📝 Multi-Label Emotion Recognition Report

📂 1. Dataset Preprocessing Steps

- **Dataset**: Utilized the **GoEmotions** dataset from Hugging Face, containing over 58k English Reddit comments labeled across **27 fine-grained emotions** + 1 neutral category.

- **Label Encoding**: Used `MultiLabelBinarizer` to one-hot encode the labels, since each input could be associated with multiple emotions.

- **Text Tokenization**: Leveraged `BertTokenizerFast` from the `transformers` library for tokenizing text with `max_length=128`, truncation, and padding.

- **Dataset Transformation**:
    - Applied `map()` function on the dataset to tokenize and attach multi-hot labels.
    - Converted processed datasets into PyTorch format for model compatibility.

🤖 2. Model Selection and Rationale

Model	Rationale
BERT (bert-base-uncased)	A transformer-based model pretrained on large English corpora; capable of capturing context, emotion, and semantics in short texts.
Custom Trainer	Built upon Hugging Face's `Trainer`, with modified loss function using **Binary Cross-Entropy (BCEWithLogitsLoss)** suitable for multi-label classification.

This combination provides strong language understanding along with multi-label learning capability.

⚠️ 3. Challenges Faced and Solutions

Challenge	Solution
Multi-label complexity	Standard classifiers fail on multi-label output. Implemented custom loss and used sigmoid activation for independent label prediction.
Imbalanced emotion labels	While not explicitly rebalanced, using BCE loss with logits and appropriate thresholding helps manage label sparsity.
Large model resource demands	Utilized Hugging Face Trainer API with optimized training arguments to manage compute efficiency.

📊 4. Results and Visualizations

Classification Report

Epoch Training Loss Validation Loss

[170/170 32:33]

	precision	recall	f1-score	support
admiration	0.09	0.97	0.17	504
amusement	0.05	0.99	0.09	264
anger	0.33	0.01	0.01	198
annoyance	0.06	0.97	0.11	320
approval	0.06	0.36	0.11	351
caring	0.00	0.00	0.00	135
confusion	0.03	0.99	0.05	153
curiosity	0.05	1.00	0.10	284
desire	0.00	0.00	0.00	83
disappointment	0.04	0.11	0.06	151
disapproval	0.05	0.10	0.07	267
disgust	0.02	0.91	0.04	123
embarrassment	0.01	0.97	0.01	37
excitement	0.02	0.68	0.04	103
fear	0.01	0.86	0.03	78
gratitude	0.07	1.00	0.12	352
grief	0.00	0.83	0.00	6
joy	0.02	0.29	0.04	161
love	0.00	0.00	0.00	238
nervousness	0.00	0.00	0.00	23
optimism	0.03	1.00	0.07	186
pride	0.02	0.06	0.03	16
realization	0.02	0.13	0.04	145
relief	0.00	0.00	0.00	11
remorse	0.02	0.02	0.02	56
sadness	0.04	0.13	0.06	156
surprise	0.03	0.28	0.05	141
neutral	0.33	0.99	0.49	1787
micro avg	0.06	0.70	0.11	6329
macro avg	0.05	0.49	0.07	6329
weighted avg	0.13	0.70	0.19	6329
samples avg	0.06	0.71	0.11	6329

📌 Summary

- **Model Performance**: BERT-based classifier trained with BCE loss effectively handles multi-label emotion recognition.
- **Best Aspects**: Strong context understanding from BERT, well-structured preprocessing, and multi-label treatment.
- **Next Steps**:
  - Implement threshold tuning for each label.
  - Try ensemble models or other transformer variants (RoBERTa, DeBERTa).
  - Add class weighting or focal loss for rare emotion categories.