

# Mufeez Ahmad

📍 Lahore, Pakistan | ✉ mufeezahmadgcu@gmail.com | 📞 +92-323-1013083 | 🌐 Mufeez Ahmad

## About Me

AI/ML Engineer with hands-on experience in developing machine learning models, RAG systems, and real-time audio AI applications. Skilled in bridging AI research with enterprise-grade software engineering using .NET and cloud technologies. Passionate about building scalable intelligent systems, optimizing performance, and applying generative AI to solve real-world problems.

## Education

**Bachelor of Science in Computer Science**, *Government College University (GCU), Lahore* 2019 – 2023

- Relevant Coursework: NLP, Machine Learning, AI, Data Mining, Parallel & Distributed Computing, Information Security, Computer Networks, Probability & Statistics, Database Systems

## Experience

**Analyst Software Engineer (AI/ML & .NET)**, *IBEX – Lahore, Pakistan* July 2024 – Present

- Developed a fully locally hosted **Retrieval-Augmented Generation (RAG)** system using PostgreSQL (pgvector) and Ollama for call center knowledge retrieval.
- Conducted hyperparameter tuning, transfer learning, and fine-tuning of pre-trained transformer models to optimize performance on custom call center knowledge retrieval tasks.
- Engineered feature extraction and pre-processing pipelines for structured and unstructured data, improving downstream model accuracy by 20–30%.
- Researched and optimized tokenization pipelines, implementing the **Flash Tokenizer** in .NET to achieve response times of 150–180ms.
- Optimized data storage and retrieval for embedding vectors by implementing custom sharding and approximate nearest neighbour (ANN) search methods (e.g. HNSW, FAISS-like strategies) to serve vector search queries under 100 ms.
- Implemented lightweight real-time inference services using ONNX or TensorFlow Lite in a .NET environment to reduce latency in deployed services.
- Integrated a pre-trained **real-time Denoiser model** into a Windows driver, creating a virtual microphone (*iMic*) to suppress background noise during live calls.
- Collaborated with cross-functional teams (data engineering, DevOps, product owners) to integrate machine learning components into production microservice architectures.
- Designed microservices for product and store management using .NET, **Ocelot API Gateway**, **IdentityServer**, **Redis**, **Docker**.
- Designed and implemented AI-powered chatbot and email automation workflows using **n8n**, significantly improving customer response efficiency.

**Software Engineer Intern**, *New Elevation – Lahore, Pakistan* Jan 2024 – Apr 2024

- Built business dashboards in Power BI for real-time analytics.
- Supported CI/CD pipelines and automated deployment tasks.
- Optimized SQL queries, improving reporting and data retrieval speeds by 50%.

## Projects

### Ibex iVoice – AI Agent Assistant

- Developed a real-time AI-powered call center assistant integrating **noise suppression (DeNoiser)** and initial research on **accent normalization** to improve call clarity and agent productivity.

### Ibex IntelliServe

- Built a fully locally hosted **Retrieval-Augmented Generation (RAG)** system using **PostgreSQL (pgvector)** for embeddings and **Ollama** for LLM integration, enabling call center agents to access real-time knowledge with high accuracy.
- Optimized vector storage and retrieval pipelines, achieving query response times under 5 seconds.

### NeuraFlow

- Designed and deployed AI-powered **workflow automations** in n8n, including a smart email automation pipeline that reduced manual handling time by 40%.

- Created chatbot workflows for customer query handling, integrating with APIs and local LLMs for real-time responses.

#### **PIGEN – AI Powered Picture Generator (Final Year Project)**

- Built a **GAN-based text-to-image generator** in PyTorch, achieving an Inception Score of 4.4 on generated images.

#### **Parkinson's Disease Predictor**

- Trained ML models (**Logistic Regression, KNN**) to predict Parkinson's disease with 95% accuracy, supporting early diagnosis research.

#### **Ibex Quartz – Internal Activity Tracking System**

- Designed a secure role-based activity tracking system in **.NET**, enabling real-time updates and predictive analytics through AI integration.

#### **Ibex TMS – Training & Employee Management System**

- Built a robust Training Management System in **.NET** with role-based access, automating onboarding & training workflows to increase HR efficiency.

### **Technologies**

---

**Languages:** Python, C#, C++ , Java, SQL, Shell scripting




**AI/ML:** PyTorch, TensorFlow, Generative AI, LLMs, RAG Systems, NLP, Transformers, CNN, RNN, LSTM, Hugging Face

**Frameworks/Tools:** .NET, Blazor Maui, Docker, Kubernetes, IdentityServer, n8n, Git/GitHub, Azure, GCP

**Databases & Storage:** PostgreSQL (pgvector), SQL Server, MongoDB, Redis, SQLite

### **Certifications**

---

- **Machine Learning Specialization**, Stanford University & DeepLearning.AI - 
- **Machine Learning with Python**, IBM - 
- **AWS Cloud Technical Essentials**, AWS - 
- **Prompt Engineering for ChatGPT**, Vanderbilt University - 