# Emotional Detection Speech Recognition System

## 1. Introduction

### What is Speech Recognition

Speech recognition is a technology that enables machines to interpret and process human speech into written text. It serves as a bridge between humans and machines, allowing for hands-free control, voice commands, and real-time transcription.

### What is an LLM (Large Language Model)

A Large Language Model (LLM) is a deep learning model trained on vast amounts of textual data to understand, generate, and manipulate human language. These models, such as GPT and BERT, can perform various tasks including translation, summarization, and question-answering.

### How are LLMs used in Speech Recognition

LLMs enhance the post-processing stage of speech recognition by refining the transcribed text, correcting grammar, understanding context, and even generating structured responses. They also support language modeling components in end-to-end speech-to-text systems.

### Speech Recognition Uses in Current Trends

- Virtual assistants (e.g., Alexa, Siri)  
- Real-time transcription (e.g., Zoom, Google Meet)  
- Voice-enabled search  
- Customer support automation  
- Accessibility tools for the hearing impaired

### Speech Recognition Future

- Context-aware recognition systems  
- Multilingual and code-switching support  
- Seamless integration in AR/VR  
- Greater accessibility in education and healthcare

## 2. Problem Statement

### What the Project Does

This project builds an end-to-end speech recognition system powered by machine learning and enhanced using an LLM. It includes a user-friendly interface via Streamlit.

### Objective

To build an efficient and accurate speech-to-text model pipeline and deploy it with a simple dashboard that also explains internal working steps such as preprocessing, model training, and evaluation.

## 3. Existing System Limitations

- Low accuracy in noisy environments  
- Poor context understanding  
- Limited support for multiple languages  
- Lack of real-time feedback and user-friendly dashboards  
- Dependency on rule-based models or shallow learning

## 4. Proposed System

- Integration of LLMs for better context handling and grammar correction  
- Use of a robust training dataset and preprocessing pipeline  
- Streamlit-based dashboard for user interaction and visualization  
- Real-time testing and evaluation capabilities

## 5. Pipeline of the System

### Module Flow Chart

Audio Input → Data Collection → Preprocessing → EDA → Model Training → Testing → Evaluation → Streamlit UI

### Data Collection

Dataset used: [Mention dataset name here, e.g., LibriSpeech, Common Voice]

### Preprocessing

Data cleaning, noise reduction, and feature extraction using MFCCs or spectrograms.

### Exploratory Data Analysis (EDA)

Analyzing audio length, speaker distribution, and label consistency.

### Training

Model training using deep learning techniques with LLM post-processing.

### Testing

Model evaluation on unseen test data for accuracy and word error rate.

### Implementation

Deployment using Streamlit for real-time interaction.

### Loss Evaluation

Cross-entropy or CTC loss metrics tracked during training.

### Coding

Python code for data pipeline, training, evaluation, and UI included.

## 6. Dashboard / User Interface (Streamlit)

Streamlit app provides an intuitive interface for uploading audio, running the speech recognition model, and viewing results interactively.

## 7. Libraries and Architecture

- Python, NumPy, Pandas, Matplotlib, Librosa, TensorFlow/PyTorch, Transformers, Streamlit  
- LLMs (like GPT or BERT) used for contextual text generation and enhancement.  
- Modular architecture with separate stages for preprocessing, modeling, and UI.

## 8. Project Limitations

- Limited to English language (extendable)  
- Dependent on quality of input audio  
- Processing time increases with longer audio files

## 9. Future Enhancement

- Support for multiple languages and dialects  
- Noise filtering using advanced models  
- Real-time voice assistant integration  
- Improved UI with analytics dashboard

## 10. Conclusion

This project presents a modern speech recognition system combining traditional deep learning with large language models for enhanced accuracy and user experience. The end-to-end system is accessible via a Streamlit dashboard, demonstrating the potential of AI in human-computer interaction.