Longcong Xu

(217) 200-1947 | lx2305@columbia.edu | https://github.com/RicoXu727 | https://www.linkedin.com/in/longcongxu/

EDUCATION

Columbia University New York, NY
Master of Science in Data Science, GPA: 3.8/4.0 Dec 2023
Courses: Advanced Algorithms, Computer Systems, Deep Learning, NLP, Artificial Intelligence, Statistical Modeling

University of Illinois at Urbana-Champaign Champaign, IL
Bachelor of Arts in Economics, Minor in Computer Science, GPA: 4.0/4.0 May 2022
Courses: Data Structures, Discrete Structures, Data Mining, Cloud Computing Applications
Honors: Dean's list all semesters, Summa Cum Laude.

SKILLS

- Coding: Python, Java, SQL, Swift, JavaScript, HTML, CSS, R
- Development: MVC Architecture, CI/CD, Flask, GCP, AWS, Docker, Linux, Git, Bash
- Data Engineering: Data ETL, OLAP, DBT, Spark, Hadoop, Kafka, Cloud Data Warehouses, PostgreSQL, MySQL
- Data Science: Regression, Classification, Time Series Analysis, LLMs, NLP, Image Processing, Tensorflow, PyTorch
- Data Visualization and Reporting: Matplotlib, Seaborn, ggplot2, Power BI, Tableau, AWS QuickSight

WORK EXPERIENCE

BlocPower Energy Services New York, NY
Data Engineer Intern May 2023 - Present
- Developed a BlocPower Chatbot using Streamlit and AWS Bedrock, focusing on fine-tuning the Command model and prompts engineering
- Created 5 API data pipelines, reducing 4+ hours of manual data fetching, resulting in enhanced customer personas
- Unified 3 TBs of energy consumption datasets using spatial joins, resulting in 8 different analytical visualizations
- Conducted thorough ETL on the merged dataset, deriving 1K bias correction coefficients archetypes between models

University of Illinois at Urbana-Champaign Champaign, IL
Research Assistant Jan 2022 - May 2022
- Leveraged Convolutional Neural Networks on 20GB image dataset, facilitating 15% improvement in staff evaluation
- Categorized 1K+ vending machines, identifying optimal ones for price elasticity experiments based on key attributes
- Streamlined drivers' routes across ten cities, achieving a 20% boost in operational inventory efficiency

People's Insurance Company of China (PICC) Hangzhou, CH
Data Scientist Intern Jul 2021 - Aug 2021
- Applied Gray modeling, LASSO regression to predict two-quarter insurance metrics, aligning with client company
- Analyzed 30 driving attributes to refine 2 models, evaluating drivers' skills and pinpointing accident risk factors

ACADEMIC PROJECTS

Information Privacy Controls for Corporate Use of Public LLMs (GE Aerospace)
- Enhanced and fine-tuned the LLama2 13B Chat model to effectively redact private information from medical datasets, achieving a notable 87% accuracy
- Successfully applied transfer learning techniques to adapt this de-identification process for aviation data, ensuring high-level data security and compliance with industry standards

Masked Topological Attention for Video-Transformers (Google DeepMind)
- Refined Video Transformers by integrating topological masking into their attention mechanisms for better structural understanding
- Modified the VideoGPT model with a Vector Quantized Variational Autoencoder (VQ-VAE), enhancing its capability to represent geometric structures in video data
- Performed experiments with the BAIR Robot Pushing Dataset, showing improvements in video generation with a new topological masking technique