YUANYUAN XIE

|| Evanston, IL || +1 7732007920 || yyuanxie1101@gmail.com || https://www.linkedin.com/in/yuanyuan-x-663145220 ||

EDUCATION

Northwestern University, Evanston, IL — Expected December 2026
Master of Science in Machine Learning and Data Science — GPA: 3.9/4.0
Emory University, Atlanta, GA — December 2023
Bachelor of Science (Double Major) in Applied Mathematics and Economics — GPA: 3.9/4.0
University of Chicago, Chicago, IL — June – July 2022
Data and Policy Summer Scholar in Data Processing/Data Analytics/R programming

TECHNICAL SKILLS

Programming: Python, R, SQL, Java, C++, SAS | MySQL, Database Design, Data Warehousing | Data Structures | ETL Workflows
Machine Learning & AI: Predictive Modeling, Classification, Clustering, Time Series | Data Mining | Deep Learning (PyTorch, TensorFlow) | NLP (Text Analysis, LLMs, RAG, LangChain), AI Agents | Computer Vision | Causal Inference, A/B Testing
Analytics & Visualization: Pandas, NumPy, PySpark | Tableau, Power BI, Plotly, Seaborn, Excel | Google Analytics, Selenium
Cloud & MLOps: AWS, Google Cloud, Databricks | MLflow, Docker, Git, Linux, CI/CD Pipeline | Model Deployment & Monitoring

WORK EXPERIENCE

Zebra Technologies, Chicago, IL — September 2025 – Present
Data Scientist (Industry Practicum with Northwestern)

- Architected and deployed a **hybrid RAG system** integrating **vector search** and **Neo4j graph retrieval**, orchestrated through a **LangGraph** workflow using **Gemini** models to route queries and synthesize troubleshooting guidance for enterprise device users.
- Partnered with cross-functional teams to develop an automated **MLOps evaluation pipeline on Google Cloud**, integrating **Phoenix** to measure RAG performance (**nDCG@5, groundedness**) and **LLM-as-Judge** scoring to drive improvements in response quality and user-facing reliability.

City University of Hong Kong Department of Management, Hong Kong SAR — February 2024 – July 2025
Research Assistant, Supervisor: Dr. Long Wang

- Engineered a **Python** and **Selenium ETL** pipeline to collect demographic and rank data for 200,000+ police officers, enabling large-scale **causal analysis** of early-career allegations on long-term misconduct outcomes using **2SLS models in Stata**.
- Designed and executed a 2×2 **factorial eye-tracking experiment** in collaboration with a local social enterprise, studying financial transparency and endorsement effects; integrated **Google Analytics** event logs with behavioral data and applied latent-variable modeling in **R** (factor analysis, SEM) to quantify cognitive and moral trust drivers that informed marketing and sales strategy.
- Developed **JavaScript pipelines** to analyze job-posting language on Boss Zhipin, quantifying urgency cues and gender-coded tone; applied **supervised NLP models** and **topic modeling** to assess effects on applicant engagement and conversion.

Emory University Department of Economics, Atlanta, GA — January 2023 – January 2024
Undergraduate Research Assistant, Supervisor: Dr. Caroline Fohlin

- Enhanced AI-driven document intelligence in **AWS SageMaker** for digitizing historical Moody's manuals; performed **A/B testing** to benchmark Amazon **Textract** and **Google Document AI** pipelines, achieving 40% higher parsing accuracy.
- Categorized 10,000+ extracted company profiles into thematic clusters using **K-Means** and built interactive **Tableau** dashboards with drill-down visualizations, streamlining background verification workflows for faster analyst review.
- Formed a robust quality control system using the **Naïve Bayes algorithm** to classify scanned document sections, filtering out low-relevance segments and raising data preparation efficiency by 20% while improving review precision.

Intellipro Group, Santa Clara, CA — September 2022 – December 2022
Marketing Analyst Intern

- Achieved 86% sentiment-classification accuracy across Latin America's IT market using **Random Forest, Support Vector Machine, and Logistic Regression** models to analyze salary perceptions and enable precise customer segmentation.
- Queried, cleaned, and segmented a 100,000+ record IT salary dataset in **MySQL** and designed an ERD to optimize table relationships, reducing dashboard refresh and data retrieval latency by 50%.
- Identified apparel purchasing trends and forecasted sales performance using **ARMA** and **SARIMA** models in R, achieving 89% forecasting accuracy and delivering data-driven insights that supported a local clothing brand's international market entry strategy.

PUBLICATIONS & PRESENTATIONS

96th International Atlantic Economic Conference, Philadelphia, PA — October 2023
Presented: "Life Insurance Reinvented: A Cross-National Analysis on Annuity Payments"

Re-imagining the Future of Forest Management -- An Age-Dependent Approach towards Harvesting — August 2023
Submitted on arXiv: DOI: 10.48550/arXiv.2308.03198