

# VIBE VOICE

## PROBLEM STATEMENT

---

Speech Emotion Recognition (SER) is a challenging task in human-computer interaction, aiming to accurately identify emotions from speech signals. Developing a robust and accurate system that automatically detects a speaker's emotional state using deep learning techniques is crucial for enhancing user experience and interaction. However, detecting an emotion is only part of the solution; the real challenge lies in translating that detection into meaningful and personalized action recommendations. This capability enables various applications where understanding the emotional context of communication is essential, improving interactions in fields such as virtual assistants, mental health monitoring, and customer service.

## PROJECT OVERVIEW

---

The project focuses on recognizing human emotions through speech signals and providing action-based suggestions to enhance user well-being. It utilizes machine learning and deep learning techniques to classify emotions such as happiness, sadness, anger, and neutrality. The system processes recorded speech, extracts relevant features, and predicts the user's emotional state. Based on the detected emotion, the system provides personalized action recommendations, such as listening to music, engaging in relaxation exercises, or performing motivational tasks. By integrating advanced artificial intelligence techniques, the goal is to create an AI-powered emotional intelligence system that improves human-computer interaction and supports mental well-being.

## SOLUTION OFFERED

---

The solution consists of three main components: the speech processing module, the emotion detection module, and the recommendation module. The speech processing module captures and processes the user's voice input, extracting key features such as pitch, tone, and intensity. The emotion detection module utilizes deep learning techniques to classify emotions such as happiness, sadness, anger, and neutrality based on the extracted features. Finally, the recommendation module analyzes the detected emotion and provides personalized action suggestions to improve the user's well-being.

## WHO ARE THE END USERS?

---

- Individuals Seeking Emotional Well-being (Regular Users)
- Mental Health Professionals (Therapists & Counselors)
- Smart Assistants & Chatbots (Alexa or Siri)
- Healthcare Applications (Doctors & Healthcare Workers)
- Call Centers & Telecommunication Services (Customer Support Teams)

## TECHNOLOGY USED TO SOLVE THE PROBLEM

---

### 1) **Speech Processing :**

- Python: Used for audio processing and deep learning model development.
- Librosa: A library for feature extraction from speech signals, such as MFCCs and spectral features.
- Pydub: Used for handling various audio formats and pre-processing speech data.

### 2) **Deep Learning :**

- TensorFlow/Keras & PyTorch: Used to train deep learning models for emotion classification.
- CNNs : Implemented for better feature extraction and sequential analysis of speech signals.
- RAVDESS Dataset: Used for training and testing the speech emotion recognition model.

### 3) **Recommendation System :**

- FFmpeg: Used for audio processing, conversion, and enhancement before emotion detection.
- Inbuilt Emotion Mapping Code: Custom-developed logic for mapping detected emotions to specific well-being recommendations.

### 4) **Web Development :**

- HTML, CSS, JavaScript: Used to create an interactive one-page website for real-time user interaction.
- Flask: Backend framework to integrate the deep learning model with the web interface.