

RANJAN SINGH

Gen AI & Product Engineer

Gurugram · India · ranjandsingh@gmail.com

[LinkedIn](#) · [GitHub](#) · [Portfolio](#)

PROFESSIONAL SUMMARY

Gen AI & Product Engineer with 8+ years shipping production AI and user-facing systems across web and mobile. Currently Lead Engineer at Telos (New York), owning product development end-to-end — took the AI Life Coach platform from prototype to thousands of active users, leading a team of 4 across architecture, delivery, and product analytics. Previously led a 3-engineer team at Pilot (Vancouver) shipping AI-powered travel planning. Active open-source contributor — built Shelly, a cross-platform terminal companion for AI coding workflows, and maintain published npm packages. Deep expertise in LangChain, LangGraph, vector DBs, Next.js (SSR / ISR), React Native, and Node.js.

TECH STACK

AI / Gen AI: LangChain, LangGraph, OpenAI API, Hugging Face, pgvector, Zep Memory, RAG pipelines, embeddings & vector search, fine-tuning, Microsoft Presidio (PII)

Frontend: React, Next.js (SSR / ISR / App Router), Redux, React Native (iOS & Android), TypeScript, Material UI

Backend: Node.js, Express, NestJS, FastAPI, REST & microservices, Socket.IO, SSE, RabbitMQ

Data: PostgreSQL (with pgvector), MySQL, MongoDB, Firestore, Redis

Cloud & DevOps: GCP (Cloud Run), AWS, Firebase, Supabase, Docker, CI/CD, Sentry

Languages: JavaScript, TypeScript, Python

PROFESSIONAL EXPERIENCE

Software Engineer II / Lead Engineer — *Telos, New York (USA)* Apr 2024 – Present

Tech Stack: Node.js, Express, React, React Native, Python, LangChain, LangGraph, pgvector, Microsoft Presidio, GCP Cloud Run, Firebase, Docker

- **Led end-to-end product development** as the lead engineer, taking the AI Life Coach platform from early prototype to production across iOS, Android, and web — now serving thousands of active users.
- **Managed a team of 4 engineers**, owning architecture decisions, sprint planning, code review standards, and delivery across the full stack.
- **Architected the multi-agent LLM system** (LangChain + LangGraph) powering journaling and conversational coaching, with long-term memory via Zep and retrieval grounded in pgvector — enabling context-aware sessions that adapt across weeks of user history.
- **Designed a privacy-by-design AI pipeline** using Microsoft Presidio to de-identify user content before any prompt leaves the platform, ensuring zero PII reaches third-party LLMs and unblocking enterprise conversations.
- **Built autonomous nudging agents** that monitor user progress and trigger personalized notifications for onboarding and exercise completion, lifting activation and retention during the post-launch scaling phase.
- **Scaled backend to thousands of concurrent users** with SSE-based streaming and GCP Cloud Run auto-scaling, keeping real-time chat responsive under load while containing infra cost.

- **Built the product analytics layer end-to-end** — instrumenting events, dashboards, and funnels across the app, then using the data to guide roadmap decisions on onboarding, retention, and feature prioritization.
- **Owned the cross-platform mobile codebase** (React Native) end-to-end — shipping rapid release cycles, dynamic content delivery, and consistent UX across iOS and Android. Also authored rn-tour-copilot, a customized walkthrough library published to npm, to power Telos's guided onboarding with custom tooltips and step-change hooks.
- **Set technical direction and mentored the team** on LLM evaluation, prompt iteration, and code quality; introduced review standards and Sentry-based observability that reduced production regressions.

Software Engineer II — *Pilot, Vancouver (Canada)*

Jun 2022 – Apr 2024

Tech Stack: Node.js, Express, React, React Native, Python, Firebase, Docker

- **Joined at early stage** and helped grow Pilot's AI-powered trip planning product through its launch and initial user scaling, leading a team of 3 engineers through rapid product iteration.
- **Owned architecture and technical direction** across the backend and mobile apps — set the core service boundaries, chose the stack, and made the scaling trade-offs that shaped the product as it grew.
- **Designed and shipped the AI itinerary generator**, turning user preferences into complete, editable travel plans — measurably reducing blank-screen drop-off and accelerating time-to-first-itinerary for new users.
- **Designed the AI-driven booking import** for transport and accommodation, auto-completing travel plans from minimal user input and improving plan completeness.
- **Built an automated email & attachment ingestion pipeline** that imports bookings directly into the planner, eliminating manual uploads and cutting setup friction for new travelers.
- **Built and owned the React Native date-range picker** powering Pilot's multi-leg trip planning UX — fully themable with custom function hooks for product-specific behavior. Published the core as an open-source npm package to fill a real ecosystem gap, while the extended internal version continued to evolve as Pilot's private infrastructure.
- **Instrumented product analytics** across onboarding, itinerary generation, and booking flows — the data directly informed prioritization decisions as the user base grew.
- **Revamped the backend for concurrency and scale** — improved response times under peak load and made the service resilient to traffic spikes during launch windows.
- **Built internal monitoring and ops tooling** that shortened debugging cycles and gave the team real-time visibility into app health.

Software Engineer — *Aeon Logiciel, Noida (India)*

Jun 2021 – Jun 2022

Tech Stack: Node.js, Express, RabbitMQ, Socket.IO · Project: Folio.law — case & client management for law firms

- **Engineered a high-throughput backend** with Socket.IO-based real-time notifications and a scalable subscription module, processing 20,000+ legal case records weekly across multiple law firms.
- **Cut backend processing delays** by introducing RabbitMQ queues, significantly improving crawl time and request throughput.

Software Engineer — *Webbasol, Patna (India)*

Sep 2019 – May 2021

Tech Stack: Next.js, React, React Native, Node.js, Express, Firestore · Projects: RevoTelecomParts (B2B marketplace), Wedcell (wedding services marketplace)

- **Architected a high-traffic e-commerce platform on Next.js** for B2B telecom parts, using a hybrid SSR + ISR rendering strategy to serve 50,000+ SKUs with fast page loads, fresh catalog data, and strong multilingual SEO — SSR for dynamic routes, ISR for product pages to balance build time against freshness.

- **Implemented auto-localization for 8 languages** using browser-based detection, removing reliance on IP-based geolocation.
- **Built the Wedcell mobile app and backend** (React Native + Node.js) with deep linking for direct vendor/venue sharing, plus APIs and microservices for third-party integrations.

Software Engineer — *Project Element, Patna (India)*

Oct 2017 – Aug 2019

Tech Stack: React, React Native, Firebase, Firestore · *Project:* Shambhavi Telecom Employee Portal — field engineer & task management

- **Engineered a GPS-based location tracking system** (React Native + Firebase) with offline-first support for low-connectivity field conditions; optimized startup time by enabling Hermes to reduce initial load latency.

PERSONAL PROJECTS & OPEN SOURCE

Shelly — Cross-platform floating terminal for AI coding workflows [GitHub](#)

- Built a ~5 MB cross-platform (macOS + Windows) terminal utility with a floating “notch” UI that toggles via global hotkey — optimized for multi-session terminal management with keyboard-driven tabs, color-coded sessions, auto-shell detection (PowerShell, zsh, bash, WSL), and persistence across restarts.
- Deep Claude Code integration: live agent status tracking (Idle / Working / Waiting / Completed), auto-open on agent attention, CLAUDE.md-aware auto-run, and IDE context detection for VS Code and JetBrains.
- Stack: Rust + Tauri v2 (native shell), React 19 + TypeScript + xterm.js (frontend), portability for live PTY sessions; fully customizable theming with direct VS Code theme import (paste JSON or URL) alongside bundled presets.

Open-Source npm Packages

- [sharp-multer](#) — Multer storage engine that uses Sharp to auto-optimize and resize uploads (JPG / PNG / WebP) on the fly. Used in production by multiple downstream projects; 11★ on GitHub.
- [rn-tour-copilot](#) — Customized React Native walkthrough library built on react-native-copilot, adding a fully customizable tooltip UI, pre- and post-step hooks, and better navigation for multi-step guided tours. Built to power the guided onboarding and contextual coaching flows in the Telos app.
- [@ranjandsingh/rn-daterange-picker](#) — Themable, customizable React Native date-range picker with custom function hooks, filling a gap in the ecosystem where existing libraries weren't themable. Public version ships the original feature set; an extended version was later taken private as internal infrastructure.
- [rn-country-code-picker](#) — Lightweight, easy-to-use country code picker component for React Native.

EDUCATION

M.Sc., Electronics — A. N. College, Patna (Patna University)

2015 – 2017

B.Sc., Electronics — A. N. College, Patna (Patna University)

2012 – 2015

CERTIFICATIONS

Artificial Intelligence A-Z 2025: Agentic AI, Gen AI, and RL — Udemy

2023

Software Architecture & Technology of Large-Scale Systems — Udemy

2023