**Task Description:**

Classify speech segments into one of eight emotional categories: **anger**, **happiness**, **sadness**, **fear**, **surprise**, **contempt**, **disgust**, and **neutral state**. The classification must prioritize **balanced performance across all classes**, ensuring that both majority and minority emotional categories are represented effectively.

**Dataset Description:**

The dataset used for this task is the **MSP-Podcast corpus**, a large naturalistic speech emotion dataset. It features diverse speakers and realistic scenarios, offering rich contextual variability. Key dataset details:

* **Clip Length:** Each audio clip ranges from **2 to 12 seconds**.
* **Labeling Process:** At least **five independent raters** annotated each clip with one of eight categorical emotional labels.

**Audio Attributes: (task2에서 acoustic features를 나중에 추가해보자)**

In addition to categorical emotional labels, each audio clip is annotated with three continuous emotional attributes:

1. **Valence (V):** Measures the positivity or negativity of the emotion.
2. **Arousal (A):** Reflects the intensity or energy level of the emotion.
3. **Dominance (D):** Indicates the degree of control or submission expressed.

The provided (V, A, D) scores for the audio clip are as follows: **( , , )**.

**Output Format:**

The model must produce the following output for each audio clip:

**Emotion Probabilities:**  
Output a vector of probabilities for the eight emotional classes, e.g.,  
[anger: 0.15, happiness: 0.40, sadness: 0.10, fear: 0.05, surprise: 0.10, contempt: 0.05, disgust: 0.05, neutral: 0.10]

**Primary Emotion:**  
Class with the highest probability: e.g., **happiness (0.40)**

**Secondary Emotion:**  
Class with the second-highest probability above a threshold (e.g., 0.10): e.g., **neutral (0.10)**

**Evaluation Metric:**

The performance of the model will be evaluated using the **F1-macro score**, which emphasizes **equal weighting across all emotional classes**, regardless of class frequency. The model must effectively handle class imbalance and perform well on both majority and minority emotional categories.

**Objective:**

Adapt the LLM to the speech-based emotion recognition task by leveraging the nuanced interplay between:

1. **Categorical emotional labels (e.g., anger, sadness, happiness)**
2. **Continuous emotional attributes (valence, arousal, dominance)**

The system should focus on capturing complex emotional dynamics in speech to improve generalization across all emotional categories, ultimately maximizing the **F1-macro score**.