

Robin K Philip

AI/ML Engineer - Immediate Joiner

 nefro313 |  robinkphilip |  robinkphilip2001@gmail.com |  +918593833872 |  Kerala, India

SUMMARY

AI/ML Engineer with **2 years of experience** building and deploying **real-world AI products in production**. Experienced in designing **Agentic AI workflows** and **RAG pipelines** using **LlamaIndex**, **LangGraph** and **Google ADK**. Implemented **MLOps workflows** with **MLflow** and **DVC**. Proficient in **Python** for model development and unit testing. Thrives in collaborative, fast-paced environments. [View Projects](#).

WORK EXPERIENCE

AI/ML Engineer, StarXAI Technology LLP

March 2025 – April 2026

- Designed and deployed **agentic chatbot** solutions for a blockchain-focused AI product, integrating **MongoDB** for dynamic user data retrieval and personalization.
- Built **confidence-scoring models** for cryptocurrency news analysis and a **CAGR-based crypto-trading backtesting engine**, enabling data-driven strategy evaluation.
- Implemented **advanced web-search augmentation** to ground agent responses in real-time information, improving analysis reliability beyond static training data.
- Streamlined **CI/CD deployment pipelines** across multiple projects and strengthened DevOps practices, improving release cycles and production system scalability.

AI Full-Stack Developer, Jardyan Inc (Remote)

March 2024 – February 2025

- Led full-stack development and deployment of **AIpublisher.org**, an AI-powered journal publishing platform built on **React.js**, **Node.js**, and **PostgreSQL**.
- Built an **AI agent-powered manuscript screening pipeline** automating quality scoring, scope matching, formatting checks, and keyword recommendations – accelerating editorial review cycles.
- Developed end-to-end platform workflows spanning article submission, peer review, publication management, secure payments, and admin controls.

PROJECTS

CorpusPilot – Multi-Domain RAG Platform for Grounded Docs Q&A [Live Demo](#) | [GitHub](#)

Tech Stack: FastAPI, React, LangGraph, Milvus, LlamaIndex, PostgreSQL, Cohere Rerank, RAGAS, LangSmith, Prometheus

- Architected a multi-domain **RAG** platform (legal, financial, healthcare, research, technical) combining hybrid semantic + lexical retrieval with domain-weighted **Reciprocal Rank Fusion** and **Cohere reranking** for citation-grounded, hallucination-resistant answers.
- Built a **SQL-RAG** pipeline that extracts structured tables from financial filings and generates safety-filtered SQL queries, delivering higher-precision numeric answers than prose-only retrieval.
- Engineered a **RAGAS-based evaluation gate wired into CI**, automatically blocking merges that regress faithfulness, factual correctness, or context recall – turning LLM output quality into a hard release gate, not a manual check.
- Instrumented full production observability (**Prometheus**, **LangSmith**) with token/cost tracking and automated anomaly detection across every query type.

- Architected an **11-agent LLM orchestration pipeline** (Google ADK + LiteLLM) that parses resumes and job postings, performs gap analysis, and auto-generates ATS-optimized, tailored application materials end-to-end.
- Designed **anti-hallucination guardrails** restricting the resume-tailoring agent to keywords verifiably supported by the candidate’s real experience, eliminating fabricated skills while cutting tailoring time from hours to minutes.
- Built a **hybrid deterministic + LLM ATS scoring system**, using the deterministic scorer to drive an automated rewrite feedback loop and the LLM scorer for user-facing before/after comparisons.
- Implemented cost-aware **tiered model routing** (Groq primary, OpenAI fallback) with full **LangSmith** tracing across 10+ sequential LLM calls per request for production-grade reliability.

SKILLS

Languages	Python, Javascript, R
AI Frameworks	LangChain, LlamaIndex, LangGraph, Google ADK, Pinecone
ML Libraries	NumPy, Pandas, Scikit-learn, XGBoost
DL Frameworks	TensorFlow, PyTorch, Keras
Visualization	Matplotlib, Seaborn
Cloud	AWS (ECR, S3, SageMaker, Bedrock, ECS/EKS), GCP (Vertex AI, BigQuery)
MLOps	MLflow, DVC, GitHub Actions, Docker, Nginx, CI/CD
Dev Tools	Visual Studio Code, Git, GitHub

EDUCATION

2020 – 2023 Bachelor’s in Computer Application with AI Specialization at **BNU**

Activities & Highlights:

- Organized and conducted **college-level hackathons**, managing end-to-end event coordination and participant mentorship.
- Active participant in the **Google Developer Student Program**, representing the college chapter and conducting developer sessions.
- Built hands-on skills in **MEAN Stack** web development (MongoDB, Express.js, Angular, Node.js) through self-directed projects.