

Muhammad Zain

AI / ML ENGINEER

muhammadzain91101@gmail.com

m-zain.com · Tartu, Estonia

+372 5361 8689



AI/ML Engineer focused on turning AI from a demo into something that runs reliably in production. Two years designing and shipping **autonomous agents**, **RAG pipelines**, and **LLM-powered systems** across multiple industries, with a current focus on **real-estate automation**: an always-on agent that searches live listings, engages sellers, and works around the clock without supervision. The work spans model fine-tuning, retrieval design, deployment, **CI/CD**, and the low-overhead infrastructure (**Firecracker microVMs**) behind it all. Currently completing an **MSc in Computer Science** at the University of Tartu.

EXPERIENCE

AI/ML Engineer, Markavy

Apr 2026 - Present

Remote · Seattle, Washington

- Building an **end-to-end, autonomous 24/7 real-estate agent** for an early-stage startup that matches buyers to properties from live **MLS feeds**, reaches out to seller agents, and schedules follow-up tasks with zero manual triggers.
- Engineered **long-term persistent memory** so the agent holds multi-session conversations and keeps working while the user sleeps: when no listing matches, it auto-schedules and continuously re-searches, then pings the user the moment one surfaces.
- Isolated each user in a dedicated **Firecracker microVM**, a per-user "AI computer" giving tenant isolation, security, and low-overhead round-the-clock execution at a **99.9% uptime** target.
- Designed a plugin-based **"domain pack" architecture** (a Real Estate plugin today, extensible to new domains) with **RAG** over bank mortgage policies and taxation rules, owning the full stack through **CI/CD**.

Associate Machine Learning Engineer, Xeven Solutions

Jan 2024 - Jul 2025 · 1 yr 7 mos

Lahore, Pakistan

- Contributed as part of the engineering team to AI solutions for **10+ clients** across healthcare, legal, and recruiting, working across data pipelines and production APIs.
- Built RAG systems and document-intelligence workflows that cut manual review time by up to **45%** in client pilots.
- Fine-tuned LLMs with **LoRA/QLoRA** and helped deploy containerized model APIs serving **thousands of requests daily**.
- Collaborated across the ML lifecycle, from data preparation and evaluation to prompt design and post-deployment monitoring.

AI Engineer & Consultant, Freelance (Part-time)

Dec 2025 - Present

Self-employed · Remote

- Delivering **production-grade AI systems**, RAG pipelines, and scalable model APIs for startups and SMEs alongside graduate study.
- Advised **2+ early-stage teams** on LLM architecture, retrieval design, and cost-efficient deployment.

SELECTED PROJECTS

Reva · Autonomous Real-Estate Agent, Markavy

24/7 autonomous agent matching buyers to live MLS listings with persistent memory and per-user Firecracker microVMs. **RAG · Agents · CI/CD**

AI Medical Encoder · Clinical Intelligence

ICD-10 coding engine on LLaMA 3.2 (3B) processing 10k+ codes; cut manual review ~45% in pilots. **Ollama · PyTorch**

AI LegalAgent · Law.co

Secure legal AI infrastructure with private LLMs, agentic workflows, and document intelligence. **LLMs · RAG · FastAPI**

AI Recruiter · Talent Pipeline

Surfaces the best-fit candidates for a role with reasoning and match stats behind each ranking. **LangChain · OpenAI**

EDUCATION

MSc Computer Science

2025 - 2027

University of Tartu, Estonia

BSc Computer Science

2020 - 2024

University of Engineering & Technology, Pakistan

Certified AI Engineer

Cert.

Corvit Systems, Pakistan

Languages

English — B2, Professional Working Proficiency

TECHNICAL SKILLS

LLMs & Agents: RAG, LangChain, LangGraph, LangSmith, Prompt Engineering, Transformers, Hugging Face, LoRA/QLoRA, Ollama, Claude, OpenAI, Llama, Whisper (STT), ElevenLabs (TTS)

Backend, APIs & Integrations: Python, FastAPI, REST APIs, Webhooks, Twilio

Vector DBs: Qdrant, Pinecone, PGVector, FAISS

Databases: PostgreSQL, MS SQL, Redis

ML / DL & CV: PyTorch, TensorFlow, Scikit-learn, OpenCV, OCR, VLMs

Infra, DevOps & Observability: Docker, Kubernetes, CI/CD, AWS (EC2, S3, Lambda), Firecracker microVMs, Git, Linux (Ubuntu), Sentry, Postman