AKSHAY K UPADHYAY

AI Engineer | LLM Applications & Backend Systems

+91 8887832311 akshay.cse23@gmail.com Gorakhpur, India · Open to on-site or remote
linkedin.com/in/akshay-kumar-upadhyay github.com/akshayds23 akshay-five.vercel.app

Summary

AI Engineer building production LLM applications. At WHIZROBO I integrate LLMs, face recognition and voice into service robots used by schools, and own the robots' configuration and licensing system. Independently built AutoApplier, a multi-user job-search platform with a rate-limit-aware multi-provider LLM pool and a serverless job queue, and a real-time computer-vision attendance system. B.S. in Data Science and Applications, IIT Madras.

Experience

AI/ML Engineer | WHIZROBO ***(AI robotics, EdTech)*** Dec 2025 - Present

- Integrate LLMs into production robot apps used by schools, a conversational assistant and an AI tutor, via OpenAI-compatible APIs.
- Work on the robots' perception and voice stack: separated face recognition from the camera pipeline and built reusable speech and camera packages shared across robot apps.
- Shipped a voice-guided AI admission kiosk used by schools on service robots: speech recognition with fuzzy answer matching and QR document upload, across **8 releases**.
- Main developer of the robot configuration and licensing system: server-driven settings with ETag caching, offline mode with safe rollback, per-customer feature control and de-duplicated usage telemetry (**110+ commits, 50 automated test files**).
- Built the licence and device admin portal with Next.js, Prisma and TypeScript.

Projects

AutoApplier - AI Job Search Platform | *Python, FastAPI, Next.js, PostgreSQL, LLMs*

- Multi-user platform that finds openings on company career sites (Greenhouse, Lever, Ashby) and job boards, scores fit with LLMs and tailors a resume and cover letter per job.
- Bring-your-own-key LLM pool across Groq, Gemini, OpenAI and Claude: key rotation, Retry-After-aware cooldowns, and deferred retries instead of blocking workers.
- Durable background-job queue with leases and time-boxed steps, so long searches run on serverless hosting (Vercel + Neon Postgres).
- Truthfulness guardrails that drop any generated resume line not grounded in the candidate's CV.

TruthLens - AI Fact-Checking Platform | *FastAPI, React, LLMs, RAG*

- Decomposes claims into verifiable sub-claims, researches them on the web and checks each against retrieved evidence with LLMs.
- RAG pipeline with vector embeddings and semantic search; fallbacks across LLM providers and search back-ends.

VisionAI Campus - Real-Time Person Identification | *Python, YOLOv11, InsightFace, OpenCV, PySide6*

- Desktop app that marks attendance and counts visitors from a live camera: YOLOv11 person detection, InsightFace face embeddings, centroid tracking with IN/OUT line crossing.
- Threaded inference pipeline processing every Nth frame to stay real-time on ordinary laptops.

Technical Skills

Languages: Python, SQL, TypeScript, JavaScript, Kotlin, Java (Android)
Applied AI: LLM APIs (OpenAI, Claude, Gemini, Groq), RAG, embeddings, vector search, prompt engineering, LangChain, document processing
Machine Learning: Pandas, NumPy, Hugging Face, feature engineering, model evaluation
Computer Vision: YOLO (Ultralytics), InsightFace, OpenCV
Backend & Data: FastAPI, async Python, REST APIs, Next.js, PostgreSQL, MySQL, MongoDB, Prisma ORM
Tools & Deployment: Docker, Traefik, Cloudflare, Linux VPS, WordPress, Git, GitHub, Vercel, CI/CD, Postman

Achievements

- Finalist (top 10 teams), Devbhoomi Cyber Hackathon 3.0 organised by Uttarakhand Police, Nov 2025.
- Built and deployed 6 public AI applications, all live online; 2,500+ GitHub contributions in the past year.

Education

Indian Institute of Technology Madras 2023 - 2027
B.S. in Data Science and Applications CGPA: 8.0
Diploma in Data Science and Diploma in Programming (2025)