

# MUHAMMAD USMAN HAIDER

AI Engineer & Full-Stack Developer

Lahore, Pakistan | +92 322 1151832 | [usman@usmancodes.com](mailto:usman@usmancodes.com)

[Github](#) | [Linkedin](#) | [Portfolio](#)

## PROFESSIONAL SUMMARY

---

Full-stack developer with 3+ years of experience and a specialization in AI-powered systems including RAG architectures, LLM evaluation pipelines, prompt engineering, and real-time voice agents. Built and shipped production AI features on platforms serving thousands of users across healthcare, sales coaching, and brand-intelligence domains. Strong backend foundation in Node.js, NestJS, and microservices, paired with hands-on experience integrating GPT-4, Claude, Gemini, and vector databases like Pinecone. Comfortable working asynchronously across distributed teams.

## PROFESSIONAL EXPERIENCE

---

**Meissasoft** Lahore, Pakistan

*Full-Stack MERN Developer* June 2024 – Present

*Software development company building scalable web applications and AI-powered solutions.*

- CloserCoach (AI sales training platform): Architect and develop the backend infrastructure using Node.js, Express.js, and Firebase, supporting an AI-powered sales coaching product with daily active enterprise users.
- Integrate VAPI for real-time voice call processing and Langfuse for prompt management and evaluation, enabling AI-driven analysis of sales conversations with sub-2-second feedback latency.
- Design and ship the admin dashboard with real-time analytics, user management, call monitoring, and billing views using Firebase Realtime Database.
- Build a webhook ingestion layer for VAPI callbacks with retry logic and idempotency handling, processing voice transcriptions and AI-generated feedback reliably.
- Develop an AI text-enhancement tool using OpenAI GPT models to improve grammar, tone, and clarity of user-submitted scripts in real time.
- Build a property booking platform with availability management, Stripe payment integration, and automated host-guest messaging.
- Implement a serverless backend using the SST framework with AWS Lambda, API Gateway, and DynamoDB to reduce infrastructure costs and scale independently.
- Build an intelligent form auto-fill bot using NLP and pattern matching to extract structured data from PDFs, emails, and spreadsheets.

**12th Spring Inc. (Phelix AI Healthcare Automation)** Lahore, Pakistan

*Backend Developer (NestJS)* May 2023 – June 2024

*IT solutions company focused on healthcare automation products.*

- Developed and maintained 3 microservices for fax processing and appointment management using NestJS and TypeScript, as part of a 20-person distributed team across the US, India, and Pakistan.
- Implemented event-driven communication between microservices using Apache Kafka for real-time healthcare document processing, significantly reducing manual triage time.
- Designed a polyglot persistence layer using MongoDB for document storage, PostgreSQL for relational data, and Redis for caching, improving query latency on hot paths.
- Built RESTful APIs and integrated with major EMR systems (Oscar, Accuro, Athena Health) to enable healthcare data exchange and automated fax triage for clinical workflows.
- Collaborated in Agile sprints with cross-functional teams across 3 time zones, consistently shipping features on time while maintaining high API uptime for enterprise healthcare clients.

## NOTABLE PROJECTS

---

### RankBee.ai — AI Visibility & Brand Optimization Platform

*Full-Stack Developer (Contract) Feb 2025 – Jul 2025*

- Built a full-stack AI optimization platform (founded by an ex-Amazon Head of SEO) tracking brand visibility and citations across ChatGPT, Claude, Gemini, and Perplexity.
- Developed the backend in NestJS with PostgreSQL and a Bull queue system for brand tracking, prompt testing, and async job orchestration over multi-LLM workloads.
- Architected the marketing site and analytics dashboard in Next.js with real-time data visualization for monitoring brand perception across LLM platforms.
- Integrated Sentry for error tracking, Datadog for APM and infrastructure monitoring, Intercom for customer support, and Clerk for authentication.
- Integrated OpenAI GPT-4, Claude, Gemini, and Perplexity APIs to deliver automated brand visibility insights, competitive benchmarking, and reputation alerts.

### RAG-Powered Document Q&A System

*Personal Project Oct 2025 – Nov 2025*

- Built a full-stack RAG system using Next.js, LangChain, and Pinecone for intelligent Q&A across personal document libraries with semantic search.
- Implemented a document ingestion pipeline supporting PDF, DOCX, TXT, and Markdown with chunking strategies and OpenAI text-embedding-3-small for vector embeddings.
- Architectural semantic retrieval over Pinecone with GPT-4-powered answer generation and source citations to keep responses grounded.
- Developed a Next.js frontend with drag-and-drop file upload, streaming query responses, and conversation history.

## TECHNICAL SKILLS

---

Languages: JavaScript, TypeScript, Python

Frontend: React, Next.js, Redux, Tailwind CSS

**Backend:** Node.js, Express.js, NestJS, FastAPI, REST APIs, Microservices, Webhooks

Databases: PostgreSQL, MongoDB, Firebase, Redis, DynamoDB

AI / LLM: OpenAI GPT-4, Claude, Gemini, Perplexity, LangChain, Langfuse, VAPI, RAG architecture, prompt engineering, LLM evaluation

Vector & Embeddings: Pinecone, OpenAI text-embedding-3-small, semantic search

Cloud & DevOps: AWS (Lambda, S3, EC2, API Gateway, CloudWatch), SST Framework, Docker, GitHub Actions, CI/CD

Monitoring & Tooling: Sentry, Datadog, Intercom, Clerk, Stripe, Postman, Jest, Git, Jira

## EDUCATION

---

**Government College University Faisalabad** Faisalabad, Pakistan

*Bachelor of Computer Science 2018 – 2022*