Raigon Kunnath Augustin

📍 Konstanz, Germany
⌂ https://github.com/raigon44/

📞 +49-163 4792657
✉ raigonkunnathaugustin@gmail.com
🔗 raigon-augustin-a10911ab/

EDUCATION

- **Universität Konstanz** — *Konstanz, Germany*
MSc, Computer and Information Science, Data Science — April 2021 - September 2023

- **Government Engineering College, Thrissur** — *Kerala, India*
Bachelor of Technology, Computer Science and Engineering — Sept. 2011 - April 2015

EXPERIENCE

- **Universität Konstanz** — *Konstanz, Germany*
Student Research Assistant — May 2022 - June 2023
  - **Argument Relation Classification:** Fine-tuned multiple pre-trained large language models such as BERT and roBERTa to accurately predict the relation between two text argument units in the QT30 dataset. *(Python, PyTorch, NLP, Research)*
  - **Video Search System:** Developed a web-based tool to perform text-based search of frames within a video. Reduced the searching time by utilizing visualization-based techniques and Open AI's CLIP model. *(Python, Flask, PyTorch, Redis, JavaScript, D3.js, Docker)*
  - Conducted experiments using the pre-trained embeddings from BERT to explore the capabilities of large language models to understand various linguistic concepts. *(Python, LLM, PyTorch, Data Mining)*

- **SAP** — *Bangalore, India*
Developer — April. 2019 - June 2021
  - **Log Classification:** Trained and deployed a machine learning model to classify error logs generated during integration tests, thereby reducing the manual effort of QA by 35%. *(Python, Flask, NLP, TensorFlow, Docker)*
  - Worked as a DevOps Engineer for various lifecycle management tools developed by the CLM team.
  - Developed a Jenkins plugin to add custom workflow steps to the CI/CD pipelines. (Java, Jenkins, Groovy, Maven)
  - Developed a Jenkins shared library in Groovy to automatically enable SAP ABAP systems to communicate with GitHub. (Jenkins, Groovy, GitHub)

- **SAP** — *Bangalore, India*
Associate Developer — Sept. 2015 - March 2019
  - **Chat Bot:** Developed a prototype chatbot to support development teams with queries related to cloud hotfixes, thereby reducing the email conversations and manual effort of Release engineers. *(Python, RASA, SAP Conversational AI, NLP, Spacy, NLU)*
  - Designed and implemented RESTful APIs using the Flask framework to serve as the backend for various applications.
  - Experienced in automating various manual tasks using Python, Groovy, and Shell scripts.
  - Worked as the production coordinator for SAP Rapid deployment solutions add-ons.

PERSONAL PROJECTS

- **Joke Generation & Rating using LLM** — ⌂ *GitHub Repository*
An interactive chatbot that can generate various categories of jokes and assess the quality of jokes.

  - Using the rjokes dataset, a GPT-2 model was fine-tuned for joke generation, and a BERT model was fine-tuned for predicting the humor level of the joke.
  - *Tools & technologies used:* Python, PyTorch, Pandas, NLP, Transformers, pytest

- **Mojo QA** — ⌂ *GitHub Repository*
A RAG-based LLM application to answer queries related to Mojo programming language

  - Extracted official Mojo documentation and created a vector store containing corresponding embeddings.
  - For the user query, most similar embeddings are retrieved from the vector store and used as context in the prompt.
  - Used the 4-bit quantized LLAMA-2 7B chat model to generate answers for the user query.
  - *Tools & technologies used:* Python, Transformers, LLM, NLP, Streamlit, DeepLake, Beautiful Soup, LangChain, Prompt engineering, Retrieval-augmented generation (RAG)

- **Deep Learning Based Image Retrieval System** — ⌂ *GitHub Repository*
An image search system powered by deep learning that retrieves similar images based on an input image.

  - Trained an autoencoder model on the CIFAR-10 dataset and used the encoder model for the task of image retrieval. Conducted rigorous sanity checks to ensure system robustness and enhanced model stability by retraining on noisy and transformed images.

– *Tools & technologies used:* Python, Keras, TensorFlow, OpenCV, Numpy, Computer Vision

- **Content Diffusion from Social Media to News Quotes (MSc Thesis)**   *Nov 2022 - May 2023*

Analyzed the impact of social media on the News Gatekeeping process in the context of social movements.

- – Extracted, transformed, and processed data from Twitter, Wikidata, Quotebank, and various other websites.
- – Generated new features such as author emotion, and stance from tweets and news articles by using state-of-the-art language models.
- – Statistical analysis conducted to understand the news selection, author selection, and objectivity in reporting tweets in mainstream news media.
- – *Tools & technologies used:* Python, PyTorch, Transformers, LLMs, Pandas, Numpy, ETL, NLP, Beautiful Soup, Data Modeling, Research, Prompt engineering

Technical Skills

Programming Languages: Python, Java, JavaScript, HTML, Groovy, SQL
Tools/Frameworks: Git, Jenkins, RASA, Docker, Flask, REST, LangChain, Streamlit
Libraries: Pandas, NumPy, PyTorch, Tensorflow, Scikit-learn, D3.js, Matplotlib, Scipy, Keras, spaCy, Huggingface
Databases & OS: MongoDB, MySQL, ChromaDB (vector database), Neo4j (basics), Linux
Soft Skills: Project Management, Agile Methodology, Stakeholder Management
Areas of Interest: NLP, Computer Vision, Data Science, Data Engineering, MLOps, Generative AI, Machine Learning
Languages: English (Fluent - written and verbal), German (Beginner), Hindi (Advanced)