YI GU

Jersey City, NJ
(713)-380-7338 ◇ Yi.Gu@u.northwestern.edu

EDUCATION

Northwestern University
Ph.D in Mathematics

Thesis Title: Generalized t-SNE and Beyond: Probabilistic Methods for Dimensionality Reduction, Combinatorial Optimization, and Machine Learning

Rice University
B.S. in Mathematics

Major GPA: 4.1/4.0, President Honor Roll

PUBLICATIONS

(Preprint) **Yi Gu**, Lingyou Pang. *CARDE: Concept-Anchored, Regime-Aligned Diffusion Embedding*

(Preprint) Lingyou Pang, Jianyu Lin, **Yi Gu**[*], *et al*. *LLM-Judge Autopilot: Analytically Grounded Reliability and Uncertainty for LLM-as-a-Judge.*

(Preprint) **Yi Gu**, Lingyou Pang, Xiangkun Ye, *et al*. *SIGMA: Scalable Spectral Insights for LLM Model Collapse.* arXiv:2601.03385

(Preprint) **Yi Gu**. *Equilibrium Distribution for t-Distributed Stochastic Neighbor Embedding with Generalized Kernels.* arXiv:2505.24311

Nan Jiang, **Yi Gu**[*], *et al*. *Learning Combinatorial Structures via Markov Random Fields with Sampling through Lovász Local Lemma.* **AAAI 2023**

Zichao Wang, **Yi Gu** *et al*. *Iterative Imputating Variational Auto-Encoders for Partially Observed Data.* **EDM 2020**

(Book) **Yi Gu**. *Introduction to Random Graphs and Models.* Research Survey in Random Graph, 2019. (PDF link available)

EXPERIENCE

Google Feb 2026 - Current
Research Data Scientist - Google Internal Developer AI/LLM *New York City, NY*

- Designed and deployed the internal causal inference agent for Google Colab, automating complex data science workflows for internal users (e.g., AI product impact and developer behavior analysis). This tool directly drives data-informed decision-making for Google's next-generation developer AI products.
- Architected and led the deployment of multiple LLM intent classification pipelines across core Google infrastructure (Critique, Buganizer, Duckie, Cider and Antigravity). Leveraged lightweight distilled models and asynchronous infrastructure to boost throughput by 200%, accelerating the iteration cycle for both internal and consumer-facing developer tools.

[*]Equal contribution, co-first author

· Delivered holistic impact analyses for internal AI products utilizing advanced causal techniques like Double Machine Learning (DML) and Difference-in-Differences (DiD). These insights provide engineering leadership with a definitive view of the development ecosystem to guide critical, high-level product strategy.

Google Sept 2024 - Feb 2026
Data Scientist - Research - Google Hardware (Pixel) *Chicago, IL*

· Independently designed and implemented a novel, end-to-end LLM sentiment analysis pipeline from scratch, integrating LLM-as-judge, SFT, and RAG to classify hardware data; achieved 99.5% accuracy, directly informing critical, company-wide hardware KPI targets and RMA analysis.
· Spearheaded an LLM-based hardware triage diagnostic pipeline that directly informed critical engineering decisions for next-generation Pixel devices. Leveraged LoRA and RLHF to train a lightweight, distilled model, boosting classification accuracy from 40% to over 90% (0.95+ recall) while increasing throughput by 300%.
· Developed a high-efficiency anomaly detection system to monitor thousands of metrics across trillion-row Pixel/Android data warehouses, using bespoke dimensionality reduction and clustering algorithms to successfully identify critical data integrity issues.
· Hired and mentored an intern during first year of employment, defining the project scope and providing direct technical guidance, leading to a successful project outcome and return offer.

Google June 2023 - Sept 2023
Data Scientist Intern (Ph.D) *Mountain View, CA*

· Designed a novel latency-tracking algorithm with millisecond-level accuracy and established a new OKR, driving a 50% latency reduction for over 90% of Pixel-related datasets and enhancing data precision for all dependent teams.
· Independently built an E2E failure prediction pipeline with 40+ high-impact features (90%+ recall, less than 10% false positive), reducing system failure rates by over 50% and saving maintenance costs.

Google June 2022 - Sept 2022
Data Scientist Intern (Ph.D) *Mountain View, CA*

· Led a time-sensitive search quality study, analyzing user surveys and query data to generate insights that resulted in the launch of a new Knowledge Panel feature.
· Researched and implemented homomorphic encryption algorithms in collaboration with Google Ads, enhancing user privacy protection for personalized advertising and ensuring compliance with new regulations.

ACADEMIC SERVICE & TEACHING

ICML	Reviewer (2026)
Northwestern University	Course Instructor, Random Matrix Theory (2019) · Designed and delivered advanced mathematical coursework focusing on foundational probability and matrix theories applicable to deep learning.
Northwestern University	Teaching Assistant · Served as the TA for various math courses ranging from undergraduate to graduate level, including multi-variable calculus, linear algebra, real analysis, complex analysis, functional analysis, probability, applied probability for econometrics, stochastic analysis, etc.
Rice University	Lead Organizer & Main Speaker, Spectral Theory Seminar (2016) · Conducted a semester-long workshop series, independently preparing and delivering 9 of 12 graduate-level lectures on Orthogonal Polynomials on the Unit Circle (OPUC).

TECHNICAL SKILLS

Programming	Python, SQL, MATLAB, Mathematica
ML/DL Frameworks	PyTorch, TensorFlow, JAX, Scikit-learn, Pandas, NumPy
Platforms/Tools	Google Cloud (BigQuery, Vertex AI), Apache Beam (Google Flume), Kubernetes (Google Borg), Docker
Expertise	Large Language Models (LLMs), Anomaly Detection, Dimensionality Reduction and Data Visualization, Statistical Modeling, Time Series Analysis, Stochastic Analysis, Pattern Recognition, Combinatorial Optimization, Supervised Fine-Tuning (SFT), Retrieval-Augmented Generation (RAG)