SANDEEP S

AI/ML Engineer | GenAI | ADAS | Robotics | Python Developer

Email: sandeepsnair1017@gmail.com | Phone: +91 8157846024 |

LinkedIn: https://www.linkedin.com/in/sandeep-s-a179ba226/

GitHub: https://github.com/SANDEEPSAJEEV

Professional Summary

AI/ML Engineer with 2 years of hands-on experience in developing Large Language Models (LLMs), Generative AI (GenAI) systems, and ADAS pipelines. Skilled in building Retrieval-Augmented Generation (RAG) applications, integrating ROS with perception systems, deploying AI on NVIDIA Jetson platforms, and leveraging LangChain, FAISS, and Azure/AWS services. Proven track record in fast-paced startup environments, delivering end-to-end ML systems with Python, PyTorch, and TensorFlow.

Skills

- **Languages/Frameworks**: Python, SQL, PyTorch, TensorFlow, OpenCV, FastAPI, Streamlit

- **AI/ML**: EDA, CNN, NLP, RNN, Transfer Learning, Drift Monitoring, Model Tuning

- **GenAI/LLM**: LangChain, FAISS, Ollama, OpenAI API, Claude, Huggingface, RAG, LoRA, QLoRA

- **ADAS/Robotics**: ROS, Autoware, SLAM, CARLA, Jetson AGX Orin, Radar & Camera Fusion

- **Cloud & DevOps**: Azure (Custom Vision, Blob Storage), AWS (Bedrock, Lambda, S3,Sagemaker), Triton Inference Server

- **Tools**: Tools: Git, Docker, VSCode, Jupyter, Arduino, MATLAB, Proteus, TI AWR1843BOOST, C++

Education

Rajagiri School of Engineering and Technology
B.Tech in Electronics & Communication (2019–2023)

Experience

Software Engineer | RoshAi private limited
(Autonomous Vehicle R&D | ADAS, GenAI, Embedded AI)   **Ernakulam, India**

Jan 2024 - present

ADAS & Robotics
- Delivered camera and radar-based emergency braking using ResNet and YOLO; deployed on NVIDIA Jetson AGX Orin with Triton Inference Server.
- Designed and implemented camera-radar fusion pipeline for robust obstacle detection and tracking under varying environmental conditions.
- Built bird's-eye view system from multi-camera fusion; validated in CARLA simulator.

AI/ML & GenAI
- Developed multi-agent LLM applications using LangChain, FAISS, Groq, and OpenAI APIs.
- Built RAG-based document QA bots and blog generators using CrewAI and Ollama.
- Deployed translation APIs using LangServe, FastAPI, and Groq's Gemma Model.

ROS & Simulation

- Integrated ROS-based pipelines for autonomous navigation, obstacle detection, PID control, and HD map generation.
- Simulated safety-critical scenarios using CARLA to validate perception models.

*IoT Intern – SFO Technologies (**NeST Digital**),* April 2022– June 2022

- Worked on ESP8266, Arduino, and real-time sensor data collection for IoT systems.

Projects

(*All code/projects available on GitHub**: **github.com/SANDEEPSAJEEV**)*

AI/ML & GenAI Projects

- **Multi-Agent Conversational AI System –**Built a scalable multi-agent chatbot framework leveraging LangGraph for agent orchestration, Groq's Gemma-2 Large Language Model (LLM) for low-latency inference, HuggingFace Transformers for embedding generation, and Astra DB (Cassandra) as a vector store. Enabled modular agent communication and enhanced real-time retrieval.

- **Multi-LLM Chatbot** – Developed a chatbot that supports dynamic model switching and session memory using LangChain and Groq; reduced user response latency and improved conversation continuity.

- **YouTube Blog Generator** – Designed an autonomous content generator using CrewAI and Ollama to extract technical content from YouTube and transform it into human-readable blogs.

- **AWS Bedrock QA Bot** – Implemented a serverless document QA bot using Claude, Titan Embeddings, FAISS, and LangChain via AWS Bedrock and Lambda; supported scalable, multi-document querying.

- **FastAPI Translator** – Built a REST API using LangServe, Groq, and gemma2-9b-it to translate input text to French; containerized for scalable deployment.

- **IMDB Sentiment Classifier** – Trained an RNN-based sentiment analysis model on IMDB movie reviews using TensorFlow and Keras; achieved high validation accuracy..

Computer Vision / Medical AI

- Diabetic Retinopathy Detector – Built ResNet-50/DenseNet-121 model; evaluated using accuracy and specificity.

Robotics & Embedded

- Radar Navigation for ADAS – Speed modulation and emergency braking using radar signals.

- Bird's Eye View System – Camera fusion pipeline with deep learning model validation.

- Navigation Pipeline – HD map creation and obstacle avoidance using custom path planning algorithms.

Awards & Recognition

- **Most Reliable Achiever Award – RoshAi Pvt. Ltd. (2025)**
  For exceptional dedication and successful delivery of AI/ADAS projects in the autonomous vehicle domain.

Strengths

Effective communicator, fast learner, and strong collaborator with leadership experience in tech events and AI development teams.

Volunteering

- Led multiple tech and cultural events as Program Committee Head, fostering strong team collaboration.
- Delivered a technical talk on "Metaverse" to a student audience, enhancing public speaking and tech advocacy.
- Member of IEEE and IEDC clubs, contributing to workshops, innovation challenges, and industrial visits.