Rahul Chauhan Generative AI & Full Stack Engineer

✉ rahul_c@me.iitr.ac.in ☎ +91-9410959369 New Delhi , India GitHub

LinkedIn X

PROFESSIONAL EXPERIENCE

Artificial Intelligence Intern - Piermind AI, Remote | (Aug 2025 - Oct 2025)

- Built and deployed **LangGraph pipelines** for multi-tenant agentic workflows.
- Collaborated with product managers on **AI integration and workflow optimization**.
- Developed **backend modules** in Python, Go, and LangChain for scalable agent deployment.

AI Engineering Intern - Nitrolens - AI(Silicon Valley-based Startup), Remote | (May 2025 - July 2025)

- Built and tested **multi-agent AI co-pilot systems** using **LangGraph, CrewAI, and LangChain**.
- Contributed to **agent orchestration and inter-agent communication** for strategy automation.
- Worked directly with founders in a **fast-paced, prototype-driven environment** to ship core features.

Tech Intern-Profcess (now kokoro.doctor), Remote | (May 2024 - July 2024)

- Shipped production features across **React, Django, and SQL**, improving usability and stability.
- Designed and optimized **REST APIs**, ensuring smooth data integration.
- Worked in a **lean startup environment**, delivering quickly under resource constraints.

PROJECTS

Open-Source Contributions — Ragas, Evaluation Framework for RAG Applications

- Fixed **caching and embedding issues**, improving reliability and performance across modules.
- Enhanced **documentation and metric tracing clarity**, ensuring smoother developer experience.
- Gained practical experience with **LangSmith integrations** and RAG evaluation workflows.

Codebase Copilot, AI-Powered Codebase Query & Documentation System (In progress)

Python · FastAPI · ChromaDB · LangChain · OpenAI APIs · Next.js

- Built a **RAG-based system** to query GitHub repos in natural language with embeddings and LLM reasoning.
- Developed backend modules for **repo ingestion, chunking, retrieval**, and **AI-powered code summaries**.
- Designed a **Next.js frontend** for chat-based interaction and auto-generated documentation.

RAG-EngineX, Modular Retrieval-Augmented Generation Framework

Python · LangChain · Pinecone · FAISS · LLMs

- Built a **professional-grade RAG pipeline** with modular components for loading, chunking, and evaluation.
- Integrated **reranking, LangSmith observability**, and **faithfulness/relevance scoring** for precise performance .
- Delivered an **interactive Streamlit UI** with configurable workflows and exportable evaluation results.

Auto-Researcher, LLM-Powered Research Assistant

Python · LangChain Agents · Web Scraping

- Built agent for **literature search, summarization, and citation extraction** from web sources.
- Integrated search, LLM summarization, and structured **PDF/Markdown report generation**.
- Showcased autonomous workflows that reduce **manual research time significantly**.

TinyLlama + LoRA, Low-Resource Fine-Tuning Proof-of-Concept

PyTorch · HuggingFace · PEFT (LoRA)

- Fine-tuned **TinyLlama-1.1B-Chat** using **LoRA adapters (~0.2% parameters)** on a 1k Alpaca subset.
- Demonstrated **low-resource LLM fine-tuning feasibility** on **Mac M2** and **Colab T4** hardware setups.
- Achieved **stable training and qualitative response improvement** with minimal compute requirements.

Reverse Supply Chain Optimizer, End-to-End ML Pipeline for Returns Management

Python · XGBoost · scikit-learn · Pandas · Optimization Heuristics

- Built a modular ML pipeline to **predict product returns, estimate costs**, and **optimize reverse logistics**.
- Developed **return-center selection and inventory decision modules** using rule-based and heuristic logic.
- Integrated all phases into a unified **Jupyter Notebook workflow** with reusable .pkl models and visual insights.

EDUCATION

Bachelor of Technology, Indian Institute of Technology , Roorkee — 2023 – 2027, India

Research Interests: *LLM Fine-Tuning · Retrieval-Augmented Generation (RAG) Evaluation · Multilingual Language Models*