Kapil

Dehradun, Uttarakhand, India
+91-9996547555 ky8050989@gmail.com LinkedIn Github

SUMMARY

I'm a Computer Science student with strong experience in building production-ready GenAI applications using LangChain, FastAPI, and Docker. Specialized in LLM fine-tuning, RAG pipelines, and full-stack AI deployment using Kubernetes and Terraform.

EDUCATION

University of Petroleum and Energy Studies **2023 – 2027**
B.Tech in Computer Science and Engineering — CGPA: 7.91/10 *Dehradun, Uttarakhand*

COURSEWORK

- Data Structures
- Machine Learning
- Generative AI
- FastAPI
- RAG
- Transformers
- NLP
- DevOps
- Computer Vision
- Deep Learning
- LangChain
- LLM

EXPERIENCE

Weazy **June 2025 – Present**
AI Intern *Remote*

- Reduced LLM response latency by 30% using request batching
- Designed and documented RESTful APIs using **Postman** and **Swagger**, improving dev onboarding.
- Integrated chat agents into SaaS dashboard with frontend team using WebSockets.

OPEN Community **Feb 2025 – Present**
Student Developer *Dehradun, Uttarakhand*

- Contributed to open-source Python and JavaScript projects: features, bug fixes, documentation.
- Led knowledge-sharing sessions on Flutter and Git workflows for onboarding contributors.
- Participated in sprints and peer-reviewed community contributions.

PROJECTS

RAG Based Assistant **LangChain, FastAPI, Docker, Kubernetes, FAISS**

- Engineered a scalable, containerized Retrieval-Augmented Generation application using LangChain, FastAPI, and FAISS vector database supporting 6+ concurrent users.
- Designed a multi-stage query pipeline prioritizing local knowledge base, with **LLM Fallback** to Wikipedia API and LLM generation, boosting answer accuracy by 35%.
- Optimized deployment workflows using **Docker** and **Kubernetes**, reducing setup time by 40% and enhancing maintainability across environments.

DeepFake Detection Model **Deep Learning, Xception CNN, Streamlit**

- Engineered a deepfake video classification system using a fine-tuned Xception CNN model on curated datasets.
- Achieved an accuracy boost from 70% to 83% by applying **dropout layers**, **early stopping**, and image augmentation
- Explored model explainability through saliency maps to identify manipulated regions in fake frames.

TECHNICAL SKILLS

Languages: Python, C/C++, Java, JavaScript, SQL, HTML5, CSS3
ML/DL: TensorFlow, scikit-learn, OpenCV, Streamlit, Xception CNN, NLTK
Frameworks: LangChain, FastAPI, Flask, Pydantic
DevOps: Docker, Kubernetes, Git, GitHub Actions, GitLab CI/CD, Jenkins, Ansible, Terraform, AWS SageMaker
Tools: Postman, Swagger, Jira, Vector DB, VSCode
Databases: PostgreSQL, MongoDB , MySQL

CERTIFICATIONS

- Introduction to Generative AI – Coursera
- Introduction to Data Science – Cisco Networking Academy

ACHEIVEMENTS

- Contributing to open-source projects as a developer in Social Summer of Code (June 2025 - August 2025), enhancing features, fixing bugs, and improving documentation.