

Umer Butt

+49 1577 3677 910 | mumertbutt@gmail.com | [Linkedin@umertariq1](#) | [Github@UmerTariq1](#) | [Website](#) | Germany

Summary

Experienced AI engineer (5+ years) specialising in NLP and Information Retrieval. Proven ability to develop and deploy AI solutions, including generative AI, with a track record of success (34% search relevance improvement in banking IR system) in both industry and research. Seeking to leverage deep learning and LLM expertise for impactful AI projects.

Work Experience

Sequire Technology GmbH, Germany

August 2025 – Present

[AI Software Engineer](#)

- Conducting R&D on large language model (LLM) security, focusing on **pen testing, vulnerabilities, harmfulness, and toxicity**.
- Developed and implemented a Retrieval-Augmented Generation (RAG) application utilizing both local and **cloud-API-based LLMs**.

IMC AG, Saarbrücken, Germany

January 2025 – June 2025

[AI Software Developer \(Werkstudent\)](#)

- Leading the development of an **LLM-powered IR** platform to enhance client support workflows, using **Python, React, Docker**, and **Grafana** for scalable search and monitoring.

DFKI, Saarbrücken, Germany

April 2022 – December 2024

[AI Research Assistant](#)

- Developed and implemented NLP-driven **automation solutions** that processed over **1000+** complex documents, reducing processing time for enterprise clients (e.g., Siemens, Beko).
- Designed and built an **AI-powered student assistant** using **RAG, LLMs**, and **OpenAI API**.
- Collaborated on **AI model deployment pipelines**, ensuring scalability and optimization for real-world applications.

Love For Data

December 2020 – September 2021

[AI Engineer](#)

- Developed an **Information Retrieval system** for the banking sector, enabling access to **10,000+** legal regulations with a **34%** improvement in search relevance using Statistical Models and **Transformer Deep Learning Models**.
- Built a **Query Recommendation** module using **Document Clustering** and **NER**, improving query accuracy by **45%** and precision by **30%**; integrated **Elasticsearch** for efficient document indexing and real-time retrieval.

Systems Limited

June 2019 – November 2020

[Software Engineer](#)

- Supervised a team on multiple projects, including customer offer promotion and delivery systems. Remodelled and optimised implementation, achieving a **28% improvement** in performance and a **7% reduction** in false recommendations.

Education

• **Master's in Language Science and Technology**, Saarland University

October 2021 - July 2025

Thesis: Created datasets for Urdu and Roman Urdu **Information Retrieval** using MT model and built a **transformer** - based transliteration pipeline. Designed **Dense Retrieval** baselines with mT5 and deep learning models, addressing **data scarcity** and enabling **low-resource IR** advancements.

• **Bachelor's in Computer Science**, FAST NUCES

July 2015 - June 2019

Thesis: Implemented a neural-based IR system with **Elasticsearch, FAISS**, and embeddings from Word2Vec, LSTM, and BERT as part of my Bachelor's thesis. Achieved 2nd position among 300 projects in a university-wide competition.

Skills

- **Languages:** Python, Java, C++, JavaScript, SQL, HTML, CSS, NoSQL
- **Frameworks & Libraries:** PyTorch, TensorFlow, Huggingface, Scikit-learn, Spacy, Pandas, Numpy, Matplotlib
- **AI & NLP Tools:** RAG, OpenAI API, Transformers, Vector Search, Elasticsearch, FAISS, Langchain
- **Infrastructure & DevOps:** Git / GitHub, Docker, CI/CD, Data Pipelines, UNIX/Linux, AWS, GCP
- **Development Practices:** Agile, Clean Code, Experiment Tracking, Model Evaluation, Data Logging
- **Soft Skills:** Communication, Problem-solving, Adaptability, Ownership, Collaboration

Publications

- Roman Urdu as a Low-Resource Language: Building the First IR Dataset and Baseline **Umer Butt**, S.Varanasi, G.Neumann RANLP @ LowResNLPT, 2025
- AIDEN: Automatic Speaker Notes Creation and Navigation for Enhancing Online Learning Experience S.Varanasi, **Umer Butt**, G.Neumann, Josef van Genabith RANLP, 2025
- Low-Resource Transliteration for Roman-Urdu and Urdu Using Transformer-Based Models **Umer Butt**, S.Varanasi, G.Neumann NAACL @ LoResMT, 2025 [🔗](#)
- Enabling Low-Resource Language Retrieval: Establishing Baselines for Urdu MS MARCO **Umer Butt**, S.Varanasi, G.Neumann ECIR, 2024 [🔗](#)
- Cross-Linguistic Processing of Non-Compositional Expressions in Slavic Languages I.Zaitova, I.Stenger, **Umer Butt**, T.Avgustinova Cog-Alex, 2024 [🔗](#)
- Spatial-Economic Analysis for Optimal Electric Vehicle Charging Station Placement A.Maurya, C.Agnes, B.Baloch, S.Anwari, **Umer Butt** (equal contrib) SASI-ITE, 2024 [🔗](#)
- AutoQIR: Auto-Encoding Questions with Retrieval Augmented Decoding for Unsupervised Passage Retrieval and Zero-shot Question Generation S.Varanasi, **Umer Butt**, G.Neumann RANLP, 2023 [🔗](#)
- Fairy Tales / Children Stories of the World: Corpus Organization and Text Analytics Y.Khan, **Umer Butt** ICLAP, 2021 [🔗](#)

Projects

- RAG LLM chatbot with external knowledge source [🔗](#) Python, Llama3, Huggingface, Streamlit
- Graph-based Cloud Troubleshooting RAG Assistant [🔗](#) Neo4j, Llama3, LangChain, Streamlit
- AI Tool for Digital Persona Classification from Social Data [🔗](#) Python, Transformers, Streamlit, OpenAI, Spacy
- Deep IR System Using Context-Aware Vector Representations FAISS, LSTM, BERT, Elastic Search
- Lyrics Generator using LLMs and its website [🔗](#) Python, OpenAI API, and Langchain API, Angular, Django
- Detailed Time Series Analysis on Solar Energy Dataset [🔗](#) EDA, Arima, SARima models, Numpy, Pandas