Rohit Makani

FOUNDING AI ENGINEER · LLM SYSTEMS · MULTI-AGENT ARCHITECTURE

Bengaluru, India | [redacted] | [redacted] | port-folio-xar7.vercel.app | [redacted] | github.com/RohitMakaniProfile

SUMMARY

Founding AI Engineer building multi-agent LLM systems from 0→1. Shipped production AI to **10,000+ daily users** at PhysicsWallah and architected a 3-agent growth intelligence platform at Ikshan (DoableClaw) — covering product design, agent orchestration, and full deployment. Specialize in cost-optimized inference, RAG pipelines, and structured output engineering that delivers consulting-grade quality at scale.

CORE SKILLS

LLM & Agents	Claude API, GPT-4, Multi-Agent Orchestration, LangChain, LangGraph, Tool Calling, Prompt Engineering, Structured Output Validation
RAG Systems	FAISS, ChromaDB, Hybrid Search (Dense + BM25), Semantic Caching (Redis), RAGAS Evaluation, Reranking, Chunking Strategies
Backend & Infra	Python, FastAPI, Docker, Redis, PostgreSQL, Git, CI/CD, Azure OpenAI, Web Scraping (Playwright / BeautifulSoup)
ML / DL	PyTorch, HuggingFace, Scikit-learn, NLI Scoring, CLIP Embeddings, Whisper ASR, Hinglish / Multilingual Pipelines

EXPERIENCE

Ikshan (now DoableClaw) — AI Business Automation Startup — Feb 2026 – Present · Remote

Founding AI Engineer • CURRENT

- **Designed and built a 3-agent pipeline from 0→1** — Website Crawler → Diagnostic Engine (Agent E) → Playbook Generator (Agent C, Claude API) — delivering a ~13-page personalized growth playbook per founder in under 5 minutes.
- **Engineered a Cross-Reference Engine** comparing founders' Q&A against live website crawl data (H1s, CTAs, trust signals) to surface specific revenue-leak contradictions — a capability no existing AI tool provides.
- **Built B2B vs D2C auto-detection** from website signals — dynamically adjusting scoring criteria, finding language, and buyer persona labels across all agent outputs.
- **Implemented strict LLM output validation** — zero hedging language, mandatory evidence citation, rule-based scoring across 5 sub-dimensions for consulting-grade quality at scale.

PhysicsWallah (PW) — India's Largest EdTech Platform — Jun 2025 – Dec 2025 · Bengaluru

AI Engineer Intern

- **Engineered end-to-end RAG pipelines** for student feedback using Hybrid Search (Dense + BM25); validated against 500-sample golden dataset — **15% accuracy lift**.
- **Built "PW Prerna" Safety AI backend** — emotional wellness chatbot serving **10K+ daily users**; dual-layer guardrails (Regex + LLM) reducing jailbreak success to **<0.1%**.
- **Optimized inference costs** via Semantic Caching (Redis) + prompt compression — **20% Azure OpenAI cost reduction** and **400ms P95 latency improvement**.
- **Automated analytics pipelines** parsing daily engagement logs; dashboards used by **200+ faculty members**.

SELECTED PROJECTS

Master Auditor — Call Intelligence Microservice — *FastAPI · Azure OpenAI · ElevenLabs · Docker · Regex*

- Production microservice auditing **1,000+ call minutes daily**; **90% cost reduction** via Fail-Fast — Regex filters fatal violations pre-LLM inference.
- Custom **"Passive Density Score"** quantifying disengagement better than sentiment; Hinglish ASR via ElevenLabs Scribe resolves code-mixed speech errors.

Luna-AI — Agentic Workflow Engine — *LangGraph · React Native · Supabase · NLI Scoring*

- Stateful multi-agent system with persistent memory; **~40% hallucination reduction** via Self-Correction node with NLI scoring; sub-second mobile delivery via Supabase Edge Functions.

EduTube Coach — Agentic Study Assistant — *FastAPI · GPT-4 · Google APIs · RAG · OAuth 2.0*

- Converts YouTube videos into flashcards & summaries; autonomous study scheduling via Google Calendar OAuth 2.0; **100% source verifiability** — RAG cites exact timestamps.

Multimodal RAG — Image + Text Retrieval — *CLIP · FAISS · FastAPI · Multilingual*

- Multilingual retrieval (English/Hindi/Spanish) using CLIP embeddings + FAISS; dynamic indexing for real-time image uploads with metadata versioning.

EDUCATION

Indian Institute of Technology, Guwahati — 2023 – Present

B.Sc in Data Science & Artificial Intelligence

ML · NLP · Data Structures & Algorithms · Deep Learning · DBMS · Statistics & Probability

ACHIEVEMENTS

- **Winner — CMRIT Power BI Hackathon (2024)** for building actionable sales & operations dashboards with business-driven insights.
- Shipped **5+ production AI systems** across 2 companies serving 10,000+ daily users. Open-Source: github.com/RohitMakaniProfile · EduTube Coach · Your-Safe-Space · Multimodal-RAG