

# Praveen Kumar Saini

Full Stack & AI Systems Engineer | Agentic LLM Systems | Healthcare AI | Real-time Apps

Gurugram, India · +91 7017430549 · praveenksaini0@gmail.com

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

## SUMMARY

Full Stack & AI Systems Engineer with 7+ years building production systems across healthcare, fintech, and enterprise. Over the last 2 years, focused on **agentic LLM workflows**, **RAG/knowledge-grounded reasoning**, and **real-time conversational systems** for clinical decision support and symptom intelligence. Comfortable in early-stage startup environments: ships fast, owns ambiguous problems end-to-end, and designs for reliability (observability, fallbacks, data integrity) from day one.

## CORE SKILLS

- Agentic AI / LLM Systems:** LangGraph · LlamaIndex · RAG pipelines · retrieval grounding · long-context reasoning · prompt/agent evaluation · tool routing · guardrails & fallbacks · embeddings · semantic search
- Backend & Data:** Node.js · TypeScript · Fastify/Express · PostgreSQL · Redis · Drizzle ORM · vector databases · event-driven systems · WebSockets · high-throughput APIs
- Frontend & Real-time:** React · Next.js · React Native · design systems · performance optimization · streaming UIs · real-time chat/voice
- Cloud & DevOps:** AWS (Lambda, S3) · CI/CD (GitHub Actions, Bitrise, AppCenter) · monitoring & alerting · serverless architecture

## EXPERIENCE

### Confidential Healthcare AI Platform (Client name withheld under NDA) — AI Systems Engineer / Full Stack AI Developer

Remote · Dec 2024 – Present (Part-time/Contract until May 2025; Full-time from Jun 2025)

Designed and delivered AI-powered clinical decision support and real-time conversational healthcare companion systems focused on reducing administrative workload and improving interactive patient engagement. Client and product identifiers withheld under NDA.

*NDA-safe: I can walk through the architecture, evaluation approach, and representative flows on a call.*

- Built **LangGraph** multi-agent workflows coordinating symptom intake → retrieval → reasoning → response composition with tool routing and safety fallbacks, supporting **~182K daily** patient chat sessions across symptom triage and engagement workflows.
- Implemented **RAG-based document intelligence** over structured EHR-like data + unstructured clinical notes/guidelines; improved retrieval grounding quality while supporting real-time reasoning across high-volume patient query workloads.
- Developed real-time conversational interfaces (chat + voice) and integrated a **MetaHuman + Pixel Streaming** experience using **Node.js + WebSockets**, powering **~30K daily** patient sessions via AI-driven virtual healthcare assistants.
- Designed backend and streaming orchestration layers capable of supporting **200K+ daily** AI-driven patient interactions with low-latency response delivery.
- Owned **cross-platform application architecture** for a React Native product spanning **iOS, Android, Windows** and **remote Android kiosk deployments** for healthcare workflows.
- Led a **3-person engineering team** (hands-on) — drove architectural decisions, sprint execution, and integration quality across high-scale real-time healthcare interaction systems.
- Created embedding pipelines and vector retrieval systems processing large-scale symptom and disease similarity queries used in triage clustering and contextual case expansion.
- Shipped backend services with reliability patterns (**rate limits, retries, timeouts, circuit breakers, observability**) enabling stable handling of **200K+ daily** AI interaction events.
- Designed horizontally scalable streaming and inference orchestration layers maintaining **sub-second** response latency targets across high-volume patient-AI interaction workloads.

**Tech:** LangGraph · LlamaIndex · Node.js · TypeScript · PostgreSQL · Drizzle ORM · Vector DBs · WebSockets · React Native · iOS · Android · Windows · Android Kiosk · Unreal MetaHuman · Pixel Streaming · AWS · Azure OpenAI

## Sushrut Health — Founder & Lead Engineer (POC; discontinued)

Bengaluru, India · Jul 2025 – Nov 2025

- Built an early-stage **Clinical Decision Support System (CDSS)** supporting pilot deployments across **5 government hospitals** and **~1000 doctors** for structured diagnosis and treatment planning workflows.
- Modeled guideline-backed medical knowledge (**WHO** + Indian treatment guidance) and integrated it into RAG/agent flows delivering evidence-grounded recommendations across multi-hospital evaluation pilots.
- Designed multi-agent architecture for documentation, diagnosis support, EHR context ingestion, and treatment validation enabling collaborative clinical case processing workflows.
- Implemented multilingual and offline-tolerant UX patterns improving usability in low-connectivity government hospital environments.

**Tech:** Node.js · TypeScript · LangGraph · LlamaIndex · PostgreSQL · Drizzle ORM · AWS · Azure

**Links:** [Website](#)

## Apollo 24/7 — Senior Software Developer

Gurugram, India · Aug 2023 – May 2025

- Improved marketing analytics reliability by resolving CleverTap/Appsflyer funnel inconsistencies and attribution gaps.
- Refactored deep linking + referral flows using AppsFlyer Universal Links; improved onboarding conversion and tracking integrity.
- Designed and implemented in-app VoIP patient-to-support communication flow, improving response workflows and user experience.
- Mentored engineers and contributed to cross-platform architecture decisions and performance initiatives.

**Tech:** React Native · Node.js · AWS · AppsFlyer · CleverTap

**Links:** [Website](#) | [Android App](#)

## Audvice — Full Stack Developer

Salzburg, Austria (Remote) · May 2021 – Jul 2023

- Built mobile + web platform from 0→1 including audio capture, notifications, and deep linking.
- Designed AWS Lambda pipelines for transcription processing and media optimization; supported high-volume asynchronous workloads.
- Implemented event-driven messaging via RabbitMQ for realtime notifications and content delivery workflows.
- Set up CI/CD for mobile (Bitrise/AppCenter) and release hygiene.

**Tech:** React Native · Node.js · PostgreSQL · Next.js · RabbitMQ · AWS Lambda · S3 · Bitrise · AppCenter

**Links:** [Website](#)

## Vested Finance — Full Stack Developer

California, USA (Remote) · Dec 2020 – May 2021

- Modernized React Native trading app architecture and improved performance via better state/data patterns.
- Rebuilt dashboard web platform in Next.js for clearer portfolio analytics and maintainable UI.
- Introduced Redis caching to reduce latency and improve API responsiveness for frequently accessed financial data.

**Tech:** React Native · Node.js · Fastify · PostgreSQL · Redis · Next.js

**Links:** [Website](#)

## CityMall — Software Developer

Gurugram, India · Dec 2019 – Dec 2020

- Built and maintained dual React Native apps for customer + vendor workflows in an e-commerce environment.
- Developed Vue/Nuxt admin tooling and backend APIs for inventory, ordering, and logistics operations.
- Implemented event-driven processing using RabbitMQ/Redis for high-volume transactional workflows.

**Tech:** React Native · Vue/Nuxt · Node.js · PostgreSQL · RabbitMQ · Redis

**Links:** [Website](#)

## rtCamp — React Developer

Pune, India · Apr 2019 – Dec 2019

- Delivered responsive React/Redux interfaces from design specs; ensured cross-browser compatibility and performance.

**Tech:** React · Redux · JavaScript

**Links:** [Website](#)

## **EDUCATION**

**B.Tech, Computer Science** — MJP Rohilkhand University, Bareilly · 2015 – 2018

**Full Stack Web Development** — Alt Campus, Dharamshala · 2018 – 2019

**Front-End Web Developer Nanodegree** — Udacity · 2018