Kavu Maithri Rao

www.linkedin.com/in/maithri-rao • +49 1636 993553

Riegelsberg, Germany • *maithrirao30@gmail.com*

Machine Learning Engineer with 3+ years of experience building and deploying data-intensive ML and deep learning systems. Strong expertise in Python, PyTorch, and GPU-accelerated training, with a proven track record of moving prototypes into scalable, containerized production systems using Docker, CI/CD, and AWS. Experienced in LLMs, computer vision, data pipelines, and model optimization.

Skills & Technologies

- **Languages:** Python, SQL, Java
- **Tools & Frameworks:** PyTorch, Transformers, scikit-learn, OpenCV, Pandas, NumPy, LangChain, Vector Databases, MLflow, Docker, Git, CI/CD (Jenkins/GitHub Actions), REST APIs, AWS, Linux, Unit Testing (pytest)
- **Expertise:** Machine Learning & Deep Learning, Computer Vision, Generative AI & LLMs, Retrieval-Augmented Generation (RAG), Data Pipelines & Feature Engineering, Model Training & Optimization, GPU-Accelerated Training, Experiment Tracking & Reproducibility, Model Deployment & MLOps, Video & Image Processing

Work Experience

ML Engineer (Research)   **German Research Centre for Artificial Intelligence**   **June 2023 - December 2024**
Saarbruecken, Germany

- **Owned and optimized end-to-end data and model pipelines for low-resource German Sign Language translation,** fine-tuning LLMs using PyTorch with GPU acceleration and quantization, achieving a 36% BLEU score improvement.
- **Built reproducible preprocessing and data augmentation workflows (Python, Pandas, NumPy),** expanding training data from 90 to 300 documents and improving model robustness in low-resource ML settings.
- **Architected multi-output transformer model with 30+ parallel prediction heads**, enabling automated generation of structured sign language data for avatar-based rendering systems.
- **Designed systematic model comparison across 3 architectures and tokenization** strategies using 10 fold cross-validation, identifying optimal hyperparameters that maximized translation quality.
- **Deployed containerized ML pipelines using Docker, AWS, and CI/CD,** enabling repeatable experiments, model monitoring, and production-grade training runs.

Data Engineer   **IDnow**   **March 2022 - March 2023**
Munich, Germany

- **Delivered high-quality annotations for 50,000+ images,** improving data quality and consistency for computer vision-based fraud detection models.
- **Implemented Python-based digital document models, converting complex ID layouts into machine-readable representations,** contributing to 85% validation accuracy in document segmentation pipelines.
- **Developed synthetic training data generators** by utilizing Python and GIMP to simulate diverse ID security features, effectively addressing data imbalance and improving pipeline robustness.
- **Collaborated with ML, QA, and product teams in a Kanban-driven CI environment,** supporting testing, validation, and deployment workflows using Git, pytest, and Jenkins.

Junior System Engineer - Data & Analytics   **Tata Consultancy Services (TCS)**   **December 2017 - September 2019**
Bangalore, India

- **Developed SQL-based analytics and Power BI dashboards for clinical datasets (100+ patient records),** transforming raw data into operational and reporting insights.
- **Optimized complex SQL queries and database procedures by refining join logic and indexing strategies,** reducing report generation time and improving the responsiveness of dashboards.
- **Ensured data integrity and pipeline reliability** by implementing unit and integration tests in Java.
- **Streamlined data mapping and technical documentation by collaborating with business analysts to translate clinical requirements into technical specifications,** ensuring precise alignment between data ingestion and final reporting.

Education

M.Sc: Computer Science, Saarland University, Germany	**May 2020 - July 2025**
Bachelor of Engineering: Computer Science, The National Institute of Engineering, India	**July 2013 - May 2017**

Certificates

LLM for Enterprise: Technical Protocols, Considerations, and Data Privacy, LinkedIn Learning	**August 2025**
Agentic AI Fundamentals: Architectures, Frameworks, and Applications, LinkedIn Learning	**September 2025**
Building AI Applications with Amazon Bedrock, LinkedIn Learning	**October 2025**

Projects & Extra

Thesis: Sign Language Video Segmentation using Temporal Boundary Identification   **December 2024**

Developed an end-to-end ML pipeline using Python, PyTorch, and video processing techniques to automatically segment long-form sign language videos and generate subtitle timestamps via temporal boundary identification, enabling scalable video-to-text alignment. Built a fault-tolerant, parallel processing system using multiprocessing, OpenCV, and SQLite. This work was published at ACL-SRW 2025 (63rd Annual Meeting of the ACL).

RAG Based Chatbots   **July 2025**

Built retrieval-augmented generation (RAG) chatbots using SentenceTransformers, LangChain, vector databases, and prompt engineering, supporting medical and open-source domains with semantic search and grounded LLM responses. Deployed containerized applications using Docker, AWS, and CI/CD pipelines, enabling reproducible builds, automated deployments, and inference.

End-to-End YouTube Sentiment Analysis   **August 2025**

Implemented a sentiment analysis pipeline using Python, TF-IDF, and LightGBM, processing thousands of YouTube comments with experiment tracking and model versioning via MLflow. Deployed the system through a Dockerized Flask REST API on AWS, enabling scalable, consistent inference and reproducible model serving.

Languages

- English: Professional Proficiency
- German: Intermediate Proficiency

References

Eleftherios Avramidis

Senior Researcher, German Research Center for Artificial Intelligence (DFKI GmbH)

Email: eleftherios.avramidis@dfki.de