

SHERAZ AHMAD

AI Engineer | AI Automation & WhatsApp Chatbots | Generative AI & Agentic Systems | GoHighLevel CRM

sherazseo51@gmail.com · +92 311 7510900 · Lahore, Pakistan

linkedin.com/in/sheraz--ahmad · github.com/25sheraz

PROFESSIONAL SUMMARY

Software Engineering graduate (BS, 2026) who builds AI automation systems, WhatsApp chatbots, and Generative AI applications for real businesses. Delivered an end-to-end WhatsApp sales and support system for a Dubai car battery company (n8n, WhatsApp Cloud API, PostgreSQL) and, at DevelMO, a GoHighLevel CRM lead-generation system that increased revenue by 30% in its first month. Hands-on with multi-agent workflows (LangGraph), RAG (Qdrant, BM25, reranking), Vision Language Models, and real-time computer vision (OpenCV, MediaPipe). Integrates LLMs with REST APIs, CRMs, and webhooks using FastAPI and n8n, and evaluates LLM applications with LangSmith.

PROFESSIONAL EXPERIENCE

AI Developer / ML Developer | Tecclub Technology

Feb 2025 – Jan 2026

- Built and deployed ML models in TensorFlow and scikit-learn for real-world use cases, owning the full pipeline (data preprocessing, feature engineering, training, and evaluation) and reaching **73% F1** on held-out data.
- Developed structured, multi-step Generative AI workflows with LangChain and LangGraph, using tool calling to connect LLM agents to external systems.
- Integrated AI/ML models with backend systems through REST APIs and webhooks, automating business workflows in n8n and saving **25 hours** of manual work per week.

Generative AI & Automation Engineer | DevelMO

Mar 2026 – Oct 2026

- Built a GoHighLevel CRM lead-generation system that **increased revenue by 30% in its first month**.
- Built DevelMO's **AI Voice Agent** on ElevenLabs with **95% accuracy** in understanding callers and responding correctly; DevelMO now uses it for its own outbound calling.
- Built n8n automation workflows and multi-step, multi-agent LangGraph pipelines connecting LLMs to business processes and third-party services via webhooks and APIs.
- Trained and fine-tuned deep learning models for domain-specific tasks, improving accuracy from **78% to 88%**.
- Traced, debugged, and evaluated LLM application quality with LangSmith, cutting failed responses by **30%**.

AI & Deep Learning Intern | Fenris Group

Apr 2023 – Jun 2023

- Built a real-time motion detection and tracking system for live video streams, combining frame differencing and optical flow with CNN-based object detection.
- Trained and evaluated CNN detection models; annotated and preprocessed video data and tuned parameters with OpenCV to improve accuracy under varying lighting and motion, improving detection accuracy from **82% to 91%**.
- Covered the full computer vision lifecycle: data collection, annotation, training, and evaluation.

AUTOMATION & APPLIED AI PROJECTS

CarBatterySquad: End-to-End WhatsApp Sales & Support Automation

Dubai, UAE

n8n · WhatsApp Cloud API · LLM · PostgreSQL (UltraHost) · Webhooks

- Automated the full customer journey on WhatsApp for a Dubai-based car battery company: the bot handles battery service and jump-start enquiries **24/7**, from the first message to a closed sale.
- Stored every conversation in PostgreSQL hosted on UltraHost, giving the bot long-term memory so it **recognises returning customers** and continues with their full history, even after a long gap.
- Built automatic team alerts: after each successful sale the bot sends the client's details to the team group, and any complaint or service issue is flagged to the team straight away.
- Added **human takeover**: a team member can take over any live chat and the bot stops responding, so staff and bot never reply over each other.

Neoline: WhatsApp AI Customer Service Chatbot

n8n · WhatsApp Cloud API · RAG · Qdrant · OpenAI Embeddings · LLM

- Built a WhatsApp chatbot in n8n on the WhatsApp Cloud API that answers any customer query about Neoline's services instantly, **24/7**, so customers get a reply without waiting for staff.
- Grounded every answer in Neoline's own service information and FAQs with **RAG**: the content is embedded with OpenAI embeddings and stored in Qdrant, and the most relevant passages are retrieved for each message before the LLM replies, so answers stay accurate instead of generic.

DevelMO: GoHighLevel CRM Lead Generation & Outreach Automation

Apollo · GoHighLevel · LLMs

- Pulled B2B leads from Apollo, cleaned the data, and stored it in GoHighLevel CRM.
- Sent AI-personalised emails from lead and company data, with automated follow-ups and tracking.
- Ran sentiment analysis on replies and booked a meeting with a developer for every positive one.
- Secured new leads and **increased revenue by 30% in the first month** after launch.

DevelMO: AI Voice Cold-Calling Agent

develmo.com/our-products/ai-voice-agent

ElevenLabs (Voice AI) · CRM API

- Built an outbound voice agent on ElevenLabs that pulls lead data from the CRM and dynamically generates a personalised opening message for each call.

- Automated the workflow end to end, from lead lookup to live conversation handling.
- Achieved **95% accuracy** in understanding callers and responding correctly.
- Packaged it as a product-based service, now listed as a DevelMO product.

AI Customer Support Chatbots: Sylink & SecureTapVPN

sylink.tecclubx.site · securetapvpn.com

n8n · RAG · Qdrant · OpenAI Embeddings · Webhooks

- Built RAG chatbots that answer FAQ and product queries in real time from each client's knowledge base, deployed on their websites 24/7, **handling 500+ queries** per month with **90%** resolved without human help.
- SecureTapVPN: built an n8n workflow with Qdrant vector search and OpenAI embeddings, with knowledge base document ingestion triggered through a webhook.
- Designed a human-in-the-loop escalation flow that routes questions beyond the LLM's knowledge to an admin panel for live agent takeover, with automatic hand-back to the bot when the agent finishes.

FoodMate: AI Meal & Nutrition Analysis Backend

foodmate.xapcoders.com

Vision Language Model (VLM) · REST API

- Built a backend that uses a VLM to identify a meal from a single photo and estimate calories, protein, fat, and carbohydrates.
- Developed an in-app AI nutrition coach bot and exposed meal analysis and coaching through an API integrated with the app's existing frontend.
- Achieved **90% accuracy** on clear images and **79% on blurry images** in meal recognition and nutrient estimation.

AI / GENERATIVE AI PROJECTS

Optivortium Alliance: AI Eyewear Platform (Final Year Project)

Python · OpenCV · MediaPipe · Dlib · scikit-learn · Flask · React · Node.js · Firebase

- Built a real-time virtual try-on (AR) module that overlays eyeglass frames on a live webcam feed using a 478-point facial landmark model and affine transformation, handling varying head poses.
- Engineered a geometry-based face shape classifier (oval, round, square, heart, oblong) from facial landmark proportions, reaching **87% accuracy on 700+ test images**, served via a Flask REST API.
- Built the React frontend and Node.js backend and integrated them with the Python computer vision pipeline to deliver personalised glasses recommendations.

Pakistan Penal Code Legal RAG Chatbot

github.com/25sheraz/RAG-CRIMINAL-CHAT-BOT

Python · OpenAI GPT-4o-mini · OpenAI Embeddings · PyMuPDF · RapidOCR · Flask

- Built a Retrieval-Augmented Generation legal assistant over the Pakistan Penal Code (PPC 1860) that advises on applicable sections and citizens' rights.
- Combined PyMuPDF text extraction with a RapidOCR fallback for scanned pages, then chunked and indexed content with OpenAI embeddings in a local vector store.
- Implemented cosine-similarity semantic search and **GPT-4o-mini** generation that returns answers grounded in retrieved text, with exact section citations and punishment ranges; web UI and streaming CLI.

Self-Learning Coding Agent

github.com/25sheraz/Self-Learning-Agent

Python · LangChain · Kimi K2.7-Code (Moonshot AI) · Rich CLI

- Built an autonomous coding agent that generates Python code, runs it in an isolated sandbox, and self-corrects from failure tracebacks, structured as a LangChain pipeline (task understanding, retrieval, generation).
- Designed a persistent memory store that distills each failure into a reusable lesson injected into future prompts to prevent repeat mistakes.
- Built an agent evaluation harness on HumanEval (164 problems) measuring pass@1, self-correction recovery rate, and memory impact, improving **pass@1 from 62% to 74%** with self-learning memory enabled.

TECHNICAL SKILLS

Automation & CRM: n8n, GoHighLevel CRM (Campaigns, Follow-ups, Tracking), CRM Integration, WhatsApp Cloud API, Human-in-the-Loop Handoff, Apollo, Zapier, Webhooks, REST APIs, Workflow Automation, ElevenLabs (Voice AI)

Generative AI & Agents: RAG, Hybrid Search (Dense + BM25), Embeddings, Semantic Search, Reranking, LangChain, LangGraph, LangSmith, Multi-Agent Workflows, Tool Calling, VLMs, OpenAI API, LLM Integration, Transformers, Prompt & Context Engineering, Agent Evaluation Harness, Sentiment Analysis

ML & Computer Vision: Python, TensorFlow, scikit-learn, OpenCV, MediaPipe, CNNs, Feature Engineering, Model Training & Evaluation, Fine-tuning

Backend & Data: FastAPI, Flask, Node.js, Express.js, React, JavaScript, C++, HTML/CSS, PostgreSQL, Qdrant, Redis, ChromaDB, MongoDB, MySQL

Deployment & Tools: Docker, UltraHost, Vercel, Git, GitHub, VS Code, Jupyter Notebook

EDUCATION

BS Software Engineering | The University of Faisalabad

2022 – 2026

CGPA: 3.61 / 4.00

Intermediate (Pre-Engineering) | Punjab Group of Colleges

2020 – 2022

Marks: 87%

Matric | Usama Bin Zaid Public High School

2018 – 2020

Marks: 82%