Multi-Label Emotion Classification System Report

# 1. Objective

Develop a system to classify multiple emotions (e.g., joy, sadness, anger) present in textual data using the GoEmotions Dataset by Google. The system should preprocess data, fine-tune a transformer model, evaluate the model, and test it on real-world textual inputs such as customer feedback or social media posts.

# 2. Preprocessing and Handling Imbalanced Data

- Loaded the GoEmotions dataset using Hugging Face `datasets` library.  
- Preprocessed text: lowercasing and tokenization using BERT tokenizer.  
- Converted emotion labels into multi-hot vectors for multi-label classification.  
- Visualized class imbalance using a bar plot.  
- Techniques to address imbalance: BCEWithLogitsLoss, class weighting, and potential oversampling.

# 3. Model Fine-Tuning

- Used Hugging Face Transformers to fine-tune `bert-base-uncased` model.  
- Modified the model head for multi-label output using sigmoid activation.  
- Tokenized inputs and padded/truncated sequences for training.  
- Used `Trainer` API for training with appropriate training arguments.

# 4. Evaluation Metrics

- Used multi-label evaluation metrics:  
 - Hamming Loss: Measures fraction of incorrect labels.  
 - Micro/Macro F1 Score: Evaluates precision and recall for multi-label classification.

# 5. Inference on Real-World Text

- Implemented inference pipeline using Gradio interface for testing on any input text.  
- Applied sigmoid on model logits and classified emotions with a threshold of 0.5.  
- Predicted labels are mapped back to emotion names.

# 6. Outcome

A fine-tuned BERT model capable of identifying multiple emotions from text. The system is useful for analyzing user feedback, social media posts, and any text data involving emotional expression.