Model Training Summary

🤠 Model Architecture

Base Model: BERT-base-uncased

- Pre-trained transformer model from Google
- 110M parameters
- 768 hidden size, 12 attention heads

Custom Wrapper: FlexibleBertModel

- Extends BertForSequenceClassification
- Automatically adapts to single or multi-intent tasks
- **Single-intent:** Uses softmax + CrossEntropyLoss
- **Multi-intent:** Uses sigmoid + BCEWithLogitsLoss

⚙ Training Configuration

Hyperparameters:

- **Epochs:** 3
- **Batch Size:** 8 (train and eval)
- **Learning Rate:** 2e-5 (default for BERT)
- **Max Sequence Length:** 512 tokens
- **Weight Decay:** Not explicitly set (uses default)

Data Split:

- **Training:** 80% (8,000 examples)
- **Testing:** 20% (2,000 examples)

📊 Training Process

1. Automatic Detection:

- Analyzes dataset to detect single vs. multi-intent
- Configures model and loss function accordingly

2. Multi-label Optimization:

- Tests thresholds [0.3, 0.4, 0.5, 0.6, 0.7]
- Selects best threshold based on F1 score
- Optimal threshold found: 0.3

3. Evaluation Metrics:

- **Single-intent:** Accuracy, F1 score
- **Multi-intent:** F1 (weighted/micro), Precision, Recall, Hamming Loss, Exact Match Ratio

🎯 Results

Performance:

- **F1 Score:** 1.0000 (perfect)
- **Precision:** 1.0000
- **Recall:** 1.0000
- **Exact Match Ratio:** 1.0000

Model Capabilities:

- Handles 1-3+ intents per email
- 8 total intent categories
- ~1 hour training time on GPU

💾 Output Artifacts

Saved Files:

- Model weights: ./results/checkpoint-3000/
- Tokenizer config: ./results/
- Metadata: ./results/metadata.pkl
- Training summary: ./results/training_summary.json