

AMIR AIJAZ

Karachi, Pk | +923233232974 | amirejaz790@gmail.com
GitHub: <https://www.github.com/amiraijaz>
LinkedIn: <https://www.linkedin.com/in/amir-aijaz-01a134233>
Portfolio: <https://amiraijaz.com>

SUMMARY

AI Engineer with two years of experience taking LLM, RAG, and agentic systems from prototype to production. I build full-stack AI in Python, FastAPI, React, and Docker, with a focus on scalable multi-tenant architecture, rigorous test discipline, and cost-efficient cloud infrastructure, drawing on vector databases, the Model Context Protocol (MCP), and Redis. Recent work includes a multi-tenant agentic commerce engine and a voice lead-generation system backed by 500+ passing tests.

EDUCATION

Bachelor of Science in Computer Science — **University of Karachi**

01-January-2021 - 31-December-2024

- **Final Year Project:** Sign Language Translator (Built a real-time sign language recognition system using MediaPipe and LSTM to classify gestures from facial, pose, and hand landmarks.)
- **Modules:**
 - Artificial Intelligence
 - Data Warehousing & Data Mining
 - Statistics & Data Analysis
 - Natural Language Processing
 - Data Structures

TECHNICAL SKILLS

- **AI & Computer Vision** (PyTorch, Hugging Face, Object Detection, Image Segmentation, YOLOv8, NLP, Transformers)
- **Generative AI & Voice Agents** (LLM Integrations, RAG, LangChain/LangGraph, Multi-Agent Systems, Vapi, ElevenLabs, Vector DBs, BGE-M3 Embeddings, Prompt Engineering)
- **Full-Stack Development** (Python, FastAPI, React.js, Next.js, Typescript, Flask, PostgreSQL, MongoDB, MySQL, Multi-Tenancy/RLS, REST APIs)
- **Cloud & Automation** (Docker, AWS, Render, Vercel, Supabase, Git, CI/CD, Pytest, n8n)

PROFESSIONAL EXPERIENCES

AI ENGINEER

Viral Tech - Remote

01-August-2025 - Present

- Designed and built a production-grade multi-tenant agentic commerce engine with Postgres row-level-security tenant isolation, a WooCommerce adapter, a PII firewall, and an owner-facing dashboard, validated against a live store of 300+ products
- Implemented semantic product retrieval using BGE-M3 embeddings on GPU and a cost-optimized LLM routing layer to keep per-conversation inference cost low at scale
- Built Knowlyst, an AI system that ingests web and video links, then scrapes, stores, and answers questions over the content using RAG
- Conducted research on emerging LLMs, benchmarking capability, latency, and inference cost for model selection
- Automated internal workflows and project delivery using n8n, Zapier, Make, and ClickUp

AI ENGINEER

TechThrone Innovations Pvt. Ltd. – Karachi, Pakistan

01-May-2025 – 31-July-2025

- Built a local LLM, similar to ChatGPT, for content and social media management
- Developed AI tools including text-to-image generation, HR screening automation, and lead-generating agents
- Cleaned and analyzed geospatial and business data from major companies like CBL and EBM
- Deployed chatbots and AI solutions to drive agency-wide AI transformation

Jr. AI ENGINEER

CodeNexo – Karachi, Pakistan

01-January-2025 – 30-April-2025

- Developed AI tools for newsletter/blog generation using Claude API, Flask, and YouTube data
- Automated workflows via Asana and deployed scalable apps on AWS
- Built internal tools for content automation and team productivity with LLMs and prompt engineering

RESEARCH & INTERNSHIP EXPERIENCE

Python and Computer Vision Development

The Disrupt Labs – Karachi, Pakistan

01-June-2024 – 1-October-2024

- Researched and Developed Python Scripts to measure 2D object distances in warehouses.
- Applied open-source depth estimation models to real-time CCTV feeds.

Artificial Intelligence & NLP

Smart City Lab (NCAI) – Karachi, Pakistan

05-December-2023 – 30-May-2024

- Built an NLP driven Arabic Medical chatbot
- Benchmarked speech models (Whisper, Wav2Vec) to evaluate transcription accuracy for Urdu.

PERSONAL PROJECTS

1. Vigil - AI Governance Dashboard:

Open-source LLM observability and safety platform with a model registry, real-time audit logging, automated PII and toxicity detection, cost tracking across 250+ models, and one-click NIST AI RMF compliance reports. Built with FastAPI, React, PostgreSQL, Redis, and Docker.

GitHub: <https://github.com/amiraijaz/ai-governance-dashboard>

2. Legal Mind AI - RAG-Powered Legal Document Assistant:

AI system that helps lawyers instantly extract accurate insights from large legal documents using natural language queries, built with Retrieval-Augmented Generation architecture.

Live: <https://legal-mind-ai-assistant.vercel.app/>

3. AI Voice Lead-Generation Agent:

Production outbound/inbound voice agent for automated phone qualification and warm transfer, built with Vapi, FastAPI, and Next.js. Includes a regulatory compliance gate, a live warm-transfer agent pool, and call-screening detection, backed by a hermetic suite of 537 passing tests.

4. Crypto Algorithmic Trading Bot:

Quantitative trading system that backtests and deploys market-neutral strategies on Binance with Python and Pandas. Integrated TradingView through a Model Context Protocol (MCP) server connected to Claude Code, and built an out-of-sample evaluation pipeline over funding-rate, open-interest, and basis history to validate funding-rate carry before live capital allocation.

PUBLICATIONS

1. **aigovkit** — Python SDK for AI governance and LLM observability, published on PyPI.
2. **governance-bench-v1** — 200-case benchmark dataset for evaluating AI governance scanners (PII, hallucination, prompt-injection resistance), published on Hugging Face under CC-BY-4.0.

Link: <https://huggingface.co/datasets/Vigil-ai/governance-bench-v1>

CERTIFICATIONS

1. **Artificial Intelligence on Microsoft Azure** – Provided by Microsoft (coursera.org/verify/U5MS6GBQ9R6G)
2. **Python for Data Science, AI & Development** – Provided by IBM (coursera.org/verify/S58MWS5MGAVD)

LANGUAGES

1. English Proficiency – IELTS (Band Score = 6.5 and CEFR Level = B2)
2. Urdu (Native)
3. Punjabi