**Emotion Recognition using Fine-tuned LLM Report**

**Objective and Model Selection**

The objective of the project is to develop an open-source large language model-based system for text emotion recognition. Emotion Recognition finds valuable applications in chatbots, sentiment analyses, and mental health monitoring. GoEmotions was chosen because it has a complete labeling of 27 emotion categories and one neutral label, hence strong for multi-label emotion classification. DistilBERT was chosen to perform fine-tuning because it strikes a good balance between computational efficiency and performance, hence befitting the available resources.

**Methodology**

Fine-tuning was performed in the following order:

1. Data preparation: This is how the GoEmotions dataset was loaded and then preprocessed using the Hugging Face tokenizer to tokenize the text and map it into numerical inputs suitable for the model.

2. Model training: Fine-tuning of DistilBERT on the training set is done using the Hugging Face Trainer API. The hyperparameters are a learning rate of 2e-5, batch size 16, and training over three epochs.

3. Evaluation: The model was then evaluated on the validation set in terms of validation loss and accuracy.

Code snippets for pre-processing, fine-tuning, and evaluation were provided. Example outputs were created to verify if the model classified emotions correctly.

**Evaluation Results**

The fine-tuned model results are as follows:

Validation loss: 1.686

Accuracy: 52.4%

This model performed reasonably but with moderate accuracy, which can prove that multi-label emotion classification is challenging. Fine-grained distinctions, for example, between disappointment and sadness, exist in GoEmotions; hence, the model cannot learn all categories as per expectations. Since DistilBERT was used, computational efficiency lowered the capacity of the model compared to even larger transformer models that may be able to perform better in capturing more.

**Challenges Faced**

The complexity of the dataset required more specialized loss functions and careful evaluation to handle overlapping emotions effectively. Computational limitations influenced the choice of a lightweight model, DistilBERT, and reduced the possibility to experiment with larger models or extensive hyperparameter tuning. Limitations in metrics also existed because accuracy alone cannot fully capture the effectiveness of multi-label classification. Evaluating metrics such as F1-score, precision, and recall would give a more detailed understanding of the performance.

**Conclusion**

This project was able to apply the fine-tuned DistilBERT model to the GoEmotions dataset for text emotion recognition. Though much challenge arose related to the complexities of the dataset and computational limitations, this model presents a very good start for emotion classification tasks. This work will be continued in trying larger models with further performance optimizations to improve system accuracy for use in real-life applications.