Script Description: train_goemotions_distilbert_colab_v2.py

This script trains a multi-label emotion classifier on the GoEmotions dataset using the DistilBERT model. It is designed to be Colab-compatible and supports multiple schema formats and output diagnostics. It can be extended to use alternative transformer models with minimal changes.

Key Features

• Supports both 'raw' (27 emotions + neutral) and 'simplified' (6 + neutral) GoEmotions configurations.

• Automatically creates a validation split if missing.

• Handles both list-based and wide one-hot label schemas.

• Uses Hugging Face AutoTokenizer and Trainer.

• Computes micro, macro, and weighted F1 scores.

• Exports metrics.json, classification_report.json, and efficiency snapshot.

• Measures trainable parameters and batch-wise latency.

Default Parameters

Parameter	Default	Description
--model_name	distilbert-base-uncased	Hugging Face model ID
--dataset_config	raw	Label schema: raw or simplified
--max_length	128	Maximum input token length
--batch_size	32	Per-device training batch size
--epochs	3	Number of training epochs
--eval_threshold	0.5	Sigmoid threshold for prediction

Adapting to Other Models

To use another transformer model (e.g., ALBERT, ModernBERT), simply set the --model_name argument to the desired Hugging Face model ID. The script will handle the rest automatically.