Vivekkumar Vishwakarma

AI/ML Engineer

+91 9328797168 • vivekvishwakarma21@outlook.com • Ahmedabad, Gujarat, India •
linkedin.com/in/viveks21 • github.com/vivekkumars21 • vivekvishwakarma.dev

Professional Summary

Final-year Information Technology student specializing in Generative AI and NLP. Built and deployed RAG pipelines and multi-agent systems using Python, LangChain, and FastAPI. Seeking an AI/ML Engineer role to build production-grade LLM applications.

Technical Skills

Programming Languages: Python, SQL, JavaScript
Machine Learning & Deep Learning: Scikit-Learn, PyTorch, TensorFlow, Keras, Model Evaluation, Feature Engineering
Generative AI & NLP: LLMs, Transformers, RAG, Embeddings, Vector Databases, Hugging Face, LangChain, AI Agents
AI Engineering & Backend: FastAPI, REST APIs, Pydantic, Model Serving, API Integration
Data & Database: NumPy, Pandas, PostgreSQL
Tools & Deployment: Git, GitHub, Docker, Linux, AWS, Postman

Education

SAL Engineering and Technical Institute, Ahmedabad 2023 - 2027
B.E. in Information Technology CGPA: 7.7/10 (through 6th semester)

- Relevant coursework: AI/ML, Data Science, DBMS, OOP, Computer Networks.

Projects

Research Lab : Multi-Agent AI Research System

- Designed a 4-agent pipeline (Search, Reader, Writer, Critic) using LangChain and Google Gemini to automate research report generation.
- Integrated Tavily search and BeautifulSoup scraping, with a Critic agent that scores source credibility before report generation.
- Exposed the system through FastAPI REST endpoints with Pydantic validation; reduced manual research time from ~15 minutes to ~3 minutes
- Tech Stack: Python, LangChain, Google Gemini API, Tavily, BeautifulSoup, FastAPI.

RAG-Based Document Question Answering System

- Built an end-to-end RAG pipeline for PDF Q&A using LangChain, Mistral AI, and ChromaDB with Hugging Face embeddings.
- Applied MMR retrieval to reduce redundant context and improve answer relevance; tested on 10 documents with 83% answer accuracy.
- Achieved average response time of 4 seconds per query.
- Tech Stack: Python, LangChain, Mistral AI, Hugging Face, Sentence Transformers, ChromaDB.

LSTM-Based Quote Generation System

- Trained an LSTM model on 8,979 words dataset achieving 74% next-word accuracy; deployed via Streamlit.

Leadership Experience

Technical Head — ISTE Student Chapter SAL | Sept 2024 – Present

- Led technical and media teams to coordinate 50+ events for 800+ participants, managing logistics and scheduling.