Yaseen Khatib

Full Stack AI Developer · MERN · LLM apps, RAG and AI agents

Hyderabad, India · yaseenkhatib04@gmail.com · +91 8208335028 · linkedin.com/in/yaseen-yk · github.com/Yaseenyk · yk.streamerosai.com

Core: MERN | TypeScript | Next.js | LLM apps | RAG | AI agents | MCP | Python and FastAPI | System design | CI/CD

Open to: Full Stack AI Developer roles · Hyderabad, Pune, Mumbai, Bangalore or remote India

SUMMARY

Full Stack AI Developer with five years shipping production web systems on the **MERN** stack with **TypeScript** and **Next.js**: hospital information systems, analytics dashboards and, since 2025, the AI tooling layer at Sparity: a hybrid **RAG** system with measured retrieval quality, an **MCP** server and an AI code-review service. I build the parts around the model that make AI usable: **RAG pipelines**, **agents** where a person approves every change, **MCP servers**, **evals**, and **cost-aware routing** between models. Made dashboards **1.5 to 2x** faster by profiling **MongoDB** aggregation pipelines; moved a team from JavaScript to **TypeScript** and **Next.js**. Shipped several products solo, each live with a public link or case study, and a client learning platform live **one day after the brief**.

SKILLS

MERN and web: **MongoDB, Express, React, Node.js, TypeScript**, JavaScript (ES2023), **Next.js**, Redux Toolkit, Zustand, **REST APIs**, WebSockets, JWT auth, Tailwind CSS, HTML and CSS

AI engineering: **LLM integration** (OpenAI, Gemini, Claude APIs), **RAG**, **embeddings and vector search** (pgvector, Vectorize, Upstash Vector), **function calling, agents, MCP servers**, prompt engineering, **LLM evals**, fine-tuning and distillation (PyTorch), Ollama

Backend and data: **Python and FastAPI**, NestJS, Java and Spring Boot, ASP.NET, **PostgreSQL**, SQL Server, Redis, SQLite, Firestore

Delivery: **System design**, unit and integration testing, Git, **GitHub Actions CI/CD**, **Docker**, **Cloudflare Workers**, Agile delivery

EXPERIENCE

Senior Software Developer, Sparity — 06/2025 – Present

Hyderabad · Hybrid

- Built a local hybrid **RAG** system over the team's web-app codebase and docs: dense plus BM25 retrieval, rank fusion and a cross-encoder reranker, incremental ingestion. Recall@5 0.95 and MRR 0.83 on a golden query set, with an eval harness that fails the build below its thresholds.
- Exposed it to Claude Code through an **MCP** server with 20 retrieval and code-lookup tools, so engineers get project context without reading whole repos; packaged so any team can vendor it into their own project.
- Built an AI pull-request review service: GitLab webhooks, Azure AD (MSAL) sign-in, internal JWTs and role-checked APIs. Led an HR resource-to-task matching pilot (FitRank) from proposal to a manager demo.

Stack: Python, Node.js, TypeScript, React, FastAPI, Chroma, Ollama, MCP, GitLab

Full Stack Developer, MSA Software — 01/2023 – 05/2025

Remote

- Made dashboards **1.5 to 2x** faster by profiling slow **MongoDB** aggregation pipelines, adding a caching layer and lazy-loading heavy panels.
- Moved the team's legacy JavaScript to **TypeScript** and **Next.js**, removing a class of type-related production bugs.
- Migrated shared state from hand-rolled Redux to **Redux Toolkit** slices.

Stack: React, Node.js, TypeScript, Next.js, MongoDB, Redux Toolkit

Software Developer (SDE 1), Manorama Infosolutions — 09/2021 – 01/2023

Healthcare software · Remote

- Built **REST** routing layers for hospital information systems on ASP.NET and **React**; **2x** payload throughput on the hot paths.
- Moved tangled request handling into MVC and hardened it for concurrent user sessions.

Stack: ASP.NET, MVC, React, REST APIs

SELECTED PROJECTS

Built and shipped solo unless marked client. Links checked on 14 Nov 2026.

Esports Hub · Tournament platform for organisers, built with Claude Code — Live · Frontend · API

Sign-ups, brackets, points tables, ID cards, live TV screens and gate check-in, with 14 free tools. Google sign-in verified against Google's public keys, email confirmation, and organisations with roles. About **640 browser tests** and about **600 API tests** run on every push.

Stack: Next.js 14, React, TypeScript, Hono on Cloudflare Workers, D1 (SQL), Playwright

SpotAI · AI gym app for Indian gyms, members and owners in one app — Android and iPhone apps · Case study

Squat form check runs on the phone: frames never leave it, only rep counts and flags reach the API. Row-level security per gym, tested with two gyms side by side. **190 API tests** and accessibility checks on every screen.

Stack: Vue 3, Capacitor, NestJS, PostgreSQL, MediaPipe, NATS

FitRank · Typed AI decisions for staffing — Live · Case study

A **23M-parameter** MiniLM model, distilled from a 9B teacher, returns a probability for every option; a manager makes every final call. Against a rules-only baseline on 35 held-out tasks: **hit@5 94%** vs 91%, and hit@1 54% vs 74%, which is why it shortlists rather than assigns. **93.5%** backend test coverage, **GitHub Actions** CI.

Stack: FastAPI, PostgreSQL and pgvector, PyTorch, React, Docker

streamerOS · Rust and Tauri 2 streaming cockpit, built with Claude Code — Website · Site and agent source

Reads chat on the streamer's own PC and switches OBS scenes when chat spikes, with no account and no cloud. **306 Rust tests**, reviewed and tested by me. The **Next.js** site and its support **agent** (**Hono** on **Cloudflare Workers**, Upstash Vector, Gemini) are public; the app repository is private.

Stack: Rust, Tauri 2, OBS WebSocket, Next.js, Hono, Upstash Vector

Path Saathi LMS · Client project, free vocational-skilling platform — Live (dev)

The brief and an MVP sketch arrived on a Monday; auth, course delivery for the client's videos and PDFs, and a self-paced flow were live on the dev environment the **next day**.

Stack: Next.js, React, Tailwind CSS, TypeScript

Also: Portfolio Concierge, a grounded RAG API and public **MCP** server that refuses to answer below a similarity threshold (try it) · IntegrateX, a React Flow and MongoDB workflow engine (sandbox) · Sable, a local-first AI finance agent where only a person commits (case study) · ShelfSight AI, FastAPI and YOLOv8, 650 tests (code). All projects: yk.streamerosai.com/projects.

OPEN SOURCE AND WRITING

Public code at github.com/Yaseenyk, including this portfolio, its RAG and MCP worker and the content pipelines. Engineering write-ups at yk.streamerosai.com/blog and dev.to/yaseenyk04.

EDUCATION

B.E. Computer Science, D. Y. Patil College of Engineering, 2021 · **Diploma, Computer Science**, D. Y. Patil Polytechnic, 2018

Certifications (Coding Ninjas): Front End Architecture, React Production Development, Advanced Backend Systems