europass

Muhammad Afnan Qasim

Date of birth: 13 Jul 2002 | **Nationality:** Pakistani | **Phone:** (+92) 3457812255 (Mobile) | **Email:** afnanqasim74@gmail.com | **Website:** https://afnanqasim.vercel.app/ | **LinkedIn:** https://www.linkedin.com/in/afnanqasim74/ | **Address:** Lahore, Pakistan (Home)

ABOUT MYSELF

AI and Machine Learning Engineer specializing in advanced Natural Language Processing (NLP) and AI-driven automation solutions. Proven expertise in developing scalable, cloud-native AI platforms using cutting-edge frameworks like LangChain, LangGraph, and Retrieval-Augmented Generation (RAG) architectures. Skilled in deploying robust AI services on Google Cloud Platform and AWS, integrating OCR, computer vision (YOLO), and NLP for complex document processing and real-time defect detection. Experienced in architecting generative AI chatbots, automated workflow orchestration, and secure data redaction systems that boost operational efficiency by over 30% and reduce manual workloads by up to 90%. Passionate about leveraging multi-agent AI systems and deep learning models to transform business processes in construction management and beyond, while optimizing resource use and ensuring compliance with data privacy standards.

WORK EXPERIENCE

AI ENGINEER – LAHORE, PAKISTAN

PREESOFT PVT LTD – NOV 2024 – CURRENT

· Developed and deployed the AI-powered Ezelogs construction management platform on Google Cloud Platform, leveraging LangGraph and LangChain for agent-based workflow automation and predictive analytics to optimize project lifecycle and resource management.
· Engineered a generative AI chatbot using advanced NLP and agent frameworks to automate appointment booking, email handling, and Google Meet scheduling, enhancing operational efficiency.
· Implemented OCR and NLP-based AI models for automated document compliance and contract review, enabling intelligent extraction and validation of key project data.
· Integrated YOLO-based computer vision models for real-time defect detection and quality assurance in field operations, improving on-site safety and compliance.
· Deployed and managed AI models and services on Google Cloud Storage (GCS) buckets and Compute Engine instances, ensuring scalable, secure, and efficient cloud-based operations.

DREAM SLEEP – CHICAGO, UNITED STATES

MACHINE LEARNING ENGINEER (LLM TRAINER) – JAN 2024 – OCT 2024

· Developed FEDml architecture to surpass GPT-3.5 in emotional intelligence benchmarks.
· Utilized PEFT to enhance AI emotional intelligence for more empathetic interactions.
· Reduced GPU costs by 60%, saving the company approximately $2 million.
· Directed agile projects with cross-functional teams, optimizing project outcomes.
· Documented advancements, boosting the company's R&D efforts.

KYAAS SOLUTIONS – LAHORE, PAKISTAN

JUNIOR ML ENGINEER – MAR 2023 – JAN 2024

· Fine-tuned Hugging Face models for project-specific needs.
· Showcased Python expertise in API development with Flask and multi-processing.
· Implemented ML algorithms using TensorFlow, Keras, and PyTorch for NLP initiatives.
· Enhanced NLP model performance, driving technological progress.
· Leveraged Kubernetes and Docker for improved project deployments.
· Enabled smoother model integration into production, boosting efficiency.
· Optimized existing models, enhancing their effectiveness.

· Utilized transformers and NLP libraries to advance data processing.

PROJECTS

15 JAN 2024 – 4 OCT 2024

LLM PEFT Training

· Developed FEDml to enhance LLM training, outperforming GPT-3.5 by 12%.
· Incorporated PEFT for richer emotional AI interactions.
· Improved AI emotional intelligence, enabling empathetic interactions.
· Cut GPU costs by 60%, saving $2 million.
· Boosted R&D through agile collaboration and innovative methodologies.

4 NOV 2024 – 2 FEB 2025

Chatbot-Rag

· Developed a Retrieval-Augmented Generation (RAG) based AI chatbot for the Ezelogs construction platform, leveraging its entire knowledge base to automate appointment booking, email handling, and Google Meet scheduling, improving operational efficiency by 35% and increasing client satisfaction scores by 25%.
· Utilized LangChain and LangGraph frameworks for multi-agent orchestration and intelligent data retrieval, enhancing contextual understanding and boosting response accuracy by 30%.
· Integrated OCR and entity recognition to extract critical information from construction documents and emails, reducing manual processing time by 40%.
· Deployed the chatbot on Google Cloud Platform using Compute Engine and Cloud Storage for scalable, secure, and resilient operations.

Link https://chatbot.ezelogs.com/

FEB 2023 – MAY 2024

AI-Powered Sensitive Data Redaction System

· Developed an AI-driven solution to automatically detect and redact sensitive information (e.g., addresses, personal identifiers) from scanned PDFs and images, ensuring compliance with data privacy regulations.
· Utilized OCR (Optical Character Recognition) to extract text and NLP (Natural Language Processing) models to identify and classify sensitive data for precise redaction.
· Applied computer vision techniques to detect and obscure sensitive content in images, supporting diverse document types and formats.
· Deployed the solution on AWS EC2 instances, enabling scalable, secure, and high-availability processing of large document volumes.
· Achieved a 90% reduction in manual redaction time, improving operational efficiency and accuracy in document management workflows.

Link https://ai.revdsm.com/

DEC 2022 – FEB 2023

Career-Prediction

- Successfully obtained semi-supervised data comprising 6 thousand rows, with 50 labeled rows.
- Cleaned the data and applied feature engineering techniques for further analysis.
- Conducted **Exploratory Data Analysis** (EDA) to gain insights and understand the data distribution.
- Employed hierarchical, **DBSCAN**, and K-means clustering models for unsupervised learning.
- Deployed the project on Streamlit, enabling users to input numeric values for 15 features.
- Utilized the K-means model to predict the user's most suitable career based on the provided inputs.

13 JUL 2024 – 15 DEC 2024

Plastic Import/Export Forecasting Platform

· Developed CommoPlast, a Django-based platform forecasting plastic import/export trends (PVC, PET, PP) using time series forecasting and machine learning.
· Integrated global trade data and built internal APIs for efficient data processing and user interaction.
· Deployed the application using Docker containers on Alibaba Cloud for scalable and secure cloud operations.
· Delivered actionable buy/sell recommendations, helping users optimize trading strategies and supply chains.
· Improved user decision-making efficiency by 30%, enabling higher profitability and streamlined operations.

Link https://commoplast.com/

SKILLS

Machine LearningAI | NLP libraries: NLTK, SpaCy | Pytorch,Tensorflow | Deep Neural Networks (CNNs, GANs) | Python - Deep Learning (tensorflow2, pytorch, transformers) | Google colaboratory | ML Model Deployment | SQL | Python | Computer Vision

EDUCATION AND TRAINING

22 JUL 2020 – 26 MAY 2024 Lahore, Pakistan

BS COMPUTATIONAL PHYSICS Univrsity of the Punjab

APR 2023 – JUN 2023 lahore, Pakistan

AMAL CAREER-PREP FELLOWSHIP Amal Academy

GENERATIVE AI WITH LARGE LANGUAGE MODELS DeepLearning.AI

PYTHON FOR DATA SCIENCE AND MACHINE LEARNING BOOTCAMP Udemy

ARTIFICIAL INTELLIGENCE (MACHINE LEARNING & DEEP LEARNING) National Vocational and Technical Training Commission NAVTTC

MICROSOFT CERTIFIED: AZURE AI FUNDAMENTALS Microsoft

HCIA-BIG DATA Huawei

LANGUAGE SKILLS

Mother tongue(s): **ENGLISH**