Himanshu Chauhan

AI Engineer | LLMs, RAG & Agents · Dehradun, India · Open to relocation (India & Malaysia)

+91 85338 60335 | himanshu.chauhan07@icloud.com | ansh.life | github.com/devhimanshuhc | linkedin.com/in/devhimanshuhc

SUMMARY

Software engineer with 2+ years of experience, including 1.5+ years shipping production **Generative AI** and LLM features: retrieval-augmented generation over clinical records, tool-calling agents for equity workflows, and agents for the browser and Mac apps. Python, FastAPI, LangChain, Vercel AI SDK, REST APIs and microservices. Builds with **Claude Code**.

EXPERIENCE

Heizen | San Francisco, CA (Remote) — Aug 2025 - Oct 2026

*Backed by **Titan Capital ($500K)**; featured on **Shark Tank India** and **Forbes Remote***

Software Development Engineer — Jun 2026 - Oct 2026

- **Evra (AI healthcare, 3k+ users):** Cut clinical data entry **40%** by shipping LLM summarization and **RAG** over patient records. MongoDB vector search and LangChain on **FastAPI**, with encrypted clinical fields, role-based auth and a React Native (Expo) app.
- **Heizen Studio (open-source AI dev tools, 5K+ users):** Owned system architecture, API contracts and delivery end to end.
- **Medurun (ambulance dispatch, pre-launch):** Architected real-time dispatch (**NestJS**, Redis geospatial search, **Socket.IO**, atomic **Lua**, BullMQ) and shipped the React Native rider and driver apps plus Next.js dashboards.

Associate Software Development Engineer — Feb 2026 - May 2026

- **DeepMechanix:** Cut ASME pressure-vessel design cycles from **weeks to hours** by automating **PV Elite** against ASME code, including the APIs, job orchestration and validation pipeline.

Software Development Engineer Intern — Aug 2025 - Jan 2026

- **EquityNav:** Built AI insight agents for vesting, exercises and liquidity on the **Vercel AI SDK** (OpenAI and Gemini), with tool calling, SQL execution, structured outputs and guardrails.
- **EquityNav:** Unified Carta portfolios and an equity-tax engine into projection, scenario and risk dashboards. Automated sync and alerts with NestJS Scheduler, Docker and GitHub Actions.

United Protocol | Auckland, NZ (Remote) — May 2023 - Jul 2025

Full Stack Developer

- Built an AI-agentic mobile browser in React Native (**Meta LLaMA, LangChain, Groq**) that runs multi-step web tasks, and shipped **Hume AI** + Twilio speech-to-speech voice agents.
- Shipped tamper-proof encrypted history on IPFS + BNB Smart Chain (Solidity), plus web and desktop apps in React, Next.js, Electron and Tauri.

WhatBytes | Bangalore, India (Remote) — Jun 2024 - Oct 2024

Frontend Developer

- Improved rendering on a Svelte + Next.js app serving **70,000–100,000+ weekly users**, and built analytics dashboards (Google Analytics, Facebook Insights).

PROJECTS

- Solo Harness (github.com/devhimanshuhc/solo-harness): From-scratch TypeScript agent harness that drives the user's own browser and Mac apps, with parallel sub-agents, replayable logs, OpenTelemetry traces and **192 offline tests**.
- Retro (retro.ansh.life, github.com/devhimanshuhc/retro): Actor-based multi-agent research system on the Anthropic (Claude) and OpenAI APIs. Isolated BullMQ worker agents write cited reports; raised hand-checked facts found in a real earnings report from **48% to 86%**. Crashed workers resume from checkpoints without duplicate reports; failed jobs replay from a dead-letter queue.
- DataSays: Plain-English questions return SQL and a chart. Hybrid retrieval over **pgvector** and Postgres full text, orchestrated with LangGraph.

SKILLS

AI: Generative AI (GenAI), LLMs, RAG, prompt engineering, vector search (pgvector), LangChain, LangGraph, tool calling, structured outputs, guardrails, OpenAI, Anthropic (Claude) API, Gemini, Vercel AI SDK, Groq, Claude Code, LLM evals (LLM-as-judge)
Languages: TypeScript, JavaScript, Python, SQL, Swift, Lua, Solidity
Backend: REST APIs, microservices, system design, FastAPI, NestJS, Node.js, MongoDB, PostgreSQL, Prisma, Redis, BullMQ, Docker, AWS (S3, ECS, ECR, ElastiCache, ALB), CI/CD (GitHub Actions), WebSockets, Playwright, pytest
Mobile & Frontend: React Native (Expo), React, Next.js, TailwindCSS, TanStack Query

EDUCATION

St. Vincent Pallotti College of Engineering and Technology | Nagpur, India — 2022 - May 2026

B.Tech, Computer Science and Engineering