

Ridwan Bello Salahuddeen, MS

Atlanta, Georgia, United States

✉ ridzy619@gmail.com ☎ +14707781174

🌐 linkedin.com/in/ridwan-salahuddeen 🏠 github.com/ridwan-salau

Professional Summary

Results-driven Data Scientist with over 10 years of experience identifying opportunities for business growth and efficiency through advanced data analysis. Proven track record of leading the end-to-end technical execution of data science projects, from defining data acquisition and analytics strategy to deploying production models on cloud platforms (AWS, Azure, GCP). Expertise in directing junior data scientists, managing stakeholder relationships, and presenting results to senior leadership. Extensive commercial experience in Python, Spark, and deep learning frameworks (PyTorch, TensorFlow) to build and scale ML solutions.

Professional Experience

ML Engineer Legal Document Processing, Apple (Contractor) Feb 2025 – Present

- Led the development and deployment of custom OCR pipeline for legal documents, with a custom deskew logic that improved Levenshtein distance by 12% and table extraction by 7%.
- Trained a document quality estimation model using MLflow for model lifecycle management, applied to over 200k legacy scanned documents, eliminating the need for hard copy of high-quality documents.
- Engineered a hierarchical, hybrid chunking and retrieval strategy using DSPy for prompt optimization and LangChain for RAG orchestration, combining small and large segments for improved RAG performance while using small-sized LLMs, reducing hallucination by 45%.
- Developed a multi-stage metadata extraction pipeline using DSPy for prompt optimization and Pydantic for schema validation, which generates metadata schema for document collections and extracts structured data from documents.
- Developed knowledge graph using Neo4j for the documents, enabling metadata-aware processing queries that go beyond regular retrieval and generation, with Redis for caching and Apache Kafka for real-time data streaming.

Co-Founder & Technical Lead, Byte Analytics August 2024 – Present

- Led the design and development of system architecture for a scalable financial reconciliation system, directing a team of engineers.
- Engineered and delivered regulatory International Financial Reporting Standards 9 (IFRS 9) compliance systems for two top banks in Nigeria, managing projects from conception to deployment and stakeholder acceptance.

Senior Machine Learning Engineer, BeeKeeperAI (Healthcare data) & RigrAI (AI-powered video processing) Jan 2023 – Feb 2025

- Led the training of a light weight Clip-based model for location estimation from images, currently adopted by Norwegian police.
- Directed the fine-tuning and deployment of a Llama model for forum summarization on darkweb data, utilizing weight orthogonalization and LoRA; currently used by some LEAs in Europe.
- Engineered scalable model serving infrastructure with FastAPI, Docker, and Nvidia Triton Inference Server, implementing comprehensive monitoring and load balancing for efficient video processing; used by Canadian police, Danish police and Interpol.
- Led the implementation and deployment of a BERT-based token classification model for PHI detection, achieving 97.3% F1 score.

Senior Data Science Engineer, Tribes.ai Jan 2022 – Nov 2022

- Led development of core Apache Airflow pipeline templates for auto-generating client-specific pipelines, reducing processing time by 45% and accelerating delivery of data products.
- Optimized large-scale batch ETL using Spark (Dataproc), BigQuery, and Datastore for improved performance, showcasing experience with distributed data processing.

Enterprise Data Science Engineer, Sterling Bank PLC

Sep 2021 – Jan 2022

- Led the end-to-end design, development, and deployment of fraud detection models, improving F1-score performance and presenting outcomes to business leaders to inform strategy.
- Architected and optimized data processing workflows using Azure Synapse and Databricks for production deployment, demonstrating experience with cloud-based data processing.

Data Science Engineer, Access Bank PLC

Mar 2017 – Sep 2021

- Developed and deployed a real-time transaction monitoring system using anomaly detection models and Spark Streaming, achieving a 46% reduction in the false-negative rate.
- Drove initiatives for database performance optimization through efficient schema design and indexing on cloud infrastructure (AWS), reducing query latency by 30% and influencing data strategy.

Research & Publications

- "Reducing Hyperparameter Tuning Costs in ML, Vision and Language Model Training Pipelines via Memoization-Awareness"
- "A Critical Appraisal of Various Implementation Approaches for Real-time Pothole Anomaly Detection: Toward Safer Roads in Developing Nations"
- Master's Thesis on efficient hyperparameter optimization for large-scale ML systems

Education

Mohamed bin Zayed University of AI (MBZUAI)

2022 – 2024

Master of Science in Machine Learning

Georgia State University

2023 – 2025

Master of Science in Statistics and Computer Science

Ladoke Akintola University of Technology

2009 – 2014

Bachelor of Technology in Electrical and Electronics Engineering

Technical Skills

- **Programming:** Python, SQL, R, Java, Shell Script
- **ML Frameworks:** PyTorch, TensorFlow, Keras, JAX, scikit-learn, Deep Learning, NLP, LLMs
- **Data Engineering:** Apache Spark, Apache Airflow, Kafka, PostgreSQL, Neo4j
- **Cloud & DevOps:** AWS, Azure, GCP, Docker, Kubernetes, CI/CD, Infrastructure as Code, System Monitoring
- **ML Infrastructure:** Model Serving (Triton Inference Server, FastAPI), Distributed Training (NCCL), Pipeline Design, Monitoring, Validation, System Performance Optimization

Certifications

- Microsoft Certified (Associates): Azure Data Scientist, Data Engineer, and AI Engineer
- Udacity: Deep Learning Nanodegree, Data Engineering Nanodegree
- End-to-End ML with TensorFlow on GCP

Projects & Publications

- "Memoization-Aware Bayesian Optimization for AI Pipelines with Unknown Costs" [ICLR 2024].
- Developed a Zero Trust AI validation solution for secure healthcare data processing.
- Built scalable recommendation systems and risk analysis models for financial services.