KALLABETTU SHAMANTH KUMAR
Hannover, Germany

📞 +49 15226237305   ✉ kshamanth49@gmail.com   in www.linkedin.com/in/k-shamanth-kumar-524194206

Profile Summary

AI engineer with expertise in creating and implementing **Large Language Model** (LLM) systems of production quality. knowledgeable about **MLOps** procedures, **FastAPI**-based backends, **LangChain/LangGraph** pipelines, **multi-agent system** design, and **Retrieval-Augmented Generation** (RAG). trained at transforming AI prototypes into practical, safe, and effective applications. focused to providing significant answers for difficult problems facing the general public industry.

Key Skills

- **Programming:** Python, Java
- **GenAI & LLMs:** GPT-4, LangChain, LangGraph, Prompt Engineering, RAG pipelines
- **Model Adaptation:** Stable Diffusion, LoRA, QLoRA, fine-tuning foundation models
- **Deep Learning:** PyTorch, Scikit-learn, Neural Networks
- **API Development:** FastAPI, Pydantic, scalable API design
- **MLOps & Deployment:** CI/CD, Docker, Kubernetes, GitLab, Bitbucket, Azure AI Foundry
- **Automation:** Multi-agent workflows, LLM orchestration, process automation
- **Evaluation:** LangSmith, RAGAS, custom QA/evaluation tools
- **Analysis & Visualization:** Numpy, Pandas, Matplotlib
- **Data Handling:** SQL, unstructured data (PDFs)
- **Documentation:** Technical writing, internal knowledge sharing

Professional Experience

Concept Development Engineer
IAV GmbH | Gifhorn, Germany | Nov 2023 – Present

- Designed and implemented **multi-agent AI** workflows for **dSPACE** simulation automation, utilizing **LangGraph**, **Pydantic**, and **GPT-4.1**, streamlining product development from scratch. Developed robust **FastAPI** backends to support scalable agent services.
- Fine-tuned and adapted **Stable Diffusion** and **GAN** models (**LoRA**) for object addition, erasing, and modification tasks, enabling new research projects in generative image manipulation for real-world data augmentation.
- Authored technical documentation and contributed to internal knowledge sharing on **Generative AI** workflows and model adaptation best practices.
- Applied **MLOps** best practices, such as **versioning**, **monitoring**, and **scalable deployment**, to optimize system performance for production use.
- Developed and automated Python and Java test workflows using TPT frameworks, ensuring robust software validation and high product quality.

Working Student – Computer Vision & AI
IAV GmbH | Gifhorn, Germany | Jun 2022 – Dec 2022

- Developed **computer vision** software and managed end-to-end data pipelines for automated measurement and testing workflows.
- Improved AI/ML workflow efficiency by implementing automated software-in-the-loop testing and result validation.

Projects

- **Multi-Agent Simulation Generator**
  - Built end-to-end pipeline using **LangGraph**, **LangChain**, **FastAPI**, and **Pydantic** for complex simulation generation from user inputs.
  - Integrated **RAG** methodology for improved information retrieval and context-rich outputs.
  - Deployed and maintained production-ready APIs with Docker and **CI/CD** workflows.
- **Pedestrian Intention Prediction**
  - Designed deep learning model with S**patio-temporal** features for autonomous driving scenarios.
- **Incremental Learning for Soft Robotics**
  - Developed an I**ncremental learning system** to predict robotic actuator trajectories.

Education

- **M.Sc. in International Mechatronics (Grade: 1.4)**
  - Leibniz Universität Hannover, Germany & Peter the Great St. Petersburg Polytechnic University, Russia
- **B.E. in Mechanical Engineering (Equivalent Grade: 1.7)**
  - NMAM Institute of Technology Nitte, India

Aditional Information

- Experienced with Jira, Confluence, GitLab, Bitbucket.
- Flexible, creative problem-solver with strong teamwork skills.
- Passionate about building trustworthy AI solutions that help organizations work more efficiently and support effective teamwork.

Languages

English - Fluent C1   **Deutsch** - Intermediate B2