

NIKHIL SHRESTHA

Software Engineer

nikhilshrestha@gmail.com | 510-929-6460 | [linkedin.com/nikhilsth4](https://www.linkedin.com/in/nikhilsth4) | github.com/nikhilsth4

SUMMARY

Full Stack Software Engineer with 5+ years of experience building scalable, production-grade web applications using React, Next.js, TypeScript, and Java/Spring Boot, with applied AI engineering experience across RAG pipelines, LangChain, vector databases, and LLM API integration. Delivered high-impact products in healthcare, HR tech, and enterprise SaaS. MS in Computer Science. Bridges modern frontend architecture with backend system design to ship performant, AI-native products.

EDUCATION

Master of Science (MS) in Computer Science

May 2024

The University of Texas Permian Basin | Odessa, TX

Bachelor of Science in Computer Science and Information Technology

December 2020

Tribhuvan University | Kathmandu, Nepal

SKILLS

Frontend Development:

React | Next.js | TypeScript | JavaScript | Redux | Context API | Tailwind CSS | Material UI | HTML/CSS/SCSS | Figma

Backend Development:

Node.js | Express | Java | Spring Boot | Python | REST APIs | GraphQL | Prisma

Claude & DevOps:

AWS (Lambda, Cognito, S3, IAM) | Docker | CI/CD | Git/GitHub

Databases and tools:

PostgreSQL | MySQL | MongoDB | Redis

AI / LLM:

RAG pipelines | LangChain | ChromaDB | OpenAI & Anthropic APIs | prompt engineering | AI-assisted development workflows (Claude Code — subagents, hooks, MCP integrations)

Practices:

Agile/Scrum | JIRA | Jest | WCAG accessibility

EXPERIENCE

Software Engineer

May 2024-Present

Desmata Inc | Irving, TX

- Architecting an agentic AI travel planning platform that autonomously plans, books, and monitors end-to-end trips, orchestrating **LLM**-driven workflows across flight (**Amadeus**), hotel (**Booking.com**), transport, weather, and maps **APIs** as agent tools.
- Designing multi-step agent workflows for itinerary generation, in-app booking, and real-time trip disruption handling (e.g., flight delay → rebook transport → adjust itinerary), with user preference memory for personalized planning.
- Building the platform's full stack with **Next.js/React** (web), **Node.js** services, **PostgreSQL** with **Redis** caching, and **Stripe** payment integration, deployed on **Vercel/AWS**.
- Architecting a full HR Management System (HRMS) in **Next.js, TypeScript**, and **PostgreSQL** — designing the **Prisma** data model, database-level **RBAC** scoping, and **S3**-backed document storage across employee lifecycle, leave, and recruitment modules.
- Designed typed server-action architecture with structured error handling and transactional workflows (e.g., leave approvals) to ensure data integrity across multi-step operations.
- Developed responsive **React/TypeScript** frontends with custom hooks for lazy loading, client-side caching, and conditional data fetching, reducing page load times by 35%; built **D3.js** dashboards with real-time updates.
- Developed supporting microservices in **Java/Spring**
- Boot** with **Redis** caching and automated **CI/CD** pipelines with **Docker**.

Software Engineer

May 2022-Aug 2022

Fourth Valley Concierge Corporation | Tokyo, Japan

- Led development of an Admin Management System using **React, Next.js, TypeScript, Redux Thunk**, and **Tailwind CSS**, streamlining HR workflows.
- Built dynamic forms with complex validation logic and real-time field dependencies using **Formik**, improving data handling for HR personnel managing employee records.
- Integrated the frontend with serverless **AWS Lambda** microservices and implemented secure authentication with **AWS Cognito**, including role-based access controls.
- Transformed **Figma** wireframes into production-ready, mobile-first interfaces following component-driven design system patterns.

Software Engineer

May 2021-Apr 2022

Nepalland IT Solutions | Kathmandu, Nepal

- Engineered frontend modules for a SaaS Hospital Management System serving 200+ facilities using **React**, **Redux Toolkit**, **TypeScript**, **Material UI**, and **React Query**.
- Built complex, data-intensive interfaces for patient records, appointment scheduling, and pharmacy inventory with optimized rendering and real-time synchronization.
- Developed interactive **D3.js** analytics dashboards for medical staff to visualize patient statistics, resource usage, and operational trends.
- Implemented caching, pagination, and optimistic updates with **React Query**, reducing unnecessary API calls and improving perceived performance.
- Developed **Java/Spring Boot** microservices for patient data, scheduling, and inventory domains, designing **REST** and **GraphQL** APIs consumed by the frontend, and resolved 500+ critical issues during the COVID pandemic, contributing to a 40% reduction in support tickets.

Junior Software Engineer

Apr 2020-May 2021

Radius Infosys Pvt. Ltd., | Kathmandu, Nepal

- Developed a full-stack restaurant management system using **React**, **Redux**, and **Laravel**, improving order-processing and inventory workflow efficiency by 20%.
- Built a real-time kitchen dashboard with **Laravel Broadcasting**, **Pusher**, and **Laravel Echo** for live order updates and push notifications, and designed **REST APIs** for menu, payment, and order management.
- Collaborated with cross-functional teams to translate restaurant operator feedback into new features on a responsive, component-based UI.

PROJECTS

Krishna AI | Next.js | FastAPI | LangChain | ChromaDB | Python | OpenAI (<https://aikrishna.vercel.app/>)

- Built a full-stack AI assistant grounded in the Bhagavad Gita, using a **RAG** pipeline to deliver contextually accurate, scripture-referenced answers to philosophical and life questions.
- Architected a **FastAPI** backend for document ingestion, chunking, and embedding generation, storing vectors in **ChromaDB** for fast semantic retrieval across the full text.
- Developed a responsive **Next.js** chat interface with real-time streaming responses and conversation history.
- Designed **LangChain** retrieval chains that surface relevant verse citations alongside AI-generated answers, grounding every response in source scripture rather than unverified generation.