Mr. Jiufeng Li

Email : lijiufeng97@gmail.com
Phone : +49 15231306943

Education

- **Universität Heidelberg** Oct 2022 – Present
  M.Sc. Data and Computer Science

- **Northeastern University** (211 & 985 projects), China Sep 2016 – Jun 2020
  B.Sc. Computer Science & Technology, ESI Rank Top 1‰ (GPA 86/100)

Experience & Projects

- **Mercedes-Benz Headquarters, Stuttgart** May 2024 – Present
  Data Science Intern, Data Science, Machine Learning, Recommender System

  DS/ML/AI: As a **Data Scientist**, I am engaged in the development and optimization of **AI-driven deep neural networks** using **PyTorch** to build **industry-leading vehicle recommender systems at scale** and roll out to **16 markets**. This involves refining **ML algorithms** to better align with **customer preferences**. Additionally, I am responsible for architecting and implementing robust **ETL pipelines** with **Apache Spark** and **Polars** hosted on **Databricks** on a **Microsoft Azure cloud server** that streamlines the processing and integration of large-scale customer data, ensuring accuracy and efficiency in data handling. For evaluating different ML models, we use **MLflow** to track the performance and deploy the best model.

- **The Academy of Science of Heidelberg, Heidelberg** Aug 2023 – Mar 2024
  Data Engineer & Software Engineer Work Study - Team ALMA

  At the Heidelberg Academy of Sciences, I spearheaded the creation of the **ETL pipeline** for our dataset and built multiple data layers including raw, bronze, and silver layers using **Apache Spark.** I also contributed to the creation of **XCC-Viewer**, an advanced tool for editing and **validating XML** with **TEI standards**, featuring user-centric highlights for key document sections.

- **BGI Genomics Shenzhen, China** Dec 2021 – Mar 2022
  BGI Genomics is the world's largest genome research organization.

  ML/AI: In my role at BGI, I pioneered the application of **Machine Learning** and **Deep Neural Networks** to refine **cell segmentation and imaging analysis**. Utilizing high-resolution imaging, I was involved in the creation of an extensive dataset for training, testing, and validating our models. The images were carefully pre-processed to enhance model intake, with segmentation into 32x32 tiles tailored for the YOLOv4 model with **PyTorch**. Through **transfer learning** techniques, the **YOLOv4** model was adeptly trained to segment and delineate cellular structures with a **15% improvement** in average accuracy.

- **Scientific Medical Articles Question-Answering Chatbot** Aug 2023 – Dec 2023
  Leveraging the **LangChain** framework, my team and I engineered a sophisticated **Question-Answering (QA)** bot, which utilizes a robust pipeline comprising advanced text processing, reading, and retrieval components. Our data (from PubMed) are housed in an **OpenSearchDocumentStore**, renowned for its **elastic search** capabilities and vector storage efficiency. To enhance the bot's interactivity and question formulation, we integrated ChatGPT, further refining the system's accuracy with the introduction of a **Retrieval-Augmented Generation (RAG).** Complementing our backend achievements, we designed an aesthetically pleasing and **user-friendly UI** using **Streamlit,** creating an engaging and seamless user experience.

Technical skills

- Programming Languages: Python, C, C++, Java
- Machine Learning: Familiar with modern DNNs architecture and networks (like ResNet & Transformer)
- Databases: SQL, MySQL, PostgreSQL, and MongoDB
- Version Control: Git, Gitlab, and GitHub
- Frameworks: Familiar with PyTorch, Mxnet, TensorFlow, and Apache Spark