SAI NIVEDH V

sainivedh26@gmail.com | 8300874166
[Github](#) | [LinkedIn](#)

Skills

Languages: Python, Java, C, SQL, JavaScript, Kotlin, Flutter, HTML,CSS, MATLAB

Technologies & Frameworks: Tensorflow, PyTorch, Azure, ChromaDB, Bootstrap, Django, Flask, scikit-learn, Langchain (Generative AI framework), LlamaIndex, LORA Quantization and Fine tuning LLM, RAG(hybrid,Reranker,merger retriever,long context retriever,sequence window etc…), Autogen framework, Neo4j(Graphdb), FastAPI, docker, CI/CD pipeline, MongoDB, pinecone, LangGraph (multi AI Agents), Unsloth

Work Experience

Binari Intelligence System Private Limited   **Nov 2024 - Dec 2024**

Software Development Intern (Hybrid)   **Salem, Tamil Nadu**

Projects

Deepfake Detection   **Mar 2024**

- Designed and implemented an application leveraging the ResNeXt-50 model for accurate identification of deepfake videos. Utilized the Celeb-DF (v2) dataset, containing 6000+ real and deepfake samples, to train the model, ensuring high reliability in detecting manipulated content.
- Integrated Intel's OneAPI Deep Neural Network Library (oneDNN) and OpenVINO toolkit to enhance inference speed and reduce latency on both GPUs and CPUs, achieving accelerated AI processing without compromising model accuracy.

Sign Language Mapping for people with Hearing Disabilities   **June 2024**

- Developed an innovative application that generates sign language videos from diverse digital media formats, including YouTube links, local video/audio files, and text articles, utilizing Google Cloud Speech-to-Text API for real-time transcription and NLP techniques to translate text into Indian Sign Language (ISL) syntax.
- Engineered a seamless video compilation pipeline using MoviePy and GIFs to map and animate words, enabling accurate sign language representation by leveraging pre-computed similarity scores and fallback mechanisms for unmatched terms, resulting in a unified, accessible video output.

Nutrino-RAG - an Full stack Health care Application   **November 2024**

- Structured an entire new pipeline for processing Complex clinical data which are in Tabular forms
- Implemented a Full stack RAG application which is capable of providing personalized diet plan recommendation Along with Accurate Food component Analysis. Created Interactive UI with React and FastAPI as backend for High speed Async performance.It's powered with Langchain framework,crawl4ai (high speed scrapping),groq(fast llm inference), OpenAI embeddings, chroma , mongo atlas and supporting services aligned to handle responses of patients in realtime and provide accurate response tailored for current user's health condition

Education

Amrita Vishwa Vidyapeetham   Aug 2023 - June 2027

B.Tech. in Artificial Intelligence   *CGPA:8.22/10*

Courses and Certificates

- **Microsoft** Azure AI-900 fundamentals
- Machine Learning with Tensorflow on **Udemy**
- MLOp's and Full stack Deep Learning **LinkedIn**