

517W. 121th St. • New York, NY 10027 • gx2150@tc.columbia.edu • 212 5181983

Summary: Ph.D. candidate in Statistics specializing in machine learning, large language models, and quantitative modeling. Skilled in dimensionality reduction, causal inference, and retrieval-augmented generation (RAG) with applications to education and financial data. Published researcher with a strong track record of deploying scalable ML solutions across education and finance domains.

Education

Columbia University

New York, NY

Ph.D. Candidate, Measurement, Evaluation and Statistics

Expected May 2026

- Doctoral Fellowship Recipient, ABD
- GPA: 4.09 (out of 4.00)
- Research: Psychometric Modeling, Machine Learning, LLM-RAG, LLM-Causal Inference

Columbia University

New York, NY

M.S., Applied Statistics

Sep 2020 – Feb 2022

- GPA: 4.09 (out of 4.00)
- Selected Coursework: Computational Statistics, Bayesian Statistics, Data Mining, Data Science Programming

University of Aberdeen

Aberdeen, UK

B.A., Economics and Finance (First Honours)

Jul 2020

- GPA: 3.71 (out of 4.00)
- Selected Coursework: Econometrics, Advanced Corporate Finance, Derivatives, Empirical Finance Methods

Publications & Working Papers

Guoliang Xu, and James E. Corter. (2025). Improving student assessments by extracting latent dimensions from multidimensional data. *Proceedings of the 47th Annual Meeting of the Cognitive Science Society (CogSci 2025)*.

Guoliang Xu, and James E. Corter. (2026, January). GraphRAG for automated short-answer grading with feedback: bridging pedagogical needs and technical capabilities. *To appear in the Proceedings of the 40th AAAI Conference on Artificial Intelligence (AAAI 2026)*.

Lawrence T. DeCarlo, and **Guoliang Xu**. Q-matrix validation methods: theoretical and computational advances. Manuscript under review.

Guoliang Xu, and Lawrence T. DeCarlo. An examination of Q-matrix validation methods: new metrics. Manuscript under review.

Selected Posters & Presentations

Guoliang Xu, and James E. Corter. (2025, July). Improving student assessments by extracting latent dimensions from multidimensional data. Poster presented at the *47th Annual Meeting of the Cognitive Science Society (CogSci 2025)*, San Francisco, CA.

Guoliang Xu, and James E. Corter. (2025, April). A framework for extracting latent dimensions from response time data. Paper presented at the *Annual Meeting of the National Council on Measurement in Education (NCME)*, Denver, CO.

Experience

Columbia University

New York, NY

PhD Researcher, GraphRAG-ASAG-F Project

Sep 2024 – Present

- Developed GraphRAG, a knowledge graph-based RAG framework for automated short-answer grading with feedback (ASAG-F), grounding LLM outputs in instructor-curated atomic facts to ensure transparency and traceability.
- Demonstrated on the Short Answer Feedback (SAF) dataset that GraphRAG achieves grading accuracy comparable to vector RAG and superior to fine-tuned LLM baselines, while providing explicit source attribution.
- Introduced rubric-aligned discretization (0.25 increments) and prompt-based length control, improving grading accuracy, feedback quality, and stability across unseen questions and answers.
- Conducted performance-cost benchmarking across LLM backends, showing that lightweight models achieve cost-effective accuracy while premium models yield diminishing returns.

Columbia University

New York, NY

Research Assistant, Education Leadership Data Analytics Project

Jan 2025 – Present

- Conducted correspondence analysis on a large-scale, curated dataset of state education records coded against 16 NASEM equity indicators, uncovering latent dimensions of how equity is represented across states.
- Applied hierarchical clustering on CA outputs to identify interpretable groupings of indicators and states, revealing systematic contrasts in equity reporting practices.
- Visualized results through two- and three-dimensional CA plots and dendrograms, providing interpretable evidence to inform equity-focused leadership and policy discussions.

Columbia University

New York, NY

Research Assistant, NSF-Funded Course Recommendation Project

Jan 2023 – Jun 2024

- Processed large NCES datasets (4,000+ variables), performing data cleaning with predictive mean matching to prepare reliable inputs for modeling.
- Reconstructed student math course pathways across four grade levels using decision trees, identifying key progression patterns and bottlenecks.
- Applied causal machine learning methods (TMLE, Causal Forests) to analyze how different students benefit from interventions and to guide personalized recommendations.
- Designed optimal treatment regimes (i.e., optimal course recommendation rules) with DynTxRegime and performed fairness optimization with Rmosek, ensuring recommendations maximize student outcomes while reducing demographic disparities.
- Developed project website on GitHub Pages to document methodology and share results.

Teachers College, Columbia University

New York, NY

Master's Researcher, Machine and Deep Learning Projects

Jan 2021 – Feb 2022

- Built predictive models for crime data, improving accuracy from 55% to 90% by applying principal component regression and model selection techniques.
- Developed NLP pipelines for sentiment analysis and topic modeling to extract themes from large text corpora.
- Constructed and analyzed classroom interaction networks, using centrality and community measures to study peer influence and group structure.

China International Capital Corporation (CICC)

Beijing, China

Quantitative Investment Intern

Aug 2020 – Nov 2020

- Implemented risk-parity allocation in Python for fixed-income portfolios; automated performance reporting.
- Applied PCA-based orthogonalization to reduce factor correlation and enhance diversification, improving portfolio stability.
- Rebalanced asset weights under risk constraints, contributing to higher risk-adjusted returns in backtests.

Professional Services & Awards

- Reviewer, Cognitive Science Society Annual Meeting (CogSci 2025)
- Reviewer, AAAI Undergraduate Consortium (AAAI-UC 2026)
- Provost's Grant for Doctoral Research (Teachers College, Columbia University 2025)
- Helen M. Walker Scholarship (MES, Teachers College, Columbia University 2025)

Professional Memberships

- American Educational Research Association (AERA), Student Member, 2024–2025
- National Council on Measurement in Education (NCME), Student Member, 2024–2025
- Cognitive Science Society (CogSci), Member, 2025
- Society for Research on Educational Effectiveness (SREE), Member, 2025
- Association for the Advancement of Artificial Intelligence (AAAI), Student Member, 2025

Technical Skills

Programming & Tools: Python (NumPy, Pandas, PyTorch), R (tidyverse, ggplot2), SQL, GitHub, SPSS

Machine Learning & LLM: LLM-RAG (GraphRAG, VectorRAG, Neo4j, Cypher), Dimensionality Reduction (weighted-MDS, CA, PCA, t-SNE), Causal Machine Learning (TMLE, Optimal Treatment Regimes)