Khursheed Ahmed

AI Engineer · Full Stack Developer

Lahore, Pakistan · +92 341 7122946 · khursheed6577@gmail.com

linkedin.com/in/khursheed-ahmed-software-engineer · github.com/khursheedahmedd

Nationality: Pakistani · Visa status: outside KSA, requires employer-sponsored work visa · Available to relocate to Saudi Arabia

Summary

AI Engineer with 3+ years of experience shipping LLM products to production: multi-agent systems, RAG pipelines and voice AI, on a full stack of FastAPI, Next.js and PostgreSQL. Founder and tech lead of Smart Hire AI, an AI recruitment platform with a paying UK customer and two proof-of-concept deployments. Microsoft Imagine Cup 2025 Global Semifinalist, and ranked 1st in Pakistan and 31st worldwide in Meta Hacker Cup 2025 Round 1. Based in Lahore and ready to relocate to Saudi Arabia.

TECHNICAL SKILLS

AI / LLM: LangChain, LangGraph, LlamaIndex, Google ADK, Model Context Protocol (MCP), RAG pipelines, multi-agent systems, vector search, Cohere and OpenAI embeddings, LLM orchestration (Azure OpenAI, Gemini, Groq, Ollama), Vapi and ElevenLabs voice agents, Hugging Face, Langfuse
Languages: Python, TypeScript, JavaScript, SQL
Backend: FastAPI, Node.js, Express.js, SQLAlchemy, Prisma, ARQ workers, REST APIs, microservices and N-tier architectures
Frontend: React, Next.js, React Native, Tailwind CSS, shadcn/ui, Radix UI
Databases: PostgreSQL, MongoDB, Redis, DuckDB, FAISS, ChromaDB, pgvector, Cosmos DB, Supabase, RBAC
Cloud & DevOps: Azure (VMs, AI Studio, Container Apps), GCP, Docker, Nginx, GitHub Actions CI/CD, Railway, Vercel, Firebase
Integrations: Meta WhatsApp Cloud API, Twilio, Stripe, Zoom Server-to-Server OAuth, OAuth 2.0, Clerk, NextAuth

WORK EXPERIENCE

Founder and Tech Lead

October 2025 – Present

Smart Hire AI

AI recruitment platform covering sourcing, screening and interviews · Next.js · FastAPI · LangGraph · Cohere · Groq · DuckDB · PostgreSQL/Prisma · Redis/ARQ · Vapi · Meta WhatsApp Cloud API · Railway

- Built the product end to end and took it to revenue: **one paying customer (NewBridgeFX, UK)** and proof-of-concept deployments with Innoworq and Seligo Consultants.
- Architected an **N-tier system**: a Next.js web app, two FastAPI services (business domain and AI/RAG) and PostgreSQL, automating recruitment from WhatsApp and voice sourcing through AI screening to interview analytics.
- Engineered **natural-language candidate search** with LangGraph workflows, **Cohere embeddings** and a re-ranked DuckDB vector index. Added circuit breakers, query caching and **hybrid local/cloud LLM routing** (Ollama and Groq) with confidence-gated fallback to keep latency stable under load.
- Built an **AI voice interview system on Vapi** with real-time transcription, MediaPipe/TF.js video proctoring and rubric-driven scoring of each answer.
- Implemented candidate outreach over **WhatsApp Business (Meta Cloud API) and voice** (Twilio and Vapi), with multi-language detection, fake-candidate scoring and opt-out handling.
- Built a **Redis-backed ARQ task queue** for bulk resume parsing, screening and re-evaluation.
- Won the **Best Final Year Project Award**, Department of Software Engineering, UMT (2026).

AI Engineer and Full Stack Developer

October 2025 – August 2026

Introva

- Designed and deployed **multi-agent systems** for task automation, data analysis, agent evaluation and workflow orchestration using LangGraph, Google ADK and the Model Context Protocol.
- Built **RAG pipelines** over large document and knowledge bases, returning context-aware answers with grounded citations.
- Built an **OCR pipeline for right-to-left, multi-column historical documents** that processed thousands of pages at **96% accuracy**, combining Google Vision API, custom layout detection and a post-processing engine backed by a 23,000-word lexicon.
- Shipped Rayfold, a **generative AI platform** that turns resumes into customisable web portfolios, with an embedded **AI code editor** that changes generated UI components from natural-language prompts.

Software Engineer *May 2023 – October 2025*
Eventus Sol, Lahore

- Delivered **multiple MERN-stack and FastAPI applications** and integrated GenAI features (RAG chatbots, semantic search, LLM-assisted content generation) into live customer products.
- Owned full-stack features end to end: **REST API design**, frontend performance, cross-browser QA and Agile sprint delivery against client deadlines, working directly with designers, stakeholders and clients.
- Built and maintained backend services on Node.js/Express and FastAPI, integrating vector databases (ChromaDB, pgvector), Stripe payments and AI providers (Azure AI Studio, Groq, OpenAI, Vapi).

Generative AI Engineer (Intern) *May 2024 – October 2024*
Circle AI, Remote

SELECTED PROJECTS

Smart Grade AI — *Co-Founder and Tech Lead* *November 2024 – Present*
AI-powered LMS for automated assessment and RAG-based mentoring · React.js · TypeScript · Node/Express · FastAPI · LangGraph · Azure OpenAI · Mistral OCR · MongoDB/Cosmos DB · Docker · Azure Container Apps

- **Microsoft Imagine Cup 2025 Global Semifinalist** and the only South Asian team to qualify; awarded **$25,000 in Azure credits** through Microsoft for Startups. Accepted for integration into the UMT learning management system.
- Built a **rubric-based AI grading engine** for text answers, code submissions and diagram-heavy papers, producing question-level marks, detailed feedback and structured JSON for analytics and RAG.
- Engineered a **per-student AI mentor** with a personal vector store that ingests every graded assessment, surfaces weakness patterns and recommends study plans.
- Built an **AI teacher assistant** for class-wide insights, struggling-student detection and remedial planning.

Other Projects

- **Vibe Terminal** *(Python · LangGraph · ChromaDB · Flask)* — a cross-platform terminal that runs shell commands from natural language through multi-step agent workflows, with voice input and output.
- **ClipSync** *(React · React Native · Kotlin · Supabase · Stripe · Socket.io)* — real-time clipboard sync, file sharing and shared workspaces across Android, web and desktop, with Stripe subscription billing.
- **StreamlyAI** *(MERN · Flask · LangChain · ChromaDB · AssemblyAI · Groq · Cohere)* — turns YouTube videos into searchable, conversational content with transcription, semantic chunking, vector search and LLM Q&A.

AWARDS

- **Meta Hacker Cup 2025:** ranked 1st in Pakistan and 31st worldwide in Round 1, among roughly 13,800 participants.
- **Microsoft Imagine Cup 2025:** Global Semifinalist with Smart Grade AI; the only South Asian team to qualify.
- **Google Gemma 2 AI Hackathon, London 2024:** Global Finalist.
- **IBM watsonx Generative AI Hackathon 2024:** 2nd Runner-Up ($3,500 in funding).
- **NASA Space Apps Challenge 2025:** Honorable Mention (local award).
- **Best Final Year Project Award 2026:** Department of Software Engineering, UMT.

EDUCATION

BS in Software Engineering *October 2022 – July 2026*
University of Management and Technology (UMT), Lahore *CGPA: 3.33 / 4.0*

- **Thesis:** Smart Hire AI, an agent-driven automated recruitment system with AI interview capabilities.
- **Teaching Assistant** for Object-Oriented Programming and Discrete Structures (Fall 2025).
- **Coursework:** Machine Learning, NLP, Data Sciences, Deep Learning and Neural Networks, Data Structures and Algorithms, System Design and Architecture, Database Systems.

CERTIFICATIONS & LEADERSHIP

- **Certified Cloud Generative AI Engineer**, PIAIC (2025)
- **Director of Management**, UMT Anti-Narcotics Society (2025)
- **Head of Volunteers**, UMT Social Welfare Society (2023)

LANGUAGES

English (professional), Urdu (native)