

Vivek Sarvaiya

Full Stack Developer | AI/GenAI Engineer | Software Developer | Backend Developer

Bengaluru, India | viveksarvaiya1873@gmail.com | +91 9322962633 | <https://www.linkedin.com/in/vivek-sarvaiya-40537a286> | <https://github.com/Vivek-hash07> | <https://viveksarvaiya.online>

SUMMARY

Full Stack Developer and AI/GenAI Engineer with ~2 years of experience building production-grade web applications and LLM-powered backend systems. Skilled across the full stack - React, Next.js, Node.js, TypeScript, PostgreSQL - and modern AI engineering, including LLM integration, Retrieval-Augmented Generation (RAG), vector search, and agentic workflows. At Quralyst, built and shipped scalable backend systems and API orchestration layers spanning 5-10 external data sources, cutting manual research effort by 30-50%. Independently created and deployed multiple full-stack SaaS products end-to-end - authentication, billing, databases, and AI-powered features. Comfortable owning problems from architecture through deployment, and targeting Full Stack Developer / AI Engineer / Software Developer roles building production-grade, LLM-integrated web applications at scale.

EXPERIENCE

Software Engineer

Quralyst Backed by Indago Research | Apr 2025 - Present

- Architected scalable, multi-tiered full-stack backend systems for an AI-powered M&A deal intelligence platform, orchestrating distributed data flows across ingestion, processing, and enrichment layers spanning 5-10 external data sources.
- Built end-to-end data pipelines in Python to ingest, normalize, and process structured and unstructured data from 5+ external sources, improving processing speed and reliability across test, integration, and production environments.
- Built an API orchestration layer connecting OpenAI and Apollo APIs to automate AI-driven data enrichment, cutting manual research effort by 30-50%.
- Developed intelligent filtering and transformation logic to extract high-quality deal insights from large datasets, surfacing higher-signal insights for investment decision-making.
- Collaborated cross-functionally with product and business stakeholders, participated in code reviews, and maintained clean, well-documented, production-grade code.
- Monitored and supported data-processing systems across test, integration, and production environments within two-week Agile/Scrum sprints (planning, standups, retrospectives), resolving defects to sustain reliable performance across 5-10 external pipelines.

Software Developer Intern

Quralyst | Apr 2025 - Dec 2025

- Contributed to backend development of an AI-powered deal sourcing platform used by investment teams, streamlining enrichment workflows across 5+ external data sources to reduce manual research effort.
- Connected external APIs and built data-ingestion pipelines for real-time enrichment, supporting the same multi-source architecture later used in production.
- Assisted in building scalable backend components to handle multi-source, mixed-format data feeding the enrichment pipeline.

Intern

Edunet Foundation | Feb 2025 - Mar 2025

- Built and evaluated multiple machine-learning models in Python and TensorFlow across real-world problem statements during a 2-month AI/ML and Green Skills internship, applying refined data-preprocessing techniques to improve model accuracy.
- Refined data-preprocessing pipelines and collaborated with fellow interns to optimize model accuracy and computational efficiency.

AI Engineer

Freudia | Nov 2024 - Feb 2025

- Built an AI-powered chatbot delivering empathetic, contextually-aware conversations for users dealing with stress, anxiety, and everyday mental health challenges, using Large Language Models (LLMs).
- Fine-tuned LLMs on domain-specific conversational data to improve response quality, tone, and emotional

relevance.

- Applied Retrieval-Augmented Generation (RAG) to ground responses in curated resources, reducing hallucination and improving factual reliability of sensitive advice.
- Leveraged NLP techniques for sentiment analysis and intent detection, enabling the chatbot to adapt tone and response style based on user emotional state.

PROJECTS

KVM AI CodeReviewer - AI-Powered Code Review SaaS

Next.js 16 TypeScript React 19 Tailwind CSS Better Auth PostgreSQL Prisma Octokit Inngest Pinecone OpenRouter Razorpay

- Built a full-stack SaaS AI code review platform with GitHub App integration (Octokit) that automatically reviews pull requests via webhook-driven triggers and posts detailed feedback on correctness, security, performance, and maintainability.
- Built an AI review pipeline using OpenRouter (LLM-based) combined with Pinecone vector search - chunking and embedding PR diffs plus synced repository code (up to 200 files/repo) for context-aware reviews.
- Orchestrated background jobs with Inngest for durable, retry-safe workflows; built GitHub OAuth authentication (Better Auth) and a usage-based billing system paired with Razorpay subscriptions.
- Built the full dashboard and database layer (Prisma + PostgreSQL) for repository sync, PR tracking, and subscription management.

AI-Powered Personal Fitness & Nutrition Coach - Full-Stack Mobile App (Solo Project)

React Native Node.js FastAPI PostgreSQL Redis AWS S3 Claude API (Vision/LLM) REST APIs JWT Authentication

- Architected and built a full-stack mobile application combining multimodal AI (vision + language models) to analyze daily physique and meal photos, generating personalized workout and diet plans in real time.
- Built a computer-vision pipeline using the Claude API to extract granular nutritional data (calories, macros, protein) from food images, enabling automatic meal logging with an interactive macro-gap-closing recommendation engine.
- Engineered a custom memory-management system using PostgreSQL and Redis with a rolling-summarization architecture (header/summary/content records), enabling long-term AI context retention without a vector database and cutting prompt token usage and API costs.
- Built a gym session tracker with automatic personal-record (PR) detection, workout adherence tracking, and progress analytics feeding back into AI-driven plan regeneration.
- Built a relational database schema (10+ interlinked tables) covering user profiles, nutrition logs, workout sessions, and AI-generated media, with a scheduled job system for automatic weekly plan regeneration.

SKILLS

Languages: Python, TypeScript, JavaScript, Node.js, Java (Learning)

Frontend: React.js, Next.js (App Router), Tailwind CSS, shadcn/ui, HTML5, CSS3, Zod

Backend: Node.js, Express.js, RESTful APIs, tRPC, WebSockets (Socket.IO), API Orchestration, Microservices

AI / GenAI: LLM Integration (OpenAI, Claude, Gemini APIs), Prompt Engineering (Zero-shot, Few-shot, Chain-of-Thought, ReAct), Retrieval-Augmented Generation (RAG), Vector Search & Embeddings, Agentic Workflows & AI Agents, NLP (Sentiment Analysis, Intent Detection), Vercel AI SDK, Inngest AgentKit

Databases: PostgreSQL, MongoDB, Redis, Prisma ORM, Vector Databases (Pinecone, Qdrant)

DevOps & Cloud: Docker, AWS, CI/CD, Git & GitHub, Vercel

Tools & Integrations: GitHub Apps/Octokit, Razorpay, Better Auth, Inngest, Postman

EDUCATION

Bachelor of Computer Applications (BCA)

Acharya Institute of Graduate Studies (AIGS) | Aug 2022 - Aug 2025

CGPA: 8.9/10 | Final Semester SGPA: 9.5

CERTIFICATIONS

- Generative AI (GenAI) - Chai Aur Code
- Web Development - Chai Aur Code
- AWS Cloud Practitioner - Stephane Maarek
- AI Agents Fundamentals - Hugging Face
- Introduction to Software Engineering - IBM