Umer Butt

linkedin.com/in/umertariq1 | +4915773677910 | mumertbutt@gmail.com | github.com/UmerTariq1 | Germany

Summary

Experienced AI engineer (4+ years) with an MSc in Language Science and Technology, specializing 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 industry and research both. Seeking to leverage deep learning and LLM expertise for impactful AI projects.

Work Experience

DFKI, Saarbrucken, Germany April 2022 – Present
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.

University Of Saarland, Saarbrucken, Germany August 2023 – Present
Software Developer

- Led the development and optimization of a **Python Django**-based linguistic experiment platform for **1000+** active users, launching **10+** new experiments and maintaining **5** existing projects with a focus on code quality and performance.

Love For Data, Karachi, Pakistan 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, Karachi, Pakistan 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.

Publications

- Enabling Low-Resource Language Retrieval: Establishing Baselines for Urdu MS MARCO **ECIR** , 2024
  Umer Butt, S.Varanasi, G.Neumann 🔗

- Cross-Linguistic Processing of Non-Compositional Expressions in Slavic Languages **Cog-Alex** , 2024
  I.Zaitova, I.Stenger, **Umer Butt**, T.Avgustinova 🔗

- Spatial-Economic Analysis for Optimal Electric Vehicle Charging Station Placement **SASI-ITE** , 2024
  A.Maurya, C.Agnes, B.Baloch, S.Anwari, **Umer Butt** (equal contrib) 🔗

- AutoQIR: Auto-Encoding Questions with Retrieval Augmented Decoding for Unsupervised Passage Retrieval and Zero-shot Question Generation **RANLP** , 2023
  S.Varanasi, **Umer Butt**, G.Neumann 🔗

- Fairy Tales / Children Stories of the World: Corpus Organization and Text Analytics **ICLAP** , 2021
  Y.Khan, **Umer Butt** 🔗

Education

- **Master's in Language Science and Technology** , Saarland University October 2021 - Present

  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, C++, JS, HTML, CSS, SQL, NoSQL.

- **Software and Frameworks:** Pytorch, Pandas, Scikit-Learn, Huggingface, XGBoost, Git, MongoDB, ElasticSearch, Tensorflow.

- **Others:** Numpy, Pandas, matplotlib, RAG, OpenAI API, Agile development, Data pipelines, Docker

- **Soft Skills:** Problem-solving skills, for data-driven challenges and pattern recognition with clear communication.

Personal Learning Projects

- RAG LLM chatbot with external knowledge source   🔗 **Python, Llama3, Huggingface, Streamlit**
- 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**
- Click through Rate and No. of Clicks Prediction   🔗 **TFIDF, Word Embeddings, Linear Regression, and RF**
- Extensive EDA and Modeling on Advertisement Dataset   🔗 **Pandas, Numpy, Linear Regression, Matplotlib**
- Detailed Time Series Analysis on Solar Energy Dataset   🔗 **EDA, Arima, SArima models, Numpy, Pandas**