Muhammad Uzair Kabeer

Machine Learning Engineer — AI Specialist

muhammaduzairkabeer@gmail.com — linkedin.com/in/uzair-kabeer — +92 314 5830771

PROFESSIONAL PROFILE

Based in Rawalpindi, I am a Machine Learning Engineer with over three years of hands-on experience designing and delivering AI/ML solutions. My core expertise lies in Retrieval-Augmented Generation (RAG) models, fine-tuning large language models, and deploying cloud-native applications. Proficient in Python, TensorFlow, and PyTorch, I leverage MLOps best practices—particularly with Azure and Docker—to ensure scalable, maintainable systems. I have successfully built and productionized AI applications across healthcare, agriculture, and HR technology domains, and I also bring a strong background in full-stack development using Node.js and React, as well as experience architecting distributed systems.

TECHNICAL SKILLS

AI/ML:	RAG, LLMs (GPT-3.5/4, Llama3), NLP, Computer Vision, LangChain, Hugging Face
Frameworks:	TensorFlow, PyTorch, Scikit-learn, Transformers, FastAPI
Cloud/MLOps:	Azure, DigitalOcean, Docker, CI/CD Pipelines, Model Deployment
Languages:	Python, JavaScript (Node.js, NextJS), SQL
Tools:	Git, GitHub Actions, Pinecone, MongoDB, MSSQL

PROFESSIONAL EXPERIENCE

AI Software Engineer *Funavry Technologies* May 2025 – On Going

- Built and deployed **smart search system** using Azure AI Search and LangChain that combines keyword and vector search, improving answer accuracy by 25% for business knowledge queries.
- Created **AI voice calling agents** for restaurant reservations and customer service with real-time conversation handling, automated booking management, and intelligent call routing.
- Developed **task management system** with natural language processing for creating tickets, tracking job status, and team assignment, plus integrated GitHub tools for developer workflow automation.

Machine Learning Engineer *Freelance* Jan 2023 – Jan 2025

- Designed and deployed **hybrid search RAG chatbots** using Pinecone and LangChain, improving query accuracy by 35% through BM25 + dense retrieval fusion.
- Fine-tuned LLMs (Llama3, Mistral) for domain-specific tasks using LoRA adapters, reducing hallucination rates by 22% via RLHF.
- Built an **AI blog generator** with CrewAI and DeepSeek, automating SEO-optimized content creation for 10+ clients (250+ articles/month).

Software Engineer *Techbaltoro - LLC* Sep 2024 – Jan 2025

- Deployed an automated server infrastructure for daily/weekly Google Drive backups, safeguarding essential company data.
- Created a predictive **NextJS** application utilizing a **Multiple Regression** model and the **KEEPA API** for advanced FBA product assessment and success rate forecasting.

AI Software Engineer *LinkHR (Tinder for Jobs)* Sep 2023 – Mar 2024

- Optimized job-matching algorithms using **cosine similarity** and **BM25**, increasing user matches by 40% through profile embedding clustering.
- Led backend development (Node.js + MongoDB) and implemented real-time analytics dashboards with WebSocket integrations.

Data Science Intern *Eziline Software House* Jun 2022 – Aug 2022

- Developed predictive models for client churn (Python + Scikit-learn), achieving 88% F1-score using XGBoost and SHAP explainability.
- Automated data preprocessing pipelines with Pandas and Dask, reducing cleaning time by 60%.

KEY PROJECTS

Plant Disease Detector *React Native, TensorFlow Lite*

- ML-powered app detecting 25+ plant diseases (90% accuracy) using ResNet50, with geolocation-based nursery recommendations via Google Maps API.

Medical Assistant AI *Flutter, Firebase, GPT-4*

- Symptom-checker app providing diagnosis/prescription drafts via GPT-4 turbo, serving 100+ monthly users.

SQL Database Chatbot *LangChain, Llama3-8B, Streamlit*

- Natural language-to-SQL translator supporting MySQL/SQLite via Zero-Shot ReAct agent, reducing query time by 70% for non-technical users.

EDUCATION

Bachelor of Science in Computer Science *National University of Modern Languages (NUML)* 2020 – 2024

- **Specialization**: Artificial Intelligence, Machine Learning, Natural Language Processing (NLP)
- **Thesis**: "SignChat – Real-Time Sign Language Translation System"
  - Developed a cross-platform application using Flutter and TensorFlow Lite to convert sign language gestures to text/speech in real-time
  - Achieved 98% accuracy on static ASL gestures using Feed Forward Neural-network
  - Nominated for "Best Graduation Project 2024" by faculty review panel

LEADERSHIP & VOLUNTEERING

Technical Team Member *Google Developer Student Clubs (GDSC)* 2022 – 2023

- Mentored 5 AI projects.

Graphic Head (ISB-Chapter) *Kaar-e-Kamal Welfare Organization* Jun 2021 – Oct 2023

- Led a 4-member design team to create 20+ promotional materials (posters, brochures) using Adobe Photoshop/Illustrator for fundraising campaigns, increasing donor engagement by 30%.
- Coordinated with NGOs to visualize impact reports, enhancing transparency for 10K+ beneficiaries.