FOUAD IKRAM

AI Product Engineer

Islamabad, Pakistan · Open to remote work and relocation
fouadikram007@gmail.com · fouadikram.com · linkedin.com/in/fouad-ikram-lxix · github.com/fouad-lxix

SUMMARY

AI product engineer working across the full stack, who designs, builds and deploys production AI systems end to end: RAG platforms, AI agents, multimodal data pipelines, analytics platforms, and voice and video automation. I lead a 3-person team, work directly with clients to turn their needs into working systems, and own delivery from the data model and AI design through the interface to deployment and security.

TECHNICAL SKILLS

AI & ML: LLM APIs, LangChain, RAG, AI agents, Tool calling, Embeddings, Vector search, Computer vision, Voice AI
Frontend: React, Next.js, TypeScript, JavaScript, Tailwind CSS
Backend: FastAPI, Flask, Node.js, Express, Go, REST APIs, GraphQL, Celery, SQL
Data & Cloud: PostgreSQL, pgvector, ChromaDB, MongoDB, Redis, Azure, AWS, Docker, Nginx, Linux VPS, CI/CD, GitHub Actions

EXPERIENCE

AI Product Engineer · TechQuest.ai, Islamabad, Pakistan *Jan 2025 – Present*

AI Inventory Intelligence Platform *Full-stack RAG application*

- Designed and built, end to end, a full-stack platform that lets teams ask questions about their private inventory data in **plain English** and get accurate answers drawn from their own records.
- Built a custom response engine that produces **live charts and analytical reports inside the AI's answers**, so users get a finished report instead of raw text.
- **Proactive risk detection** flags at-risk stock early, such as items running low or items with safety concerns, before anyone has to go looking.

Multimodal AI Billboard Data Platform *Deployed on Azure*

- Users upload messy data in every format, including **Excel, CSV, PDF files and web links**, describing billboards and their photos; the platform cleans and structures it automatically.
- Uses **vision AI to match each photo to its exact billboard** and to detect photos that aren't billboards at all, replacing slow manual review.

Stack: Azure Function Apps, Azure VM, Azure PostgreSQL Flexible Server, Key Vault, multimodal LLMs

Location-Based AI Research & Video Engine *Led a 3-person team*

- **Worked with the client in requirement meetings** to shape the end product, then translated their needs into the system design.
- A user picks a target area; the system researches it across **Google, YouTube and Reddit** to surface local trends, market gaps, what people are talking about and the questions they can't find answers to.
- AI turns those findings into video scripts, then the system produces **AI avatar videos with lip-sync**, renders them as parallel jobs with automatic retries, stitches the pieces together and publishes to social platforms automatically.
- Led the build **end to end, from architecture to production**, and handled the deployment myself: set up the client's VPS, connected the domain and configured Nginx to route traffic.

Stack: Python, LLMs, HeyGen avatars and lip-sync, Redis rate limiting, parallel job processing, Linux VPS, Nginx

Multi-Channel Ad Campaign Tracking Platform *Full-stack, deployed on Azure*

- Built and deployed a full-stack platform for a high-traffic media company that brings together **spend, impressions, locations, billboards, line items and creative assets from 5 ad platforms** into one consistent view.
- Used daily to track campaigns across billboards, mobile and TV screens, **processing millions of impressions a day**; most of the engineering went into normalizing data that each platform reports differently.
- **Worked directly with the client's team** to pinpoint slow processes and reshape workflows and the UI around how they actually work.
- Treated security as a core feature: application-level checks, firewall rules, **Defender for Cloud, Key Vault**, logging and alerts on traffic spikes and failures. **Passed an independent third-party security audit.**

  Stack: Azure Web Apps, Azure Function Apps, Azure PostgreSQL Flexible Server, Key Vault, Defender for Cloud

AI Video Production Plugin for Claude *Loom-style walkthroughs, fully automated*

- A Claude plugin that produces narrated product walkthrough videos end to end: it reads the project, **writes a script for approval**, plans what to show on screen, then records and renders a polished MP4.
- Narration from **ElevenLabs voice AI is synced word by word to on-screen actions**, with automatic zoom, chapter cards and a title card (Node.js, Playwright, ffmpeg).

AI Engineering Intern · TechQuest.ai, Islamabad, Pakistan *Jun 2024 – Aug 2024*

Autonomous AI Moderator for Discord

- An AI moderator that does the full job of a moderation team, including **warnings, kicks, bans and member management**, reading conversations in real time and escalating step by step, around the clock.

AI Sales Outreach Platform

- Learns a business from its website, documents and links, including **what it sells and its brand colors, tone and style**, then writes personalized outreach emails in the company's own voice.

PROJECTS

Propla *AI SaaS for client proposals · built and launched solo · github.com/fouad-lxix/Propla*

- Turns a short brief into a complete, client-ready proposal using **structured LLM tool use** for reliable output, delivered as a tracked web page with e-signing and payments.
- **Section-level engagement analytics** show the sender exactly which parts of a proposal the client read. I owned everything from database design to Stripe billing and deployment.

Spelunk *Autonomous codebase Q&A agent · github.com/fouad-lxix/spelunk*

- Answers questions about an unfamiliar codebase **with no index or embeddings**: it searches, reads a slice and decides what to read next, citing the exact files and lines it used.
- A **verification ledger** checks every citation against what the agent actually read and rejects any it can't back up, so it cannot cite a file it never opened.
- Measured on live runs: correct after reading **under 2% of a 28,000-line app**, at about $0.09 per question.

Seatbelt *AI API gateway in Go · github.com/fouad-lxix/Seatbelt*

- Sits between an app and the Claude and GPT APIs (adopting it takes **one URL change**) and enforces a **hard monthly spend cap** that survives restarts, queues bursts with **per-provider worker pools and a hand-written token-bucket rate limiter**, and gives an immediate, clear answer when the queue is full.
- Retries temporary failures with backoff while returning real errors immediately, supports streaming (stopping when the client disconnects), and serves a **live stats page**.

EDUCATION

BS Computer Science · National University of Computer & Emerging Sciences (FAST-NUCES) *2021 – 2025*

Relevant coursework: Artificial Intelligence, Cloud Computing, Information Security, Blockchain & Cryptocurrency.

Languages: English (Full Professional) · Urdu (Native) · Arabic (Elementary)