Abhishek Harkar

+91 9373581139 | abhishekharkar@gmail.com | https://www.linkedin.com/in/abhishek-harkar-595647276 | https://github.com/AKIS-4

EDUCATION

SPPU University   Baramati, Pune
Bachelor of Engineering in Computer Science   *Dec 2021 - May 2025*

EXPERIENCE

GenAI Full Stack Intern   Oct 2025 – Present
GenAIKit Software Solution Private Limited   *Remote*
- Built and shipped Generative AI applications using FastAPI, LangChain, LangGraph, and Azure OpenAI, improving overall project turnaround time by 25% through efficient system design and automation.
- Designed and deployed RAG pipelines, ML microservices, and real-time AI chatbots, reducing client manual workload by 40% through intelligent automation and streamlined knowledge retrieval.

DevOps Intern   July 2025 – Aug 2025
PearlThoughts   *Remote - **Certificate***
- Contributed to the deployment and management of client projects, gaining hands-on exposure to production workflows, release processes, and cloud-based application delivery.
- Deployed Strapi and Medusa headless backends using Terraform and GitHub Actions, and managed infrastructure through AWS CodeDeploy and CloudWatch, ensuring reliable CI/CD and system monitoring.

PROJECTS

AI Interview Platform | *FastAPI, Next.js, LiveKit, Azure OpenAI*   **Link**
- Built a real-time voice-to-voice interview assistant that uses RAG over candidate resumes to deliver context-aware guidance and responses during interviews.
- Reduced recruiter preparation time by 40% by automating resume analysis, summarization, and real-time conversational support.

Live Video Streaming Platform | *EC2, S3, Route53, ReactJs, NodeJs, Nginx Rtmp*   **Link**
- Developed a live video streaming platform using the MERN stack and deployed it on AWS EC2 with Nginx RTMP, enabling scalable real-time video delivery and user interaction.
- Architected and configured RTMP ingest, FFmpeg-based transcoding, and HLS segmentation servers, implementing an efficient end-to-end video processing pipeline optimized for low latency and multi-resolution streaming.

LLM Fine-tuning – Healthcare Summarizer | *Mistral 7B, LoRA, FastAPI, Gradio*
- Fine-tuned the Mistral 7B model on 10k+ healthcare documents to build a domain-specific summarization system, improving accuracy by 25% through targeted dataset curation and optimized training workflows.

TECHNICAL SKILLS

Languages: Python, C/C++, SQL (Postgres), MongoDB, JavaScript, HTML/CSS
Frameworks: ReactJs, NodeJs, FastAPI, LangChain, LangGraph, Autogen
DevOps Tools: Git, Docker, AWS, Azure, Github Actions(CI/CD), Terraform, Kubernetes, Linux
Libraries: pandas, NumPy, Matplotlib, Scikit-learn
AI and Deployment: Azure OpenAI, Hugging Face, Ollama, Unsloth, Retrieval-Augmented Generation (RAG), LoRA Fine-tuning, LLMOps, Prompt Engineering, MCP, Vector Databases (Chroma, FAISS)

ACHIEVEMENTS AND CERTIFICATIONS

- ***LeetCode*** — 1975 (Knight), ***CodeChef*** — 1778 (3∗)
- ***AWS Certified Cloud Practitioner***