

Zaraar Malik

LinkedIn: Zaraar Malik linkedin

GitHub: Zaraar125

Email: zaraarmalik5303@gmail.com

Portfolio: Personal Portfolio

SUMMARY

AI Engineer with hands-on experience in GenAI, RAG pipelines, and multimodal AI systems. Skilled in designing and deploying scalable, production-grade AI applications using Python, FastAPI, and modern ML/DL frameworks. Experienced in data engineering, automated ETL pipelines, and model optimization for high-performance AI systems. Passionate about building robust NLP and CV architectures for real-world impact.

EDUCATION

- **FAST National University of Computer and Emerging Sciences** Islamabad, Pakistan
Bachelor of Science in Artificial Intelligence; CGPA: 3.56 08/2021 – 07/2025

EXPERIENCE

- **Apexbeat** Manchester, United Kingdom
Machine Learning Engineer, Remote 01/2025 – Present
 - **Multi-Modal Medical Chatbot:** Developed a production-grade Multi-Modal Medical Chatbot enabling users to query and search across multiple medical textbooks with context-aware retrieval and answer synthesis.
 - **Automated Data Pipelines:** Engineered automated document ingestion and data generation pipelines for large-scale knowledge base construction, including data cleaning, embedding creation, and metadata indexing.
 - **Cloud Deployment:** Deployed RAG pipelines and inference services on AWS, serving over 2,000 active users through FastAPI.
 - **Database Integration:** Integrated MongoDB for dynamic storage of user interactions, embeddings, and chat histories to enable persistent and personalized AI responses.
- **Systems Limited** Islamabad, Pakistan
AI Intern, Onsite 06/2025 – 08/2025
 - **Model Fine-Tuning:** Fine-tuned transformer-based LLMs for enterprise GenAI use cases, enhancing response precision and inference speed.
 - **Feasibility Analysis:** Conducted end-to-end feasibility studies of AI solutions, analyzing scalability, accuracy, and cost-performance trade-offs.
- **FSM – FAST School of Management** Islamabad, Pakistan
AI Research Assistant, Onsite 09/2024 – 03/2025
 - **Financial ML Experimentation:** Built and evaluated predictive models on financial datasets to analyze churn, credit risk, and investment patterns.
 - **RAG-based Investment Assistant:** Developed a Retrieval-Augmented Generation chatbot for personalized investment recommendations across Pakistani regions, integrating structured and unstructured financial data.
- **AIO – Silicon Valley Stealth Startup** Islamabad, Pakistan
AI Intern, Onsite 06/2024 – 08/2024
 - **BERT-Based Entity Extraction:** Developed a domain-adapted BERT model for extracting structured entities from restaurant reviews. Created evaluation metrics and optimized model layers for improved precision.
 - **Data Preparation:** Collaborated with the data engineering team to curate, clean, and format datasets for fine-tuning and evaluation pipelines.
- **Artificial Intelligence & Machine Learning Lab** Islamabad, Pakistan
AI Intern, Onsite 06/2023 – 08/2023
 - **AI Video Dubbing System:** Built an end-to-end multilingual dubbing pipeline using Whisper for transcription, Tortoise TTS and Coqui for voice generation, and Wav2Lip for lip-syncing.
 - **Audio Style Transfer:** Implemented CNN-based Neural Style Transfer to transfer acoustic features from one sound source to another for creative media use.

PROJECTS

- **Snap Shop – GenAI Fashion Synthesizer:** Created a web-based try-on platform using LoRA-tuned diffusion models for garment synthesis and fashion e-commerce visualization.
- **Multi-Agent Chatbot for Medical Reimbursement:** AI assistant automating medical receipt reimbursements, enabling agent collaboration, validation checks, and error handling to reduce manual effort.
- **PhishNet-Detection:** A comprehensive machine learning solution for detecting and classifying malicious URLs (phishing, malware, spam, defacement) using XGBoost classification.
- **TinyLLM RAG Chatbot:** Experimented with lightweight LLMs for RAG, optimizing chunking strategies and evaluating trade-offs between efficiency, accuracy, and deployment cost.
- **Procedural Game Level Generation:** Trained DCGAN and WGAN architectures on Mario levels to generate playable maps and assist automated game design workflows.
- **Hidden Object Detection:** Fine-tuned YOLOv8 and Faster R-CNN on iMaterialist and Object365 datasets to detect objects in cluttered environments, with an interactive web-based frontend.
- **Multimodal Meme Classifier:** Built a dual-branch neural network combining visual and textual inputs for meme sentiment classification.

PROGRAMMING SKILLS

- **Languages:** Python, C++, JavaScript, HTML, CSS, SQL
- **Frameworks & Deployment:** PyTorch, Scikit-Learn, FastAPI, Flask, Docker, Vercel, LangChain
- **Databases & Pipelines:** MongoDB, QDRANT, FAISS, Pinecone, Chroma
- **Deep Learning:** LLMs, RAG, LSTMs, BERT, GANs, ViTs, Transformers, YOLO, Diffusion, Wav2Lip

HONORS AND AWARDS

- **Data Quest, Nascon 03/2023:** Winner of Pakistan's largest AI Hackathon.
- **AIO Hackathon, FAST 06/2023:** 2nd Place in Generative AI hackathon by AIO.
- **Infyma Hackathon, FAST 03/2024:** 2nd Place in AI-focused hackathon by Infyma.
- **Dean's List:** Recognized four times for outstanding academic performance.