1. **Pipeline.ipynb**

 Pipeline函數是Transformer用來呼叫大型語言模型(LLM)最簡單的方式

 指定『任務』，pipeline會自動幫你搭配最適合的LLM，自動幫你完成任務

 如沒有指定，Pipeline會自動挑選適合的LLM

1. **Transformers.ipynb**

 LLM進行訓練是很複雜且困難，必須特定的模型、斷詞工具(Tokenizer)

 Transformer除了提供各種LLM與斷詞工具，還協助使用者自動挑選，例如AutoModel、AutoTokenizer

 提供簡化訓練過程，例如Trainer、TrainingArguments

1. **Finetune\_Chinese\_Weibo.ipynb**

 利用中文微薄的評價資料，微調一個可以針對評價判斷是正面/負面的模型

 以Bert為訓練模型(1.1億參數，夠小的模型才有辦法微調訓練)

 Weibo資料有119988筆評價資料(為簡化訓練只採用20000筆)

 使用wandb紀錄訓練過程

PEFT效率微調:

LLM已經大到無法進行微調

GPU記憶體不足(Nvidia 4090 GPU: 24G ram)

PEFT(Parameter-Efficient Fine Tuning)

* 透過降低權重精準度(32bit->4bit)
* 只微調外加Adapter層
* Lora、Qlora、unsloth

 訓練時只改動adapter，沒有動到基礎模型

 使用時，只需讀入adapter+基礎模型

AdapterHub

儲存PEFT訓練好的Adapter

下載Adapter+基礎模型，就可以進行PEFT模型使用(Inference)

1. **Peft\_with\_Adapter.ipynb :** 利用PEFT進行遊戲內容推論
2. **Classic\_Chinese\_Inference\_with\_Adapter.ipynb :**利用PEFT進行文言文轉換推論