

Nikunj Patel

Backend AI Engineer

Navsari, Gujarat, India • [+91 87802 70959](tel:+918780270959) • nikuni2977@gmail.com • [LinkedIn](#)

[Claude Certified Architect – Foundation](#), Anthropic (2026)

PROFESSIONAL SUMMARY

Backend AI Engineer with 3 years of experience building production-grade Generative AI applications, RAG pipelines, and large-scale Flask backends. Anthropic-certified (Claude Certified Architect – Foundation), with hands-on experience across vector stores (pgvector, Qdrant, Milvus, Pinecone, ChromaDB), AWS cloud architecture, and custom LLM integration (OpenAI, Anthropic Claude, Google Gemini). Early adopter of AI-assisted development tools (Cursor, Claude Code) since 2023.

PROFESSIONAL EXPERIENCE

Backend AI Engineer, [Pragnakalp Techlabs](#)

July 2024 – Present | Gujarat, India

- Designed and deployed production-grade GenAI applications (OpenAI, Anthropic Claude, Google Gemini), chatbots, RAG systems, and voice agents across client verticals; see Key Projects.
- Built and hosted backends on AWS (Lambda, EC2, RDS) and GCP/Azure; deployed voice AI with Twilio, Gemini, and Django; built NLP/CV pipelines and automated workflows (n8n, Zapier).
- Early adopter of AI coding tools (Cursor, Claude Code, GitHub Copilot) across all development workflows since 2023.

Software Development Intern, [Pragnakalp Techlabs](#)

January 2024 – June 2024 | Gujarat, India

- Built foundational RAG/chatbot systems and a Copper CRM to GoHighLevel (GHL) data migration pipeline; see Key Projects.
- Built core skills in Flask, LangChain, and OpenAI integration, carried forward into full-time work.

TECHNICAL SKILLS

AI & LLMs:

OpenAI, Anthropic Claude, Google Gemini, LangChain, RAG pipelines, prompt engineering, vector search, voice AI agents, NLP/CV (YOLO, CLIP, OCR)

Backend & Cloud:

Python, Flask, FastAPI, Django, AWS (Lambda, EC2, RDS, API Gateway), GCP, Azure, Celery, Redis, Docker, CI/CD

Data & Vector Stores:

PostgreSQL, MySQL, MongoDB, Elasticsearch, pgvector, Qdrant, Milvus, Pinecone, ChromaDB

Security, Automation & Integrations:

Strix (AI pentesting), AST-based SAST, GitHub Actions, n8n/Zapier, Selenium, Salesforce, HubSpot, Copper CRM API, GoHighLevel API, Bubble.io, Twilio, Cursor, Claude Code

KEY PROJECTS

[LeaderMind, AI-Powered Leadership Analytics SaaS Platform](#)

Tech: Python, Flask (38+ Blueprints), SQLAlchemy, Celery, Redis, Elasticsearch, PostgreSQL, AWS EC2/RDS, Google Gemini, JWT

- Designed and built the backend for LeaderMind, a 38+ blueprint modular Flask platform (SQLAlchemy, Celery/Redis, Elasticsearch), deployed on AWS EC2/RDS behind uWSGI/Nginx.
- Migrated the production database from MySQL to PostgreSQL and overhauled query patterns (eliminating N+1 loads, pushing filters to the DB layer), cutting the slowest endpoint from ~11s to ~2–3s (70–80% latency reduction) and unlocking more flexible indexing for future scale.
- Implemented an 8-role RBAC system with WordPress-based OAuth/SSO for a unified user base, plus the DIAMONDS trait-rating and TMT analysis modules.

Road Semantics, Edge-to-Cloud Road Data Platform

Tech: Python, Flask, pandas, rdfliib, Raspberry Pi, Arduino

- Built the edge pipeline on vehicle-mounted Raspberry Pi units collecting gyroscope, ultrasonic, camera, and GPS sensor data, in a multi-client/multi-server architecture (Pis as workers, fixed servers as collectors).
- Designed and implemented the complete backend, then validated it in the field, mapping 20 km of real-world roads to detect speed breakers and surface defects.

No-Code RAG Chatbot Platform

Tech: LangChain, Milvus, OpenAI, FastAPI, PostgreSQL on Amazon RDS, AWS, Telegram & Slack APIs

- Built a no-code platform turning documents and web URLs into RAG chatbots, deployable to an embeddable web widget, Telegram, and Slack via a FastAPI backend on AWS.

AI Security Scan-and-Remediate Pipeline

Tech: Strix, Claude, GitHub Actions, AST-based SAST

- Built a closed-loop pipeline where Strix finds and validates exploits and Claude auto-generates fixes, wired into GitHub Actions as a merge-blocking gate alongside a custom AST-based SAST tool.

Production RAG & AI Backend Systems

- Designed, built, and deployed the backend for 4 production RAG/AI systems (LangChain, OpenAI, MongoDB/ChromaDB/Pinecone, Bubble.io), hosted on AWS EC2 via uWSGI/Nginx.

EDUCATION

Bachelor of Computer Engineering (BE), Government Engineering College, Gandhinagar | April 2019 – June 2024