1. **Transformer 採用什麼損失函數？是如何運作使模型能夠學習到上下文關係和文法規則？**

Transformer 模型採用cross-entropy來進行訓練，使模型學習預測下一個正確的單詞。Transformer核心機制是自注意力（self-attention），使模型能夠考慮序列中的所有位置，學習到上下文關係和語法規則。

1. **Attention 機制如何改善 RNN 與 CNN 的缺點？具體如何改善？**

Attention 機制透過權重分配，使模型能夠專注於當前任務中。

在RNN 中，幫助模型捕捉長距離依賴，解決梯度消失問題；在 CNN 中，增強模型對局部特徵的理解，透過統合卷積層來提高模型對非局部上下文的感知能力。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述