Hamza Ali Khan

AI Engineer

Karachi, Pakistan | hamzasofficial1@gmail.com
GitHub | LinkedIn | Portfolio

Objective

Computer Science graduate with a Data Science and Analytics specialization and hands-on experience building machine learning, NLP, and LLM applications through internships and end-to-end projects. Skilled in Python, scikit-learn, TensorFlow/Keras, PyTorch, RAG, embeddings, vector databases, and AI agent workflows. Built and deployed projects including an AI-assisted university admissions platform, a multi-agent forensic audit system, and a crisis intelligence platform. Interested in developing reliable AI/ML systems and evaluating their performance in real-world workflows.

Technical Skills

Programming: Python, SQL, JavaScript/TypeScript
Machine Learning: scikit-learn, TensorFlow/Keras, PyTorch; classification, regression, clustering, model evaluation, feature engineering
NLP & Generative AI: LLM applications, prompt design, RAG, embeddings, semantic search, information extraction, LangChain, Google ADK
Vector Search & Data: ChromaDB, FAISS, Supabase pgvector, PostgreSQL, Firebase
Backend & Deployment: FastAPI, REST APIs, Docker, Google Cloud Run, Vercel, Render
Databases: PostgreSQL, Supabase, pgvector, SQLite, ChromaDB, FAISS

Education

BS Computer Science | Hamdard University, Karachi **2022 - 2026**

Specialization: Data Science – CGPA (3.81/4.0)

Experience

Data Science Intern | 10Pearls | Remote **Apr - July 2026**

- Built a Karachi air quality prediction project using machine learning models and a feature store/model registry workflow.
- Worked with Random Forest, Ridge, XGBoost, and LSTM approaches, comparing results using regression metrics.
- Developed an hourly data and prediction pipeline using Hopsworks and GitHub Actions.
- Investigated model performance and drift during experimentation.

Data Science Intern | DevelopersHub Corporation | Remote **May - Jun 2025**

- Built a news summarization pipeline on the CNN/Daily Mail dataset using spaCy for extractive summarization and Hugging Face Transformers for abstractive summarization.
- Tuned generation parameters and evaluated summaries against human-written references using ROUGE metrics and manual review.
- Developed a Random Forest model to predict employee attrition, applying data preprocessing, feature engineering, and SHAP explainability to interpret predictions and inform retention recommendations.

Projects

Ledger Sleuth | Multi-Agent Forensic Audit System | GitHub

- Developed a multi-agent workflow for investigating sales data anomalies and producing executive reports.
- Developed the FastAPI backend and deployed the containerized application to Google Cloud Run.
- Implemented anomaly detection, trend-break analysis, severity scoring, and root-cause hypotheses.

SafarAI — AI-Powered Travel Assistant | GitHub

- Built an AI-powered travel assistant that generates personalized itineraries using Google Gemini.
- Integrated Google Maps, Google Places, Supabase, and live weather data to support trip planning and place discovery.
- Designed a community reporting feed for sharing real-time travel conditions, incidents, and local recommendations and deployed it on Google Cloud Run.

Dakhla | AI-Assisted University Admissions Platform| GitHub

- Built an admissions recommendation system using Random Forest, trained on a dataset of 5,000 student records.
- Improved recommendation accuracy from 78.5% to 82.7% through GridSearchCV tuning across 108 configurations with five-fold cross-validation.
- Developed a RAG-based admissions chatbot with document ingestion and a FastAPI backend.
- Deployed the platform using Vercel and Render.

Certifications

- Python for Data Science, AI & Development — Coursera
- Google Data Analytics — Coursera
- Data Visualization in Power BI — Datacamp
- SnowPro Associate: Platform — Snowflake