

Jemima Paulraj

Boston, MA | (978) 930-7312 | jemimapaulraj@gmail.com

LinkedIn: linkedin.com/in/jemimapaulraj | GitHub: github.com/JemimaPaulraj

Professional Summary

Data Scientist with hands-on experience delivering enterprise AI/ML solutions, now specializing in LLM, RAG, and agentic systems. Excellent in building scalable, optimized Generative AI applications, including multi-agent workflows using LangChain, LangGraph, and CrewAI. Experienced in integrating AWS services to deliver efficient, low-latency, and production-ready AI solutions

Technical Skills

- **Generative AI :** LLMs, RAG, Agentic AI, Prompt Engineering, Embeddings, Vector databases, Fine-Tuning & Optimization, Human feedback loops, Guardrails, MCP, Orchestration framework, Evaluation, LLMOps
- **Generative AI Frameworks :** LangChain, LangGraph, LlamaIndex, CrewAI, Hugging Face
- **Databases & Vector DBs :** MySQL, MongoDB, BigQuery, FAISS, Pinecone
- **Programming :** Python, SQL
- **Cloud Services:** AWS (Bedrock, SageMaker, Lambda), Azure OpenAI, GCP Vertex AI
- **Libraries :** Pandas, NumPy, REST API, Flask, FastAPI, Scikit-learn, PySpark, PyTorch, Matplotlib, Streamlit, NLTK
- **Machine Learning :** Regression, Classification, Clustering, Forecasting, Deep Learning, NLP
- **DevOps & Tools :** CI/CD, Docker, Kubernetes, Git, Databricks, Airflow

Professional Experience

Data Scientist Intern

Jan 2024 - Apr 2024

Scotiabank

Toronto

- Deployed an AI forecasting model achieving 90% accuracy in predicting call volumes and staffing needs, improving customer satisfaction and reducing operational costs by 25% at Scotiabank's Peru call center.
- Performed in-depth EDA on 5+ years of call center data (1M+ records) from multiple domains to uncover demand patterns, seasonality, and operational bottlenecks, driving data-informed model design.
- Elevated the platform with an LLM-enabled experience, shifting analytics from static reports to interactive, conversational intelligence.

Data Scientist

Aug 2022 - Mar 2023

MindSprint

India

- Fine-tuned and deployed classification and object detection models on GCP (Vertex AI, AutoML Vision), achieving 95% accuracy and reducing manual labeling effort.
- Preprocessed 50k+ video frames for handling noise, class imbalance and annotation gaps to build high-quality training dataset.
- Built an end-to-end automated pipeline on GCP, processing wildlife videos from ingestion to predictions stored in BigQuery, enabling clients to showcase biodiversity and secure funding.
- Implemented a production-grade inference system with monitoring and retraining hooks, improving inference latency in real-world environments.
- Evaluated multiple YOLO variants, optimizing accuracy-latency tradeoffs to improve detection.

Data Engineer Intern

MindSprint

Aug 2021 - Jul 2022

Chennai

- Performed comprehensive data analysis on complex log database to generate actionable predictions and uncover critical operational patterns.
- Designed an interactive Tableau dashboard to visualize any unusual patterns.
- Proof of concept: Developed an end-to-end Android application integrating AutoML (GCP) for real-time leaf disease detection and automated remedy recommendations, streamlining field diagnostics for agricultural users.

Education

- **Post Graduate Diploma** : Big Data Solution Architecture May 2023 - Apr 2024
Conestoga College, Kitchener, Canada
- **Master of Technology** : Artificial Intelligence and Data Science Jul 2020 - May 2022
Amrita Vishwa Vidyapeetham, India
- **Bachelor of Engineering** : Computer Science Jul 2016 - May 2020
RMD Engineering College, India

Projects and Publications

- **Generative AI : Multi-functional AI application** with Streamlit UI integrating chatbot with memory, RAG, text summarization, SQL/NoSQL query handling, image generation, and image captioning. Utilized LangChain, LangGraph, and AI Agent concepts to build efficient, context-aware workflows.
Web App Link : <https://generative-ai-application-jemimapulraj.streamlit.app>
- **Deep Learning : Covid-19 prediction using CT-Scan**: Segmenting (U-Net) and classifying Covid-19 in Lung CT Scan Images using CNN and various Transfer learning algorithms like VGG-16, VGG -19, DenseNet-169 and DenseNet-201 and performance enhancement by Ensemble based approaches like model averaging, weighted average ensemble, majority voting.
Publication : https://www.philstat.org.ph/special_issue/index.php/MSEA/article/view/299
- **Machine Learning : Used Vehicle Sale prediction using Python**: Finding sales price using various machine learning techniques such as Multiple Linear Regression, Support vector regression, Random Forest Using Python.
Publication : Published in ICWMR2020 paper 55

Certifications and Awards

- AWS Certified AI Practitioner
- Certified in "Applied Generative AI Specialization" from Purdue University Online
- HackerRank Python Assessment Certification
- Awards: Received Quarterly Awards from MindSprint for being instrumental in setting up the AI pipeline for the Bioverse project which involves data modelling and model building