Rochak Shrestha

AI Agent Engineer: autonomous LLM agents in production, backend and mobile

Kathmandu, Nepal (UTC+5:45) · rochakshresthawork@gmail.com · +977 9860305454
linkedin.com/in/rochak-shrestha-a23a09210 · github.com/Rochak69

Portfolio, live apps and demos: rochakshrestha.com

AI engineer who designed and built Kai, the autonomous AI coaching agent at the center of Livio, a US weight loss app live on iOS and Android. Kai runs the whole coaching conversation on Claude: it calls 30+ tools, remembers each member over time, acts on their behalf with human confirmation, and is kept in check by evals and cost controls I built. I lead a team of four, built version 2 of the product from scratch, and have 5+ years in software, three of them shipping Flutter apps. AI coding agents are part of my daily workflow.

EXPERIENCE

Thankz Global Staffing, Lead Engineer (AI, Backend and Mobile) June 2024 – Present

Livio: AI Weight Loss Coach · iOS · Android

Remote, full-time. A US subscription app where an autonomous AI coach, Kai, is the whole product experience.

- Designed and built Kai, an autonomous Claude-based agent that runs every coaching conversation: it streams its replies, decides which of its 30+ tools to call, and reads and updates a member's food, weight, sleep and workout data on its own.
- Technical lead for the product. Lead a team of four engineers, run sprint planning and code reviews, and work directly with the US founders on the roadmap.
- Architected and built version 2 from the ground up: backend, AI system, web app and mobile, in Python, FastAPI and PostgreSQL on AWS.
- Made the agent safe to trust with real data: every change it proposes is confirmed by the member with one tap, and numbers are computed in code, never by the model.
- Built long-term memory so Kai remembers each member: facts from past conversations, a coaching profile it updates daily, and retrieval (RAG) over a health-document library.
- Set up the evaluations that decide which model ships, including a blind LLM-as-judge test for coaching quality, and used them to stop a switch to a cheaper model that gave worse answers.
- Cut the AI cost per message by routing simple requests to a small model and caching prompts, and added failover between AWS Bedrock and the Anthropic API.

Kata Jaam, Founder and Engineer February 2026 – Present

katajaam.com · iOS · Android. My own product: a two-sided events marketplace with 2,400+ users, run by an AI operations agent.

- Built an autonomous AI operations agent on Claude Code that runs the marketplace each day: it works the Instagram DMs and email inboxes, finds new events on Instagram and other sources and imports them, onboards and sells to organizers, prepares the social posts and writes a daily report, handing focused jobs to subagents along the way.
- Built the event-discovery pipeline behind it: a small model filters venue posts, Claude extracts the event details, and duplicates are caught before human review.
- Founded and built the product alone, shipping the first version in two weeks with AI coding agents; the site ranks first on Google for "events in Kathmandu".

Contract Work, Freelance Mobile Engineer — August 2024 – Present

Alongside my full-time role, for agencies with US, Australian and Nepali clients.

- **Hyperce (US and Nepal).** Built the ordering and loyalty apps for two US restaurant chains, Piccadilly and K&W Cafeterias, and My Mart Now, a grocery delivery app for Kathmandu.
- **Uno Technology (Australia).** Built four apps from scratch for Australian clients: Hooked Up, Become Secure, Akuna Tec and The Empowered Man.
- **Trifecta Digi Media (Nepal).** Built NC Janasamparka, the official app of the Nepali Congress party.
- **Webpoint Solutions (US).** Built One Accord: Scanner, a ticket-scanning app for event check-in.

E-Signature, Senior Mobile Application Developer — October 2023 – June 2024

- Led mobile development of Baideshik Rojgar (Sajha Jobs), the app Nepali migrant workers use to find jobs abroad, with over a million installs on Google Play.

Varosa Technology, Mobile Team Lead — January 2023 – October 2023

- Led a mobile team of six in two-week sprints and built the calling engine for Calilio, a business phone system with numbers in 100+ countries.

Codware Lab, Mobile Application Developer — June 2022 – January 2023

- Built the monitoring app used by 5,000+ officers of the Election Commission of Nepal during the 2022 elections.

25 Hours, Mobile Application Developer — July 2021 – June 2022

- Built Fish Producer and Fish Buyer, a pair of Android marketplace apps.

SHIPPED APPS

Livio iOS · Android
Calilio iOS · Android
Hooked Up iOS · Android
Akuna Tec iOS · Android
One Accord: Scanner Android
Kata Jaam iOS · Android · Web
Baideshik Rojgar iOS · Android
K&W Cafeterias iOS · Android
The Empowered Man iOS · Android
My Mart Now iOS · Android
Piccadilly Cafeterias iOS · Android
Become Secure iOS · Android
NC Janasamparka iOS · Android

SKILLS

Agents and LLMs: Claude (Anthropic API, AWS Bedrock), OpenAI, tool use, streaming agent loops, RAG, embeddings, context engineering, prompt caching, long-term memory, model routing, human-in-the-loop

Evaluation and operations: golden datasets, LLM-as-judge, tracing, cost metering, provider failover, guardrails

Backend: Python, FastAPI, SQLAlchemy, PostgreSQL, pgvector, Redis, WebSockets, pytest, AWS (EC2, RDS, Bedrock)

Web, mobile and tooling: TypeScript, React, Next.js, Flutter, Swift, Kotlin, Claude Code, MCP servers

Leadership: team leadership, Agile sprint planning, code review, mentoring

EDUCATION

Bachelor of Computer Engineering, Himalaya College of Engineering, Tribhuvan University, Nepal, 2021