Johnny Jonathan Jeyakumar

Wembley, London, UK

📱 +44 7378 650219 • ✉ johnnyjonathan008@gmail.com

🌐 https://www.zoomi.ca/portfolio • **in** johnny941 • ⓞ jjjonathan96

Profile

Machine Learning Engineer with over **2+ years** of expertise in **Large Language Models (LLMs), Retrieval-Augmented Generation (RAG), and Deep Learning**. Experienced in **computer vision, recommendation systems, and real-time inference optimization**. Skilled in **building scalable ML pipelines, CI/CD, optimizing models for deployment, and processing large datasets**. Passionate about AI research and deploying state-of-the-art models in production environments.

Experience

Zoomi Technologies, Canada

Machine Learning Engineer *Apr 2022 – Oct 2023*

- Developed scalable AI solutions for **LLM, RAG, AI Agent, Computer Vision, XGBoost** leveraging **GPT2, LLAMA**, YOLOv8, and UNet, improving model metrics by **15%**.
- Architected and optimized cloud-based AI pipelines in **AWS** (EC2, SageMaker, Lambda, S3), reducing training runtime by **10%** and improving inference speed by **35%**.
- Led a team, Designed, trained, and deployed AI models for self-checkout systems (Amazon Go style) using YOLOv5 and real-time tracking, increasing object detection accuracy (F1-score **+20%**).
- Developed and deployed end-to-end MLOps workflows, utilizing Docker, Kubernetes, MLFlow, and CI/CD pipelines, improving model reliability and versioning.
- Built **multimodal** Foundation models, integrating vision-language models (CLIP, Transformers) for product recommendation, sentiment analysis, and personalized **LLM**-driven marketing.

Effective Solution, Sri Lanka

Associate Electronic Engineer Intern *May 2021 – Dec 2021*

- Designed and developed AI-powered recommendation models for e-commerce and agriculture, optimizing classification accuracy and increasing CTR (Click-Through Rate) by **10%**.
- Implemented **LLM-based** text processing pipelines for automated customer sentiment analysis, improving email marketing engagement.
- Developed and deployed AI APIs using Django and JavaScript, optimizing backend performance by **25%** and enhancing customer experience by **15%**.
- Integrated Segment Anything Model (SAM) and computer vision pipelines, reducing manual annotation time and improving dataset quality.
- Managed large-scale datasets , applying PCA, UMAP, and t-SNE for feature engineering, enhancing model interpretability by **10%**.

Education

Teesside University London **Distinction**

Master's in Data Science *2024*

University of Moratuwa, Sri Lanka

BSc (Eng) in Electronic and Telecommunication Engineering *2017–2022*

Technical Skills

Machine Learning: LLMs (GPT), CLIP, BERT, Transformers, YOLO, UNet, ViT, XGBoost, LightGBM, CatBoost

Big Data & MLOps: PySpark, MLFlow, CI/CD, Docker, Apache Airflow

Cloud: AWS (SageMaker, S3, EC2, Lambda, RDS), Docker, Kubernetes, GCP, Azure

Languages: Python, JavaScript, Bash

Databases: MySQL, MongoDB, Vector DB

Frameworks: PyTorch, TensorFlow, Scikit-learn, Hugging Face, FastAPI, Flask, OpenCV

Projects

Multimodal E-Commerce LLM Recommendation Engine

Python, PyTorch, LLM, NLP   *2024*

○ Built a multimodal recommendation system integrating LLMs (GPT, BERT), vision models (CLIP), and temporal embeddings, improving personalization efficiency by **18**%.

○ Leveraged Transformer-based sequence modeling to enhance user-item interaction predictions, improving recommendation accuracy and click-through rates.

○ Applied LLM-powered text analysis on Amazon review dataset, extracting sentiment, trends, and user intent for context-aware recommendations.

○ Conducted seasonality-based product trend analysis using time-series forecasting, increasing engagement and retention for personalized promotions.

○ Optimized fine-tuning of pre-trained LLMs with LoRA (Low-Rank Adaptation) for efficient training and inference, reducing computational cost by **25**%.

AI-Powered Retail Checkout

Python, OpenCV, TensorFlow, Jetson Xavier   *2023*

○ Developed a 3D perception system for checkout-free shopping using computer vision and AI, integrating multiple cameras for real-time customer tracking.

○ Leveraged NVIDIA Jetson Xavier for efficient processing and optimized object detection using YOLO models, enhancing tracking accuracy by **15**%.

○ Applied LLM-powered text analysis on Amazon review dataset, extracting sentiment, trends, and user intent for context-aware recommendations.

○ Conducted seasonality-based product trend analysis using time-series forecasting, increasing engagement and retention for personalized promotions.

○ Pioneered AI agent development for autonomous machine perception, integrating real-time AI inference with IoT edge devices (Jetson Xavier).

Certifications

Amazon Bedrock: Getting Started with Generative AI (Coursera)

AWS: Foundations of Prompt Engineering (Coursera)

LangChain: LLM Application Development (Coursera)

RAG: Building Multimodal Search and RAG (Coursera)

Awards

2024: Best Overall Project in Computing - Teesside University London

2020: HackX - 2nd Place — SLIoT, Brain Strom, IEEE Innovation Nation - 1st place — SL EXSTO - Merit award — Hackstat 2.0, Yarl Geek – participate

2017: Mahapola Scholarship for best result in G.C.E Advance Level Mathematic Stream.

2016: Gold Medal for Entering University of Moratuwa for Engineering.

Leadership

2024: Course Representative, MSc Data Science, Teesside University London

2020: Led a team in a Biomedical ML competition

Availability

Eligible to work full-time in the UK (No visa sponsorship required, Post-Study Work visa) Expires 2027.