

---

# How to train?

台大電機計算機訓練班

Teacher: Prof. Hung-yi Lee

TA: Kuang Ming Chen

---

哪個步驟最耗費人力?



# 課程大綱

1. 測試
2. 準備資料
3. 處理資料
4. 選擇模型
5. 怎麼判斷模型好壞
6. DEMO

有問題的話可以在上面留言！！



Slido 問答

密碼: #6186721



即時提問  
#6186721

# 測試



即時提問  
#6186721

# 測試

- 又稱為inference
- 在大型語言模型中，我們可以用不同的generate策略來測試並輸出model的output
  - max\_length
  - num\_beams
  - temperature
  - Top-k
  - Top-p

指令：跟我講解總結記憶這項技術

回答：

隨機存儲是一種技術，其中數據被存儲在一個或多個位元組（位元）上，這些位元被稱為“行”或“列”。這個技術在計算機和其他計算設備中非常常見。它允許快速存儲和檢索數據。在某些情況下！它也被用於創建密碼學和數據加密。隨機數據通常以二進制形式存儲，這意味著它可以用於存儲任何兩個數字。這使得它非常適合存儲大量數據的情況。它也非常有用於在數據庫中存儲數據；這些數據可以是文本、數字、圖像、音頻和視頻數據等等。此技術的一個應用是創建安全的數據存儲解決方案。這可以包括使用隨機位元存儲密碼或使用



即時提問  
#6186721

## max\_length

- 顧名思義，就是輸出的長度
  - 生成tokens的最大長度
  - 通常一個中文字會是1到3個tokens
- 當你遇到LLM輸出有被截斷的現象，有可能就是輸出的長度不夠！

指令：有機電致發光材料及裝置

回覆：

本研究所提供的有機發光二極體及其相關裝置，是一種使用有機材料（通常為硅）進行光電效應的光電元件及其應用。該裝置具有以下特點：

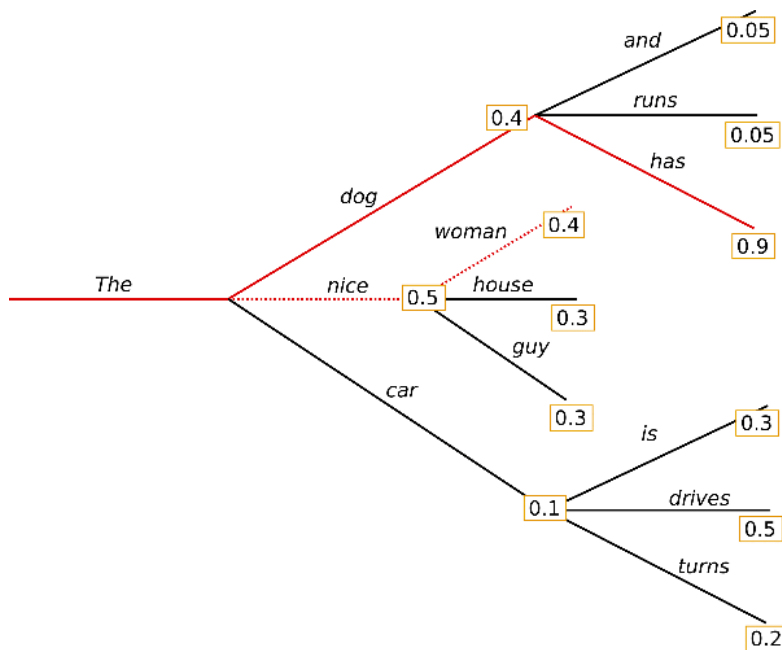
1. 光電導性；
2. 具有高光電容率與低光電阻率；
3. 能夠產生高頻率的光，並且具有較低的光損失率。
4. 可以應用於各種應用領域、如顯示器、顯示屏、通訊裝置、照明裝置等。
5. 對環境的影響較小。
6. 適用於低光環境下的應用，如夜間環境、低光照明環境等。

研究結果顯示：使用這種裝置可以有效地提高光學顯示技術的效能（如電視、電



# num\_beams

- 生成文本時要保留的擴展搜尋分支的數量
  - 較高的num\_beams會導致模型保留更多的生成序列，生成的文本可能會更加多樣，但也會增加計算量

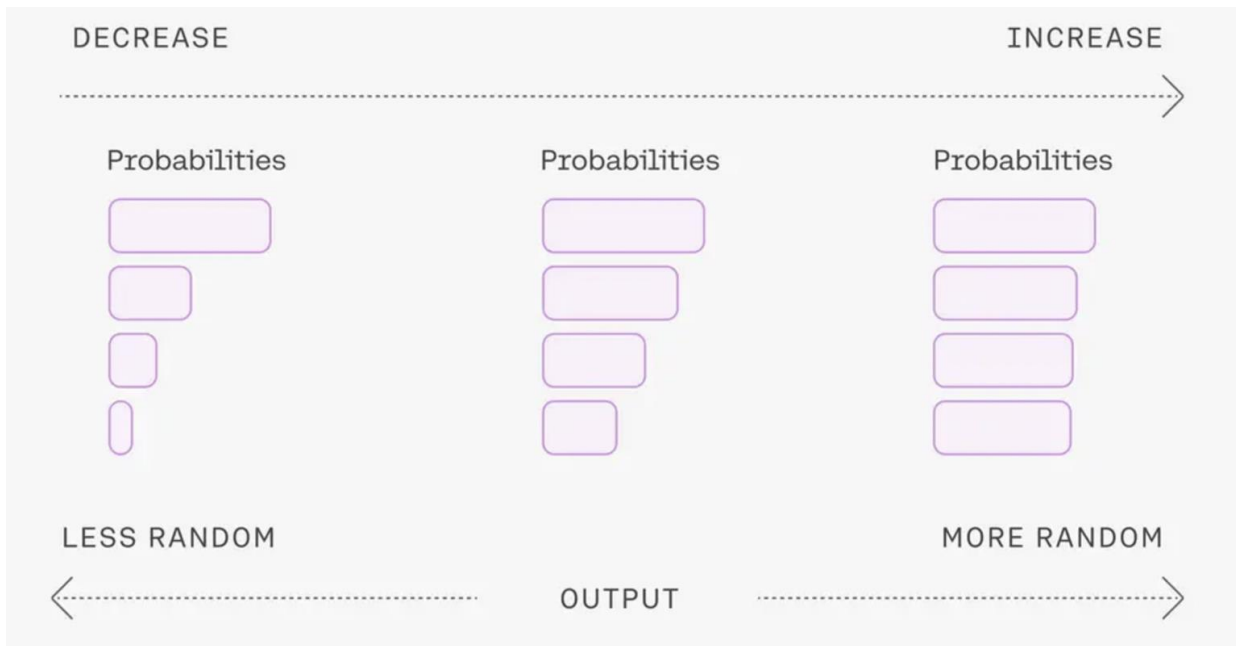






# temperature

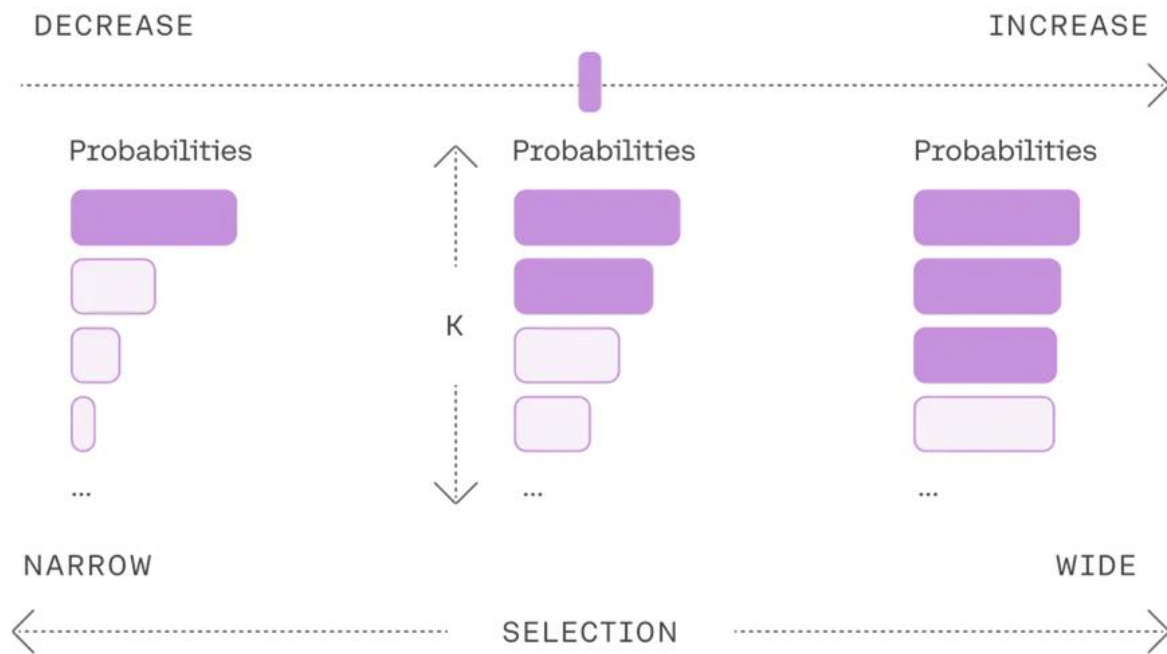
- temperature設定較低時，模型會更加保守
- temperature設定較高時，模型會更多樣性





即時提問  
#6186721

# Top-K





# Top-P

1. Consider only the top tokens whose likelihoods add up to 15%. Ignore all others.

2. Sample from them based on their likelihood scores.

United + Netherlands  $\approx 15\%$





即時提問  
#6186721

## 準備資料



即時提問  
#6186721

## 準備資料

- 中華民國專利資料資料



專利檢索

操作說明



檢索設定

不限欄位

公開/公告日



yyyymmdd



~

yyyymmdd



檢 索



清 空

欄 位 代 碼



即時提問  
#6186721

## 處理資料



即時提問

#6186721

## 處理資料

- Instruction format

```
[  
  {  
    "instruction": "this is a description of patent, give a title of it",  
    "input": "本invention設計磷光錯合物，其具有分子內H鍵性質以防止鄰近分子去質子化。 ",  
    "output": "有機電致發光材料及裝置"  
  },  
  {  
    "instruction": "this is a description of patent, give a title of it",  
    "input": "一種具有集成領襯的鞋類物件。鞋類物件具有鞋面，鞋面包括上表面以及下表面，上表面形成鞋類  
    "output": "鞋類物件以及用集成領襯製造鞋類物件的方法"  
  }  
]
```



即時提問  
#6186721

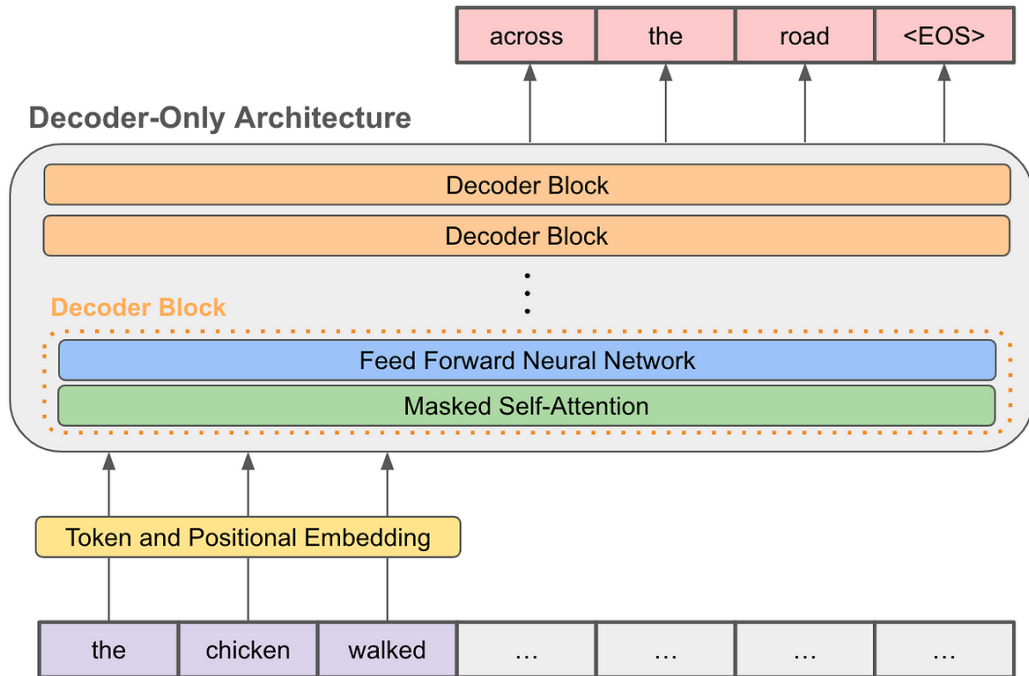
# 選擇模型





# 選擇模型

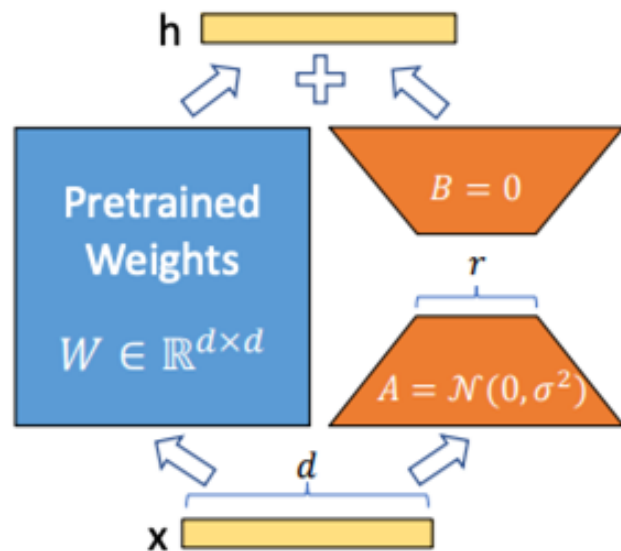
- Taide 1.3B
  - OPT模型架構





# 選擇模型

- LoRA
  - Low-Rank Adaptation
- 好處
  - 用到較少記憶體 ( 可以在24G ram run )
  - 訓練時間也會減少 ! (只需要15分鐘以內)
- 如果不使用 LoRA 的話，在 colab 會遇到 out of memory



$$\begin{aligned} 100 * 100 &= 10000 \\ 100 * 20 + 20 * 100 &= 4000 \end{aligned}$$



即時提問

#6186721

# 怎麼判斷模型好壞



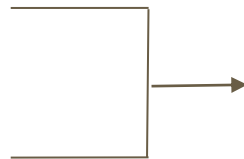
即時提問  
#6186721

# 怎麼判斷模型好壞

input : 請問台灣大學的地址是？

output : 台北市大安區羅斯福路一段

groundtruth : 台北市大安區羅斯福路四段1號



與正確答案存在誤差，在模型訓練時稱為 loss



目標就是要讓模型的 output 結果更接近正確答案



即時提問  
#6186721

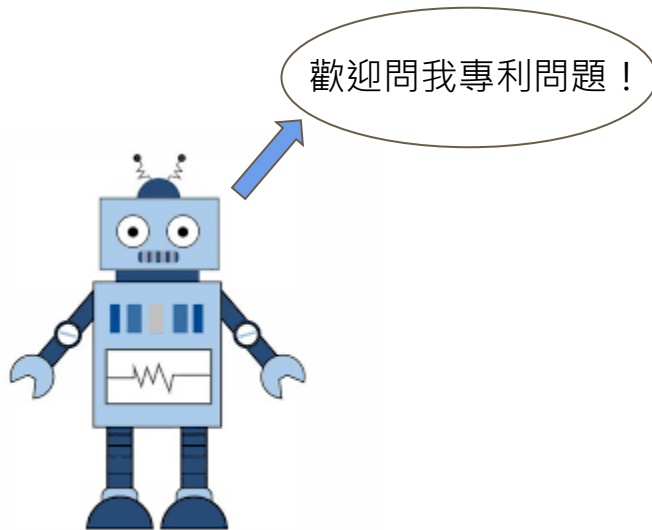
# Demo



即時提問  
#6186721

# 目標

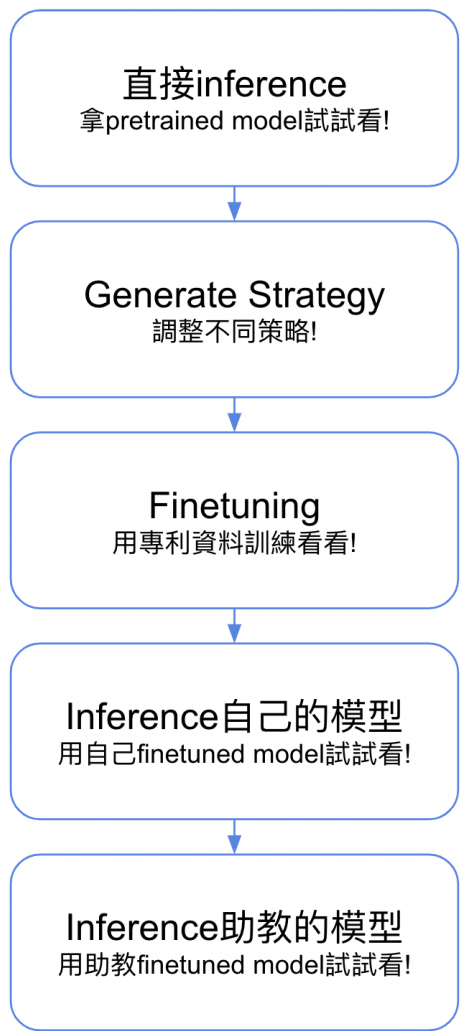
- 讓 Taide 1.3B 能夠了解專利資料，並透過測試來問他專利的說明！



# Colab pipeline



即時提問  
#6186721



## Other Ref.

<https://github.com/ntunlp/plab/traditional-chinese-alpaca>

<https://twpat.tipo.gov.tw/>

<https://arxiv.org/abs/2106.09685>

<https://www.nstc.gov.tw/folksonomy/detail/f094b57a-204e-4114-99d3-a9412ae42d7a?l=ch>