

Daniyal Hasan Syed

✉ daniyalhasan2003.28@gmail.com · +92 335 2775957

🌐 Daniyal Hasan · 🐙 DaniyalHasan089

Skills

- **AI/ML:** LLMs, SLMs, RAG, LangChain & LangGraph Agents, Prompt Engineering, Voice Pipelines (STT/TTS), OCR & Document Processing, NLP, Computer Vision, YOLOv8
- **Backend/API:** FastAPI, LangChain, LangGraph, HuggingFace, WebSockets, Streamlit, Flask, Gradio, Docker, Celery, Redis, SSE
- **Vector DBs:** ChromaDB, FAISS, Qdrant, PGvector
- **Data & Integrations:** Pandas, NumPy, REST APIs, Web Scraping, Firebase, Pydantic, PostgreSQL

Professional Experience

- **AI/ML Engineer** – Naturetech Software Solutions *Jul 2026 – Present*
 - Developed Track & Trace and industrial monitoring solutions using AI for the Federal Board of Revenue (FBR).
 - Designed computer vision pipelines using the YOLOv11 framework for product detection, tracking, and inspection across paper manufacturing, packaging, and beverages — including a Track & Trace demo at NBL Peshawar (Pepsi bottling) hitting 99% mAP@50, counting bottles at 50K+/hour and identifying SKUs.
 - Engineered an AI email agent supporting 500+ users, using LangGraph that summarizes threads, extracts actionable insights, and autonomously drafts and sends replies.
- **Associate AI Engineer** – Tafsol Technologies *Oct 2025 – Jul 2026*
 - Independently owned and led 7+ end-to-end AI projects, taking full responsibility for architecture and development.
 - Architected voice-to-voice pipelines integrating STT, LLM reasoning, and TTS over WebSockets, enabling real-time conversational AI applications.
 - Engineered RAG pipelines using ChromaDB, FAISS, and pgvector, improving response accuracy and context relevance across LLM applications.
 - Developed document intelligence solutions combining OCR and LLM-based extraction to automate client data workflows.
 - Built scalable Python backends powering RAG, STT/TTS, agentic workflows, OCR, and document manipulation pipelines for diverse use cases.
- **AI Engineer - Internship** – Lenaar Digital *Feb 2025 – May 2025*
 - Built LangChain-based chatbots and retrieval systems that improved information access and automated repetitive queries, smoothing client interactions and boosting demo engagement by 25-30%.
 - Designed and deployed FastAPI-powered AI services, including speech-to-text pipelines and full AI workflows, enhancing product demos and system performance.
 - Developed robust web-scraping and data-processing pipelines that streamlined data collection, enabling faster experimentation and reliable downstream AI/ML performance.

Education

- **Software Engineering** – Sir Syed University Of Engineering And Technology, Karachi *Oct 2021 – Aug 2025*

Projects

- **InvestFusion - AI-Powered Real Estate Platform**
 - **Problem:** Evaluating property investments required hours of manual due diligence, data aggregation, and risk assessment.
 - **Solution:** Built a FastAPI backend to scrape and aggregate listings from Zillow, Realtor, and LoopNet, and engineered an LLM-based investment analysis engine to evaluate cap rates, rental yields, and environmental risks.
 - **Results:** Automated the generation of prioritized, client-ready property reports with AI insights, completely replacing manual due diligence.
- **EasyRevie - AI Powered Document Extraction & Manipulation Pipeline**
 - **Problem:** Exam papers were image-based PDFs (not searchable text), containing MCQs, short/long answers, answer keys, and illustrations, making structured extraction slow and error-prone.
 - **Solution:** Built an OCR pipeline with PaddleOCR for text and Docling for illustrations; aligned figures to questions using Y-axis position and page number, then used an LLM to normalize the extracted content into structured JSON.
 - **Results:** Delivered an end-to-end workflow that converts complex scanned exam papers into clean, machine-readable JSON with questions, answers, and matched illustrations.
- **Tarjumaah - Real-Time Quran Transcription/Translation**
 - **Problem:** Translating classical Arabic religious content (prayers, khutbah) required a low-latency, context-aware system capable of handling continuous speech.
 - **Solution:** Architected a bidirectional WebSocket-based voice-to-voice pipeline integrating STT, a prompt-tuned LLM for classical terminology, and TTS.
 - **Results:** Achieved real-time, continuous audio streaming with highly accurate and culturally appropriate translations between the client and AI backend.