Xuanmao Li

+1-323-942-4804 | xl979@cornell.edu

Xuanmao Li | Xuanmao Li

Education

- **Cornell University** *Sep. 2024 - Jan 2026*
  Master of Engineering in Computer Engineering Ithaca, New York, USA
- **Huazhong University of Science and Technology** *Sep. 2020 - Jun. 2024*
  Bachelor of Science in Automation Wuhan, China
  – GPA: 3.66/4.00

Skills

- **Programming Languages:** Python, SQL, Matlab, R, C
- **Tools:** PyTorch, PostgreSQL, Pyspark, AWS, Tableau, MySQL, scikit-learn, Seaborn, Plotly

Research Experience

- **Arizona State University** *Jun. 2023 - Sep. 2023*
  Research Assistant - Automatic Data Transformation Using LLMs Tempe, Arizona, USA
  – Designed and developed SQLMorpher, a data transformation tool using large language models (LLMs) to automate data transformations.IEEE Big Data 2023: arXiv:2309.01957.
  – Used Python to build a tool that interprets and restructures multi-source datasets for consistent analysis, demonstrating the tool's effectiveness in streamlining data workflows in real-world scenarios.
  – Benchmarked the tool using building energy data, creating a large dataset and validation framework, which helped standardize evaluation metrics for assessing data transformation tools.
  – Conducted data analysis and visualization using Matplotlib and Seaborn to showcase performance improvements and accuracy of the transformation processes compared to traditional methods.
- **Purdue University** *Apr. 2023 - Nov. 2023*
  Research Assistant - NeRF Dataset Collection and Object Analyzer West Lafayette, Indiana, USA
  – Contributed to build a large-scale dataset DL3DV-10K with over 47.2 million frames, improving data availability for object detection and scene rendering. CVPR 2024: arXiv:2312.16256.
  – Employed advanced machine learning techniques, specifically CNNs and transfer learning, to refine data collection and model tuning, significantly surpassing baseline object detection accuracy.
  – Automated the data processing pipeline using Python and Box to handle and preprocess vast amounts of image data, ensuring efficient and scalable data management.

Projects

- **Streaming Data Analytics Platform** *June 2024 - September 2024*
  Project Developer Personal Project
  – Engineered a real-time analytics platform using **Apache Spark** and **Kafka** on **AWS EC2**, leveraging advanced statistical methods like ensemble methods for optimizing data stream processing.
  – Created dynamic **Tableau** dashboards to visualize user activity and transactions, integrating statistical analytics to improve decision-making and system resilience through comprehensive error monitoring.
  – Utilized K-Means clustering for segmenting transaction data, enabling targeted user engagement strategies based on behavior patterns identified in the analysis.
- **AWS-Integrated Data Transformation Pipeline** *June 2024 - September 2024*
  Project Developer Personal Project
  – Developed an end-to-end ETL pipeline with **AWS Glue** and **AWS S3**, automating data workflows with **AWS Lambda** for enhanced data handling and responsiveness.
  – Implemented **AWS QuickSight** dashboards for real-time data insights, using L-Nearest Neighbors algorithms to refine analytics and provide precise operational metrics.