NIRANJAN HEBLI

niranjanhebli77@gmail.com | Ponda, Goa | LinkedIn | GitHub | CodeChef | Google Skill Boost

SUMMARY

Solutions-driven AI Engineer specializing in building scalable AI applications, Vision-Language Models, and intelligent automation workflows. Proven ability to translate complex business requirements into robust AI solutions, with a strong foundation in machine learning and scalable product development.

EXPERIENCE

Jr Software Developer 08/2024 – 08/2025
Creative Capsule Verna, Goa (Hybrid)

- **Architected and deployed a stateful conversational agent** on Azure Cloud, utilising LangGraph to orchestrate complex, multi-step workflows and manage continuous conversational memory.
- **Developed and integrated custom tools** into the agent state graph, expanding the system's autonomous capabilities and dynamic data retrieval.
- **Optimised LLM system prompts** and instituted strict object-oriented Python design patterns, reducing hallucination rates by 70% and maintaining 8/10 Pylint code quality scores across the deployment pipeline
- **Leveraged LangSmith** for system observability, monitoring agent trajectories, analysing execution traces, and optimising latency across the pipeline.

ORGANISATIONS

Google Developer Students Clubs 2022 – 2023
GDSC Lead Farmagudi, Goa
• **Mentorship & Training:** Designed and conducted hands-on training sessions on TensorFlow, Google Cloud, and Firebase, mentoring junior peers and guiding them in technical execution and best practices.
• **End-to-End Event Ownership:** Spearheaded the end-to-end delivery of a campus-wide hackathon, managing project timelines and driving cross-functional collaboration.

EDUCATION

Bachelor of Technology (BTECH) in Computer Science and Engineering 05/2020 – 05/2024
National Institute Of Technology, Goa Cuncolim, Ponda - Goa
CGPA : 8.16

Post Graduate Diploma in Agentic AI and ML 02/2026 – 07/2026
Indian Institute Of Technology - Gandhinagar Gandhinagar, Gujarat
CGPA : 10

Master Of Technology(MTECH) in Computer Science and Engineering 07/2026
Indian Institute Of Technology - Kharagpur Kharagpur, West Bengal

SKILLS

Technical Skills
Data Structures and Algorithms, C, C++, Java, Golang, Python, Machine Learning, RAG, LLM Finetuning, LangChain, LangGraph, Agentic AI, Multi-Agent System, CrewAI, AutoGen, MCP, NestJS, TypeScript, ReactJS, MySQL, Linux and Shell, RHEL, Git and GitHub, Docker, Jenkins, AWS, Google Cloud, Azure, PyTorch, OpenCV, YOLO, CNNs, Image Embeddings, and Vision-Language Models

PROJECTS

Operations Assistant 06/2026 – 06/2026

- **Architected a CrewAI multi-agent workflow** via MCP, autonomously synthesising operational data into structured reports 5 times faster than manual processes, processing 60+ documents per minute.
- **Engineered dual MCP servers (SSE & Stdio)** to expose custom execution tools for real-time document retrieval, database queries, and web scraping.
- **Implemented LLM observability** pipelines via OpenTelemetry and Langfuse, integrating a Human-in-the-Loop (HITL) gate to ensure factual accuracy.

Drive Thru Voice Assistant 05/2026 – 05/2026

- **Built a voice-first AI** ordering assistant for QSRs using Llama-3, Faster-Whisper, Edge-TTS, and FastAPI.
- **Implemented conversational memory** to track order history and maintain context across multiple turns.
- **Engineered deterministic LLM** tool-calling to strictly validate user requests against structured JSON menus.

NCERT based RAG System 04/2026 – 04/2026

- **Built a RAG-powered** AI study assistant using Python and Groq (Llama 3.1) for an ed-tech non-profit problem statement.
- **Implemented Hybrid Retrieval (BM25 & ChromaDB)** to accurately process noisy, real-world text inputs.
- **Engineered strict grounding prompts** to eliminate hallucinations and enforce textbook-only answers.
- **Automated evaluation pipelines** using Ragas, validating the system against 30 test queries to achieve a **1.0 faithfulness** score and a 0.98 answer relevancy metric.

- **Integrated a OCR workflow** to parse unstructured text from diagrams and scanned PDF pages, overcoming the limitations of pypdf4llm and ensuring comprehensive data ingestion for the RAG knowledge base.