

Osama Ansar

Senior Software Engineer — AI/RAG Systems & Full-Stack Platforms

📍 Lahore, Pakistan 📞 +92 348 4108105 ✉ osama.ansar.pk@gmail.com
🌐 linkedin.com/in/osama-ansar 📄 github.com/OsamaAnsar 🌐 osama-ansar.netlify.app

SUMMARY

Senior Software Engineer with 7+ years shipping scalable, high-performance web applications across SaaS, healthcare technology, logistics, streaming, and enterprise platforms — now specializing in TypeScript-based RAG systems, agentic pipelines, and LLM evaluation infrastructure. Delivered production AI features including a Model Context Protocol server and LLM-powered document extraction, alongside a self-directed portfolio of from-scratch RAG and vector-search implementations and an open-source contribution accepted into Weaviate's official TypeScript client.

TECHNICAL SKILLS

AI / RAG	Retrieval-Augmented Generation, Vercel AI SDK, LangChain.js, BM25, HNSW & vector search, LLM evaluation, prompt engineering, Model Context Protocol (MCP), OAuth 2.1 / PKCE
Languages	TypeScript, JavaScript, Python, SQL
Frontend	React, Next.js, Angular, Vue, RxJS, Tailwind CSS, SCSS
Backend	Node.js, Koa.js, Hono.js, Express.js, ASP.NET Core Web API
Data	PostgreSQL, Sequelize, Drizzle ORM, Entity Framework Core, MongoDB
Architecture	Microservices, microfrontends, REST APIs, real-time systems, distributed systems, SaaS
Cloud / DevOps	AWS, GCP, Cloudflare Workers, Docker, Kubernetes, Terraform, GitHub Actions, GitLab CI/CD, Jenkins
Payments / Streaming	Stripe, Montonio, Mollie, Mux, AWS IVS, Dolby, DRM

PROFESSIONAL EXPERIENCE

- MenuLogic K12 — Senior Software Engineer** Nov 2025 – Present · United States
 - Built the product's Model Context Protocol (MCP) server, exposing recipe, ingredient, and district tools to AI agents (Claude in Chat/Cowork) via a custom OAuth 2.1 + PKCE authorization-code flow and tenant-scoped tool authorization that mirrors the REST API's own role system.
 - Build and maintain full-stack features using Angular, standalone components, RxJS, and Kendo UI on the frontend with ASP.NET Core Web API and Entity Framework Core on the backend.
 - Develop USDA-compliance reporting features with Telerik Reporting, including report data pipelines, EF Core queries, REST data-source endpoints, and parameter-driven report definitions.
 - Implement cross-application SSO between a legacy MVC shell and an Angular SPA using JWT token handoff via iframe messaging.

- Fix UX and data-integrity issues across list views, form validation, and menu, recipe, and ingredient management workflows.

Truckpedia — Software Engineer

Jan 2024 – Oct 2025 · United States

- Led frontend development using React and Vue across a logistics platform.
- Integrated AI-driven features including OpenAI/LLM-powered prompt engineering and knowledge-base chat for intelligent, context-aware interactions.
- Optimized scalability and user performance and implemented real-time updates across multiple devices.
- Collaborated across teams, led feature implementation, mentored junior engineers, and maintained code quality.

Omnivex Corporation — Software Engineer

May 2022 – Dec 2023 · Canada

- Developed custom UI components for the Ink platform using Angular, React, and Vue.
- Built real-time user experiences with SignalR and improved component reusability while reducing code duplication.
- Led bug triage for ASP.NET and Angular frontend issues.

Contrive Solutions — Senior JavaScript Developer

Jan 2020 – Jun 2022 · Pakistan

- Built Node.js automation for S4 Salesforce bug reporting.
- Developed the GoOrder food-delivery application using Vue/Vuex and the Penmi calendar/CRUD features using Angular.
- Built reusable UI components across seven products and shared Angular microfrontend component libraries for Finlex.

Webicosoft — JavaScript Developer

May 2019 – Nov 2019 · Pakistan

- Developed full-stack features for the Blunk job-matching platform using React, Redux, Semantic UI, Node.js, and MongoDB.

AI / RAG ENGINEERING – INDEPENDENT PROJECTS

rag-eval-harness

github.com/OsamaAnsar/rag-eval-harness

LLM-judge evaluation harness for RAG pipelines — faithfulness, answer relevance, context precision, and context recall — with a CI regression gate that catches planted hallucinations in a benchmark dataset. Pluggable Judge interface (real OpenAI judge plus a dependency-free FakeJudge for CI).

agentic-rag

github.com/OsamaAnsar/agentic-rag

Multi-hop agentic RAG pipeline: decomposes complex questions into sub-questions, retrieves evidence with a from-scratch BM25 implementation, checks answer sufficiency, and synthesizes a cited final answer.

hnsvector-index

github.com/OsamaAnsar/hnsvector-index

HNSW (Hierarchical Navigable Small World) approximate-nearest-neighbor index implemented from scratch, no ANN library — recall@k climbs from 66% to 100% against a brute-force baseline as efSearch increases.

generative-ui-chat · doc-chat-citations

Two Vercel AI SDK tool-calling chat assistants: one streams live, interactive React components (charts, tables, booking cards) instead of plain text; the other grounds every answer in from-scratch BM25 retrieval with clickable, source-linked citations.

OPEN SOURCE

weaviate/typescript-client

github.com/weaviate/typescript-client

Diagnosed and fixed a silent data-loss bug in Weaviate's official TypeScript client — `data.ingest()` accepted the documented input shape but stored it empty — with regression tests; submitted upstream.

SELECTED CLIENT PROJECTS

GoStork

AI-powered fertility services marketplace

Built an AI-powered PDF extraction pipeline using GPT-4 to transform unstructured agency documents into standardized, searchable donor and provider profiles — structured data mapping, cost/compensation analysis, image analysis, and summary generation. Built real-time messaging infrastructure; integrated PostgreSQL/Sequelize, HubSpot, Stripe, Google Calendar, Outlook, Twilio, and SendGrid on Docker/Kubernetes/GCP.

Streamflow / LiveArena

Pay-per-view live streaming platform

Built DRM-protected streaming and an event-based pricing engine for an enterprise PPV platform designed for up to 500,000 concurrent users. Implemented real-time payment celebrations and dynamic watermarking via Pusher, resumable S3 uploads with TUS/Uppy, and seamless player switching during buffering. Stack: Next.js/React, Hono, Cloudflare Workers, PostgreSQL, AWS Lambda/SQS, Stripe, Mux/AWS IVS.

EDUCATION

BS Software Engineering

University of Gujrat · 2015–2019 · CGPA 3.7/4.0

Artificial Intelligence · Machine Learning · Data Structures & Algorithms · Database Management Systems · Software Engineering

AWARDS

- Best Performance Award — Contrive Solutions
- 1st Place — Speed Programming Competition
- Merit Scholarship Recipient
- UOG Topper — 1st Position
- Selected for Business Incubation Program