Balaragavesh G M

balaragavesh | @balaragavesh | balaragavesh | balaragavesh@gmail.com | 9498423751

EXPERIENCE

Edsols Innovations Private Limited *Bengaluru, India*

AI Engineer *Sep 2024 – Present*

- Fine-tuned **OpenAI Whisper** and **Wav2Vec2** on Indian-accented speech data, achieving a **Validation Loss of 0.16** and maintaining a **Word Error Rate (WER) < 20%** for phoneme-level pronunciation analysis.
- Built a **custom tokenizer** and speech processing pipeline to adapt standard Transformer models for detecting **fine-grained pronunciation errors** in a clinical therapy setting.
- Optimized ASR inference by converting the fine-tuned **Whisper** model to **Faster-Whisper (CTranslate2 backend)**, enabling **low-latency, real-time transcription**.
- Developed **backend APIs** using **Python** and **FastAPI** to integrate the ML model with the application and managed therapy session data and speech analytics in **PostgreSQL** for clinical tracking
- Built an efficient **audio streaming pipeline** using **voice activity detection** to stream data to the backend only during active speech, reducing processing overhead and latency
- Designed **scalable and well-structured PostgreSQL schemas** to manage user session data and speech analytics, enabling **efficient querying** and reliable long-term progress tracking.

Graduate Engineer Trainee *July 2023 – Aug 2024*

- Developed a gesture-controlled game using hand coordinates from MediaPipe, enabling intuitive and touch-free interaction.
- Mapped **real-time hand movements** to game controls using pyautogui and also integrated a Wii Nunchuk, translating inputs into keyboard commands.
- Deployed the system on NVIDIA's Jetson Xavier NX, utilizing **CUDA-accelerated** OpenCV and running MediaPipe models on the GPU for improved performance, **achieving 20 FPS**.

PROJECTS

Whisper finetuning

- Fine-tuned OpenAI's Whisper model using Hugging Face Transformers to predict phoneme sequences from raw audio input.

W2vIndia

- Pre-trained a self-supervised Wav2Vec2 Base model from scratch on multilingual Indian speech data.

Hybrid Search RAG Assistant

- Built an RAG assistant with Phi-2, ChromaDB, and hybrid search to generate context-aware responses from documentation.

Hand gesture recognition

- Developed a real-time hand gesture recognition system using MediaPipe and OpenCV with customizable gestures.

EDUCATION

Bachelor of Technology in Information Technology *Coimbatore, India*

Sri Ramakrishna Engineering College

- CGPA: 8.39/10

SKILLS

Programming Languages: Python, JavaScript, TypeScript, Java

Machine Learning & AI: Deep Learning, NLP, Computer Vision, Audio Processing, Speech Recognition (ASR), Model Fine-tuning

Frameworks & Libraries: React, FastAPI, Uvicorn, PyTorch, TensorFlow, Transformers, OpenCV, MediaPipe, YOLOv5, Scikit-Learn, Whisper, Wav2Vec2, CUDA, Electron, TailwindCSS

Databases: PostgreSQL, SQLite

Tools & Platforms: E2E Cloud (GPU Infrastructure), Docker, AWS (EC2, RDS), Linux, NVIDIA Jetson, Git

CERTIFICATIONS

- Artificial Intelligence Foundation, Nasscom
- Building Video AI Applications at the Edge on Jetson Nano, NVIDIA
- Getting Started with AI on Jetson Nano, NVIDIA
- Statistical Thinking for Data Science and Analytics, edX
- Machine Learning for Data Science and Analytics, edX