

Muhammad Zain

📍 Tartu, Estonia

📞 +372 53618689

✉️ muhammad.zain@ut.ee

🌐 linkedin.com/in/ml-m-zain

🐙 github.com/muhammadzain727

Professional Summary

AI Engineer with hands-on experience building and deploying LLM-based systems in production environments. Strong background in retrieval-augmented generation (RAG), unstructured data processing, and backend integration of AI workflows. Experienced in developing reliable AI solutions for regulated and document-heavy domains, with a focus on robustness and practical impact.

Technical Skills

Programming: Python

Backend & APIs: FastAPI, REST APIs

Databases: PostgreSQL, MS SQL, Redis

Vector Databases: Qdrant, Pinecone, PGVector, FAISS

LLMs & Agentic Systems: Transformers, RAG Pipelines, Prompt Engineering, LoRA/QLoRA Fine-tuning, LangChain, LangGraph, LangSmith, HuggingFace, Ollama, OpenAI, Llama

Computer Vision: OpenCV, OCR, VLMs

Speech Processing: Speech-to-Text (STT), Text-to-Speech (TTS)

Data Processing: Pandas, NumPy, PDF and text data parsing

Machine Learning & Deep Learning: Scikit-learn, PyTorch, TensorFlow

Infrastructure & Tools: Docker, AWS (EC2, S3), Git, Linux (Ubuntu)

Work Experience

Associate Machine Learning Engineer

Seven Solutions – Lahore, Pakistan

Jan 2024 – Aug 2025

- Designed and implemented AI-driven systems automating workflows in healthcare, legal, recruitment, and marketing domains.
- Built and deployed retrieval-augmented generation (RAG) pipelines for large unstructured document collections.
- Developed data ingestion and normalization pipelines to process noisy text and PDF-based information into structured formats.
- Integrated LLM workflows with PostgreSQL and vector databases for scalable backend services.
- Improved model reliability through systematic evaluation, error analysis, and iterative refinement.
- Collaborated with cross-functional teams to translate business requirements into working AI solutions.

Selected Projects

AI Medical Encoder

- Developed an LLM-assisted system mapping unstructured clinical notes to standardized ICD-10-AM codes.
- Fine-tuned LLaMA 3.2 (3B) on ICD-10-AM coding datasets to improve classification accuracy.
- Designed validation workflows and deployed the system as an API integrated into enterprise intake processes.

Legal AI Agent

- Built a retrieval-augmented system generating structured summaries from large legal document corpora.
- Implemented semantic search for relevant legal cases using vector similarity to retrieve and ground responses in precedents.
- Designed document chunking and indexing strategies to improve retrieval accuracy and response reliability.

AI Recruiter

- Engineered an automated candidate evaluation system with structured scoring logic.
- Integrated backend database workflows to ensure consistent and auditable screening.

SEO & Marketing Agent

- Developed an LLM-based content generation system with structured prompting and evaluation pipelines.
- Implemented batch generation and API-based deployment for scalable usage.

Education

Master of Computer Science

University of Tartu, Estonia

Sep 2025 – Present

Relevant Coursework: Machine Learning, Deep Learning, Transformers, Distributed Systems, Research Methods

Bachelor of Computer Science

University of Engineering and Technology, Pakistan

Sep 2020 – Aug 2024

Relevant Coursework: Artificial Intelligence, Information Retrieval, Database Systems, Data Structures and Algorithms

Certifications

Artificial Intelligence (On-Campus Program)

Corvit Institute, Lahore

Sep 2023 – Dec 2023

Completed structured training in machine learning fundamentals and applied AI problem solving.

Languages

English – B2

Estonian – A1