Wenyue (Suzie) Xi

wenyue.xi001@gmail.com | suziexi.github.io | (413) 923-8842 | New York, NY

EDUCATION

New York University *Sep 2021 - May 2023*
Master of Science, Computer Science, Courant Institute of Mathematical Sciences New York, NY
Smith College *Sep 2017 - May 2021*
Bachelor of Arts, Computer Science, Art History Northampton, MA
Davis United World College Scholars Program, Forté Fellowship Program

RESEARCH INTERESTS

Methodological Interests: Statistical Modeling, Causal Inference, Difference-in-Differences, Machine Learning, NLP
Substantive Interests: Digital Platforms, Social Media, Generative AI, Large Language Models(LLMs), Algorithmic Bias

PUBLICATIONS AND WORKING PAPERS

1. **Wenyue Xi**, Oscar Wan, João Sedoc, "Generative AI in Recruitment: A Comparative Study of Llama2 and ChatGPT for Resume Screening", [abstract]
- Presented at **2024 INFORMS Annual Meeting**
- Methods: Prompt Engineering, Chain-of-Thoughts, Unstructured Data Processing, Word Embedding, BERTopic
- Literature: Generative AI, Natural Language Processing in Recruitment, Model Transparency and Explainability

2. Siqi Pei, **Wenyue Xi**, Hongshen Sun, "Mobile Payments and Urban Mobility: Assessing the Influence of Mobile Payments on Metro Ridership and Environmental Enhancement", [abstract]
- Presented at the 46th **INFORMS Society for Marketing Science (ISMS) Conference** in 2024
- Methods: Causal Inference, Difference-in-Differences (DiD), Statistical Analysis
- Literature: Mobile Payment, Urban mobility, Sustainability, Transportation Economics, Social Welfare

3. Chuang Tang, **Wenyue Xi**, "Decoding Visual Emotions: Impact on Sales from Yelp Photo Review", [abstract]
- Methods: Multimodal Data Analysis (numerical, text, image), Web Scraping, Regression, Sentiment Analysis
- Literature: Visual Sentiment Analysis, Online Reputation, E-commerce, Multimodal Data Processing

4. **Wenyue Xi**, Zhongheng Cheng, Mark Turner, "Empirical Benchmark of Generative AI for Frame Blending Generation", Submitted to the 17th **International Cognitive Linguistics Conference**, [abstract]
- Methods: Multimodal Data Analysis, Human-in-the-loop Machine Learning
- Literature: Cognitive Linguistics, Multimodal Learning, Language Pattern recognition, Frame Semantics

5. **Wenyue Xi**, Ruiying Liu, Rui Cui, "Volunteer Future Collaboration Prediction with Topology-based Ensemble Models", Proceedings of 33rd **IEEE Workshop on Machine Learning for Signal Processing**, 2023, [pdf]

6. Peter J. Whitehouse, Kristin Bodiford, Patrik Standar, Arthur Namara Aarali, Sylvia Asiimwe, Vanessa Vegter, **Wenyue Xi**, Paloma Torres-D´avila, "Intergenerative Transdisciplinarity in "Glocal" Learning and Collaboration"**,** Transdisciplinary Journal of Engineering & Science, 2021, [pdf]

RESEARCH EXPERIENCE

NYU Stern School of Business *Oct 2023 - Present*
Research Assistant, Supervisor: Dr. João Sedoc New York, NY
- Developed algorithms in Python to evaluate Generative AI models (Llama-2 and ChatGPT) for unstructured data parsing, matching resumes with job descriptions, addressed concerns of accuracy, transparency, and explainability
- Built and managed the experiment dataset, applied advanced techniques such as word embedding, BERTopic, and LlamaIndex to evaluate the models' performance in resume parsing and job matching
- Used prompt engineering techniques, like Chain-of-Thoughts and integrated gradient to enhance model explainability

MIT Sloan School of Management & Shanghai University of Finance and Economics *Oct 2023 - Present*
Research Assistant, Supervisor: Dr. Siqi Pei Cambridge, MA

- Collected and analyzed large-scale metro ridership data from over 20 cities across China, using diverse and complex data sources to quantify the effect of mobile payments on public transportation usage
- Validated and processed 100,000+ data entries to conduct statistical analysis to assess metro ridership changes
- Utilized Difference-in-Differences methodology to capitalize on the varied adoption timelines of mobile payments across cities, providing robust evidence of their role in boosting metro ridership and contributing to sustainability

Peking University HSBC Business School *Oct 2023 - Present*
Research Assistant, Supervisor: Dr. Chuang Tang Remote

- Analyzed 1 million Yelp and Amazon entries to investigate how visual sentiment in review images influences user engagement, measured by additional reviews and visitor traffic
- Executed data collection and web scraping, built and managed large-scale databases, and applied multimodal algorithms with Python and TensorFlow to process numerical, text, and image data
- Conducted statistical analyses including linear regression, logistic regression, fixed and random effects, and random forest to evaluate the influence of image sentiment on user numbers

Case Western Reserve University Red Hen Lab *Jun 2020 - Aug 2020*
Research Assistant, Supervisor: Dr. Mark Turner Remote

- Developed representation and multimodal learning techniques from scratch to detect frame-blending, a form of cognitive operation in language usage
- Built semantic distance algorithms using word embedding methods and statistical models with Berkeley FrameNet
- Improved linguistic pattern detection accuracy by 32.7% by adapting Semafor parser and NLTK framework
- Proposed and mentored a follow-up project evaluating Large Language Models' ability to generate frame blends using prompt engineering and techniques like zero-shot, few-shot, and Chain-of-Thoughts

The University of British Columbia Industrial Automation Lab *Jun 2019 - Aug 2019*
Research Assistant, Supervisor: Dr. Teng Li Vancouver, Canada

- Implemented Encoder-Decoder Seq2seq model using Python and Stata to increase the accuracy of satellite cloud imagery classification and abnormal detection by 27%
- Improved accuracy of predicting wildfire by 34% for satellite imagery data from NASA Earth database
- Leveraged knowledge in Deep Learning models, including CNN, RNN, LSTM, and Transformer

PROFESSIONAL EXPERIENCE

Bank of America Merrill Lynch *Jul 2023 - Present*
Quantitative Analyst New York, NY

- Implemented advanced data architecture tools to enhance large-scale financial data storage and lineage tracing
- Deployed quantitative models and analytical tools for comprehensive assessment of market and climate risk data
- Executed rigorous quantitative experiments and financial modeling, analyzing market data, risk metrics, and time series to derive actionable insights and escalation recommendations.
- Performed evaluations of risk model outputs using SQL, Python, and R, ensuring reliability of financial assessments

Google *Jun 2020 - Aug 2020*
GSoC Student Developer Mountain View, CA

- Developed and launched a full-stack website utilizing a range of technologies, including C++, Python, Java, React, HTML, CSS, and JavaScript
- Automated annotation processes with FrameNet and Red Hen Rapid Annotator, boosting efficiency by 53.4%
- Applied expertise in full-stack development, incorporating knowledge of data visualization with Tableau, Natural Language Processing, Machine Learning, and Data Mining

TECHNICAL SKILLS

Programming Languages: Python, R, STATA, SAS, SQL, Java, C/C++, MATLAB, KDB+/Q, C#, JavaScript, LaTeX
Statistical Programming: Classification, Clustering, Regression Analysis, Time Series Analysis, Optimization, A/B Testing, Linear Models, Bayesian Statistics, Cross Sectional Model, Causal Inference, Propensity Score Matching
Visualization: Tableau, Power BI, Adobe Analytics, matplotlib, ggplot2, Seaborn