

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。

2. 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。

```
• (base) guanweiliiao@cluster-ai:/data/guanweiliiao/113.08.12/transformer_class1/src$ conda activate transformer
• (transformer) guanweiliiao@cluster-ai:/data/guanweiliiao/113.08.12/transformer_class1/src$ python model.py
Initialized model with random weights.

Final output size: torch.Size([2, 5, 10])
恭喜你的模型通過測試！
```