Yuhan (Victoria) Nian

Houston, TX | +1-832-955-9956 | vn23@rice.edu | Personal Website

Research Interest

My research interests lie in uncertainty quantification, bayesian inference, trustworthy machine learning, and causal inference, with applications in biostatistics, computational social science, applied statistics, and natural language processing.

Education

- **Rice University** *Aug 2022 - May 2026*
  BS in Statistics and BA in Asian Studies; Minor in Data Science Houston, TX
  - GPA: 3.86/4.00
  - Relavent Coursework: Linear Algebra (A+), Linear Regression (A), Probability (A), Bayesian Inference (A), Causal Inference (A), Statistical Inference (A), Real Analysis (A, transferred from UIUC), Time Series Analysis (A-)

Publications & Preprints

C=Conference, S=In Submission, T=To be submitted

[C.1] Abramson, C. M., Turner, K., Arteaga, I., Hernández de Jesús, A., Ginn, B., **Nian, Y.**, & Dohan, D. (2025). *"Pragmatic Sensemaking: Mapping the Cultural Work of Living with Dementia."* American Sociological Association Annual Meeting — Computational Approaches to Culture and Cognition (Section on Social Psychology). Chicago, IL.

[C.2] Abramson, C. M., Turner, K., Arteaga, I., Hernández de Jesús, A., Ginn, B., **Nian, Y.**, & Dohan, D. (2025). *"Pragmatic Sensemaking: Semantic Maps of Dementia Narratives."* ARS'25: Tenth International Workshop on Social Network Analysis — Networks in Culture, Culture in Networks. Naples, Italy.

[S.1] Abramson, C. M., Turner, K., **Nian, Y.**, Arteaga, I., Hernández de Jesús, A., Ginn, B., & Dohan, D. (2025). *"Pragmatic Sensemaking: Mapping the Cultural Work of Living with Dementia."* Manuscript under review at *Social Science and Medicine*.

[S.2] Abramson, C. M., & **Nian, Y.** (2025). *"The Cultural Mapping and Pattern Analysis (CMAP) Visualization Toolkit: Open Source Text Analysis for Qualitative and Computational Social Science."* Manuscript under review at *Journal of Open Source Software*. https://doi.org/10.48550/arXiv.2510.16140

[T.1] Fernandez, R., Leon-Pena, A., **Nian, Y.**, Li, M., Salan-Gomez, M., D'Silva, E., Mattar, A., Elsenousi, A., Garcha, P., Liao, K., Shafii, A., & Loor, G. (2025). *Contemporary lung transplant outcomes with portable ex-vivo lung perfusion: a single center analysis.* To be submitted to *American Journal of Transplantation.*

[T.2] **Nian, Y.**, Liu, J., Li, M.(2025). *Trustworthy Online Conformal Prediction by Super Learner Ensembling.* To be submitted to *ICML 2026*

Research Experience

- **Contemporary Outcomes of Portable EVLP in Lung Transplantation (Rice University)** *Sep 2025 – Present*
  Research Assistant | Advisors: Dr. Meng Li & M.D. Gabriel Loor Houston, TX
  - Collaborated with physicians at Baylor College of Medicine to conduct a single-center analysis on portable EVLP outcomes.
  - Performed propensity score–adjusted analyses comparing portable EVLP and standard cold preservation using overlap and inverse probability weighting, with standardized mean difference plots to assess covariate balance.
  - Evaluated primary (PGD3, 90-day mortality) and secondary outcomes including postoperative ECMO use and renal failure, finding no significant differences between groups.

- **Neuron Redistribution for Multi-Layer Transformer Interpretability (Rice University)** *Mar 2025 – Present*
  Research Assistant | Advisor: Dr. Hanjie Chen Houston, TX
  - Fine-tuned Qwen3-1.7B, LLaMA3.1-8B, and GPT-2 baselines on classification (SST-2, IMDB), text generation (WritingPrompts), and math reasoning (GSM8K) tasks in PyTorch.
  - Implemented cross-layer regularization on high-attention tokens to promote monosemantic representations and on correlated token pairs to manage polysemantic interactions.
  - Achieved average improvements of 25% in key-token detection and 4% in model accuracy over fine-tuned baselines.

- **Cultural Mapping Visualization Toolkit (Rice University)** *Aug 2024 – Oct 2025*
  Research Assistant | Advisor: Dr. Corey M. Abramson Houston, TX
  - Designed an open-source toolkit for computational text analysis in social science and health policy research.
  - Integrated RoBERTa, TF-IDF, and Cosine/Jaccard similarity to extract semantic associations from user transcripts.
  - Developed interactive semantic network, heatmap, and t-SNE visualization modules using Plotly and Dash.

- **Online Conformal Prediction by Super Learner Ensembling (Rice University)** *May 2024 – Oct 2025*
  Research Assistant | Advisor: Dr. Meng Li | Manuscript Available upon Request Houston, TX
  ◦ Implemented online conformal prediction models using ARCH and LSTM base learners in Scikit-learn, Statsmodels, and TensorFlow, maintaining ∼90% empirical coverage in sequential forecasting.
  ◦ Constructed a Super Learner ensemble in NumPy/Pandas using convex weighting of calibrated predictors, reducing prediction interval width by ∼30% compared to the best single model.
  ◦ Benchmarked performance against AdaHedge majority aggregation on jump–diffusion simulations, achieving ∼14% interval width reduction while preserving nominal coverage.

Professional Experience

- **Xiamen C&D Paper & Pulp Group Co., Ltd.** *Apr 2025 – Jun 2025*
  Data Engineer Intern Chongqing, China
  ◦ Engineered data preprocessing pipelines using pandas, NumPy, and SQLAlchemy to standardize paper product codes and grades, reducing data mismatches by 12%.
  ◦ Developed routing optimization with Google OR-Tools, cutting shipping costs by 32%; created a delay classifier using XGBoost for the logistics department.
  ◦ Deployed real-time dashboards in Streamlit for cost estimation, route planning, and sales tracking, reducing decision turnaround time from 3 hours to 30 minutes.
- **Xiamen C&D Inc. (Fortune 500)** *May 2024 – Jul 2024*
  Data Engineer Intern Xiamen, China
  ◦ Indexed MySQL tables, reducing key mismatches between warehouse and client sales systems by 35%.
  ◦ Built a Python–SQL ETL pipeline processing 5M+ nightly sales rows with automated validation for duplicates and nulls, ensuring reliable and consistent data loads into the warehouse.
  ◦ Containerized the pipeline with Docker and orchestrated runs using Airflow, automating deployment across development, testing, and on-prem environments, cutting rollout time from days to hours.
- **China Merchants Bank** *Jul 2023 – Aug 2023*
  Financial Analyst Intern Chongqing, China
  ◦ Analyzed 20K+ transaction records in SQL covering customer demographics, behaviors, and product adoption trends; implemented K-Means clustering (pandas, scikit-learn) to segment customers for cross-selling strategies.
  ◦ Built Tableau dashboards to automate weekly performance reporting, improving review efficiency by 40%.

Leadership Experience

- **Rice ASA (American Statistical Association) Student Chapter** *Sep 2022 - Present*
  Undergraduate Co-President
  ◦ Organized ASA student activities and assigned responsibilities to officers.
  ◦ Led the undergraduate division and promoted engagement in data initiatives.
  ◦ Collaborated with Co-Presidents to foster a supportive and connected Rice data community.

Teaching

- **STAT 425: Bayesian Inference (Rice University)** *Aug 2025 – Present*
  Teaching Assistant | Instructor: Dr. Marina Vannucci Houston, TX
  ◦ Graded weekly assignments for 30 students, analyzed common errors, and provided feedback to head TA.
- **BUSI 395: Data Analytics (Rice University)** *Aug 2025 – Present*
  Teaching Assistant | Instructor: Dr. Yueyang Liu Houston, TX
  ◦ Graded weekly assignments for 60+ students, analyzed score distributions, and presented findings in TA meetings.
- **STAT 410: Linear Regression (Rice University)** *Jan 2024 – May 2025*
  Teaching Assistant | Instructor: Dr. Roberto Bertolusso Houston, TX
  ◦ Graded exams and assignments for 300+ students; wrote weekly reports summarizing common mistakes and provided feedback to the professor.
  ◦ Led tutorials and prepared supplementary materials to reinforce conceptual understanding and application of linear regression in R.
- **STaRT@Rice Workshop (Rice University)** *Oct 2025*
  Teaching Assistant | Session: "Visualizing Text Data" with Dr. Corey M. Abramson Houston, TX
  ◦ Tutored 30+ students in Python-based text analysis and visualization during the Statistical Training and Research Techniques Workshop.

Skills

- **Programming Languages:** Python, R, SQL, C/C++, Java, JavaScript, HTML
- **Tools & Frameworks:** React, Figma, Excel, Tableau, Power BI, Linux, CUDA
- **Data Science & Machine Learning:** scikit-learn, PyTorch, TensorFlow, NumPy, Pandas, Apache Spark, Docker, AWS, Hadoop, RoBERTa, Gensim, NLTK, A/B Testing
- **Languages:** Native in Chinese, Proficient in English, Intermediate in Japanese