YOUSSEF SALEM

✉ youssefkhaled02@gmaill.com   📞 +4917663241140   📍 Berlin, Germany   🔗 Youssef Salem   ⊙ youssefmazenn

PROFESSIONAL EXPERIENCE

AI Research Assistant   03/2025 – 09/2025
University of the Bundeswehr Munich (UniBW) — Technical University of Munich (TUM)   Munich
- Designed **human-centered AI evaluation pipelines** for **LLM-driven conversational agents** in **decision-support scenarios**.
- Architected a **4×2 within-subjects factorial experiment** to **benchmark multiple LLMs** under **persona-based prompting**.
- Built a **full-stack experimentation platform** using **Python**, **FastAPI**, and **Next.js** for **scalable data collection**.
- Engineered **LLM persona prompting strategies** manipulating **epistemic certainty** for **comparative behavioral analysis**.
- Performed **statistical analysis** using **repeated-measures ANOVA** and **qualitative thematic analysis**.
- Analyzed **model-level vs. prompt-level effects**, highlighting **model-specific limits of persona control**.
- Conducted research under **cross-institutional supervision (UniBW & TUM)**; **presented at TUM** and **co-authored a submitted paper**.

Application specialist   09/2024 – 05/2025
Alliance Healthcare GmbH   Berlin
- Developed **Linux/Unix automation scripts** for **production data workflows** and **job scheduling**, reducing processing time by **35%** and saving **10+ hours per week**.
- Supported and optimized **business-critical data pipelines**, improving **system reliability** and operational efficiency in a **production environment**.
- Re-architected **technical documentation** for automated systems (**50+ pages**), reducing **support tickets by 20%** and improving **new-hire onboarding**.

Database software engineer   07/2024 – 08/2024
Middle East Oil Refinery (MIDOR)
- Built **Python-based ETL automation scripts**, reducing **manual data processing** by **10+ hours per week**.
- Implemented **enterprise-scale data validation**, detecting **150+ data inconsistencies** and improving **data integrity**.
- Collaborated with **engineering teams** to translate **business data requirements** into **robust, maintainable pipelines**.

SKILLS

Generative AI & ML: RAG • LangChain • LLM Fine-tuning • Hugging Face Transformers • PyTorch • Scikit-Learn • NLP **Data Science & Analytics:** Python • SQL • Pandas • NumPy • Matplotlib • Statistical Modeling • EDA **Backend & Software Engineering:** FastAPI • REST APIs • Backend System Design • Modular Architectures • Databases (SQL & NoSQL) • Linux / Unix Scripting • Automation **MLOps & Cloud:** Docker • Kubernetes • AWS • CI/CD • Git • Model Packaging & Deployment **AI Systems & Evaluation:** LLM System Design • Persona Prompting • Human-Centered AI Evaluation • Experimental Design • A/B Testing • Statistical Analysis

PROJECTS

Germany RAG Assistant — Production AI System 🔗
Technologies: Python, FastAPI, LangChain, ChromaDB, Sentence Transformers, OpenAI API, Docker, Kubernetes, Git, Makefile
- Designed and deployed a **production-grade Retrieval-Augmented Generation (RAG) system** using **LangChain, Sentence Transformers, ChromaDB, and FastAPI**, enabling **citation-grounded LLM responses** over real German government documents and minimizing hallucinations.
- **Containerized** and orchestrated the **AI backend** with **Docker** and **Kubernetes**, implementing **persistent vector storage (PVCs), Secrets & ConfigMaps, health probes, resource limits, and Horizontal Pod Autoscaling (HPA)** to deliver a **scalable, cloud-ready AI service**.

Conversational AI Agents in VR
Technologes: Unity, C#, FastAPI
- Created **immersive**, **voice-driven VR NPCs** by integrating **speech-to-text (Whisper)** and **text-to-speech (OpenAI TTS)** pipelines into **Unity** using a **Meta Quest 3** headset.
- Defined a **modular AI dialogue architecture** for **VR NPCs** leveraging **locally hosted LLMs**, achieving **<300 ms latency** for **real-time inference** and enhancing **conversational interactivity**.
- Developed a **Python-based backend** solution using **FastAPI** that processed **200+ requests per second** for **low-latency AI inference**, supporting **continuous, uninterrupted real-time interactions**.

EDUCATION

M.Sc. Computer Science for Digital Media   04/2026 – Present
Berliner Hochschule für Technik   Berlin, Germany

B.Sc. Computer Science and Engineering   10/2021 – 09/2025
German International University in Berlin   Berlin, Germany

LANGUAGES

English
C1, IELTS 7.5

German
C1, Telc C1 Hochschule