Xuanzhang(Lucas) LIU

(404) 263-7022 | **xuanzhangliu@gatech.edu** | Atlanta, GA | **LinkedIn**

Education

Georgia Institute of Technology ***Atlanta, GA***
M.S. Computer Science - GPA: 4.0/4.0 *Aug 2023 - May 2026(Expected)*

Zhejiang University ***Hangzhou, CN***
B.E. Systems Engineering - Chu Kochen Honors Degree (Top 1%) *Sep 2019 - Jun 2023*

Skills

Programming Languages:	Python, Golang, SQL, R, C++, C#, Java, Javascript, Node.js, Perl, CSS, Scala
Frameworks & Libraries:	Spring Boot, MyBatis, Rest API, Flask, React, Angular
Databases & Cloud Services:	MongoDB, Pinecone, AWS, Azure, Spark, Hive, Hadoop, MapReduce, Kafka
DevOps & Operating Systems:	Docker, Kubernetes, Git, CI/CD, Agile, Linux/Unix
AI & ML:	PyTorch, Tensorflow, CUDA, TensorRT, LlamaIndex, Hugging Face

Internship Experience

Alibaba Group ***Hangzhou, CN***
Software & Data Engineer *Jul 2024 - Sep 2024*

- Developed a backend system for real-time predicted LTV computation using device ID, touchpoint attribution, with data stored in HBase for high concurrency and low latency, enhancing marketing attribution accuracy.
- Refactored projects using Java, Spring Boot, and MyBatis, reducing system coupling and enhancing maintainability.
- Engineered ETL pipelines with SQL and Python, boosting data loading by 30% and streamlining data cleaning processes.
- Integrated microservices architecture with HSF and implemented Tair for distributed caching, enhancing system stability and enabling high concurrency with effective RPC and service governance.
- Automated deployment and service management with Docker and Kubernetes, improving service reliability. Monitored and visualized system health with Prometheus and Grafana, reducing downtime by 15%.

Desay SV ***Singapore, Hybrid***
Software Engineer *May 2024 - Jul 2024*

- Developed a Flask RESTful API backend, integrating MongoDB for chat history and Pinecone for vector embeddings. Optimized the 7B model using TensorRT to enhance performance on resource-constrained systems.
- Built a responsive React frontend and deployed the application on AWS with Docker and GitHub Actions for CI/CD. Integrated real-time updated knowledge sources and designed a scalable RAG server with LlamaIndex.
- Improved chatbot accuracy and efficiency using advanced agents and optimization techniques, reducing input prompt size by 50%. Implemented security measures to prevent abuse and ensure responsible API usage.

NIO Energy Co. ***Shanghai, CN***
Data Scientist *Mar 2023 - Jun 2023*

- Conducted Time Series Analysis on 80 million dataset using Python and Spark, tested the novel pricing strategy by A/B testing, and achieved an 18.6% increase in station utilization and a 21.5% rise in revenue.
- Extracted battery swapping order data using SQL, utilized hypothesis test to evaluate the effectiveness of improved battery inventory strategies designed with the Engineering team, and achieved a 24.2% boost in process efficiency.

Projects

Deep Learning for Sleep States Detection *Oct 2023 - Dec 2023*

- Led a team to develop a robust Deep Learning model with PyTorch for predicting sleep stages.
- Crafted several RAM-friendly, efficient data-loading strategies to process 128 million data, reduced memory usage by 60%.
- Designed powerful model architecture by comparing the Random Forest, XGBoost, BiGRU and Transformer-based models.
- Increased accuracy by 48%, and discovered the insight of different models handling time-series sequential data.

BuzzDB - Advanced Cpp database implementation *Sep 2023 - Dec 2023*

- Developed a concurrent buffer manager with 2Q strategy, including page ID handling and page-locking operations.
- Combined buffer management with C++ templating, ensuring dynamic adaptability while emphasizing compile-time efficiency and node resolution without direct pointers.
- Implemented a robust B+-Tree index, incorporating key functions like lookup, insert, and erase, etc.
- Utilized CMake for building within a Docker Linux environment and performed memory checks using Valgrind.

Video Streaming System *Nov 2023 - Jan 2024*

- Developed back-end functionalities using Java and SpringBoot, while employing Python for the recall and sorting modules.
- Built video uploading, playback, liking, and favoriting functions utilizing FastDFS as the file server.
- Leveraged collaborative filtering and deep learning techniques to achieve multi-channel recall, resulting in accuracy of 74%.
- Designed a hybrid ranking model by leveraging TensorFlow and Keras, resulting in a top-10 accuracy of 77%.