

Ila Rehman

+91-8447354282 | ila.rehman.khan@gmail.com | linkedin.com/in/ila-rehman | github.com/IlaKhan17 | ilarehman.com

Summary

MCA graduate and self-taught AI engineer with a physics background, focused on shipping LLM-powered products that work in production. Built an evidence-grounded AI sales agent and a retrieval-augmented (**RAG**) study tool, and currently build **AWS** microservices and **LLM**-powered agents as a software engineer intern.

Experience

Manrke

April 2026 – Present

Software Engineer Intern

Remote

- Design and build the backend for an end-to-end document-automation platform on **AWS (Lambda, ECS, Step Functions, SQS/SNS, DynamoDB, RDS)**, engineered for reliable, exactly-once processing at scale.
- Integrate **LLMs** into production pipelines for document extraction, classification, and natural-language generation, with guardrails and deterministic fallbacks that keep outputs reliable and costs controlled.
- Built a multi-tenant WhatsApp travel-booking agent (**FastAPI, LangChain, Twilio, Duffel**) with strict per-client data isolation and a memory design that keeps conversations accurate without inventing prices or details.
- Made the agent production-reliable by diagnosing a **56% LLM** tool-call failure rate from traces and eliminating it with a provider-fallback chain and date-grounding fixes.

Projects

AI SDR (Davis) [\[code\]](#) [\[live demo\]](#)

Next.js, React, TypeScript, FastAPI, LangGraph, OpenAI, Supabase, PostgreSQL, Pinecone, Redis, Braintrust

- Built an evidence-grounded AI sales-development agent (**LangGraph, FastAPI, OpenAI**) that researches prospects via web search and **Pinecone** retrieval; restricted drafting context to a human-approved claims table so unverified personalization is unreachable by construction, not merely discouraged by prompt.
- Made lead scoring explainable and reproducible by separating **LLM** signal-extraction from a deterministic **Python** scorer over **6** weighted ICP dimensions, turning an untestable model output into unit-tested logic where identical inputs always score the same and a rep sees exactly why a lead ranked where it did.
- Shipped it as a safe, multi-tenant product: an approval-first send pipeline with a **6**-rule policy guard and DB-enforced idempotency making duplicate sends structurally impossible, a **12**-intent reply classifier gated in CI at **1.0** unsubscribe recall, and **PostgreSQL row-level security** leaving **0** tenant tables unprotected.

AdaptQuiz API [\[code\]](#) [\[live demo\]](#)

Python, FastAPI, LangChain, FAISS, GPT-4o, Docker, Railway

- Built and deployed a **RAG**-powered study tool that turns any document into an adaptive quiz, generating every question from the source material so nothing is hallucinated.
- Added an **LLM-as-judge** grader (**GPT-4o**) that scores free-text answers against a **3**-criterion rubric with partial credit and pinpoints each learner's specific knowledge gaps, closer to a human tutor than a right/wrong checker.
- Ran the full retrieval pipeline on local embeddings (**FAISS**), so the product runs at **\$0** external AI-API cost.

Technical Skills

Languages: Python, SQL, JavaScript / TypeScript

AI & LLM: LLM APIs (OpenAI GPT-4o, Groq, Anthropic Claude), RAG (Retrieval-Augmented Generation), LangChain, LangGraph, AI Agents, Prompt Engineering, Structured Outputs, Tool / Function Calling, LLM-as-Judge Evaluation, Embeddings (sentence-transformers), Semantic Search, FAISS, Pinecone

Backend: FastAPI, REST API Design, Pydantic, SQLAlchemy / Alembic, PostgreSQL, Supabase, DynamoDB, Redis, Event-Driven Microservices, Multi-Tenant Architecture, Row-Level Security, Next.js

Cloud & DevOps: AWS (Lambda, ECS, Step Functions, SQS/SNS, S3, DynamoDB, RDS, ECR, CloudWatch), Docker, Railway, Vercel, Git & GitHub, CI/CD

Tools & Foundations: pytest, Braintrust, OpenAPI / Swagger, NumPy, Pandas, ML Evaluation Metrics, Linear Algebra, Statistics

Education

Master of Computer Applications (MCA)

2022 – 2025

Indira Gandhi National Open University (IGNOU), New Delhi

Coursework: Software Engineering, Data Science, Cloud Computing, Artificial Intelligence & Machine Learning

B.Sc. (Hons) Physics with Mathematics

2017 – 2020

Jamia Millia Islamia, New Delhi

First Class Honours