Robert Lewis

Email: contact@robertlewis.dev

Los Angeles, CA robbylew robbylew Cell: +1 (323) 886-9730

Summary

Motivated Data Engineering and Machine Learning intern candidate pursuing an M.S. in Computer Science at Georgia Tech. Proficient in **Python**, **SQL**, cloud platforms (**AWS, Azure, GCP**) and big data frameworks (**Spark, Hadoop, Kafka**). Demonstrated experience in building scalable pipelines and ML systems through hands-on projects. Eager to leverage technical skills and drive data-centric solutions.

Education

Georgia Institute of Technology (Georgia Tech) Remote
M.S. in Computer Science **Expected June 2026**

- **Relevant Coursework**: Advanced Database Concepts & Design, System Design for Cloud Computing, Data Analytics & Security, Computer Networks, Artificial Intelligence, Machine Learning for Trading

University of California, Los Angeles (UCLA) Los Angeles, CA
B.S. in Statistics and Data Science **Graduated June 2023**

- **Relevant Coursework**: Probability Theory, Deep Learning, Neural Networks, Bayesian Statistics, Computer Vision, Data Warehousing, Data Mining, Data Visualization, Data Structures & Algorithms

Technical Skills

Programming & Scripting: Python, Go, SQL, JavaScript, TypeScript, R, Bash

Machine Learning & Data Science: Pandas, NumPy, TensorFlow, PyTorch, scikit-learn, Matplotlib

Cloud & MLOps: AWS, Azure, GCP, Docker, Kubernetes, Terraform, Airflow, Jenkins, GitLab CI/CD

Data Engineering & Pipelines: Kafka, Hadoop, Spark, Hive, dbt, Apache Beam

Databases & Storage: MySQL, PostgreSQL, MongoDB, Redis, S3, Elasticsearch, Cassandra

Projects

Song Recognition System

- Built an end-to-end **Golang** pipeline re-implementing Shazam's algorithm to extract audio fingerprints.
- Integrated **Spotify** and **YouTube** APIs for metadata and downloads; used **FFmpeg** for preprocessing.
- Developed scalable ingestion and matching with **SQLite** and **MongoDB** for efficient fingerprint retrieval.

LLM-Driven Document Summarization Platform

- Streamlined retrieval across **millions** of documents to accelerate decision-making.
- Orchestrated LLM prompts with **LangChain** and indexing via **Pinecone**; containerized with **Docker**.
- Deployed on **AWS Fargate** for serverless scaling, reducing latency by **30%**.

End-to-End MLOps Pipeline for Image Classification

- Implemented a **Kubeflow** pipeline for data ingestion, training, and validation; provisioned **GPU**-enabled **Kubernetes** clusters with **Terraform**.
- Reduced manual intervention by **40%** and inference latency by **30%** through automation.
- Automated ETL decreased data prep time by **35%** while ensuring high integrity.