

Om Singh

LLM / AI Engineer | Data Scientist | Building Intelligent Learning Systems

Bengaluru, India | +91-7977019708 | ossingh6@gmail.com | linkedin.com/in/omsingh06

SUMMARY

AI Engineer specializing in **LLMs** and intelligent learning systems. Proven expertise in **fine-tuning models**, architecting **high-performance RAG pipelines**, and building **evaluation frameworks** for educational AI. Focused on transforming data into adaptive, context-aware learning experiences that drive measurable impact.

EXPERIENCE

ALLEN Digital — *Data Scientist I*

Apr 2024 – Present | Bengaluru

- **Spearheaded** fine-tuning using **SFT and DPO** techniques for academic doubt classification, **boosting accuracy by 22%** through cross-functional collaboration with content and product teams.
- **Architected** a scalable **RAG pipeline** leveraging **hybrid search and re-ranking**, optimized for high concurrency to support 50K+ daily student interactions.
- **Led** development of multi-agent LLM workflows using **LangGraph** with prompt evaluation and reinforcement loops, **improving response quality by 30%**.
- **Revamped** AnswerGenerationService to agentic architecture using **Google ADK**, enabling modular tool orchestration and **adaptive task** planning across teams.
- **Designed** automated LLM evaluation framework benchmarking **faithfulness and relevancy** for A/B testing, **reducing iteration cycles by 40%**.
- **Implemented** end-to-end observability using **OpenTelemetry**, enabling real-time distributed tracing, latency monitoring, and performance debugging across the AI pipeline.
- **Extended** PromptBreeder research paper to build an automatic prompt generation system, scaling **evolutionary optimization** use cases org-wide.

ALLEN Digital — *Data Science Intern*

Oct 2023 – Mar 2024 | Bengaluru

- **Partnered** with engineering teams to design and deploy **FastAPI-based** microservices and data pipelines for real-time analytics and model integration.
- **Built** graph-based learning paths using **Neo4J and Cypher**, collaborating with content teams to optimize knowledge graph retrievals.
- **Automated** data preprocessing workflows with data teams, **reducing preparation time by 35%** and enhancing dataset integrity for downstream training.

PROJECTS

CardioGuard — *Heart Disease Classification App*

- **Engineered** end-to-end ML pipeline using **XGBoost and scikit-learn**, achieving **96% accuracy** and **0.98 ROC-AUC** on UCI dataset.
- **Designed** clinician-facing UI with **Streamlit**, reducing data entry time by 30% and achieving 95% user satisfaction.

CataractDetectX — *Deep Learning for Eye Health*

- **Developed** CNN-based cataract detection model using VGG19, achieving **95% accuracy** on 10,000+ eye images.
- **Implemented** image augmentation pipelines, improving model generalization by 10%.

EDUCATION

Vellore Institute of Technology — *B.Tech in Computer Science*

2020 – 2024 | CGPA: 8.2/10

University College London (Summer School) — *Data Science & Big Data Analytics*

2023 | Grade: A

SKILLS

Core: Machine Learning, Deep Learning, NLP, LLM Fine-Tuning, RAG, Multi-Agent Systems | **Stack:** Python, TensorFlow, PyTorch, Transformers.js, Hugging Face, LangChain, LangGraph, Google ADK, CrewAI, Neo4J, FastAPI, OpenTelemetry | **Data Science:** Feature Engineering, Model Evaluation, Hypothesis Testing, Tableau | **Deployment:** Git, REST APIs, Docker, AWS, GCP, MLOps, CI/CD