

Jai Kumar

+91-7440826857 | jaikumar.dev03@gmail.com | github.com/jaikumar2406 | linkedin.com/in/jai-kumar2406

Professional Summary

AI/ML Engineer specializing in Agentic AI, LLM Systems, RAG Architectures, and MLOps. Built and deployed multiple production-grade AI applications — voice agents, phishing detection systems, and multi-agent orchestration platforms — actively used by real users.

Technical Skills

Languages: Python, SQL
LLM & Agentic AI: LangGraph, LangChain, RAG, MCP, Transformers, Prompt Engineering
Machine Learning & Deep Learning: PyTorch, TensorFlow, Scikit-learn, Pandas, NumPy
LLMOps & Evaluation: LangSmith, DeepEval, RAGAS, AI Guardrails
Computer Vision: YOLO, OpenCV, Vision Transformers, CNNs
Backend & APIs: FastAPI, REST APIs
Databases & Vector Stores: PostgreSQL, Pinecone, Redis
MLOps & Deployment: Docker, MLflow, CI/CD, Prometheus, Grafana
Tools: Git, Jupyter Notebook, N8N

Experience

- AI Voice Agent System** 2025
Freelance AI Developer | Client: AI Automation Labs
 - Built a scalable AI voice agent handling concurrent inbound/outbound calls using Deepgram (STT), LLaMA, and ElevenLabs (TTS), achieving ~700ms end-to-end response latency.
 - Tech Stack: Python, FastAPI, LLaMA, Deepgram, ElevenLabs, Twilio, Docker
- WhatsApp Ordering Automation** 2025
Freelance AI Developer | Client: Yaari Circle Restaurant
 - Developed a WhatsApp-based ordering system with LLM-driven conversational flow, integrated payment gateway and backend APIs, reducing manual order handling effort by 70%.
 - Tech Stack: Python, FastAPI, LLMs, WhatsApp Business API, Payment Gateway APIs
- Data Science Intern** Jul – Aug 2024
Evoastra Ventures OPC Pvt. Ltd.
 - Built an end-to-end CNN-LSTM image captioning pipeline on 8K+ images, achieving 20% improvement over baseline with automated preprocessing, training, and evaluation.
 - Tech Stack: Python, TensorFlow, Keras, OpenCV, Jupyter

Projects

- NexusFlow** | Multi-Agent LLM Orchestration Platform | GitHub May 2026
 - Built a production-grade multi-agent LLM orchestration platform using LangGraph with dynamic routing, self-reflection, confidence scoring, and RAG integration, improving answer accuracy by 28% and reducing hallucination rates by 35% during benchmark evaluations.
 - Developed a FastAPI + SSE streaming backend supporting real-time agent communication, evaluation pipelines, and automated performance monitoring for scalable deployment.
 - Tech Stack: Python, LangGraph, LangChain, FastAPI, Groq, PostgreSQL, Docker, RAG, Tavily
- ShieldSense** | AI Phishing Detection Platform | Live Demo Aug 2025
 - Processing 10K+ URLs/day at 96.2% accuracy using ensemble ML (XGBoost, Random Forest) with LLM verification and threat intel integration.
 - Live with 100+ users; deployed via Docker + CI/CD, monitored with Prometheus & Grafana, reducing API costs by 60%.
 - Tech Stack: Python, LangChain, FastAPI, XGBoost, Docker, PostgreSQL, Prometheus, Grafana
- ChatBookLLM** | Multi-Agent Document Intelligence Platform | GitHub June 2026
 - Built a multi-agent document intelligence platform using LangGraph that converts uploaded documents into grounded answers, PDF reports, PPT decks, quizzes, and mind maps via a RAG pipeline with cross-encoder reranking and source attribution.
 - Engineered a fault-tolerant LLM gateway with key rotation, exponential-backoff retries, and automatic Groq→Gemini fallback, cutting LLM calls by 67% and input tokens by 89% via combined-planner optimizations.
 - Tech Stack: Python, LangGraph, LangChain, FastAPI, RAG, ChromaDB, PostgreSQL, Redis, Docker, Kubernetes, CI/CD, LangSmith, AI Guardrails

Achievements & Leadership

Winner (1st) — Hackomania 2K25: Built VisionVoice, an AI-powered multi-agent voice assistant for visually impaired users, deployed to production.

Runner-Up (2nd) — HackBIOS 2K25: Developed an autonomous agent for phishing detection and a full-scale cyber threat intelligence ecosystem.

Runner-Up (2nd) — Algostrom 1.0, Amity University: Delivered ShieldSense, featured in the Kavach cybersecurity initiative.

Problem Statement Winner — Code Utsava 9.0: Built an AI-powered diagnostic ecosystem for dermatologists, enabling intelligent skin condition analysis.

3rd Place — Hack-Sprint 2026 (DCGC & GFG): Multimodal AI for acute clinical deterioration prediction using LSTM/Transformer + clinical notes.

3rd Place — 4th Semester Project Competition: Among 50+ AI/ML projects at department level.

Founder & Manager — BLAZE E-Sports: Launched and managed an e-sports gaming startup, growing the community and overseeing competitive operations end-to-end.

Organized and led technical events, increasing student participation in tech activities by 40%.

Education

Rungta College of Engineering and Technology

Expected: May 2027

B.Tech in Computer Science & Engineering (AI/ML)

- Coursework: Machine Learning, Deep Learning, Generative AI, NLP, Data Science