# Muhammad Azeem

Faisalabad, Pakistan muhammad.azem03@gmail.com +92 322 2619000 muhammad-azeem
Muhammadazeem-eng Google Scholar

## Profile Summary

AI Engineer with 4+ years of experience building production-grade ML systems across healthcare, HR automation, and enterprise AI. Delivered 15+ end-to-end projects spanning LLM fine-tuning, RAG pipelines, computer vision, and multi-agent orchestration. Published 4 peer-reviewed papers and deployed scalable NLP solutions integrating 9+ language models for real-time inference. Combines applied research with full-stack ML engineering to ship high-performance, cost-optimized AI products.

## Education

**Harbin Institute of Technology, China** *2021 – 2023*
*Master's in Electronics and Information Engineering*
- Official Website

**The University of Faisalabad** *2015 – 2019*
*Bachelor's in Electrical Engineering*
- Official Website

## Experience

**AI Engineer** *Remotely, Pakistan*
*CyberGen USA* *Feb 2022 – Present*
- Architected and deployed 9+ LLM-based systems integrating OpenAI and Hugging Face models for enterprise AI orchestration.
- Developed and fine-tuned transformer-based NLP models using LoRA and PEFT across healthcare and enterprise use cases.
- Built scalable RAG pipelines supporting multi-source ingestion (PDFs, YouTube, web) with vector databases (Pinecone, ChromaDB).
- Engineered 5+ domain-specific AI solutions including Clinical Trial Matcher, CV Screening, and Service Desk Automation.
- Optimized deep learning inference pipelines using CNN and transformer architectures to improve throughput and deployment efficiency.
- Designed automated data pipelines handling structured and unstructured datasets for ML training workflows.
- Conducted exploratory data analysis on large-scale datasets to support AI product development decisions.

### Key Projects at CyberGen

- **LLM Infrastructure & Orchestration:** Architected AI Bus System integrating 9+ LLMs for enterprise prompt routing and multi-model orchestration. Built ARGUS monitoring framework enabling real-time observability and anomaly detection for production AI pipelines. Implemented policy enforcement layers (Rego Guardrails, Cortex Shield) to ensure compliant AI outputs.
- **RAG & Knowledge Systems:** Engineered Notebook LM aggregating YouTube, websites, and PDFs with contextual embeddings, indexing 10K+ documents. Constructed RBAC-enabled Knowledge Graph RAG chatbot with context-aware retrieval achieving 92% relevance scores. Shipped 3 domain-specific chatbots (Flyzone, Hedu, Ordea) for travel, education, and support verticals.
- **Healthcare AI:** Assembled transformer-based Clinical Trial Matcher for EHR-to-trial patient matching across 500+ active trials. Delivered LLM-powered Clinical Note Summarizer for radiology and progress notes, reducing documentation time by 45%. Produced UMLS-aligned Medical Entity Extractor normalizing 200+ clinical entity types.

- **HR & Enterprise Automation:** Engineered CV Screening system with resume ranking and parsing, processing 1,000+ applications per cycle. Launched Strategic Staffing Platform with AI-driven JD generation and salary recommendations. Established Service Desk Automation with ticket classification, web search, and auto-resolution, resolving 30% of tickets without human intervention.
- **Computer Vision:** Trained YOLOv8 models for real-time road defect detection from dashcam footage, achieving 89% mAP. Constructed YOLOv4-based acne detection system with custom dataset of 5K+ images and front-end interface. Integrated helmet detection system with CCTV for construction site safety compliance across 3 sites.
- **Productivity & Content Tools:** Fine-tuned GPT Writing Assistant with tone control and summarization supporting 8 writing styles. Orchestrated Multi-Agent System for automated content generation with content calendars and email journeys. Crafted Coding Editor Agent for real-time code suggestions and debugging, reducing developer review time by 25%.

**Freelance Python Developer** *Remote (Freelance)*
*Jul 2021 – Jan 2022*

- Analyzed 20+ client datasets using **pandas**, **NumPy**, and **Matplotlib** to extract insights and identify revenue-impacting trends.
- Automated data cleaning, transformation, and validation tasks with custom Python scripts, reducing manual processing time by 60%.
- Produced interactive data visualizations using **Plotly** and **Seaborn** for business reporting across 10+ client engagements.
- Conducted exploratory data analysis (EDA) for e-commerce, marketing, and logistics clients, informing in budget reallocation decisions.
- Deployed Jupyter Notebook-based reports for clients ensuring transparency and reproducibility across 15+ deliverables.
- Trained small-scale machine learning models using **scikit-learn** for classification and regression tasks, achieving 85%+ accuracy.
- Delivered well-documented Python code and comprehensive analysis reports tailored to client requirements, maintaining a 5-star satisfaction rating.

## Publications

**Microcontroller-Based Hybrid Power Electronics Trainer using PWM Techniques** 2020

10.14741/ijcet/v.10.6.4
*Role:* Led trainer hardware development and PWM-based control system using microcontrollers. Handled circuit modeling, control calculations, and hardware testing procedures.

**An Interdigital Microwave Sensor Based on Differential Structure for Dielectric Constant Characteristics Measurement** 2023

10.3390/s23146551
*Role:* Master's thesis work. Independently designed the sensor, performed CST simulations, conducted experimental validation using a vector network analyzer (VNA), and formulated mathematical models for dielectric constant estimation.

**High-Sensitivity Liquid Dielectric Characterization Differential Sensor by 1-Bit Coding DGS** 2023

10.3390/s23010372
*Role:* Led sensor design and optimization using adaptive genetic algorithms and Debye framework modeling. Executed simulation analysis and regression-based permittivity estimation.

**Enabling Sensor-Free Emission Monitoring for SMEs With Vision-Language Appliance Detection and Operational Context Inference** 2025 (Accepted)

*Accepted to AINA 2026.*
*Role:* Designed and fine-tuned the YOLOv8-based appliance detection model, integrated CLIP-based operational context inference, and conducted performance evaluation for emission monitoring feasibility.

## Certifications

**Python for Data Science, AI & Development**

*Authorized by IBM*
View Certificate

**Machine Learning Specialization**

*Authorized by DeepLearning.AI & Stanford University (Coursera)*
View Certificate

**Supervised Machine Learning: Regression and Classification**

*Authorized by DeepLearning.AI*
View Certificate

**Advanced Learning Algorithms**

*Authorized by DeepLearning.AI*
View Certificate

**Unsupervised Learning, Recommenders, Reinforcement Learning**

*Authorized by DeepLearning.AI*
View Certificate

**Introduction to Computer Vision and Image Processing**

*Authorized by IBM*
View Certificate

## Technical Skills

- **Languages:** Python, C++, SQL
- **ML/DL Frameworks:** PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn
- **LLM & NLP:** LangChain, LlamaIndex, OpenAI API, RAG pipelines, prompt engineering, fine-tuning (LoRA, PEFT)
- **Computer Vision:** YOLOv4/v8, OpenCV, CLIP
- **Infrastructure:** Docker, Git, Linux, Google Cloud, Azure
- **Databases:** MongoDB, MySQL, vector databases (Pinecone, ChromaDB)

## Languages

**Urdu:** Native proficiency

**English:** C1 – Reading, Writing, Listening, Speaking