Shivani Murukannaiah

+1-(959)274-8338 | shivanimurukannaiah@gmail.com | Hartford, CT, 06103, USA | LinkedIn

Summary

Data Scientist & Machine Learning Engineer with experience across healthcare research, enterprise AI, and large-scale analytics consulting. Strong background in ML model development, LLM-based systems, cloud-native data pipelines, and regulated data environments. Proven ability to productionize analytics and AI solutions that support decision-making, compliance, and operational efficiency. Seeking data-driven roles in Machine Learning, Applied AI, or Advanced Analytics.

Skills

Programming & Querying: Python, R, SQL, Java, C, Bash
Machine Learning & AI: Regression, Classification, Clustering, Feature Engineering, Model Evaluation, Random Forest, XGBoost, LightGBM, Neural Networks, CNNs, RNNs, LSTMs, Transformers
Generative AI & NLP: GPT-4, LLaMA, BERT, Prompt Engineering, LoRA / QLoRA, Retrieval-Augmented Generation (RAG), Text Classification, Embeddings, Tokenization, Topic Modeling, Whisper API, LangChain, Hugging Face
Data Engineering: ETL / ELT, Data Pipelines, AWS S3, EC2, Lambda, Step Functions, SageMaker, PostgreSQL, MongoDB, DynamoDB
Visualization & Analytics: Power BI, Tableau, Matplotlib, Seaborn, Plotly
MLOps, Cloud & Infrastructure: AWS, Docker, Git, MLflow, FastAPI, Flask, Jupyter, HPC Computing (SLURM), API Development.

Work Experience

UConn Health **Jun 2025 – Present**

Research Data Analyst – Machine Learning and Genomics

- Engineered and **architected end-to-end ML and bioinformatics pipelines on AWS and HPC (Xanadu) to process 750GB+ RNA-seq datasets, reducing total runtime by ~40% through parallelization and workflow optimization.**
- **Developed and validated tumor risk prediction models (Logistic Regression, Random Forest, XGBoost), improving the identification of high-risk tumor profiles by 27% and supporting downstream clinical research decisions.**
- **Designed reproducible ML-ready data pipelines including alignment, normalization, DESeq2-based differential expression, batch-effect correction, and feature engineering.**
- **Performed GO/KEGG pathway enrichment and clustering analysis in R (clusterProfiler) to identify biological drivers linked to tumor growth and inflammation. Built interactive Power BI and Tableau dashboards integrating genomic and clinical metadata to accelerate hypothesis testing for oncology teams. Implemented HIPAA-compliant, cloud-native workflows using AWS S3, EC2, Lambda, and Step Functions to support secure NIH and cross-institution collaborations.**

MasterCard **Jan 2025- May 2025**

AI Engineer

- Developed Contributed to the development of **LLM-powered internal AI tools** for document understanding, knowledge retrieval, and summarization, supporting operational and compliance workflows.
- Built **Retrieval-Augmented Generation (RAG)** pipelines using LangChain, vector databases, and embedding models, improving internal document retrieval accuracy by **30–40%** compared to keyword-based search.
- Designed and deployed **FastAPI and Flask-based AI microservices** to serve LLM inference endpoints, reducing response latency by **~25%** through optimized request handling. Assisted in implementing **AWS-based deployment workflows** (S3, EC2, Lambda) to support scalable and controlled AI model access across internal teams.
- Performed **prompt engineering, response evaluation, and quality benchmarking**, reducing hallucinated or irrelevant responses by **~20%** across tested use cases. Partnered with data engineers, product managers, and compliance stakeholders to ensure AI solutions aligned with **governance, auditability, and enterprise security standards**.
- Documented model behavior, limitations, and evaluation results to support **responsible AI adoption and internal review processes**.

Accenture **May 2022- May 2024**

Data Analyst

- Analyzed and validated **large enterprise operational and financial datasets (100K records)** during ERP and reporting modernization initiatives, improving downstream data accuracy by **15–20%** post-migration.
- Built and maintained **Power BI dashboards and KPI reports** used by client operations and leadership teams, reducing manual reporting effort by **35–40%** and improving reporting turnaround time.
- Automated recurring data preparation and reporting tasks using **SQL, Excel (Power Query), and Python**, cutting analyst effort by **6–8 hours per week** per reporting cycle. Collaborated with business analysts, architects, and offshore teams to translate **business requirements into analytical outputs**, supporting delivery across multi-team, multi-time-zone engagements.
- Participated in **UAT cycles, defect triage, and stakeholder reviews**, strengthening understanding of enterprise delivery lifecycles, governance, and client-facing analytics.

Project

Genomic Variant Insight Engine – LLM-Powered Genomics Interpretation Platform

- Built an end-to-end **genomics ML pipeline** using RNA-seq preprocessing (alignment, normalization, **DESeq2**) and XGBoost/Random Forest models to identify high-risk gene signatures.
- Developed an **LLM-powered RAG system** (GPT-4/LLaMA + NCBI/ENSEMBL) to generate clinician-friendly variant insights and deployed scalable inference with **AWS Lambda + API Gateway**, supported by Power BI/Tableau dashboards.

Enterprise AI Compliance Analyst – LLM + RAG System for Document Understanding

- Built an **LLM-powered RAG system** using GPT-4/LLaMA, LangChain, and Pinecone to automate compliance document search, summarization, and policy comparison via a FastAPI service. Deployed with **Docker + AWS EC2**, added Whisper for voice queries, and created dashboards to track document insights and compliance gaps.

Education

Master of Science in Data Science
University of Connecticut Storrs, CT