Apurv P Chudasama

Ahmedabad, India
+91 9409339698
apurvchudasama.edu@gmail.com

GitHub · LinkedIn · Portfolio

Skills

Programming Languages: Python, C++

Generative AI & LLMs: Prompt Engineering, LLM FineTuning, RAG pipelines, LangGraph, LangChain, LangFuse, MultiAgentSystem

Computer Vision: Object detection, Image Classification, YOLO, OpenCV, Segmentation

Machine Learning & Deep Learning: Classification, Regression, Random Forest, SVM, Transformers, CNNs, RNNs, LSTMs, XGBoost, AdaBoost, NaiveBayes

CS Fundamentals: Data Structures and Algorithms, Computer Networking, Mathematics

Tools & APIs: FastAPI, REST APIs, WebSockets, Server-Sent Events (SSE), Docker, Git, Linux, LanceDB, FAISS, PostgreSQL, JetsonNano, ArduinoUno

Soft Skills: Navigating challenges, bringing people together, flexible thinking, thoughtful communication, creating meaningful social connections.

Experience

Junior AI Engineer — Codiste Pvt Ltd — June 2025 – Present

- Designed a production grade AI agent using LangGraph and RAG pipelines for dynamic context retrieval.
- Implemented persistent state management and conversational memory using PostgreSQL.
- Handles low-latency real-time AI voice Agent (STT - LLM - TTS) handling 1K+ calls with a sub-700ms end-to-end response time.
- Handled real-time telephony integration (Twilio, Telnyx) using WebSockets to ensure continuous voice streaming achieving 99.5% uptime..
- Collaborated with product, frontend, and infra teams to design scalable agent workflows and production-ready APIs.

Lead — AI/ML Club — June 2024 – June 2025

- **Coordinated 3+ technical workshops and 2 hands-on coding sessions**, gained experience in planning and organising while ensuring smooth coordination of teamwork.

Chair Person — IEEE Computer Society Student Chapter — September 2024 – April 2025

- Enhanced ability to organise events through chapter activities and increased member participation by 15%.

Key Projects

Dialora.ai

- Handling a low-latency AI calling platform integrating STT, LLM reasoning, and TTS.
- Implemented real-time automatic language detection, enabling the agent to dynamically switch and converse in multiple languages mid-call.
- Designed dynamic, graph-based agent workflows with seamless tool execution for Google Calendar, Cal.com, and GoHighLevel.
- Integrated diverse LLMs (Gemini, Grok) and specialized speech models (Sarvam STT/TTS, Inworld TTS) with Elision telephony for robust call routing.
- Skills: AI Voice Agent, Telephony Integration, API Integration

PodGenix — Multi-Agent AI Podcast Generator

- Architected a multi-agent podcast generation platform using FastAPI, orchestrating the end-to-end pipeline from user instructions to final audio output.
- Designed a central orchestration engine to manage specialized AI agents for dynamic segment planning, detailed script writing, and multi-speaker dialogue partitioning.
- Integrated real-time data ingestion via News APIs to provide up-to-date context for AI-generated podcast content.
- Implemented a dedicated Text-to-Speech (TTS) service pipeline to synthesize partitioned scripts into natural-sounding, multi-speaker audio files.
- Technologies: FastAPI, Multi-Agent Orchestration, Prompt Engineering, TTS APIs, REST APIs, Python.

Theft Detection System Using NVIDIA Jetson

- Engineered a low-latency, real-time theft detection system deployed entirely on the edge using NVIDIA Jetson hardware.
- Architected a hybrid deep learning pipeline combining 3D-CNNs for spatial feature extraction and LSTMs for temporal sequence analysis across live video frames.
- Optimized model inference to maintain real-time performance, successfully reducing false positive alarms by 15% while triggering immediate security alerts.

Achievements

1. **Winner of DoseHack AI Hackathon** (1st Prize: 25,000 INR) — **Dosepeakers Meditab Group of Companies.**
2. **Top 12 Teams in Codesahastra'11 Hackathon** — **DJCSI, Computer Society of India.**
3. Conducted a **2-day workshop on NVIDIA Jetson Nano** across **5 colleges**, focusing on AI and edge computing, engaging **200+ participants**.
4. Conducted **10 hands-on peer-to-peer learning sessions** on ML and Deep Learning for the Department of CSE.

Publications

Lumbar Spine Degenerative Classification Model

- Architected a multi-modal YOLOv11x computer vision pipeline to classify degenerative lumbar spine conditions across complex MRI scans (Sagittal T1, T2/STIR, Axial T2).
- Achieved state-of-the-art diagnostic accuracy with an F1 score of 0.92 and mAP@0.5 of 0.925, leading to a publication in the AIJR Proceedings.

Entropy-Gated Retrieval for Clinical RAG

- Engineered a training-free retrieval architecture that uses Shannon entropy and contrastive decoding to filter redundant context and solve the "lost in the middle" phenomenon in LLMs.
- Reduced token consumption by 32.45% on the MedQA benchmark—significantly lowering latency and API costs—while hitting a peak Recall@K of 87.0% on BioASQ using Llama-3.

Education

CSPIT – CHARUSAT **June 2022 – Present**
Bachelor of Technology in Computer Science *CGPA: 9.64*

GSHSEB **May 2020 – May 2022**
HSC in Physics, Chemistry, & Mathematics *83%*

GSEB **May 2020**
SSC *89%*

Certifications

1. Machine Learning Specialization — Stanford Online
2. Introduction to Deep Learning — NVIDIA
3. Getting Started with AI on NVIDIA Jetson Nano — NVIDIA
4. Fundamentals of Red Hat Linux — Red Hat Academy
5. Database Management System — NPTEL
6. Machine Learning — NPTEL, IIT Madras
7. Deep Learning — NPTEL, IIT Ropar