

JACOB BINU MANCHERIKALAM

MLOps Engineer | Scalable Pipelines | Cloud Expertise | Model Deployment & Monitoring |

+44-7880490227 jacobbinu4488car@gmail.com github.com/Jakee4488 linkedin.com/in/jacob-binu-4a0596205 97 Holyhead Road, CV1 3AD

PROFILE

MLOps and AI Data Engineer specialising in the end-to-end lifecycle of scalable intelligent systems. I have experience in developing and deploying robust, production-ready machine learning solutions, including large language models (LLMs) and agentic AI solutions across AWS, Microsoft Azure, Google Cloud Platform (GCP) and Databricks. My expertise includes utilising Docker, Kubernetes, and CI/CD pipelines. I excel at bridging the gap between data science and software engineering by architecting full-stack applications with React and optimising RESTful APIs. I am skilled in transforming experimental machine learning models into production-grade, user-facing applications using React, Django, and Flask/FastAPI. Furthermore, I also have experience in training deep learning models and orchestrating cloud-native deployments in enterprise environments. I am passionate about building automated, scalable pipelines that deliver high-performance AI solutions while providing seamless user experiences

WORK EXPERIENCE



MLOps Engineer

Coventry Building Society

01/2025 - Present Coventry CV3 2TQ

https://www.coventrybuildingsociety.co.uk/

- Implement MLOps workflows for model versioning, reproducibility, and CI/CD across development and production environments, improving deployment reliability for regulated credit risk models
- Implement of AI agentic capabilities with a focus on explainability, governance, and responsible AI, leading Proof-of-Value initiatives to evaluate enterprise AI platforms
- Deploy critical Credit Risk models (IRB & IFRS9) within the analytics platform using Python, ensuring high-performance execution of regulatory financial standards
- Engineer serverless ETL pipelines using AWS Glue, automating complex data manipulation and loading into Amazon Redshift to support high-throughput analytics
- Architect and maintain scalable data frameworks using Python, PySpark, and SAS, optimising data flow between Data Science and Engineering teams



IT Engineer

Tata Motors Design Tech Centre

08/2022 - 08/2023 Coventry CV4 7AL

https://www.tatamotorsdesigntech.com/

- Engineered an autonomous Windows 11 deployment pipeline, automating software provisioning and system configuration.
- Automated deployment processes, reducing manual intervention by 80% and saving 30 man-hours per cycle, which enhanced the efficiency and scalability of machine learning operations.
- Optimised IT infrastructure and incident management workflows using PowerShell and Bash automation, increasing ticket resolution by 50% and achieving a 100% closure rate through streamlined Active Directory administration.



Assistant IT Engineer

Mancherikalam Supermarket

04/2019 - 07/2021 Kottiyam

- Provided Level 1/2 IT support for 50+ Windows endpoints, managing operating system updates, patching, hardware troubleshooting, and software deployments while minimising user downtime.
- Developed and deployed a supplier and inventory tracking solution that enhanced asset tracking accuracy and reduced reordering lead time by 30%, streamlining operational processes and improving IT asset management efficiency.

PROJECTS

TracerLensAi

https://github.com/Jakee4488/TracerLensAi

TracerLensAi is a cloud-native AI chat interface powered by Google Gemini. It provides an interactive workspace for evaluating agentic workflows with causal reasoning, live code execution, and web search all backed by a production-grade GCP deployment pipeline.

Deep Agents from Scratch

https://github.com/Jakee4488/deep-agents-from-scratch

Created AI agent framework with LangGraph & Claude, showcasing advanced architectures. Implemented task planning, virtual file systems, sub-agents for multistep reasoning & workflow orchestration. Features ReAct loops, Pydantic state management, Tavily web search, async Python orchestration.

Knowledge Graph RAG System

https://github.com/Jakee4488/Graphrag-flaskapp

Built GraphRAG system with Flask, integrating Ollama API and local models for enhanced context-aware queries and reasoning using graph structures.

End-to-end MLOps with Databricks

https://github.com/Jakee4488/MLOps_Databricks_Project

End-to-end MLOps on Databricks: minimal, production-ready ML workflow including data prep, model training and tracking with MLflow, model registration, and batch inference/serving, showcasing core MLOps stages.

Financial AI MLOps

https://github.com/Jakee4488/Financial_AI_MLOps

Enterprise-grade MLOps on Databricks for real-time anomaly detection in finance. Features include streaming data ingestion, automated model retraining, and deployment for continuous, production-ready performance.

CNN for Plant Disease Classification

https://github.com/Jakee4488/PlantDiseases_CNN_Classification

Built a scalable MLOps system for classifying plant diseases (Healthy, Powdery, Rust) with CNNs, transforming research into an automated ML pipeline for deployment and lifecycle management.

Twitter Sentiment Analysis for Financial Markets

https://github.com/Jakee4488/Twitter_Sentiment_Analysis

Created a Flask app for real-time tweet sentiment analysis from top financial influencers using NLP, to track market sentiment, spot trends, and display financial opinions via a dashboard.

CERTIFICATIONS

CCNA: Enterprise Networking, Security, and Automation

CCNA: Switching, Routing, and Wireless Essentials

Introduction to LangGraph: AI Agentic workflow with LangGraph

Advanced Machine Learning with Databricks

Advanced Machine Learning Operations with Databricks

Learning Kubernetes: Kubernetes for container autoscaling

Learning Docker: Docker for containerization

Generative AI: Working with Large Language Models

EDUCATION



B.Sc. Computer Science with AI

Coventry University

2021 - 2024

Coventry

• Modules: Artificial Neural Networks & Intelligent Agents, Machine Learning and Related Applications, Advanced Algorithms & Mathematics for CS, Big Data Programming, Security, Theory of Computation, Software Engineering.

• Dissertation: Contextual Intelligence in LLMs via RAG Framework.

• Engineered an end-to-end Retrieval-Augmented Generation (RAG) pipeline using OpenAI and Meta LLMs with vector databases to enable efficient semantic search.

• Optimised prompt engineering and document chunking strategies to improve the accuracy of context-aware information retrieval.

• Evaluated performance trade-offs between proprietary (GPT) and open-source (Llama) models across real-world query scenarios, informing model selection based on performance, cost, and latency.

Grade

First Class / 4.0

LEADERSHIP & COMMUNITY



Currently leading the implementation of the MLOps pipeline for machine learning models at Coventry Building Society, while delivering Proof-of-Value initiatives for the Data Intelligence Platform



Led software development projects in Python and C++, mentoring a team of five members and delivering solutions 15% ahead of schedule for University Group Projects



Collaborated as a member of the Formula Student AI (FSAI) team, contributing to autonomous racing software development, technical documentation, simulation testing, and on-track validation



Served as Lead Team Member of the Google Student Developer Club, organising three technical workshops attended by 100+ students

KEY ACHIEVEMENTS



Strengthened Model Reliability

Cut deployment failures by and improved model versioning consistency through stronger MLOps controls.



Enhanced Data Automation

Achieved increase in analytics throughput by optimizing AWS Glue pipelines.



Optimised IT Workflows

Achieved a 100% incident resolution rate by optimizing Active Directory processes, ensuring seamless integration and operational reliability essential for maintaining machine learning systems in production environments.



Automated Software Provisioning

Saved 30 man-hours per cycle with automated Windows deployment pipeline.

LANGUAGES

English

Proficient

Malayakam

Native

Hindi

Intermediate

Tamil

Intermediate

SKILLS

AI & GenAI

PyTorch

TensorFlow

Keras

Scikit-learn

OpenCV

Hugging Face Transformer

LlamaIndex

LangChain

CewAI

LangGraph

RAG Architectures

LLMs (OpenAI, Claude.Llama)

Open Claw

FAISS

Pinecone

Google ADK

MCP

Vertex AI

BigQuery

MLOps & Cloud

AWS

Azure

GCP

Docker

Kubernetes

Azure DevOps

CI/CD Pipelines

Terraform

Databricks

MLflow

Git

Data Engineering

PySpark

Pandas

NumPy

Kafka

Redis

MongoDB

MySQL

SQLite

AWS Glue

SAS EG

SAS DI

SAS MG

Airflow

Glue

Deep Learning

CNNs

RNNs

LSTMs

CUDA/cuDNN

Transformers and other architectures

AI development Tools

VS Code

Claude Code

Anti-Gravity CLI

Cursor

Replit

Lovable

Full Stack Development

React

Flask

Django

HTML5

CSS3

RESTful APIs

FastAPI

Operating Systems

Linux (Ubuntu/CentOS)

Windows

macOS

Programming Languages

Python

SQL

C++

JavaScript

Bash

SAS

YAML