

JIN XIE

Research Data Scientist
Google, Core ML Data
Mountain View, CA

Phone: (650) 646-8252
Email: jinnxyxie1024@gmail.com
Website: <https://jinnxyxie1024.github.io/>

EDUCATION

Stanford University

Sep 2016 – Dec 2021

Ph.D. in Computational Mathematics, GPA: 3.97/4.3

Advisor: Prof. Peter Glynn, Prof. Jose Blanchet

Ignite Program, Certificate in Innovation and Entrepreneurship in Graduate School of Business.

Coursework: Natural Language Processing with Deep Learning, Data Mining, Parallel Computing, Optimization, Numerical Linear Algebra, Advanced Topics in Reinforcement Learning

Peking University

Sep 2012 – Jul 2016

B.S. in Applied Mathematics, GPA: 3.78/4.0

Double Major, Economics.

PUBLICATIONS

- [1] *Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities.* <https://arxiv.org/pdf/2507.06261>, Google Gemini team (2025).
- [2] *Surgical Scheduling via Optimal Transport*, to be submitted to Operations Research, 2024+
- [3] *Asymptotic Analysis of the Profile Function in Distributionally Robust Optimization*, submitted to AAAI, 2023+
- [4] *Simulation: address the limitation of consideration of non-controllable part of scheduling*, under major revision by Healthcare Management Science, 2023
- [5] *Optimal transport and healthcare operations.* Ph.D. Thesis (2022).
- [6] *Interactive Model to Estimate Bed Demand for COVID-19-Related Hospitalization Developed by Stanford Medicine-Engineering Partnership.* Health Management, Policy and Innovation, 2020.
- [7] *A model to estimate bed demand for COVID-19 related hospitalization.* medRxiv preprint, 2020.

WORKING EXPERIENCES

Google, Core ML Data

Jan 2025 – Present

Research Data Scientist

Mountain View, CA

Google Gemini - Core ML Data Team, data Science Hub to improve data quality, model quality, and infrastructure quality for LLM teams across Google.

Projects:

- Gemini post-training and model evaluation for long context PDF understanding (2025 I/O launch, Gemini 2.5 pro core contribution acknowledged in <https://arxiv.org/pdf/2507.06261>)
- Gemini Frontier Safety: User-level long-term trajectory analysis; high-recall high-efficiency L1 autorater training and evaluation for misuse detection pipeline (Production launch coming in 2026 Google I/O),
- Prompt bank and post-train eval refactor (2026 Deepmind Data Science Breakthrough Team Award),
- Automated loss pattern analyzer for multimodal long context PDF understanding (Demo presented in 2025 Google DataVerse Summit).

Grail, LLC

Dec 2022 – Oct 2024

Machine Learning Scientist

Menlo Park, CA

Grail IPO in June 2024: <https://grail.com/press-releases/grail-to-begin-trading-on-the-nasdaq-stock-exchange/>

Team: Core Machine Learning Team, using the richest biological datasets in the world to find signals for early cancer detection

Project: Multi-modal data fusion; Improving Ensemble Diversity; and Classifier calibration for cancer signal predictor.

Uber Technologies

Nov 2021 – Dec 2022

Data Scientist, Driving Pricing

San Francisco, CA

Team: Driver Offer Pricing, leading marketplace-wide launch of algorithmic driver pricing to 75% of Uber's global business

Project: The team launched a driver pricing ALG in South America and improved revenue by 4%. Led the AS project for launching the ALG for global Uber Premium Services.

Facebook Research

June 2020 — Sep 2020

Research Intern, Core Data Science

Menlo Park, CA

Project: Auto-Segmentation for Machine Learning Model Calibration

Stanford School of Medicine, Stanford Children's Hospital

June 2019 — Dec 2021

Research Consultant

Stanford, CA

Manager: David Scheinker, SURF

Project: Surgical Scheduling Simulation and Online Optimization Tools; Interactive Demo Building

Media Coverage: Wall Street Journal, Stanford Daily

Fusion Fund

June 2019 — Sep 2019

Investment Analyst Intern

Palo Alto, CA

Project: Data-Driven Graph-Based Venture Capital Relationship Detection; Dual Diligence in Healthcare Industry

Microsoft

June 2018 – Sep 2018

Research Intern, Research Program, Bing Ads

Sunnyvale, CA

Project: Auction Arbitration with Contextual Bandits and Deep Learning.

PROFESSIONAL SERVICES

Community Services:

- **Vice President** of Business Development, Stanford Chinese Entrepreneur Organization
- **President**, SIAM (Society for Industrial and Applied Mathematics) Student Chapter, Stanford
- **Board Member**, Women in Computational Sciences and Engineering, Stanford
- **Vice President**, Stanford Oriental Zither & Qin Association
- **Vice President**, Teenage Volunteers Association School of Mathematical Sciences, Peking University

Reviewer:

- Management Science, Mathematical Programming, Annual Conference on Neural Information Processing Systems (NeurIPS), Artificial Intelligence and Statistics (AISTATS), IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE Conference on Decision and Control (CDC), The Design Theory and Methodology (DTM)

TECHNICAL SKILLS

Programming Languages: C/C++, Python, SQL, Matlab

Agent/deep Learning Platforms: Tensorflow, Scikit-learn

HONORS AND AWARDS

- [1] First Place in Stanford Citadel Datathon (1% of candidates), 2019
- [2] Bronze Medal in the Sixth S.T.Yau College Student Mathematics Contest (1% of candidates), 2016
- [3] National Innovation Research Funding for Undergraduate (1% of candidates), 2015
- [4] Founder scholarship of Peking University (5% of candidates), 2014
- [5] First Prize in National College Student Physics Olympics Competition, China (1% of candidates), 2014
- [6] May-Fourth scholarship of Peking University (5% of candidates), 2013
- [7] First Prize in National Physics Olympics Competition, China (1% of candidates), 2011