

HAMZA ALI

I-11/1 Islamabad, Pakistan

📞 +923303053300 ✉ [Gmail](#) [in](#) [LinkedIn](#) [GitHub](#) [Portfolio](#) [ResearchGate](#)

Summary

Accomplished Machine Learning Engineer with a proven track record in researching, designing and developing SOTA machine learning models. Possess specialized skills in product architecture and MLOps. Proficient in developing and deploying projects on prominent cloud platforms such as AWS and Azure. I have designed and implemented Machine Learning Pipelines that streamline experimentation and facilitate the hosting of ML Models on these platforms.

Experience

K2X Tech

March 2023 – Present

Machine Learning Engineer

Peshawar, KPK

- Developed and deployed ML models with machine learning pipelines, for seamless data flow and model integration.
- Proficient in Python backend development with FastAPI for API integration and testing.
- Skilled in Data Engineering, MLOps, AWS/Azure Cloud Platforms and Docker for containerization of services.
- Extensively used ML tools and frameworks like MLflow, DVC, and DagsHub for model tracking and experimentation.
- Experienced in Generative AI, Stable diffusion and GAN based models for image and video editing and generation.
- Proficient in shell scripting, Linux administration, and network configuration to enhance system reliability and performance for ML operations.

AntonX

July 2023 – Feb 2024

Machine Learning Engineer

Peshawar, KPK

- Collaborated with app and web development teams at AntonX to seamlessly integrate AI into various company products
- Skilled in fine-tuning LLMs and applying Retrieval-Augmented Generation (RAG) techniques for tailored use cases.
- Engineered a low-latency, real-time recommendation system using content-based filtering for optimized performance.

Projects

Electronic Media Monitoring System | *EasyOCR, Kafka, AWS, Docker, NGNIX, Airflow, PySpark*

- Developed a real-time Machine Learning system for KPK Government to monitor news channels and social media.
- Finetuned EasyOCR for urdu text recognition and integrated facial recognition.
- Created a pipeline to preprocess Twitter data using PySpark, with Airflow orchestrating jobs for real-time monitoring.
- Used Docker to containerize services, optimized APIs, and reduced latency through analysis and tuning across channels, enhancing real-time reliability.

Real-time Videos Recommendation Engine | *MongoDB, ChromaDB, Transformers, Azure Services, CI/CD Pipeline*

- Architected a real-time recommendation system for a social media app, leveraging transformers and ChromaDB.
- Orchestrated different services using docker-compose while ensured a scalable and fault tolerant containerization.
- Deployed the whole system to Azure Web Apps using a CI/CD pipeline with Azure Container Registry, ensuring seamless scalability and performance.

AI-Driven Image Transformation and Generation | *FastAPI, GFPGAN'S, U2Net, Stable Diffusion, Docker*

- Developed an AI image processing system in FastAPI, including background removal and addition, image enhancement and video generation
- Used different models like U2Net, GFPGAN, and Real-ESRGAN, transformers and stable diffusion techniques.
- Successfully deployed the entire application on a local GPU server, providing scalable endpoints for mobile app integration for fast inferencing with ONNX optimization.

Metabloic AI HealthCare Coach | *Backend Development, Qdrant, OpenAI API, Twilio, MongoDB, AWS*

- Developed AI-powered Coaches that assist users in losing weight and improving their lifestyles by sending reminders via calls and SMS.
- Built a knowledge base system where agents utilize user-provided documents and web links to offer personalized guidance stored in Qdrant
- Deployed the system with Docker on AWS, and integrating a CI/CD pipeline for continuous delivery via AWS ECR.

Customer Support Chatbot | *Llama, RAG, Langchain, ChromaDB*

- Fine-tuned a customer support chatbot for a food delivery app using Llama, enhancing response and relevance.
- Implemented RAG with Langchain agents and vector databases Qdrant for handling customer queries.

- Optimized chatbot performance with LLAMA2 prompt engineering, ensuring quick and precise responses for being customer support.
- Forecastly.AI** | *XGBoost, ARIMA, Prophet*
- Engineered AI-driven sales forecasts for bakeries, restaurants, and amusement parks using a hybrid stack of XGBoost, Prophet, and ARIMA.
 - Applied MTSL techniques to manage complex sales data, enhancing the precision of daily, monthly, and yearly forecasts and used MAPE, RMSE, and MAE metrics for evaluation.

Technical Skills

- **Programming Languages:** Python, C++, MATLAB - **Libraries & Frameworks:** Scikit Learn, TensorFlow, PyTorch, OpenCV, PySpark - **Databases:** SQL, MySQL, MongoDB, Redis - **Data Analytics:** Pandas, NumPy, SciPy, Matplotlib, Seaborn, Web Scraping, Tableau - **Machine Learning:** Regression, Classification, Clustering, Dimensionality Reduction, Forecasting, XGBoost - **Deep Learning:** ANN's, CNN's, RNN's, LSTM's, GAN's, GRU's, Transformers - **NLP and Computer Vision:** NLTK, spaCy, Gensim, MediaPipe, OpenCV, RoboFlow - **Cloud Computing:** AWS and Azure (ML and AI Services for Deployment and Monitoring) - **MLOps:** Docker, Kubernetes, GitHub, CI/CD Pipelines, MLflow, Airflow, Kafka, Nginx - **Backend and API Development:** FastAPI, Flask, Postman - **Generative AI:** LLMs, Prompt Engineering, RAG, LangChain, Hugging Face - **Image Processing and Video Streaming:** OpenCV, FFmpeg - **Vector Databases:** ChromaDB, Pinecone, Qdrant - **Networking:** Network Configuration and Interface Management

Education

University of Engineering and Technology <i>Bachelor of Science in Computer Systems Engineering (CGPA: 3.21)</i>	Oct. 2020 – Jul 2024 <i>Peshawar, KPK</i>
IIUI Schools and Colleges <i>Intermediate(Pre-Engineering) (Grade: A+)</i>	Oct. 2018 – Jul 2020 <i>Mansehra, KPK</i>

Courses and Certifications

IBM Data Science Professional Certificate <i>Coursera</i>	March 2022 – June 2022 <i>IBM Data Science Certificate</i>
Machine Learning and Deep Learning <i>YouTube</i>	July 2023 – November 2023 <i>CampusX by Nitish Singh</i>
Generative AI with Large Language Models <i>Coursera</i>	October 2023 – December 2023 <i>DeepLearning.AI</i>
Machine Learning in Production <i>Coursera</i>	February 2024 – April 2024 <i>DeepLearning.AI</i>

Research and Publications

Detection of Holes in Point Clouds using Statistical Techniques <i>International Journal of Innovations in Science and Technology</i> Hamza Ali , Zain ul Abideen,Muhammad Sajjad .etl	March 2024
Combatting Illegal Logging with AI-powered IoT Devices for Forest Monitoring <i>International Journal of Innovations in Science and Technology</i> Abdullah Khan, Hamza Ali , Nasru Minullah .etl	April 2024