ANKITH REDDY SUBHANPURAM

Gainesville, Florida, United States, 32608
ankithreddy653@gmail.com • +1 (352) 642-5515
LinkedIn : https://www.linkedin.com/in/ankithreddys/ GitHub : https://github.com/ankithreddys

PROFESSIONAL EXPERIENCE

University of Florida – Biomedical AI Collaborative (IC3) **Gainesville, Florida, United States**

Research Assistant, August 2025 – Present

- Engineered synthetic clinical (EHR) datasets with frontier models from clinically validated data for hospital deployment.
- Implemented an embedding-based deduplication pipeline leveraging FAISS and Jaccard similarity, reducing dataset size by **36%** and improving model accuracy by **4%** compared to similar open-source architectures.
- Fine-tuned open-source LLMs through Axolotl on HiPerGator's 4 NVIDIA B200 GPUs with FSDP2 for efficient training.
- Developed an LLM-as-judge evaluation framework with **14 judges**, validating **110K EHR samples** and assessing clinical reasoning, diagnostic accuracy, and treatment quality for fine-tuned state-of-the-art systems.
- Will orchestrate end-to-end MLOps pipeline with PyTorch, HuggingFace, MLflow, and Docker on NVIDIA DGX systems for scalable model optimization and clinical avatar deployment.

Scholarship Auditions **Nashville, TN, United States**

Artificial Intelligence Engineer Intern, May 2025 – August 2025

- Architected a Retrieval-Augmented Generation (RAG)-based chatbot utilizing ChromaDB as the vector store and OpenAI API as the base language model to enable dynamic, context-aware user interactions.
- Contributed to deep learning (Vision) model development for classifying music transcription images, focusing on domain-specific challenges in music education.
- Deployed automated web scraping pipelines employing scrapy and Selenium to compile structured datasets.

Centre For Artificial Intelligence and Robotics (DRDO) **Bengaluru, Karnataka, India, 560093**

Artificial Intelligence Research Intern, October 2023 – March 2024

- Engineered ASR models based wav2vec2 on DGX GPU, utilizing distributed data parallel in PyTorch for efficient model creation. Incorporated Byte Pair Encoding for optimized vocabulary handling in multilingual settings.
- Enhanced system performance for low-resource languages, Assamese and Bengali, focusing on accuracy and resource efficiency. Applied fine-tuning and hyperparameter tuning to elevate model robustness and scalability.
- Executed advanced data pre-processing techniques, including Voice Activity Detection (VAD) and noise reduction, to improve the quality of input audio data, leading to better ASR performance across various languages and environments.
- Evaluated and refined speech recognition systems across diverse metrics, such as word error rate (WER), achieving an 8.3 WER and boosting model robustness by 20% in low-resource and noisy environments.

SKILLS

Programming: Python, C

Professional Skills: Machine Learning, Artificial Intelligence, Data Science, Data Analytics, DSA, Linux, NLP, LLMs, RAG systems, LLM Finetuning, MLFlow, Docker, Git, Apache Airflow, NVIDIA DGX, Distributed Data Parallel (DDP), Selenium, BeautifulSoup, Scrapy, Web Scraping, Deep Learning, Speech Recognition, Audio Processing, Computer Vision, Video Processing.

Frameworks: PyTorch, Scikit-Learn, TensorFlow, Pandas, NumPy, NLTK, Spacy, Keras**,** Librosa, HuggingFace

EDUCATION

University of Florida **Gainesville, Florida, 32608**

Herbert Wertheim College of Engineering

Masters in Artificial Intelligence Systems, August 2024 – May 2026

Cumulative GPA: 3.83/4.00

Vignana Bharathi Institute of Technology **Aushapur, Telangana, India, 501301**

Bachelor of Technology: Computer Science and Engineering (AI & ML), May 2020 – May 2024

Cumulative GPA: 8.76/10.0

PROJECTS

MSAIS Program Assistant, Independent Project

Designed a conversational AI assistant for UF's MSAIS program implementing a Retrieval-Augmented Generation (RAG) pipeline with LangChain, FAISS, and HuggingFace. Scraped and structured details from the official website with Scrapy, deployed on HuggingFace Spaces with a UF-branded Gradio interface, and powered inference via Llama-3.3-70B through the NaviGator Toolkit API.

Personal AI Assistant, Independent Project

Built an autonomous personal assistant powered by LangGraph agents and OpenAI, capable of executing tasks from natural-language prompts. Implemented automated workflows to draft and send emails, create calendar events, and publish posts on X (Twitter). Embedded action-taking agents for real-time API execution, enabling end-to-end task automation.

Virtual Surya Namaskar Coach

Designed an advanced tracking system with OpenCV to analyze participant movements during online yoga classes; findings helped identify the most common errors made by practitioners, leading to enhanced instructional support materials. **Technologies Used:** Python, OpenCV, TensorFlow.

LEADERSHIP & VOLUNTEER EXPERIENCE

EpsilonPI, ML Club in VBIT, ***ML Associate, July 2023 – May 2024, Alumni*** **VBIT, Hyderabad, Telangana**

Led a team of 4 members in spearheading research and application development focused on harnessing the power of AI and ML for enhanced university benefits.