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 / NLP — April 2021 - August 2023
- **Government Engineering College, Thrissur** — *Kerala, India*
  Bachelor of Technology, Computer Science and Engineering — Sept. 2011 - April 2015

Experience

- **Universität Konstanz** — *Konstanz, Germany*
  Research Assistant — May 2022 - June 2023
  - **Video Search System:** Implemented a video search system by developing a web-based tool to perform a 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)*
  - **Argument Relation Classification:** Fine-tuned multiple large language models such as BERT, roBERTa, etc. to accurately predict the relation between two text argument units in the QT30 dataset and best-performing model had comparable F1-Score with the benchmark Student Essay dataset. *(Python, PyTorch, NLP, Research)*
- **SAP** — *Bangalore, India*
  Developer — Sept. 2015 - 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)*
  - **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)*

Personal Projects

- **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 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
- **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, LLM
- **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 unstructured data from Twitter, Wikidata, Quotebank, and various other websites.
  - Multiple features generated from text data fields from tweets and news articles by utilizing various NLP methods such as author emotion detection, stance detection, sentiment analysis, etc.
  - 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, LLM, Pandas, Numpy, ETL, NLP, Beautiful Soup, Data Modeling, Research

Technical Skills

Programming Languages: Python, Java, JavaScript, HTML, Groovy, SQL
Tools/Frameworks: Git, Jenkins, RASA, Docker, Maven, Robot Framework, Flask, REST, Jupyter Notebook
Libraries: Pandas, Numpy, PyTorch, TensorFlow, scikit, D3.js, Matplotlib, Scipy, Keras
Databases: MongoDB, MySQL
Soft Skills: Project Management, Agile Methodology, Stakeholder Management
Areas of Interest: NLP, Computer Vision, Data Science, Data Engineering, MLOps, GenAI
Languages: English (Fluent - written and verbal), German (Beginner), Hindi (Advanced)