

GARIMA SWAMI

AI ENGINEER

Garimaswami646@gmail.com

https://garimaswami.vercel.app/

https://github.com/workgarimaswami

Cottbus, Germany

http://linkedin.com/in/garimaswami

§ 16b Abs. 1 AufenthG



PROFIL

AI engineering graduate student with practical experience in training, evaluating, and deploying machine learning systems in cloud-based environments. Worked on speech-based AI systems, large-scale NLP pipelines, and production-ready ML deployments using Docker and AWS. Strong interest in audio intelligence systems, model benchmarking, and building end-to-end AI solutions from research prototype to scalable deployment

WORK EXPERIENCE

07.2024 – 04.2025 Bengaluru, India	<div><div>Software Developer, CodingPro</div><ul style="list-style-type: none">Supported training and fine-tuning of transformer-based NLP models using PyTorch/TensorFlow, including dataset preparation and validation workflows for production use.Built and maintained ML pipelines for preprocessing large unstructured datasets (interview transcripts, text responses) used for model training and evaluation.Contributed to model benchmarking by monitoring key KPIs such as validation accuracy, response quality, and inference latency.Supported deployment of inference services in Docker-based containerized environments, ensuring scalable and production-ready integration.Developed a React-based proof-of-concept web interface to demonstrate AI model capabilities in real-time interview simulations.Optimized inference workflows to improve latency and overall system stability.</div>
02.2024 – 05.2024 Bengaluru	<div><div>Software Developer, Project Trainee, Indian Space Research Organisation</div><ul style="list-style-type: none">Built a speech-based AI system for employee information retrieval, enabling audio-driven attendance and automated performance analysis using NLP.Processed and validated large structured and unstructured datasets for model training, evaluation, and scalable system design.Developed FMECA(Failure Modes, Effects, and Criticality Analysis) automation software to analyze spacecraft failure modes and compute risk prioritization metrics.Designed dynamic reporting and validation workflows to ensure accurate, engineering-grade reliability assessments.</div>
06.2023 – 07.2023 Delhi, India	<div><div>Full Stack Intern, Municipal Corporation of Delhi</div><ul style="list-style-type: none">Developed smart urban solutions for efficient waste management, utilizing technologies such as Java, NLP, IoT, geotagging, GIS, and cloud computing.</div>

EDUCATION

10.2025 – heute Germany	<div>M.Sc. in Artificial Intelligence, Brandenburgische Technische Universität</div>
2021 – 2024 Bengaluru, India	<div><div>B.Sc. in Computer Science, M.S. Ramaiah University of Applied Sciences</div><div>GPA: 1.9 / 4.0 (equivalent grading scale)</div></div>

PROJECTS | GITHUB

Enterprise Knowledge Assistant (RAG System), LLMs • LangChain • FastAPI • Vector DB • Docker

- Built retrieval-augmented generation pipeline with semantic indexing and hybrid search.
- Implemented document ingestion, chunking strategies, and embedding-based retrieval.
- Designed FastAPI backend for real-time querying.
- Containerized full-stack system for scalable deployment.
- Evaluated retrieval accuracy and optimized similarity thresholds.

[End-to-End MLOps Pipeline | AWS • Docker • CI/CD \(demo\)](#) [↗](#)

- Integrated logging and alerting systems for proactive maintenance and error tracking
- Collaborated with data science and engineering teams to enhance model performance
- Conducted thorough testing across multiple stages to ensure reliability and accuracy
- Leveraged infrastructure as code (IaC) for seamless environments setup and scalability
- Developed comprehensive documentation for codebase and deployment procedures

SKILLS

Cloud & Deployment

Docker (Containerized Deployment), Kubernetes (Basic), OpenShift (Exposure), AWS (EC2, S3, IAM), CI/CD Pipelines (GitHub Actions), REST API Deployment (FastAPI/Flask), Model Monitoring & Logging

Data handling & Databases & Vector Search

Large Dataset Processing, Data Preprocessing & Feature Engineering, Structured & Unstructured Data Pipelines
SQL, MySQL, FAISS, ChromaDB

Machine Learning & Model Development

PyTorch, TensorFlow, Scikit-learn, JAX (Basic Exposure), Model Training & Fine-Tuning, Transformer Architectures, Performance Evaluation & Model Benchmarking, Hyperparameter Optimization

Audio & NLP Systems

Speech Recognition Systems, Semantic Analysis, Transformer-based NLP Models, Large-Scale Text & Audio Data Processing

Programming Languages

Python, SQL, JavaScript, HTML5, CSS3, Bash

Data Engineering

Handling Large Structured & Unstructured Datasets, Data Preprocessing & Feature Engineering, ETL Pipelines, Pandas, NumPy

Frontend Development

React (Proof-of-Concept Interfaces), API Integration for ML Applications, Responsive UI Development (HTML, CSS, JavaScript)

Development Environment & Tools

Linux, Vim, Git, GitHub, VS Code, Jupyter Notebook, Postman

PUBLICATIONS

02.2025 [Early detection of dyslexia Using Multimodal Analysis](#), [↗](#), INCOFT

12.2023 [WeSAFE : Enhancing Human Security Mobile Application](#) [↗](#), IJIRCC

SPRACHEN

English: C1 - Fluent

German: A2 - B1 (Currently pursuing)