Prasanjit Dubey (he/him)

Ph.D. Candidate, Machine Learning · Graduating May 2027

Contact Information

Email: dubey.prasanjit@gmail.com
pdubey31@gatech.edu
Website: pdubey96.github.io
Google Scholar: Profile

LinkedIn: in/prasanjit-dubey
George Tower 1608
55 5th St NW, Atlanta, GA 30332

Research Interests

Federated learning and knowledge distillation for language models, communication-constrained learning and inference, robustness to faulty and adversarial sources, evaluating and monitoring AI systems, anytime-valid inference, multiple hypothesis testing, conformal prediction.

Education

Georgia Institute of Technology, Atlanta, GA, USA
Ph.D. in Machine Learning | GPA 3.85/4.00 — Aug 2021 – May 2027 (expected)

- H. Milton Stewart School of Industrial and Systems Engineering (ISyE), Minor in Mathematics
- Advisor: Prof. Xiaoming Huo

Indian Institute of Technology Bombay, Mumbai, India
M.Sc. in Applied Statistics & Informatics | Class Rank 1 of 38 | GPA 8.96/10.00 — 2017 – 2019

- Department of Mathematics
- Thesis: *Financial Data Mining Using Multivariate Copulas*

University of Calcutta, Kolkata, India
B.Sc. in Statistics, First Class Honors | Grade 92.5% — 2014 – 2017

- Department of Statistics, Minor in Mathematics & Computer Science

Publications and Preprints

Titles link to the papers. Author order is by contribution. ∗ marks an award or a presentation associated with the paper.

[1] **BOOST: Power-Optimal Strong-FWER Testing for Block-Structured Multiplicity.**
Prasanjit Dubey & Xiaoming Huo.
Conference on Neural Information Processing Systems (NeurIPS), 2026.

[2] **Certified Adaptive Refresh: Anytime-Valid Monitoring for Federated Conformal RAG.**
Prasanjit Dubey & Xiaoming Huo.
Under review at the International Conference on Learning Representations (ICLR), 2027.

∗ Accepted at the NeurIPS 2026 Second Workshop on ML×OR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making

[3] **Byzantine-Robust Federated RAG via Aligned Calibration and Fixed-Membership Conformal Prediction.**
Prasanjit Dubey, Aritra Guha, & Xiaoming Huo.
Under review at the International Conference on Learning Representations (ICLR), 2027.

[4] **Matching Rates and Optimal Allocation for Federated Probe-Logit Distillation under Heterogeneous Bandwidth Budgets.**
Prasanjit Dubey & Xiaoming Huo.
Invited submission under review at the Journal of the American Statistical Association: Theory and Methods, 2026.

∗ Poster and Travel Award, Conference on Statistical Foundations of AI (CSFAI), University of California, Santa Barbara, 2026
∗ Poster, Statistics and Trustworthy AI for Cross (X)-Domain Acceleration (STAI-X), Harvard University, 2026

[5] **Federated Language Models Under Bandwidth Budgets: Distillation Rates and Conformal Coverage.**
Prasanjit Dubey & Xiaoming Huo.
Under review at Transactions on Machine Learning Research (TMLR), 2026.

[6] **Federated Multiple Testing with Quantized *p*-Values under Family-Wise Error Control.**
Prasanjit Dubey & Xiaoming Huo.
Under review at Bernoulli, 2026.

∗ Accepted at the NeurIPS 2026 Second Workshop on ML×OR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making

[7] **Pooling Sequential Evidence Across Hypotheses: Rate-Optimal Multiple Testing at a Fixed Horizon.**
Prasanjit Dubey & Xiaoming Huo.
Under review at the Electronic Journal of Statistics, 2026.

[8] **Most Powerful Tests with Exact Family-Wise Error Rate Control: Necessary Conditions and Fast Computation for General K.**
Prasanjit Dubey & Xiaoming Huo.
Revision under review at the Journal of Computational and Graphical Statistics (JCGS), 2025.

[Companion paper] *Most Powerful Test with Exact Family-Wise Error Rate Control: Necessary Conditions and a Path to Fast Computing*
∗ Invited talk, Optimisation for Statistics and Learning session, 2025 INFORMS Annual Meeting

[9] **LLmFPCA-detect: LLM-powered Multivariate Functional PCA for Anomaly Detection in Sparse Longitudinal Texts.**
Prasanjit Dubey, Aritra Guha, Zhengyi Zhou, Qiong Wu, Xiaoming Huo, & Paromita Dubey.
Preprint, 2025.

∗ 2nd Place Paper Award, Student Research Showcase, 2025 INFORMS Regional Analytics Conference (IRAC)

[10] **Generalization Bounds for Magnitude-Based Pruning via Sparse Matrix Sketching.**
Etash Kumar Guha, Prasanjit Dubey, & Xiaoming Huo.
International Conference on Learning Representations (ICLR) Workshop on Bridging the Gap Between Practice and Theory in Deep Learning, 2024.

Invited Talks & Presentations

Matching Rates and Optimal Allocation for Federated Probe-Logit Distillation
Poster, Conference on Statistical Foundations of AI (CSFAI), University of California, Santa Barbara 2026
∗ Travel Award
Poster, Statistics and Trustworthy AI for Cross (X)-Domain Acceleration (STAI-X), Harvard University 2026

Most Powerful Test with Exact Family-Wise Error Rate Control: Necessary Conditions and a Path to Fast Computing
Invited talk, Optimisation for Statistics and Learning session, INFORMS Annual Meeting, Atlanta, GA 2025
Poster, Georgia Statistics Day, University of Georgia, Athens, GA 2025

LLmFPCA-detect: LLM-powered Multivariate Functional PCA for Anomaly Detection in Sparse Longitudinal Texts
Poster, Student Research Showcase, INFORMS Regional Analytics Conference (IRAC), Atlanta, GA 2025
∗ 2nd Place Paper Award

Awards & Honors

Top Reviewer, Conference on Neural Information Processing Systems (NeurIPS) 2026
Travel Award, Conference on Statistical Foundations of AI (CSFAI), University of California, Santa Barbara 2026
2nd Place Paper Award, Student Research Showcase, INFORMS Regional Analytics Conference (IRAC), Atlanta 2025
Stewart Topper Fellowship Award, Georgia Tech (first-year Ph.D. academic performance) 2021 – 2022
Dr. P. V. Sukhatme Memorial Award, IIT Bombay (M.Sc. rank 1) 2019
All India Rank 26, Graduate Aptitude Test in Engineering (GATE), Statistics 2019
Pfizer Scholarship, IIT Bombay (M.Sc. first-year rank 1) 2018
All India Rank 7, IIT Joint Admission Test for M.Sc. (IIT JAM), Mathematical Statistics 2017
Swami Asaktananda Award, University of Calcutta (highest B.Sc. academic distinction) 2017
Summer Research Fellowship, Indian Academy of Sciences (one of 65 in 2016 and 72 in 2017, in mathematics) 2016, 2017

Research & Professional Experience

Georgia Institute of Technology, *Atlanta, GA, USA* Jan 2026 – Sep 2026
Graduate Research Assistant, Georgia Department of Transportation Sponsored Research
Data Sufficiency at Continuous Count Stations · PI: Prof. Xiaoming Huo [Code]

- Developed a **statistical framework** to establish **how many valid days a traffic-monitoring station needs** in a month with incomplete data before its monthly average can be published.
- Applied the framework to the entire network of **233 traffic-monitoring stations** across Georgia, delivering the results to transportation planners as the **Data-Sufficiency Explorer**, an interactive dashboard.

Bloomberg L.P., *New York, USA* Jun 2025 – Aug 2025
Quant PhD Intern, Quantitative Risk Analytics

- Built a **Kalman-filter hazard-curve model** for pricing credit instruments by estimating dynamic CDS term structures from **latent Nelson–Siegel factors and macroeconomic covariates**; **integrated the model into Bloomberg Terminal's** Credit VaR engine to forward-simulate risk-neutral hazard rates.
- Implemented a **Credit Rating Migration module** for the Credit VaR engine using **macro-conditioned multinomial logit and intensity-based models** to simulate obligor- and index-level rating paths, capturing systematic and idiosyncratic risk.

AT&T Labs, Inc., *New Jersey, USA* May 2024 – Aug 2024
Research Scientist Intern, Data Science & AI Research

- Developed a novel framework for analyzing **sparse longitudinal customer-journey text** by integrating **large language models (LLMs)** with **multivariate functional data analysis**; work resulted in a manuscript (paper [9]).
- Used **LLM-guided dynamic keyword profiling** on real-world data (Amazon reviews) to surface customer pain points and actionable insights, **outperforming state-of-the-art baselines**; shipped a **production application** (Emotional AI) for behavioral anomaly detection and targeted interventions.

Australia and New Zealand Banking Group (ANZ), *Bengaluru, India* Jun 2019 – Aug 2021
Senior Quantitative Analyst, Wholesale Credit Risk Modeling

- Built the **Financial Ratio Surveillance Tool (FRST)**, a production **R Shiny** stress-testing application designed to forecast default probabilities and assess the predictive power of financial ratios.
- Led the bank's **statistical COVID-19 impact assessment** across commercial portfolios, flagging high-volatility sectors for stakeholders as part of **APRA regulatory** requirements.

Walmart Labs, *Bengaluru, India* May 2018 – Aug 2018
Research Intern, Data Science (International Pricing & Demand Forecasting)

- Built a **two-stage price-elasticity model** to forecast sales across product categories from item-, region-, and covariate-specific **price elasticities**, using high-dimensional, large-scale data.
- Combined *k*-**means clustering** with an **ensemble of ridge, lasso, and elastic-net models** fit within each cluster, **reducing mean absolute percentage error (MAPE) by 42.4% relative to Walmart's incumbent methods**.

Teaching Experience

Georgia Institute of Technology, *Atlanta, GA, USA* 2021 – 2025
Head Teaching Assistant and Teaching Assistant, H. Milton Stewart School of Industrial and Systems Engineering

- **ISyE 7406, Data Mining and Statistical Learning** (graduate level, 300+ students)
  Teaching Assistant: Fall 2024, Spring 2025, Fall 2025
  Instructor: Prof. Xiaoming Huo
- **ISyE 6414, Statistical Modeling and Regression Analysis** (graduate level, 400+ students)
  Head Teaching Assistant: Fall 2022, Spring 2023 | Teaching Assistant: Summer 2022
  Instructor: Prof. Nicoleta Serban
- **ISyE 2027, Probability with Applications** (undergraduate level)
  Student Assistant: Fall 2021 · Led tutoring sessions

Student Mentoring

- **Kareena Mehta**, undergraduate, Georgia Institute of Technology (Computer Science) Summer 2026
- **Kevin Li**, undergraduate, Washington University in St. Louis (Finance and Data Science) Summer 2026, Fall 2026

Professional Service

Conference Reviewer: International Conference on Learning Representations (ICLR) 2023 – present
International Conference on Machine Learning (ICML)
Conference on Neural Information Processing Systems (NeurIPS)
∗ Top Reviewer, NeurIPS 2026

Technical Skills

Programming:	Python, R, R Shiny, C++, SQL, MySQL, MATLAB
Machine Learning:	LLMs, PyTorch, scikit-learn
Computing & Tools:	High-performance computing (Slurm), GPU computing, Apache Hadoop (MapReduce, Streaming), Git, LaTeX
Relevant Graduate Coursework:	Mathematical Foundations of Machine Learning, High-Dimensional Probability, High-Dimensional Statistics, Statistical Inference, Time Series Analysis, Stochastic Processes, Convex Optimization, Stochastic Optimization, Computational Data Analysis

Last updated: October 2026