YU XIA

+1 (917)-684-0341 | xiayu6500@gmail.com | LinkedIn | Personal Page

EDUCATION

New York University, New York 09/2018 - 05/2022
The Courant Institute of Mathematical Sciences
B.A. in Mathematics and Data Science

- ***GPA:*** 3.854 / 4
- ***Coursework:*** Honor Analysis, Probability & Statistics, Numerical Analysis, Partial Differential Equation, Stochastic Calculus, Causal Inference, Machine Learning, Reinforcement Learning

PUBLICATIONS

- A. Mueller, **Y. Xia**, and T. Linzen. 2022. Causal Analysis of Syntactic Agreement Neurons in Multilingual Language Models. In Proceedings of the 26th Conference on Computational Natural Language Learning (CoNLL), pages 95–109. Association for Computational Linguistics.

PREPRINTS

- **Y. Xia**, C.H. Wang, J. Mabry, G. Cheng. 2024. Advancing Retail Data Science: Comprehensive Evaluation of Synthetic Data. arXiv. https://arxiv.org/abs/2406.13130. Accepted in KDD 2024 GenAI Evaluation Workshop.
- **Y. Xia**, S. Narayanamoorthy, Z. Zhou, J. Mabry. 2024. Simulation-Based Benchmarking of Reinforcement Learning Agents for Personalized Retail Promotions. arXiv. https://arxiv.org/abs/2405.10469
- **Y. Xia**, A. Arian, S. Narayanasmoorthy, J. Mabry. 2023. RetailSynth: Synthetic Data Generation for Retail AI Systems Evaluation. arXiv. https://arxiv.org/abs/2312.14095. Accepted in INFORMS Business Analytics Conference 2024.

TALKS

Woman in Data Science Worldwide, Upskill Workshop – Probabilistic Programming 101 06/2023

- Introduced Numpyro framework with a simplified Bayesian hierarchical linear regression example in an interactive coding session.

RESEARCH EXPERIENCE

UCLA Trustworthy AI Lab, Remote 05/2024 - Present
Research Assistant, Advisor: Prof. Guang Cheng, Dr. Chi-Hua Wang

- Cleaned open Dunnhumby dataset, Complete Journey, to extract customer transactions from retail stores on a weekly level; fitted generative models to create synthetic datasets using AutoGAN, CTGAN, TabDDPM, LLM, etc.
- Proposed an evaluation framework for the performance of generative models on:
    - Fidelity: compared marginal and joint distributions' similarities to the real data in customer demographics, purchase behavior, basket size, etc.
    - Utility: tested synthetic datasets on downstream ML tasks against the benchmark of the real data, like premium customer classification, product association, etc.
    - Privacy: computed differential privacy metrics, distance to the closest record, etc.

- First-authored paper accepted by KDD 2024 GenAI Evaluation Workshop.

NYU Stern School of Business, New York 05/2021 - 05/2022
Research Assistant, Advisor: Prof. Zhengyuan Zhou

- Led a student team of 5 to build an article recommendation system:
    - Served as a product manager to refine user stores, create a product roadmap, prioritize features, maintain backlog, facilitate sprint planning and reviews, etc.
    - Served as a tech lead to build article crawler, backend API, frontend, and vector database; developed bandit algorithms, like UCB, Exp3, to adjust recommenders.
- Expanded Exp3 algorithm to an adversarial bandit setup with clustered arms:
    - Imitated the case when an e-commerce site categorizes items into groups and observes rewards for this whole group from user interaction with a single item.
    - Proved a sublinear regret bound and ran statistical simulations in both the synthetic environment and recommendation platform.
- Wrote white paper for proof of concept of retail marketing digital twin, following the Partially Observable Markov Decision Processes framework; built the digital twin and demonstrated the capability to recover parameters of synthetic customer purchase lifecycle.

NYU Center of Data Science, New York 10/2021 - 04/2022
Research Assistant, Advisor: Prof. Tal Linzen, Dr. Aaron Mueller

- Created a dataset with English phrases with various syntactic structures, like simple agreement, across prepositional phrases, across object relative clauses, etc.
- Customized BERT, XGLM, GPT2 models to enable intervention in certain neurons at once.
- Evaluated natural indirect effect (NIE) to quantify the causal mechanism of syntactic agreement, on model layer-wise NIE contours, syntax neurons shared across structures and languages, and neuron sparsity pattern across structures and languages.
- Set up experiments on the HPC cluster and analyzed monolingual and multilingual models.
- Paper accepted by 26th Conference on Computational Natural Language Learning.

INDUSTRY EXPERIENCE

Bain & Company, San Francisco 09/2022 - Present
Machine Learning Engineer Intern / Machine Learning Engineer

- Research IP
    - Built a personalized demand model using Bayesian Hierarchical Linear Regression; derived price elasticity per customer per product in a closed analytic form; simulated revenue and customer lifetime value under different marketing scenarios. Paper accepted by INFORMS Business Analytics Conference 2024.
    - Wrapped previous demand model in a retail digital twin to evaluate customized pricing strategies with bandit and reinforcement learning algorithms, like LinUCB, Neural Boltazmann, PPO, DQN, etc. Paper in preparation.
- Retail & Supply Chain
    - Ran customer analysis on a large-scale transaction dataset for price sensitivities, the causal relationship between customer features and purchases; Built forecasting models by customer segments by matrix factorization.

    - Leveraged multimodal models to build an OCR pipeline to extract carrier and shipper information from delivery documents, which replaces the manual effort of scanning at the wholesale warehouse.
    - Built route optimization toolkit to schedule delivery jobs under time, cost, and distance constraints using linear programming.
- Healthcare
    - Conducted digital transformation of an insurance company's data and analytics pipeline to enable large-scale experiments for data-driven decision-making.
    - Integrated LLM to understand transcripts from phone calls and chatbot conversations to summarize patient needs quickly; leveraged LLM embeddings to predict patient satisfaction from transcripts and reviews.

JPMorgan Chase & Co., New York 06/2022 - 08/2022
Quantitative Research Intern

- Evaluated Commodities portfolio P&L with full revaluation pricing and sensitivity pricing.
- Conducted a pricing accuracy test and constructed a pipeline to flag weak approximates.
- Analyzed pricing accuracy in supply, book, and instrument levels to generate reports.

TEACHING EXPERIENCE

Teaching Assistant at NYU Stern, New York 09/2022 - 12/2023

- Served as the teaching assistant and grader on the course, Decision Models and Analytics, for both undergraduate and MBA students in two fall semesters.

Learning Assistant at NYU University Learning Center, New York 09/2019 - 05/2024

- Provided 1:1 tutoring sessions on mathematics, statistics, and computer science undergraduate courses, including calculus, linear algebra, probability and statistics, data structure, etc.
- Led group sessions of about 20 students to review and practice algebra and calculus.

Volunteer Tutor at the City Tutors Program, New York 09/2018 - 12/2018

- Assisted the teacher in a pre-calculus course with 10-15 students and helped them prepare for the High School Equivalency (HSE) exam.

Honor & Awards

- Graduated with magna cum laude, 2022
- Received Dean’s List for four years, 2018 - 2022
- Outstanding Winner at Interdisciplinary Contest in Modeling (top 1%), 2020

SERVICE

Ad-hoc reviewer for Expert Systems With Applications, IEEE Control Systems Letters, IEEE Control Systems Society

TECHNICAL SKILLS

Python, R, SQL, Git, AWS, GCP, Docker, Latex