Hassan Abbas

+92 313 3722595 hassanabbas2975@gmail.com linkedin.com/in/hassan-abbas-arain github.com/hassan-abbas-arain

Professional Summary

Results-driven AI engineer with hands-on experience in building LLM-integrated applications, RAG pipelines, and production-grade AI services. Proficient in Python, FastAPI, Flask, and containerization with Docker and CI/CD. Skilled in LangChain, Hugging Face Transformers, and vector databases for scalable retrieval-augmented generation and chatbot systems. Experienced in deploying secure APIs with JWT, implementing AES encryption, and integrating guardrails and prompt injection protection. Strong background in NLP, OCR (PaddleOCR), BioBERT, and real-time multi-agent systems using YOLOv8 and ArcFace. Passionate about responsible AI, model evaluation, and designing intelligent solutions that deliver real-world impact.

Education

FAST NUCES Karachi **Aug. 2021 – May. 2025**
Bachelor in Computer Science *Karachi, Pakistan*

Technical Skills

- OOP
- SQL
- Python
- RESTful APIs Development Integration
- FastAPI
- LangChain
- LLM
- Retrieval-augmented generation (RAG)
- OpenAI
- Hugging Face
- Deep Learning
- YOLO
- Paddle-OCR
- PyTorch
- Torch
- OpenCV
- Vector Embeddings Search (FAISS)
- CI/CD deployment pipelines
- Docker

Experience

AI Engineer at Teknoloje Solutions **Oct. 2024 – Present**
Karachi, Pakistan

AI-Powered HR Chatbot & Analytics Dashboard

* Built a secure **Flask backend** with **JWT authentication** and RESTful APIs for employee login, analytics, and chatbot access.
* Fine-tuned a **DistilBERT transformer model** on a custom HR policy dataset to classify employee queries by policy type.
* Implemented chatbot logic to generate dynamic responses based on the **employee database** (e.g., leave or loan eligibility).
* Parsed HR policy **PDF documents** using `PyPDF2` and extracted relevant sections with `RapidFuzz` for fuzzy keyword matching.
* Created interactive API endpoints for HR dashboards visualizing employee **performance**, **attrition**, and **department-wise statistics**.
* Secured sensitive data using **AES encryption (CBC mode)** and implemented session-based route protection.
* Saved and deployed the transformer model along with tokenizer and label encoder using `Joblib` for persistent predictions.

TekIndoor - Verification System

* Built an intelligent document verification pipeline using **PaddleOCR** to extract and validate IBAN and cheque numbers from scanned cheque images provided by NIFT.
* Integrated a **Flask-based service** with dynamic rule-based matching and **regex parsing** to cross-verify OCR results with structured financial data from the NIFT database.
* Enhanced document processing accuracy by tuning OCR engine parameters and implementing logic to handle multiple test cases and data inconsistencies.
* Enabled secure and automated cheque validation for internal bank systems, reducing manual errors and accelerating processing time.

Smart Surveillance – Multi-Camera Face Recognition and People Tracking System

* Developed a real-time surveillance system using **YOLOv8** and **ArcFace (InsightFace)** for person detection and face recognition across multiple video streams.
* Built a multi-threaded processing system with individual trackers for each camera using `OpenCV` and `cvzone`.
* Implemented **cosine similarity** on ArcFace embeddings to match detected faces against known identities.
* Designed polygon-based area logic for counting entries and exits and visualizing tracked movement using bounding boxes and IDs.
* Integrated Flask APIs to control camera monitoring endpoints and manage tracking state across sessions.
* Ensured thread-safe recognition using `queue.Queue` and `threading.Lock` for consistent face labeling across frames.

Projects

CureBridge – AI-Powered Telehealth Platform **Aug. 2024 – May 2025**

Final Year Project – FAST NUCES, Karachi

* Designed and implemented an AI-integrated telehealth platform combining real-time video consultations, transcription, and intelligent diagnostic support.
* Developed a real-time transcription pipeline using **Deepgram Speech-to-Text API**, enabling live capture of doctor-patient conversations.
* Built a symptom extraction engine using a fine-tuned **BioBERT NER model** to identify medical terms from live transcriptions.
* Engineered a **stacked ensemble model** (Random Forest, SVM, Decision Tree with Logistic Regression meta-learner) for disease prediction from extracted symptoms.
* Integrated **XLNet-based sentiment classifier** to detect toxic or unprofessional language during consultations; flagged segments were stored and reviewed via admin dashboard.
* Implemented secure user authentication using **JWT tokens** and enforced **role-based access control** across patients, doctors, and admins.
* Coordinated real-time video sessions using **Room API and WebSocket (Socket.IO)** for low-latency streaming and session management.
* Containerized backend microservices and Python-based AI modules using **Docker** for scalable deployment.

Certificates

Python OOP + Data Structures **Feb 2023 – Mar 2023**

Certificate via Course

Astera Bootcamp: Data Integration **Mar 2023**

Certificate via BootCamp

Data Cleaning – DataCamp **Sep 2024 – Oct 2024**

Certificate in Python Cleaning

Supervised Machine Learning **January 2024 - March 2025**

ML, KNN, Visualization

Advanced Learning Algorithms **January 2024 - March 2025**

Tensorflow, NN Models

Soft Skills

- Proactive Issue Resolution
- Troubleshooting
- Open-Mindedness
- Peer Support
- Versatility
- Growth Mindset
- Schedule Optimization