025 – Present

- Built and optimized LLM-based internal search engine; improved top-k document relevance by 35% and latency by 40%
- Designed and deployed LLM-powered recommender system for semantic ad matching, boosting CTR by 20%

UNIVERSITY OF SOUTHERN CALIFORNIA

Los Angeles, CA

Research Assistant to Professor Rahul Jain and Professor Bhaskar Krishnamachari

Aug 2024 – Present

- Developed Multi-agent General-Sum and Zero-Sum game algorithms that can be used in Large Scale Autonomous System
- Built reasoning circuits thru autoencoder to interpret LLM reasoning to complex math problems and decision-making tasks

LLM STEALTH STARTUP (Accepted by MIT Momentum Accelerator)

Boston, MA & San Francisco, CA

Founding Engineer (AI Branch, LangChain, LangGraph)

Aug 2023 – Aug 2024

- Post-trained LLM with 100+ users and 1000+ documents for a social platform to foster friendships and dating experiences
- Developed personalized AI Agent for tailored event recommendation using LangGraph and increased user retention by 500%

AMERIPRISE FINANCIAL

Boston, MA

Data Scientist (TensorFlow, Keras, Pytorch, SKlearn, Deep Learning, Reinforcement Learning)

Feb 2023 – Aug 2023

- Built predictive models to forecast geographical location growth (500m grid level) within Europe for real estate investment
- Employed DNNs to increase the testing accuracy by 50%, and employed Boosting Algorithms for interpretability
- Researched and developed data-specific training metrics to enhance model performance

MIT OPERATIONS RESEARCH CENTER

Boston, MA

Research Assistant to Professor Dimitris Bertsimas (Python, Julia: Gurobi, IAI, Pytorch)

Sep 2022 – Feb 2023

- Proposed Integer programming approximation algorithms that performs better than state-of-the-art solver for some problems
- Developed ML scheme to maximize cancer patient survival time for various treatments and improved AUC to 80%

ALPHA SQUARE GROUP

New York, NY

Quantitative Research Intern (Python, Java, TensorFlow, Pytorch)

Jun 2022 – Aug 2022

- Created NLP model to summarize alternative data including customer review and quantify keywords
- Constructed smart scoring system for primary market investment based on 500+ historical investment data

PROJECTS AND PUBLICATIONS

- Generalized Quantal Response Equilibrium: Existence and Efficient Learning (NeurIPS'25)
- Fine-Tuning LLMs for Strategic Multi-Agent Reasoning (AISTAT'25)
- Episodic Zero-Sum Game Learning (IEEE Conference on Decision and Control'25)
- End-to-End Learning Framework for Solving Non-Markovian Optimal Control (ICML'24)
- Vision Based Automatic Coastal Erosion Detection (Marine Geology)
- Vision based Image Contamination Detection (Journal of Physics)
- Optimized Computation for Quantum Random Number Generator (Physics Review Letter)
- Actor Critic Reinforcement Learning Recommendation System Survey (MIT EECS Poster Session, Workshop'23)
- Deep-learning models for refugee relocations in Columbia (MIT & United Nations Analytics Conference, Best Paper)