

Haseeb Raza

Budapest, Hungary
Email : haseeb.javed715@gmail.com
Mobile : +36 70 586 6179

Portfolio : haseebraza.netlify.app
Github : github.com/haseebraza715
LinkedIn : linkedin.com/in/haseebraza715

Summary

AI Engineer and Computer Science student building Python-based LLM applications, RAG pipelines, agentic workflows, and responsible AI evaluation systems. Hands-on with LangChain, LangGraph, LlamaIndex, vector search, CI/CD, Docker, AWS, and monitoring.

EXPERIENCE

Infineon Technologies

Software / AI Engineering Intern

Cegled, Hungary
Aug 2025 – Present

- Built a production multilingual LLM extraction pipeline in **Python**, converting support conversations into **10+ structured fields** across **4+ languages** for internal ticket-routing services.
- Packaged the pipeline as a reusable module and added **Docker**, **GitHub Actions**, logging, and run-level metrics to improve deployment, monitoring, and operational reliability.

Agentic Systems Lab, ETH Zurich

Research Assistant, Agentic AI

Zurich, Switzerland (Remote)
Dec 2025 – Present

- Architecting a multi-stage LLM peer-review pipeline with evidence extraction, rubric-based critique, evaluation loops, and fairness benchmarking, modularized so stages can be swapped or rerun without touching orchestration.

RC2S2, ELTE

Responsible AI Research Intern

Budapest, Hungary
Sep 2025 – Present

- Quantified framing bias across **5 LLMs** and **15,552 ADL-derived personas** using **Python**, **OLS regression**, and lexicon-based indicators. Presented at STS NL 2026; accepted at ESA Warsaw 2026.

ELTE, AI Department

Machine Learning Research Intern

Budapest, Hungary
Jul 2025 – Dec 2025

- Operationalized **ACOM**, an improved Ant Colony Optimization Mapping method for 2D document embedding, outperforming PCA, t-SNE, and UMAP by **+11.6%** neighborhood preservation and **+3.8%** trustworthiness on 20 Newsgroups.

PROJECTS

Agent Autopsy | GitHub

- Built a local-first agentic debugging toolkit normalizing LangGraph, LangChain, and OpenTelemetry traces, auto-detecting **10+ failure classes** such as retry storms, hallucinated tools, and context overflow, and surfacing fix guidance via CLI and MCP server.

FailSafeRAG | GitHub

- Designed a failure-aware OCR-to-RAG pipeline with confidence-gated parsing, chunk-level recovery, vector indexing, and fallback routing so degraded inputs are flagged instead of silently corrupting retrieval.

AuditLens | GitHub

- Packaged a Python fairness-auditing library (pandas, scikit-learn, FastAPI) detecting class imbalance, subgroup outcome gaps, and sensitive-attribute correlations, with optional LLM interpretation and PDF/markdown reports.

SKILLS

Languages : Python, SQL, Bash/Linux, TypeScript, Java, FastAPI, Flask

AI Dev Tools : Claude Code, Codex, GitHub Copilot, Cursor, prompt engineering

AI / LLM : LangChain, LangGraph, LlamaIndex, RAG, FAISS, pgvector, semantic search, tool calling

Safety / Evaluation : LLM evaluation, guardrails, bias analysis, hallucination mitigation, monitoring

ML & Data : pandas, NumPy, PyTorch, scikit-learn, OLS regression

Infrastructure : Docker, GitHub Actions, CI/CD, PostgreSQL, MongoDB, AWS (ECS, App Runner, S3)

EDUCATION

Eötvös Loránd University (ELTE)

BSc in Computer Science, Stipendium Hungaricum Scholar

Budapest, Hungary
Feb 2023 – Jul 2026 (Expected)

- **Thesis** : EngageMind – RAG System with Stateful Memory, Adaptive Training, and Hallucination Mitigation