

Naveen Reddy

AI Data Engineer

✉ naveenreddyvdk@gmail.com ☎ +91 9391855881 🔗 [linkedin.com/in/naveen-reddy-813a651b9](https://www.linkedin.com/in/naveen-reddy-813a651b9)

📍 Hyderabad, Telangana, India

PROFESSIONAL SUMMARY

AI Data Engineer with 2.5+ years building production-grade AI applications. Specialised in RAG pipelines, agentic AI systems, and LLM-powered tools using Python, FastAPI, LangGraph, and Claude Code. Experienced with OpenAI & Gemini SDKs, vector storage (Qdrant, PostgreSQL), and deploying scalable workflows on Google Cloud via Kubernetes and Argo Workflows. Independently built and deployed SadhanAI (sadhanai.com) — an AI-powered coding practice platform — solo, from concept to production.

EXPERIENCE

AI Data Engineer

Apr 2025 – Present

- ▶ Built agentic RAG pipelines for public sector persona development — scraping PDFs from government websites, cleaning and extracting data using LLMs, chunking, generating embeddings, and serving via a LangGraph-powered stateful chatbot.
- ▶ Developed a live WhatsApp scam-detection chatbot using LangGraph and Qdrant vector storage, enabling real-time retrieval and conversational AI over extracted scam data PDFs.
- ▶ Designed and deployed internal Claude Code agents with custom skills.md files for automated code review, workflow log analysis, code correction, and auto-raising PRs — reducing manual engineering effort significantly.
- ▶ Built a reusable internal Pydantic validation library used across multiple data pipelines for schema enforcement, data quality checks, and error handling.
- ▶ Orchestrated daily Argo Workflows on GKE for automated scraping, processing, validation, and database uploads — keeping all production applications updated with fresh data every 24 hours.
- ▶ Managed vector storage and semantic search optimisation using Qdrant and pgvector across multiple production AI applications.
- ▶ Developed and exposed a REST API (FastAPI) for the scam detection dataset, integrated as the backend for a WhatsApp chatbot with a dedicated UI.

Data Engineer

Jan 2024 – Mar 2025

- ▶ Scraped and ingested large volumes of government PDFs spanning 2000–2026 across multiple projects — handling both monthly and weekly publication cycles using Scrapy and Python.
- ▶ Built and maintained ETL pipelines to scrape, clean, transform, and structure raw government data using Pandas, storing processed files in Wasabi object storage and PostgreSQL/MongoDB databases.
- ▶ Built an internal workflow library that allows anyone to create and schedule Argo Workflows (daily, weekly, monthly) using just a config file — no workflow or Kubernetes knowledge required.
- ▶ Developed a CRUD-based internal tool with FastAPI enabling 10+ non-technical team members to view PDFs, manage records, and update database tables without engineering support.
- ▶ Managed multiple simultaneous data pipelines with independent scraping schedules, validation rules, and storage targets — ensuring data quality using schema validation and automated error handling.

NOTABLE PROJECT

Sadhanai.com — Founder & Solo Developer

2026 – Present

- ▶ Built and deployed a production-grade coding practice platform solo — from concept to live product, owning the full product lifecycle.

- AI generates custom coding problems with test cases from a single prompt, powered by a LangGraph agentic workflow.
- Debug mode — AI generates intentionally broken code, users identify and fix it.
- Real-time AI assistance while coding with post-submission code analysis.
- Users can generate problem bundles with one prompt and share with friends.
- Built an Organisation model — teams create batches with assigned instructors who generate and share assignments using a single prompt; admins track time taken, completion rates, and student progress per assignment.
- Implemented Google Authentication via Better Auth for secure user management.
- Full stack: Next.js · FastAPI · LangGraph · PostgreSQL · pgvector · Better Auth · Google OAuth · Docker · Google Cloud Platform

EDUCATION

Vardhaman College of Engineering (VCEH)

B.Tech (Minors) – Data Science

2022 – 2024

- - Completed a specialisation in Data Science covering machine learning, data pipelines, and hands-on AI project work alongside core ECE degree.

B.Tech – Electronics & Communications Engineering

2020 – 2024

- Electives focused on Artificial Intelligence, Machine Learning, and Python programming — which sparked interest in AI and Data Science.
- Active participant in IEEE events and technical societies.

TECHNICAL SKILLS

AI / LLM: RAG Pipelines, Agentic AI, LangGraph, OpenAI SDK, Gemini SDK, Claude Code, Prompt Engineering, Vector Search, Embeddings

Backend: Python, FastAPI, Pydantic

Data & Storage: PostgreSQL, MongoDB, Qdrant, Wasabi, ETL Pipelines, Pandas, Scrappy

Cloud & DevOps: Google Cloud Platform (GCP) (Basic), Kubernetes (Basic), Argo Workflows, Docker

Frontend: Next.js (Vibe Coded)

Auth: Better Auth, Google OAuth

Other: Git, Linux, Data Cleaning & Analysis

INTERESTS

- trips to Manali, Shimla, Delhi, Lonavala and 500km+ rides around Hyderabad
- Gym & fitness
- Sci-fi series
- Building AI-first products by vibecoding