Atharva Ajit Kate

Pune, India
+91 9404275801 | atharvakate25@gmail.com | https://www.linkedin.com/in/atharva-kate2001
https://github.com/AtharvaKate2001 | https://atharva-kate.netlify.app

Aspiring AI Engineer

AI Engineer with 1.5+ years of experience developing and deploying Generative AI solutions. Skilled in Large Language Models, RAG systems, multi agent workflows, workflow automation, and Python development. At NVIDIA, resolved 450+ prompt and data quality issues, improving model accuracy by 25% and reliability by 30%. Experienced in building AI applications using LangChain, LangGraph, FastAPI, Docker, and end to end machine learning pipelines, with a strong interest in MLOps, agentic AI, and advanced LLM technologies.

Work Experience

Associate Prompt Engineer | NVIDIA (Payroll: Randstad) Aug 2024 – May 2025

- Engineered and optimized Python based LLM pipelines for production Generative AI applications.
- Resolved 450+ prompt and data quality issues using prompt engineering and instruction tuning, improving model precision by 25 percent and reliability by 30 percent.
- Developed evaluation and monitoring frameworks to detect prompt drift, data inconsistencies, and inference failures.
- Integrated Large Language Models with computer vision workflows including image segmentation, object detection, and human action recognition.
- Contributed to multi agent systems through tool routing, workflow orchestration, message passing, and state management.
- Led onboarding and quality assurance for 15+ AI evaluators and earned Top 3 Contributor recognition for three consecutive weeks

Data Analyst Intern | TUV SUD Pvt. Ltd. Jan 2024 – Jul 2024

- Built Python based data pipelines to clean, transform, and structure large scale government datasets, enabling reliable analytics and supporting end to end data engineering workflows.
- Developed machine learning models and interactive Power BI and Tableau dashboards for trend analysis, resource planning, and data driven decision making.

Skills

Python, FastAPI, SQL, OpenAI API, Anthropic Claude, Hugging Face Transformers, Ollama, Prompt Engineering, Instruction Tuning, LangChain, LangGraph, AutoGen, CrewAI, LlamaIndex, RAG, FAISS, Vector Search, LLM Reranking, Multi Agent Systems, n8n, REST APIs, Workflow Automation, Docker, MLflow, CI/CD, Model Evaluation, Monitoring, Linux, Kubernetes, Scikit learn, Machine Learning, Feature Engineering, EDA, Power BI, Tableau, Excel, Data Visualization.

Personal Projects

ClinIQ: Autonomous Multi-Agent Clinical Intelligence Platform | GitHub **|** *Python, LangGraph, FastAPI, ChromaDB, Groq, Ollama*

- Designed and deployed a clinical intelligence GenAI application using LangGraph with a stateful multi agent architecture, conditional routing, workflow checkpointing, and specialized agents for intake, knowledge retrieval, risk assessment, and report generation.
- Implemented hybrid BM25 and semantic vector retrieval with Reciprocal Rank Fusion, and developed a risk scoring engine combining rule based drug interaction checks with LLM driven clinical reasoning for accurate analysis.
- Integrated Langfuse for LLM observability, including token usage, latency, cost tracking, and prompt versioning, and deployed the solution using FastAPI, Docker, and Streamlit with support for multiple LLM providers.

Finance Intelligence Agent | GitHub | *Python, LangChain, FAISS, FastAPI, Docker, Ollama*

- Built a production-ready GenAI application integrating OpenAI API with LangChain, featuring RAG architecture with FAISS vector search, LLM reranking, and agentic function calling for multi-step financial intelligence reasoning.
- Implemented observability layers tracking response quality, hallucination risk scoring, and token utilization, demonstrating end-to-end ML lifecycle thinking from problem definition through production deployment and monitoring.

Sentiment Analysis on Public Data | GitHub | *Python, NLTK, TF-IDF, YouTube Data API, Multilingual Translation, NLP*

- Applied NLP and machine learning techniques including TF IDF, multilingual text processing, tokenization, and sentiment classification to analyze over 900 YouTube comments and extract actionable insights from unstructured data.
- Converted user generated content into quantified sentiment trends, demonstrating expertise in text analytics, document classification, and customer feedback analysis for data driven decision making.

EDUCATION

Imarticus Learning, Pune 2023 – 2024
Post Graduate Diploma in Data Science and Analytics | Grade A, 89%

Pimpri Chinchwad College of Engineering, Nigdi 2019 – 2023
Bachelor of Engineering in Mechanical Engineering | CGPA 8.46