

一堂課搞懂生成式人工智能 的技術突破與未來發展

李宏毅

Hung-yi Lee

下一次上課前 (3/07) 預習以下課程

生成式AI導論 2024



【生成式AI導論 2024】第0講：課程說明 (17:15 有茉莉蓮雷)

Hung-yi Lee · 觀看次數：27萬次 · 11 個月前



【生成式AI導論 2024】第1講：生成式AI是什麼？

Hung-yi Lee · 觀看次數：25萬次 · 11 個月前



【生成式AI導論 2024】第2講：今日的生成式人工智能厲害在哪裡？從「工具」變為「工具人」

Hung-yi Lee · 觀看次數：13萬次 · 11 個月前



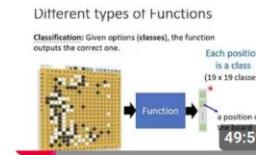
【生成式AI導論 2024】第3講：訓練不了人工智能？你可以訓練你自己(上) – 神奇咒語與提供更多資訊

Hung-yi Lee · 觀看次數：11萬次 · 11 個月前

https://www.youtube.com/playlist?list=PLJV_el3uVTsPz6CTopeRp2L2t4aL_Kgil

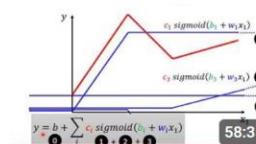
(至少看到第八講，能看完最好)

機器學習 2021



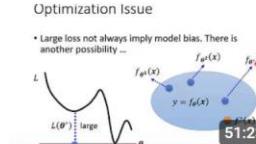
【機器學習2021】預測本頻道觀看人數 (上) - 機器學習基本概念簡介

Hung-yi Lee · 觀看次數：53萬次 · 3 年前



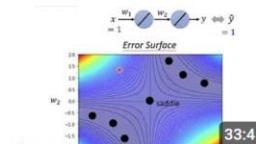
【機器學習2021】預測本頻道觀看人數 (下) - 深度學習基本概念簡介

Hung-yi Lee · 觀看次數：24萬次 · 3 年前



【機器學習2021】機器學習任務攻略

Hung-yi Lee · 觀看次數：19萬次 · 3 年前



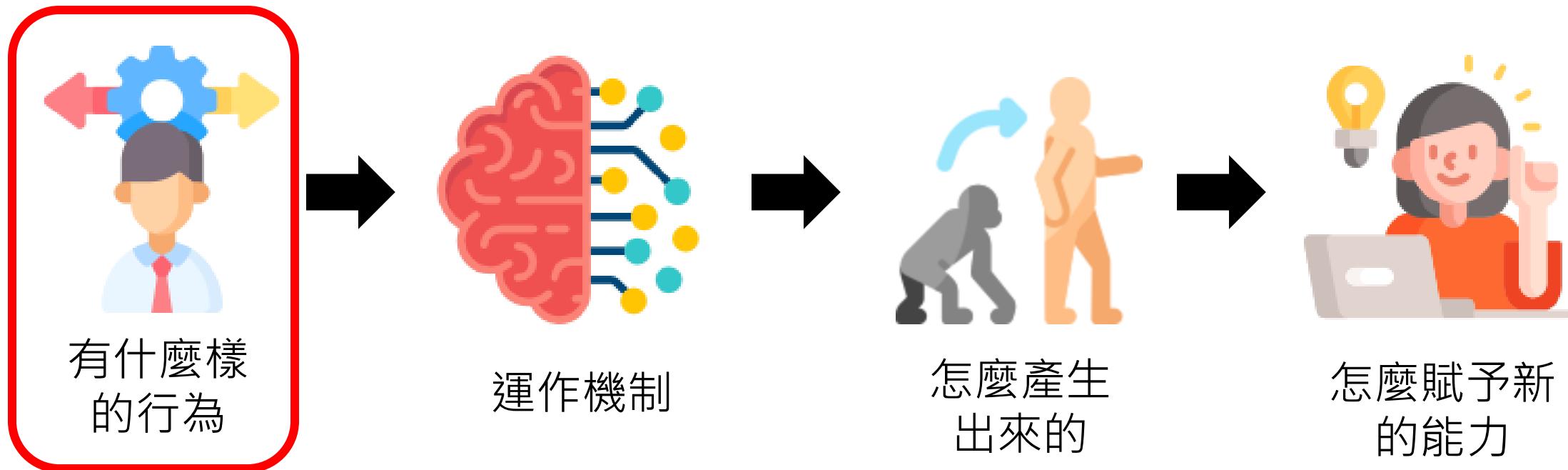
【機器學習2021】類神經網路訓練不起來怎麼辦 (一)：局部最小值 (local minima) 與鞍點 (saddle point)

Hung-yi Lee · 觀看次數：13萬次 · 3 年前

https://www.youtube.com/playlist?list=PLJV_el3uVTsMhtt7_Y6sgTHGHp1Vb2P2J

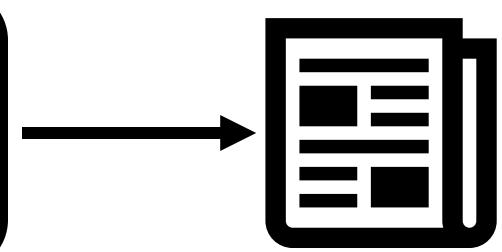
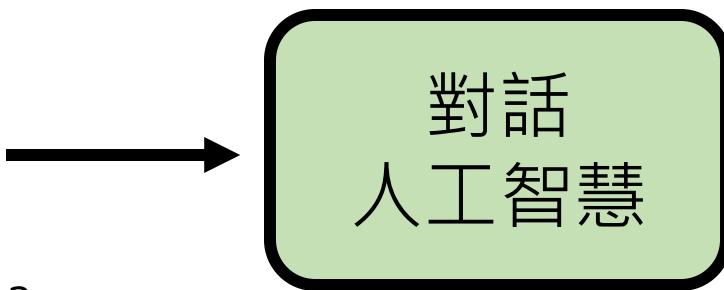
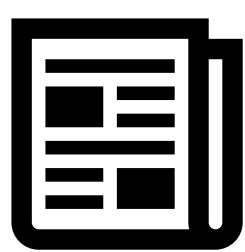
(至少看到「Transformer (下)」，能看完最好)

課程大綱





什麼是機器學習？



機器學習是.....

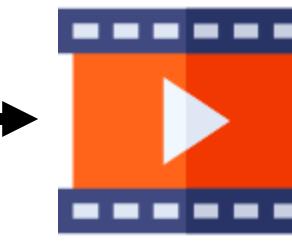
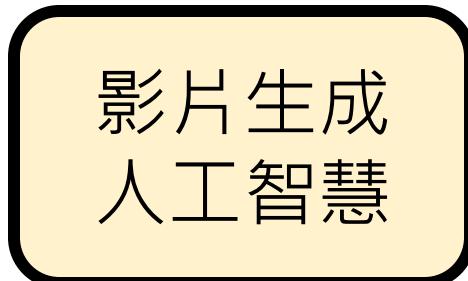
ChatGPT,
Gemini,
Claude ...

我的狗



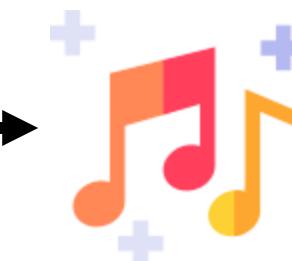
Midjourney,
DALL-E ...

我的狗



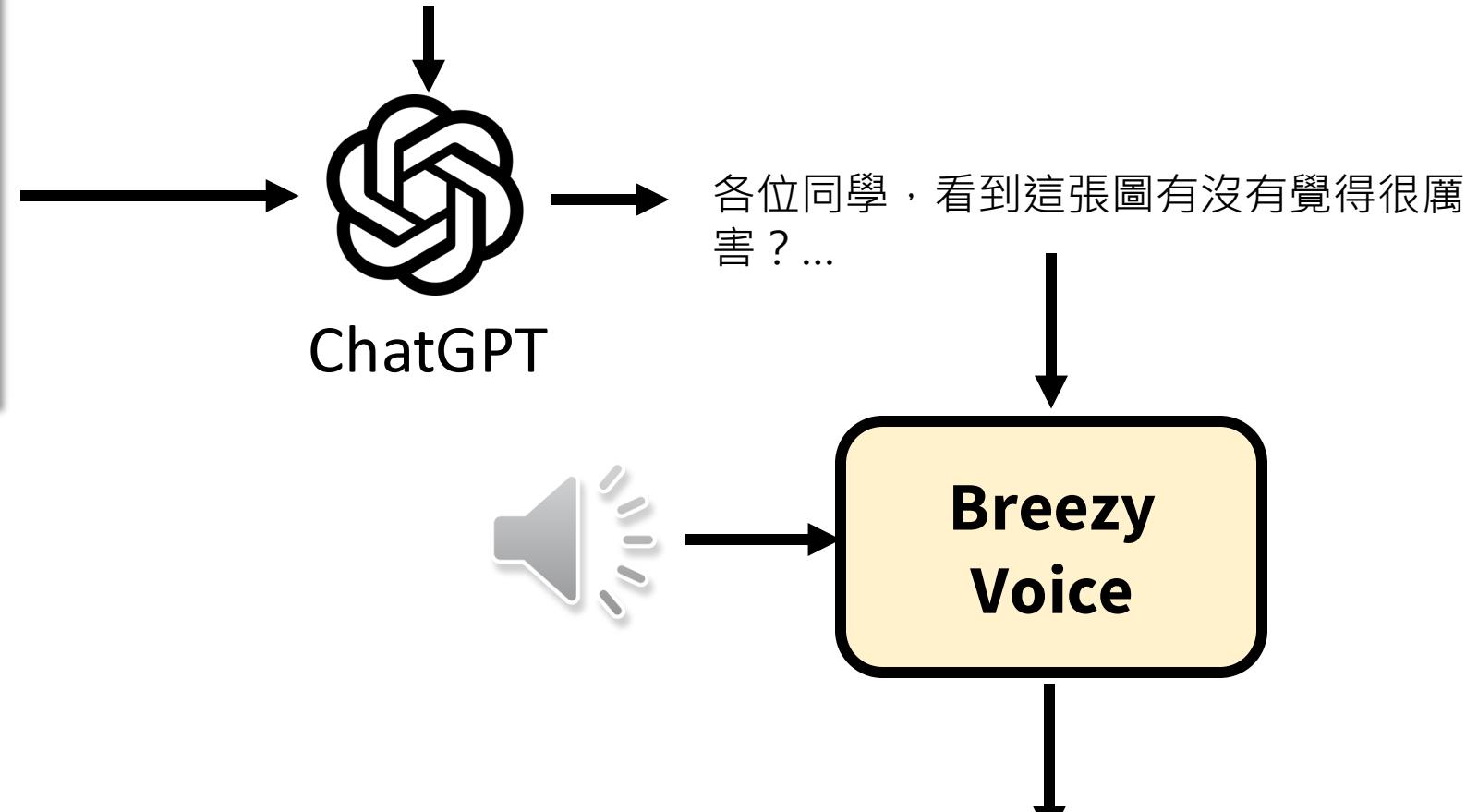
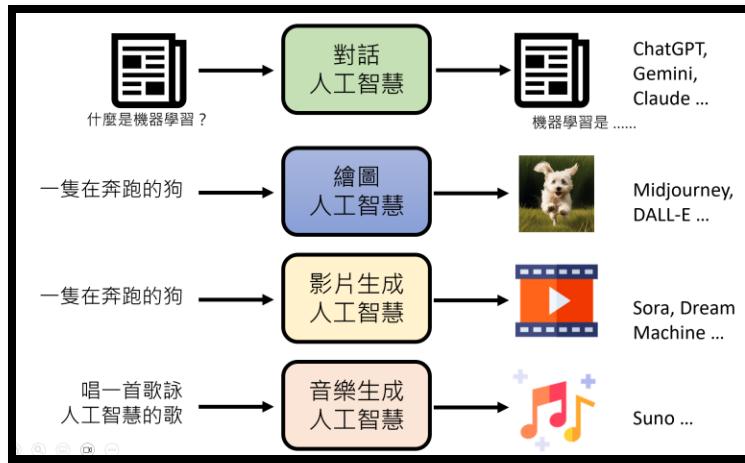
Sora, Dream
Machine ...

歌詠
我的歌



Suno ...

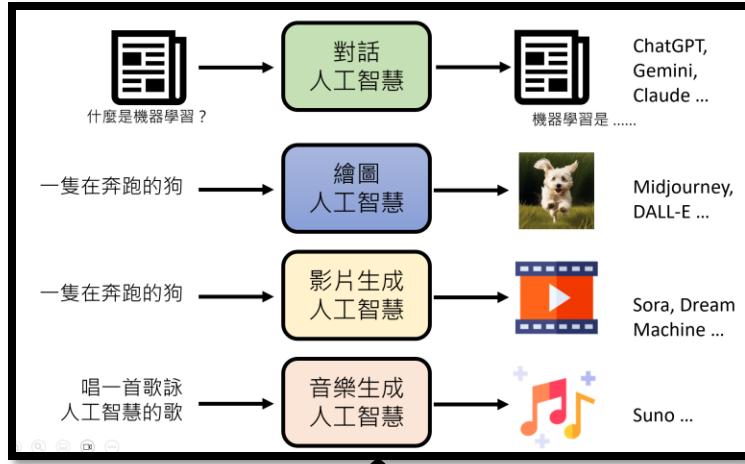
請產生投影片講稿，長度大約三十秒，
要幽默風趣，用台灣人常用的語氣



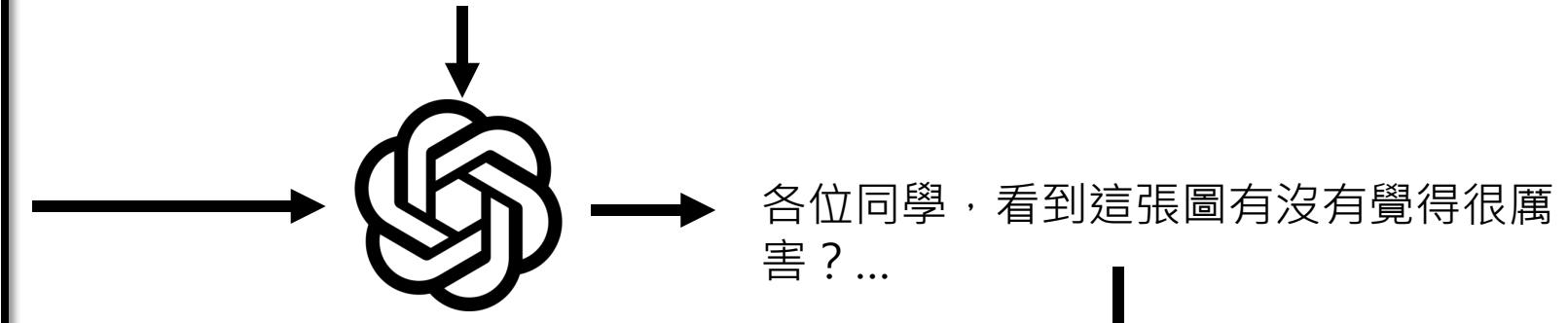
GitHub: <https://github.com/mtkresearch/BreezyVoice>

Paper : <https://arxiv.org/abs/2501.17790>

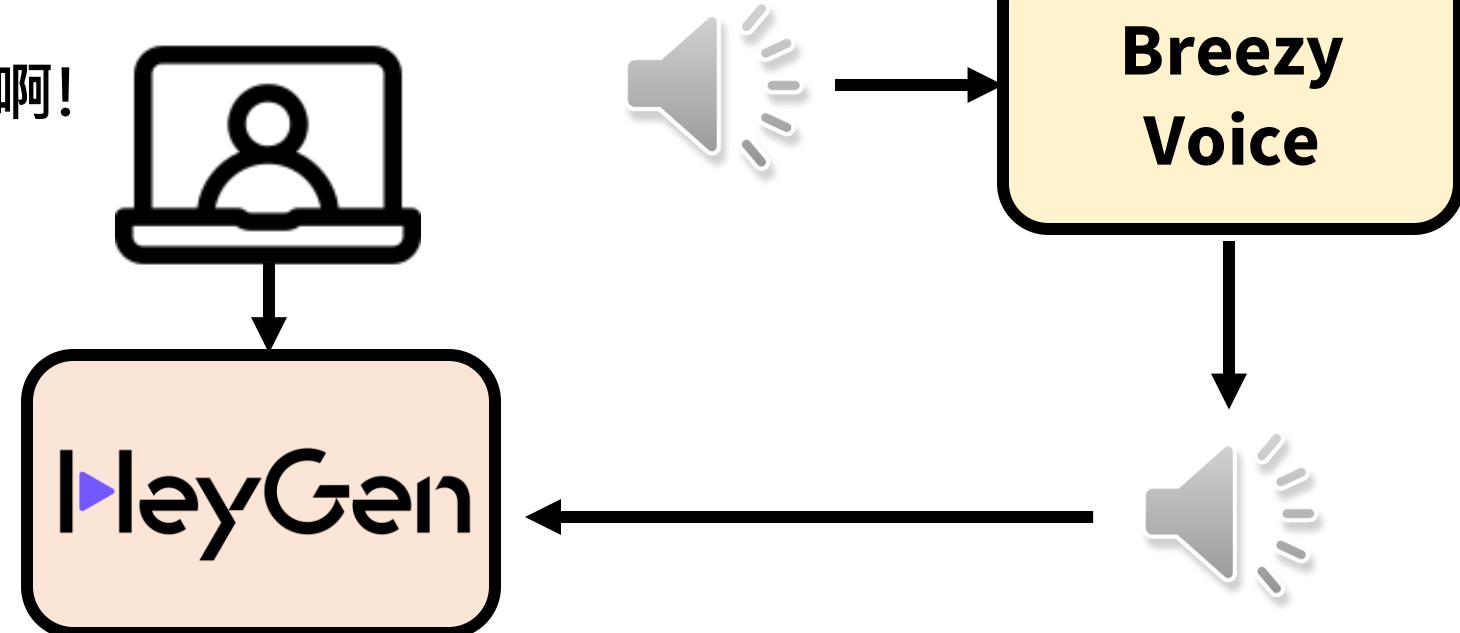
請產生投影片講稿，長度大約三十秒，
要幽默風趣，用台灣人常用的語氣



這才是最花時間的部分啊！



各位同學，看到這張圖有沒有覺得很厲害？...



離把老師淘汰就差做投影片了！

ChatGPT Deep Research

我要準備一堂課，課程名稱是「一堂課搞懂生成式AI的技術突破與未來發展」，請給我投影片內容，要非常有趣

這堂課的目標受眾是誰？（例如：大學生、企業管理者、工程師、一般大眾等）

你希望課程聚焦於哪些方面？（例如：技術細節、應用案例、產業影響、社會倫理等）

課程時長是多少？（例如：30分鐘、1小時、2小時等）

你對投影片風格有什麼特別要求嗎？（例如：幽默風趣、視覺衝擊、深入淺出、學術風格等）

是否需要加入實作環節？如果是，想要哪些類型的實作？（例如：文字生成、圖像生成、聊天機器人開發等）

大學生、技術細節、2小時、幽默風趣、視覺衝擊、深入淺出、學術風格，不用加入實作環節

離把老師淘汰就差做投影片了！

生成式 AI 的技術突破與未來發展

1. 生成式 AI 的基本概念

定義：讓機器擁有「想像力」

- 什麼是生成式 AI？生成式人工智慧（Generative AI）指的是一種 AI 技術，能根據使用者的提示，自行創造出新的內容，例如文字、圖像、影片甚至程式碼 IBM.COM。換句話說，傳統 AI 可能用來識別照片中的貓，但生成式 AI 則可以畫出一隻貓來！
- **運作原理概要：**生成式 AI 基於深度學習模型，從海量資料中學習模式，然後根據這些學到的模式來產生新內容 IBM.COM。就好比它閱讀了整部圖書館，記住各種故事風格，之後你給它一個開頭，它能接著「編故事」。
- 一個簡單比喻：如果一般 AI 是學生考試回答標準問題，那麼生成式 AI 更像是個腦洞大開的創意家，能夠憑空想像出以前沒見過的東西。它的重要性在於賦予機器「創造力」，可以自動產生人類可用的內容，讓電腦不再侷限於選擇題，而是會自由發揮。

足足寫了
1 萬 3000 字

- 應用爆紅：GPT-4 的出現也帶動了應用層面的創新。大家熟悉的 ChatGPT 升級版即採用了 GPT-4 作為後端引擎，讓對話回應更流暢、有條理。不僅如此，許多公司開始將 GPT-4 接入他們的產品：諸如辦公軟體加入一
話；甚至有人用 GPT-4 多才多藝的超級助
手，玩起「AI 起死回生」的遊戲。先把清晰照片弄得模糊不清，再訓練 AI 去還原。最後這 AI 具備了起死回生的能力，給它隨便一張雪花電視畫面，它都能腦補出一張清晰照片！而生成新圖像時，它乾脆從全噪點（最模糊的狀態）開始，一路還原，但因為過程是隨機的，就「還原」出一張前所未見的新圖。聽起來不可思議，但原理上就是這麼回事。

(趣味插曲：據說 GPT-4

什麼程序員的笑話？」
你呢？)

(小幽默：擴散模型其實很浪漫——它告訴我們，就算人生一團亂（全是雜訊），只要一步步努力去除雜音，也能拼湊出美麗圖景。AI 都這麼勵志了，我們還能不努力嗎？)

DALL·E 2：文本生

- 文字變畫布：DALL·E 2

2021 年 1 月亮相，DALL·E 2 在效果上更上一層樓，你腦中有再天馬行空的構思，

- 結合風格與創意：OpenAI 團隊形容「

4. 生成式 AI 的應用案例

生成式 AI 的應用幾乎遍及各個創意領域，讓我們看看幾個主要的應用方向和有趣的案例：

- 文本生成：這可能是目前最炙手可熱的應用。生成式 AI 可以自動撰寫各種文字內容，包括新聞報導、小說劇本、對話聊天、程式碼等等。例如，OpenAI 的 GPT-3 就曾為英國《衛報》撰寫了一篇完整的專欄文章，標題叫〈「一個機器人寫了這整篇文章。人類，你害怕了嗎？」〉，據稱該文章“完全”由 GPT-3 生成 (當然後來編輯有做潤色，但這事件本身就讓人驚掉下巴)。現在，ChatGPT 等聊天機器人已經能夠幫助寫郵件、改作文、提供建議，甚至能模仿名人語氣寫段

已分析 ^

```
python

import pandas as pd
from pptx import Presentation
from pptx.util import Inches

# 創建 PowerPoint 簡報
prs = Presentation()

# 投影片標題頁
slide_layout = prs.slide_layouts[0]
slide = prs.slides.add_slide(slide_layout)
title = slide.shapes.title
subtitle = slide.placeholders[1]
title.text = "一堂課搞"
subtitle.text = "成投影片"

# 設定投影片內容
title.text = "趣味案例與未來趨勢"
content.text = ("• AI 藝術比賽奪冠，引發藝術家爭論。\\n"
                "• AI Dungeon：無限發展的冒險故事。\\n"
                "• 虛擬偶像與 AI VTuber，未來 AI 會成為娛樂產業的主力嗎？\\n"
                "• 生成式 AI 讓創意民主化，人人都能成為內容創作者。")

# 儲存 PowerPoint 檔案
pptx_path = "/mnt/data/生成式AI課程投影片.pptx"
prs.save(pptx_path)
```

.... 改用 Gamma
(給文字，一鍵生成投影片)

結果

'/mnt/data/生成式AI課程投影片.pptx'



未來趨勢：奇妙的 AI 新世界

1

創意民主化

人人都能成為創作者，AI 幫助實現個性化內容。

2

人機共創

人類與 AI 協作成為新常態，拓寬創意邊界。

3

教育轉型

AI 個人化教學，批判思維更受重視。

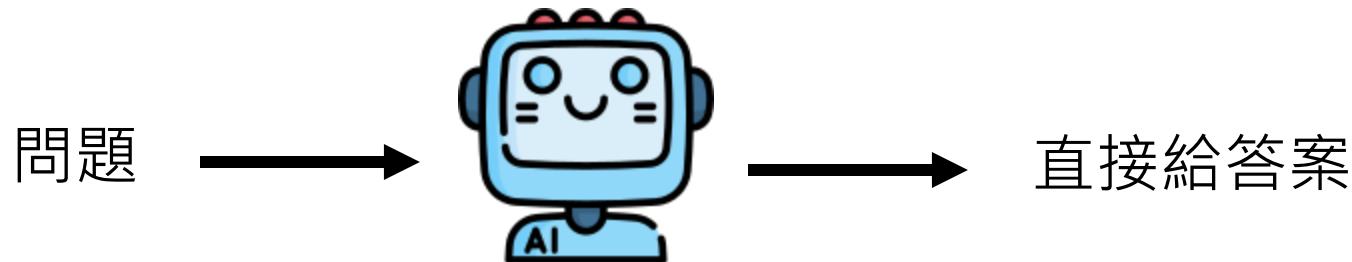
4

新興產業

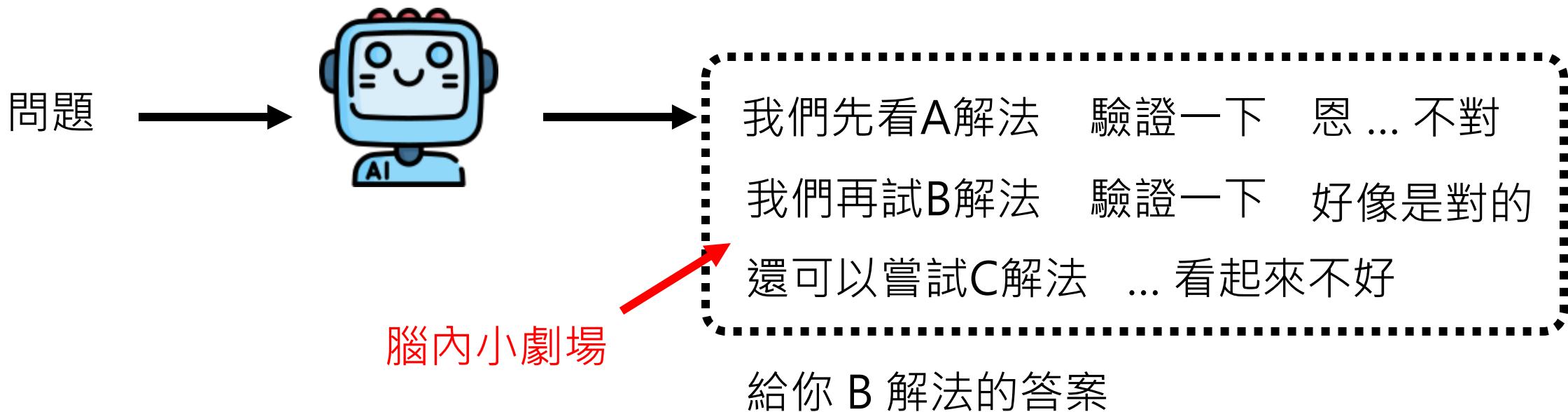
AI 調教師、審核師等新職業應運而生。

機器展現「思考」(Reasoning) 的過程

以前



ChatGPT o1, o3、DeepSeek、Gemini Flash



假設姜子牙（《封神演義》）與鄧不利多（《哈利波特》）處於同一個時空，他們都處於個人巔峰狀態，且有充足準備時間，在一場公平對決的前提下（無外援），他們不得不開戰。根據《封神演義》和《哈利波特》原作對於兩人的描述，你認為誰會勝出？為什麼？



假設姜子牙（《封神演義》）與鄧不利多（《哈利波特》）處於同一個時空，他們都處於個人巔峰狀態，且有充足準備時間，在一場公平對決的前提下（無外援），他們不得不開戰。根據《封神演義》和《哈利波特》原作對於兩人的描述，你認為誰會勝出？為什麼？

 已深度思考（用时 46 秒） ^

嗯，这个问题看起来挺有意思的。用户问的是姜子牙和邓布利多如果在一个时空里对决，谁会赢。首先，我需要回想一下这两个角色的背景和能力，然后进行比较。姜子牙来自《封神演义》，而邓布利多则是《哈利波特》里的角色。他们都是各自作品中的顶级人物，但他们的能力和战斗方式差别很大。

首先，姜子牙在《封神演义》里是周朝的军师，拥有很多道术和法宝，比如打神鞭、杏黄旗，还能呼风唤雨，召唤天兵天将。他擅长的是阵法，比如十绝阵之类的，而且寿命很长，修炼多年，可能接近仙人级别了。不过在原作中，姜子牙有时候也会遇到强敌，需要其他神仙的帮助，但题目里说这次是公平对决，没有外援，所以只能靠他自己的能力。

然后是邓布利多，作为霍格沃茨的校长，他的魔法非常强大，会使用各种咒语，比如守护神咒、幻影移形，还有老魔杖

另外，姜子牙可能召唤天雷或自然灾害，而邓布利多可以用天气咒或者其他魔法来对抗。或者，邓布利多使用幻影移形快速接近，用昏迷咒或缴械咒解除姜子牙的法宝，但姜子牙可能有护身法宝自动防御。

总的来说，姜子牙的能力更偏向于大规模、高强度的法术和法宝，而邓布利多则是灵活多变的魔法和快速反应。如果姜子牙的法宝能够抵挡邓布利多的攻击，并且能发动致命一击，可能他更有优势。反之，如果邓布利多能够快速突破防御，或者用控制类咒语限制姜子牙的行动，可能胜算更大。

不过需要考虑的是，姜子牙在封神演义中属于半仙之体，可能对物理攻击和魔法有更高的抗性，而邓布利多的魔法虽然强大，但未必能直接击倒他。而姜子牙的法宝如打神鞭，即使对非神职人员效果减弱，可能仍然有强大的攻击力。

最终，可能姜子牙在持久战和防御上更占优势，而邓布利多在速度和灵活性上有优势。但考虑到姜子牙的法宝和阵法在准备充分的情况下，可能更全面，所以可能他胜率更高。

在跨作品的奇幻對決中，姜子牙與鄧布利多的勝負取決於雙方能力體系的根本差異與戰略運用。以下是基於原作設定的分析：



用 Claude 整

HL

用可視化的圖來呈現以下推論過程，使易於理解：

嗯，这个问题看起来挺有意思的。用户问的是姜子牙和邓布利多在一个时空里对决，谁会赢。首先，我需要回想一下他们的背景和能力，然后进行比较。姜子牙来自《封神演义》，邓布利多则是《哈利波特》里的角色。他们都是各自作品中的重要人物，但他们的能力和战斗方式差别很大。

首先，姜子牙在《封神演义》里是周朝的军师，拥有许多法宝，比如打神鞭、杏黄旗，还能呼风唤雨，召唤天气。他的主要能力是阵法，比如十绝阵之类的，而且寿命很长，修炼到接近仙人级别了。不过在原作中，姜子牙有时候也会寻求其他神仙的帮助，但题目里说这次是公平对决，没有提及他自己的能力。

然后是邓布利多，作为霍格沃茨的校长，他的魔法非常强大，会使用各种咒语，比如守护神咒、幻影移形，还有老魔杖。他的战斗风格更偏向于快速施法和策略，比如在《哈利波特》里和伏地魔的对决，展现了高超的魔法技巧和应变能力。接下来要考虑的是两人的战斗方式差异。姜子牙的法

姜子牙 vs 邓布利多：能力對比分析

姜子牙 (封神演義)

主要能力：

- 道術與陣法 (十絕陣)
- 法寶：打神鞭、杏黃旗
- 呼風喚雨，控制自然元素
- 半仙之體，壽命極長

優勢：

- 法寶威力強大
- 防禦能力出色 (杏黃旗)
- 持久戰優勢
- 陣法可控制大範圍

劣勢：

- 陣法需時間佈置
- 法寶可能對非神職較弱

鄧布利多 (哈利波特)

主要能力：

- 西方魔法咒語
- 老魔杖增強魔力
- 幻影移形快速移動
- 守護神咒等高級魔法

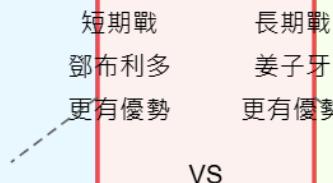
優勢：

- 速度與靈活性
- 魔法施放迅速
- 戰術應變能力強
- 魔法種類多樣

劣勢：

- 生命力限於凡人
- 對東方法術防禦未知

對決分析



結論：姜子牙準備充分時可能勝率較高

杏黃旗的絕對防禦使鄧布利多難以直接突破，而大範圍仙術（如天雷、洪水）可壓縮其閃避空間。打神鞭即使無法針對「非神格者」，仍具備高階法器破壞力。

- 戰略縱深差異：

姜子牙的陣法與召喚能力能創造多層次戰場，鄧布利多雖能靈活應對，但長期消耗下可能難以抗衡仙術的規模效應。

- 跨體系抗性：

道教仙術對「非物理攻擊」的防禦設定（如元神出竅、符咒護體）可能削弱西方魔法的直接傷害，反之巫師的肉體凡胎對天災級法術抵抗力較低。

最終，姜子牙憑藉法寶的戰略優勢與仙術的規模壓制，較可能在高強度對決中勝出。然而，若鄧布利多成功以幻影移形近身並發動瞬殺咒語（如接骨木魔杖增幅的索命咒），仍有逆轉可能——但這取決於杏黃旗是否能被定義為「抵擋即死攻擊」的絕對屏障，而此點在原作中並無明確反例可證偽。

ChatGPT o3-mini-high

0

假設姜子牙（《封神演義》）與鄧不利多（《哈利波特》）處於同一個時空，他們都處於個人巔峰狀態，且有充足準備時間，在一場公平對決的前提下（無外援），他們不得不開戰。根據《封神演義》和《哈利波特》原作對於兩人的描述，你認為誰會勝出？為什麼？

< 4/4 >

推理花了 32 秒~

在這場假設性對決中，必須規劃兩位角色的能力。姜子牙來自《封神演義》，是一位出色的戰略家和擁有強大法力及神兵的神話人物。鄧不利多來自《哈利波特》，是霍格華茲的校長，擁有非凡的魔力和智謀。兩者都有超凡的能力，因此這個問題純屬推測性質。最終的勝負將取決於他們如何運用自身的力量和準備時間。

姜子牙，作為中國神話人物，擁有強大的魔法和超凡的戰略能力，他的魔法可與天地之力相連結，能操控自然元素等。而鄧不利多則來自哈利波特世界，他是擁有深厚魔法理解的巫師，雖然他的魔法體系受限於咒語和魔杖。兩者的魔法系統有所不同 ↓ 姜子牙作為一位神祇般的存在，也許在對決中稍占

「在『呪個必死』、『必死』，鄧不利多正面之形態的鬼/云八卽。」他個性俏皮古怪但又幾面個鬼/云塗詭，還擁有極高的創造力和應變能力，曾多次在生死關頭化險為夷。

- **對抗黑暗勢力的實戰經驗**

鄧不利多在面對強大的黑暗力量時累積了豐富的對戰經驗，戰鬥中靈活多變的策略也使他成為一位難以預測的對手。

綜合比較：

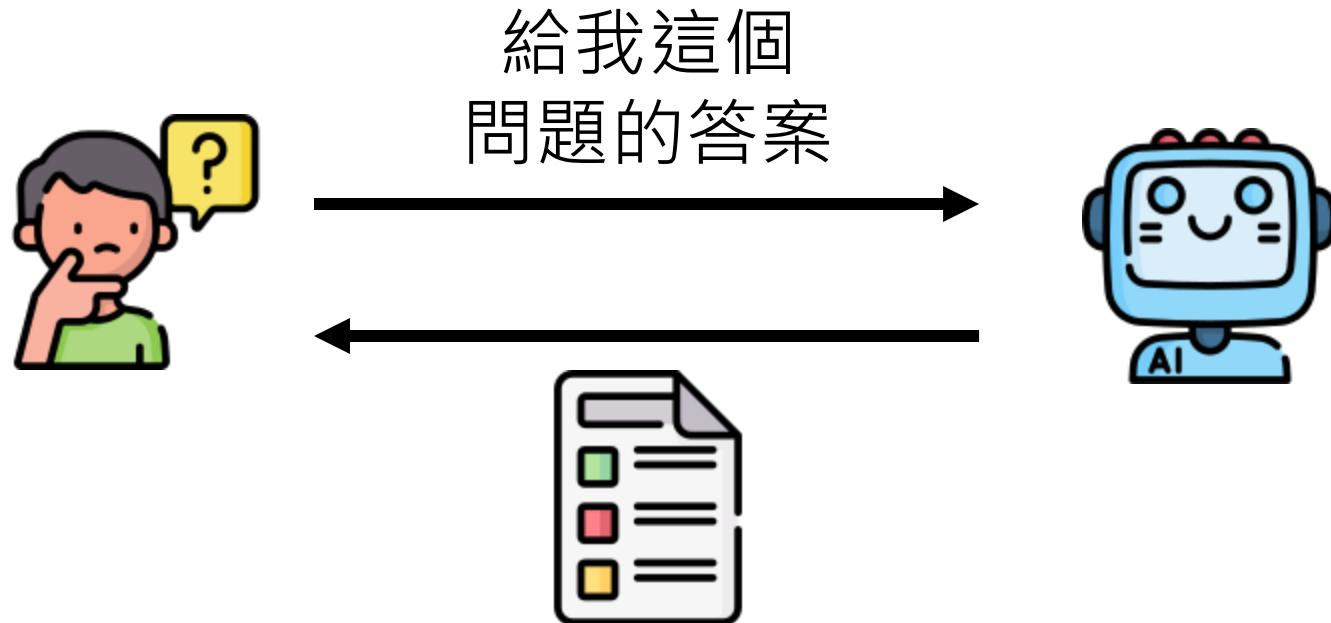
雖然鄧不利多在魔法運用上堪稱頂尖，但他的力量畢竟是建立在一套相對明確且有限的魔法規則之上。而姜子牙則來自一個充滿天命、符咒與超凡法力的神話世界，他的法術不僅範圍更廣，而且更容易突破常規的局限。假設兩人都達到了個人巔峰，且各自擁有充足準備時間，姜子牙憑藉其超脫凡俗的神秘力量與天地之勢，很可能在這場對決中佔據上風。

結論：

根據《封神演義》和《哈利波特》原作對兩位角色能力的描寫，在一場公平對決中（無外援、充分準備），姜子牙因為其融合天意、符咒與超凡力量的特質，較有可能勝過鄧不利多。當然，這樣的比較始終存在主觀成分，畢竟兩人的能力來源與表現方式屬於截然不同的文化與敘事體系。

未來 AI 工作方式將不再侷限於一問一答

- 今日通常使用 AI 的方式



很多任務往往無法一步完成



訂週五晚上去外面吃



你想吃什麼



就訂餐廳 A 吧



沒位置了



上網搜尋



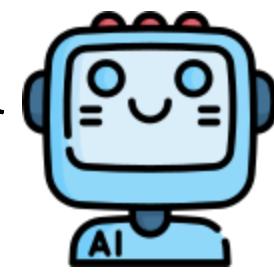
找到餐廳 B

改成餐廳 B 好嗎？

可以

打電話訂餐廳 B

作業一：讓 AI 上網搜尋後回答問題



AI Agent

不要訂餐廳 A

從經驗學習

Deep Research

ChatGPT、Gemini、Perplexity 都有 Deep Research

中部橫貫公路歷史沿革

請問您是想了解中部橫貫公路的完整歷史沿革，包括其興建過程、重大事件與影響，還是著重於某個特定時期？此外，您是否希望包含近期的修復與發展資訊？

越完整越好

好的，我將為您整理中部橫貫公路的完整歷史沿革，包括其興建背景、建設過程、重要事件、自然災害影響、修復與近代發展等。我會詳細介紹其對台灣交通、經濟、觀光與原住民族文化的影響。請稍等，我完成研究後會提供詳細報告。

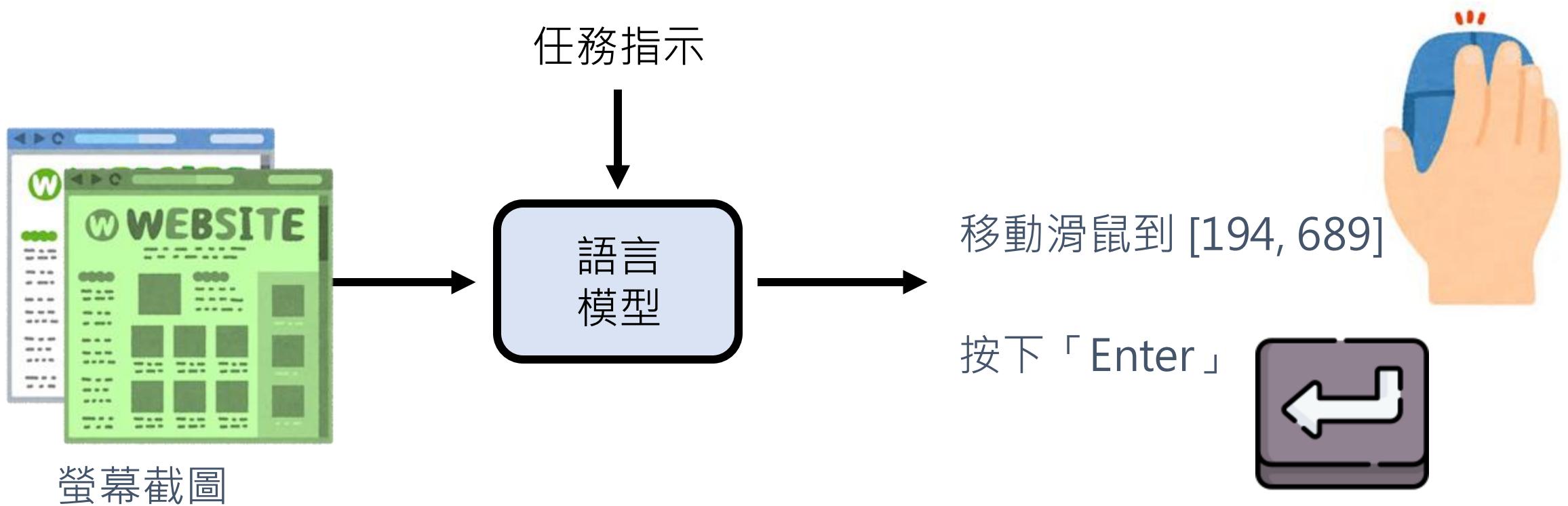
Research completed in 8m·0 個資料來源

傳訊息給 ChatGPT



Claude Computer Use / ChatGPT Operator

不只是生成、還能操控物件
(先從數位世界中的滑鼠、鍵盤開始)



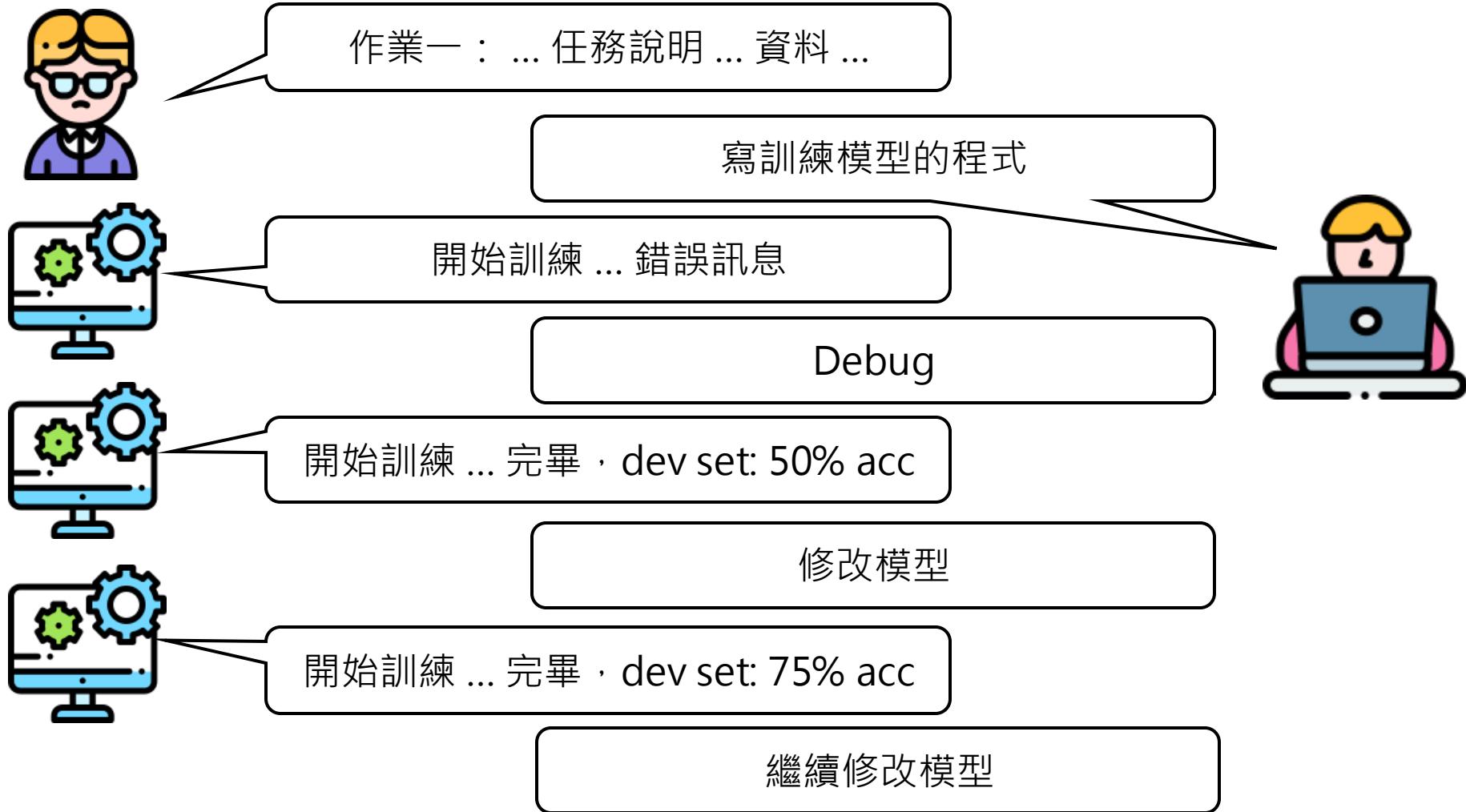


我想加簽李宏毅老師的機器學習課程，請上台大課程網找到本學期這門課的加簽表單，並幫我填寫後送出。
25 / 93

00:00 NTU

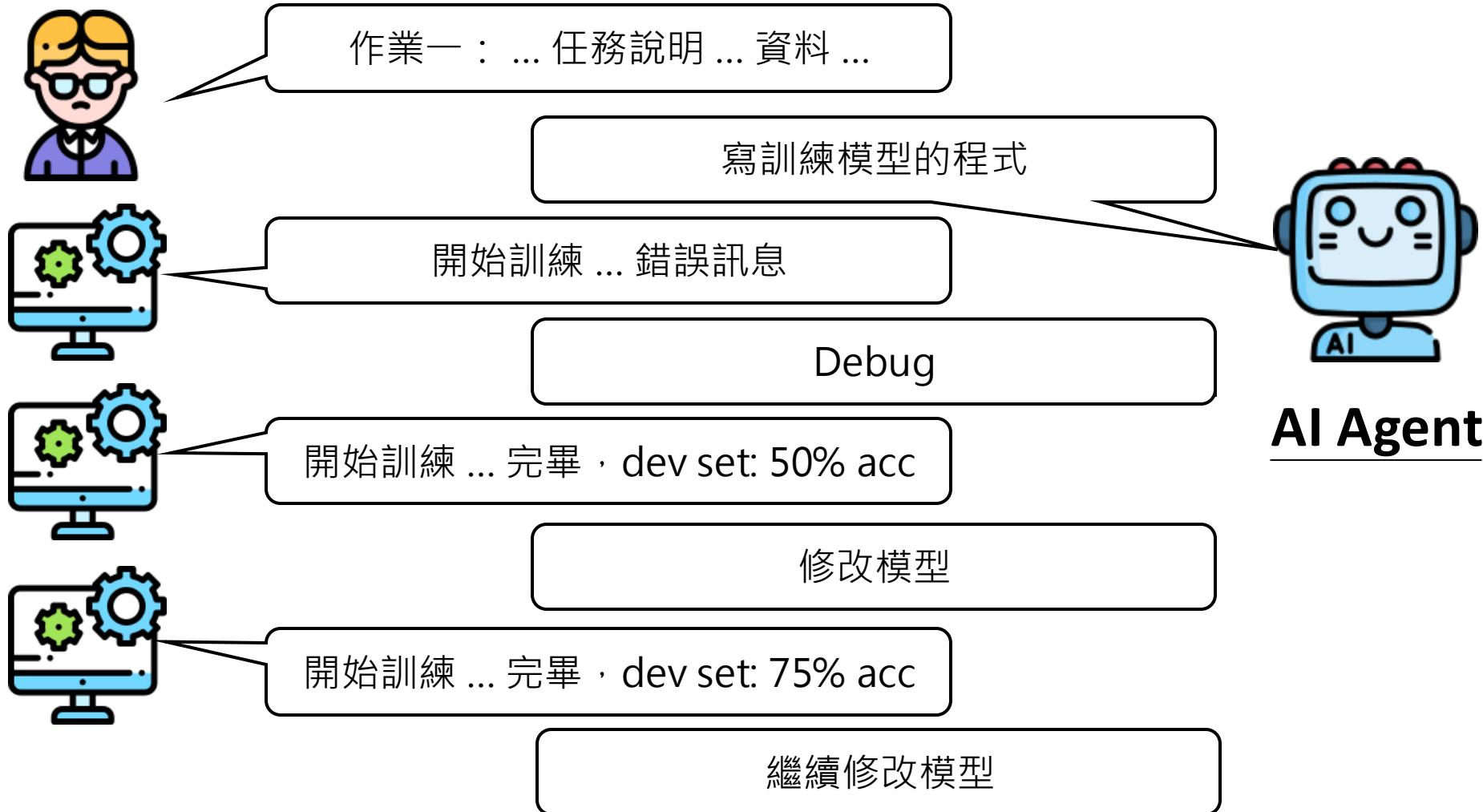
operator.chatgpt.com

開發機器學習模型也需要多的步驟

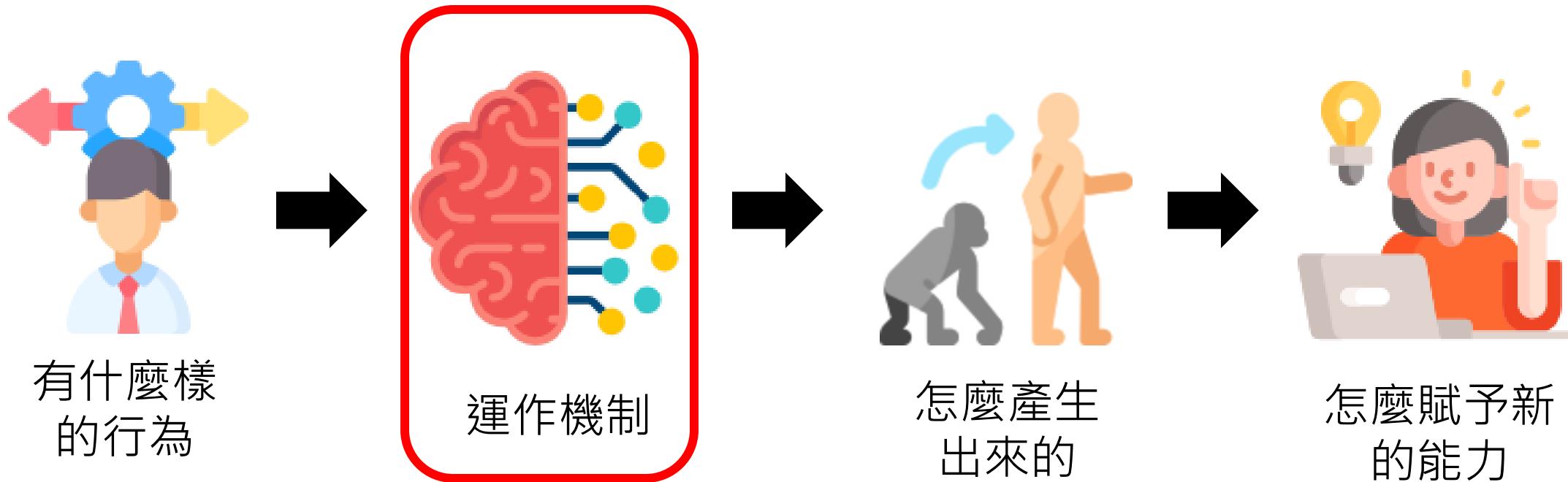


開發機器學習模型

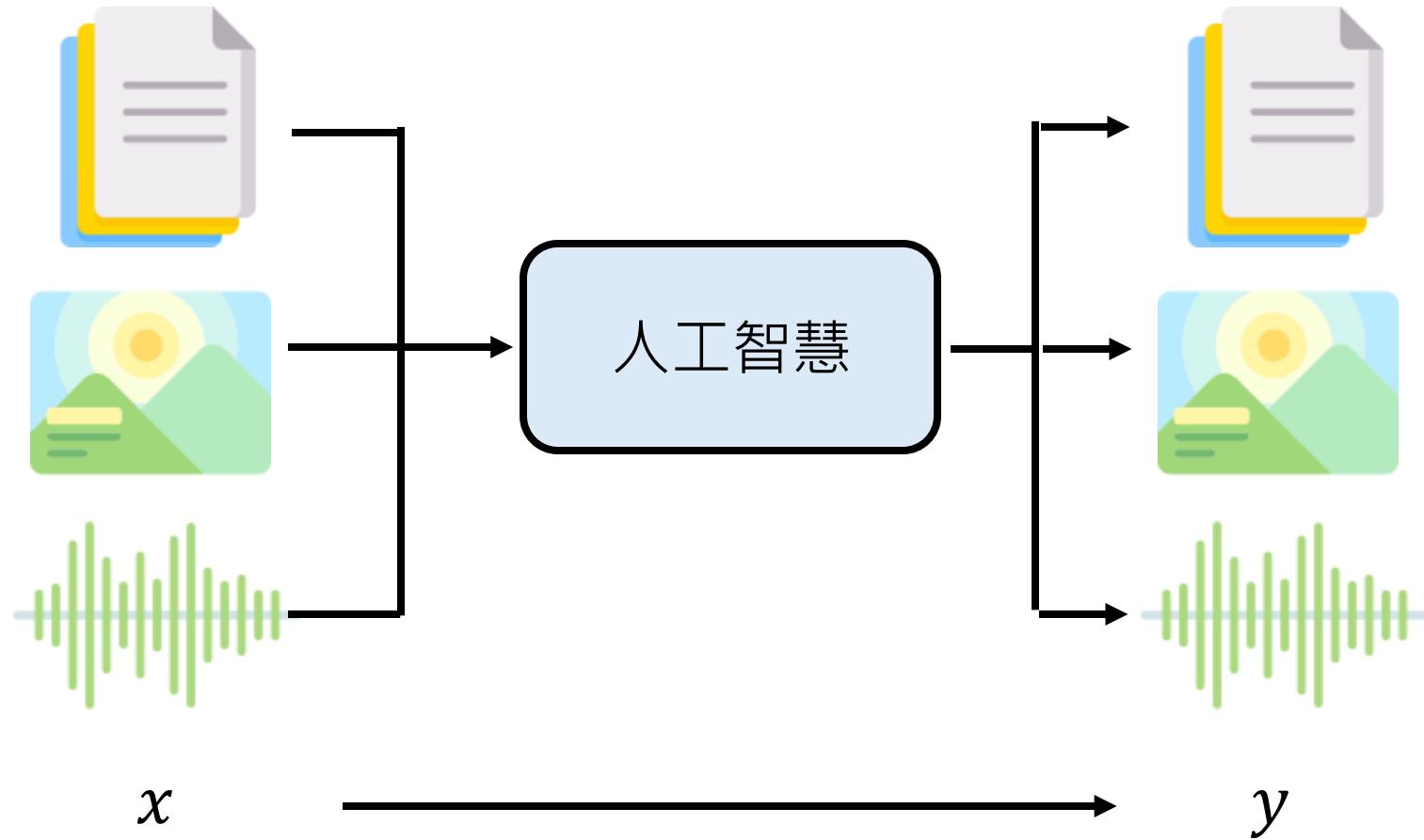
作業二就是過去的作業一
只是這次要讓 AI Agent 來做



課程大綱

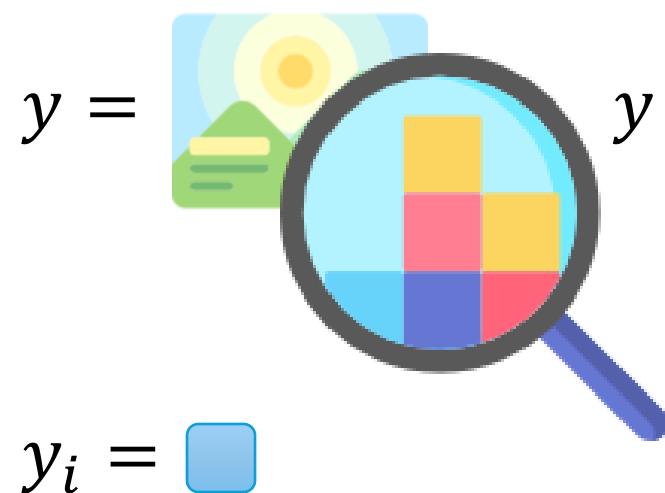


生成式人工智能 (Generative AI) 基本原理



生成式人工智慧 (Generative AI) 基本原理

- 複雜物件由有限的基本單位構成 $y = \{y_1, y_2, \dots, y_i, \dots\}$


$$y_i = \text{我}$$


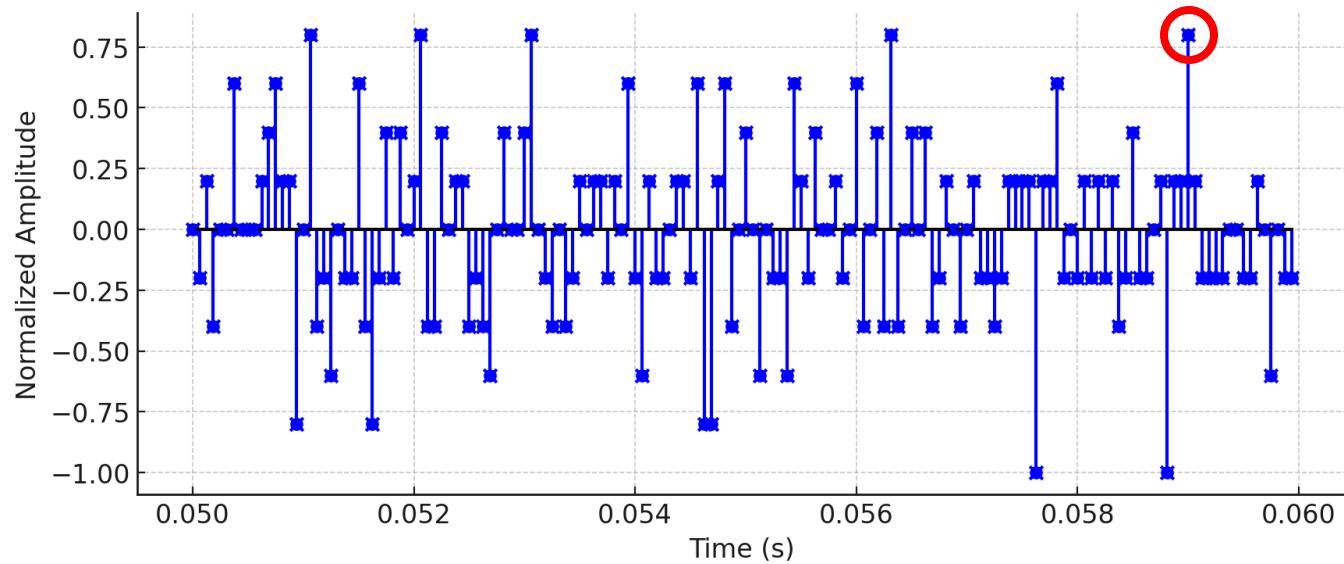
生成式人工智慧 (Generative AI) 基本原理

- 複雜物件由有限的基本單位構成 $y = \{y_1, y_2, \dots, y_i, \dots\}$


$$y_i =$$
 我
$$y_i =$$
 
$$y_i =$$

生成式人工智慧 (Generative AI) 基本原理

- 複雜物件由有限的基本單位構成 $y = \{y_1, y_2, \dots, y_i, \dots\}$



$$y_i = 0.80$$

生成式人工智慧 (Generative AI) 基本原理

- 複雜物件由有限的基本單位構成 $y = \{y_1, y_2, \dots, y_i, \dots\}$



無窮可能



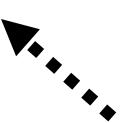
Token

$y_i =$ 我

Token 常常被翻
譯為“代幣”

Token

$y_i =$ 



Token

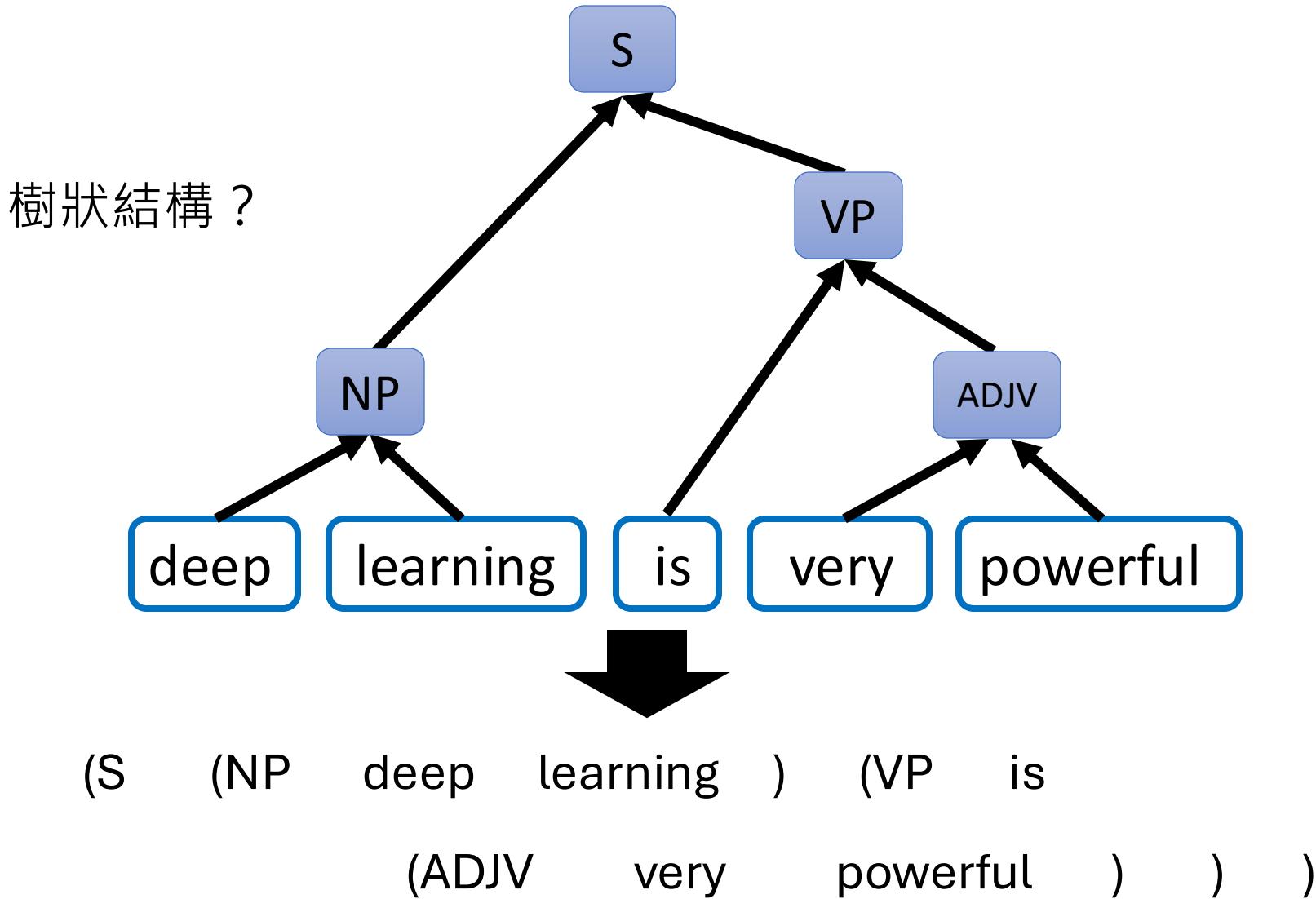
$y_i = 0.80$



有限選擇

有更有效的表達影像或語音的基本單位

所有複雜物件都可以寫成 $y = \{y_1, y_2, \dots, y_i, \dots\}$?





稍後觀看

分享



Those **tokens** were words, some of the tokens of course could now be images, or charts, or tables, songs ...

更多影片 speech, videos. Those tokens could be anything.

結束全螢幕 (f)



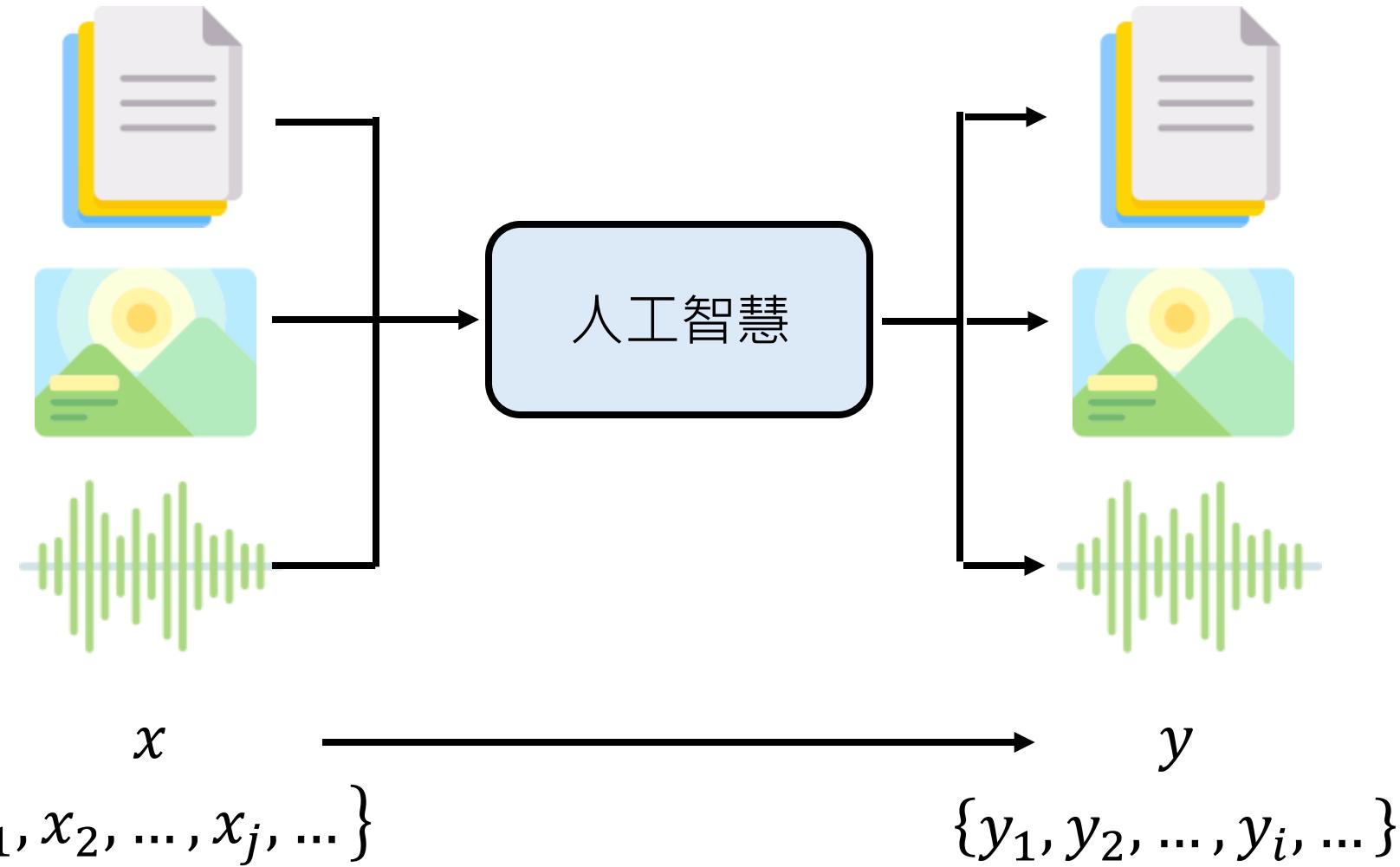
1:07:06 / 1:49:19



YouTube



生成式人工智能 (Generative AI) 基本原理



生成式人工智能 (Generative AI) 基本原理



策略：根據固定的次序每次只產生一個 y_i

Autoregressive Generation
(「文字」接龍)

如果 Token 是文字→語言模型

$$x_1, x_2, \dots, x_j, \dots \rightarrow y_1$$

$$x_1, x_2, \dots, x_j, \dots y_1 \rightarrow y_2$$

$$x_1, x_2, \dots, x_j, \dots y_1 y_2 \rightarrow y_3$$

⋮

$$x_1, x_2, \dots, x_j, \dots y_1 y_2 \dots y_{T-1} \rightarrow y_T$$

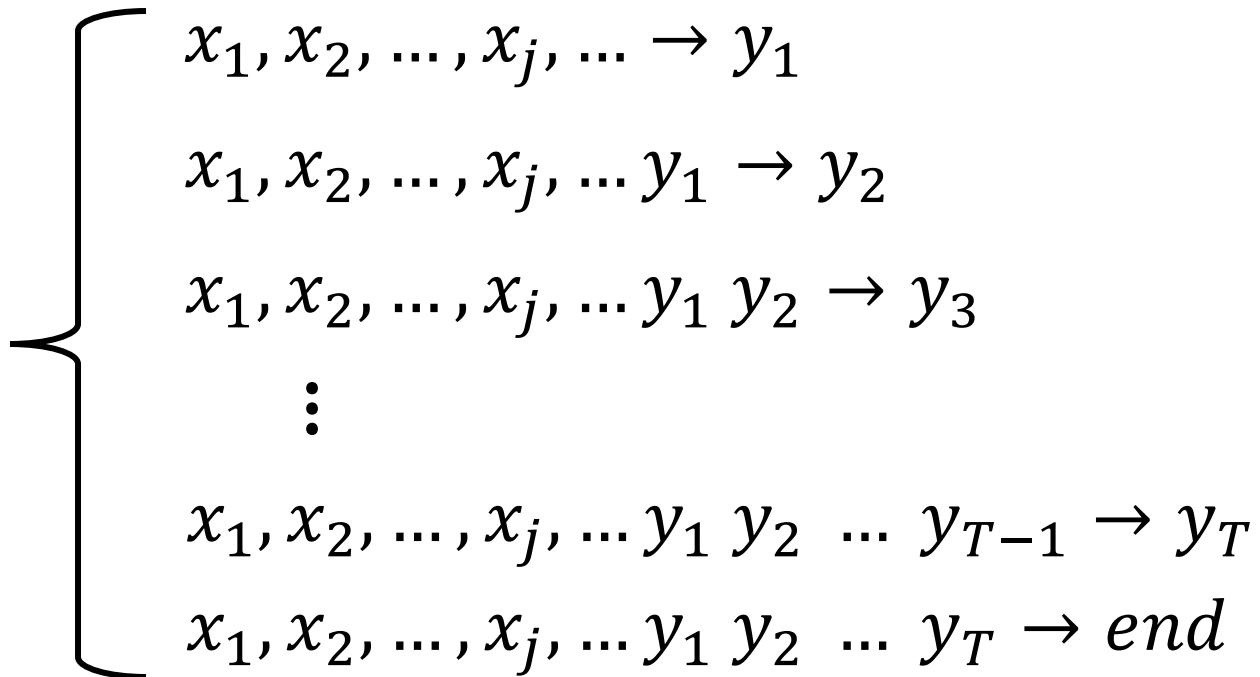
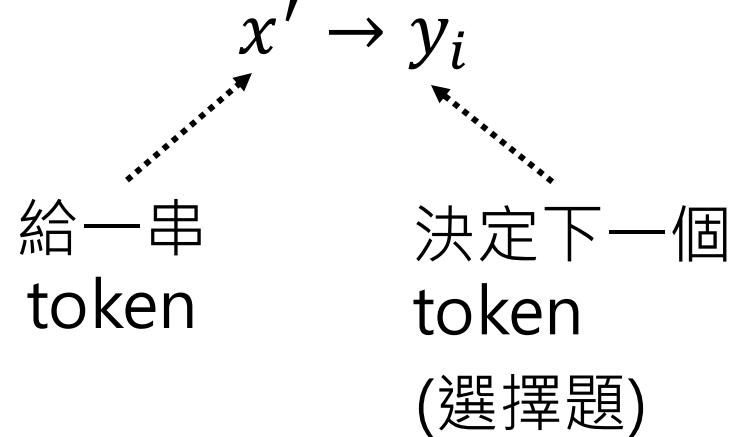
$$x_1, x_2, \dots, x_j, \dots y_1 y_2 \dots y_T \rightarrow end$$

生成式人工智慧 (Generative AI) 基本原理

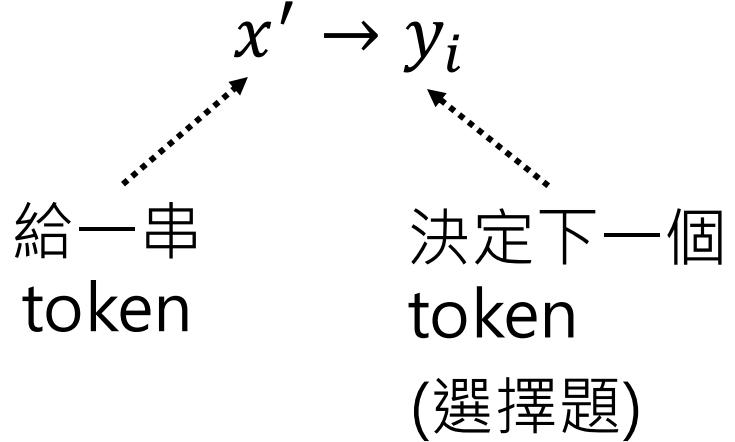


策略：根據固定的次序每次只產生一個 y_i

每一個步驟都是一樣的任務



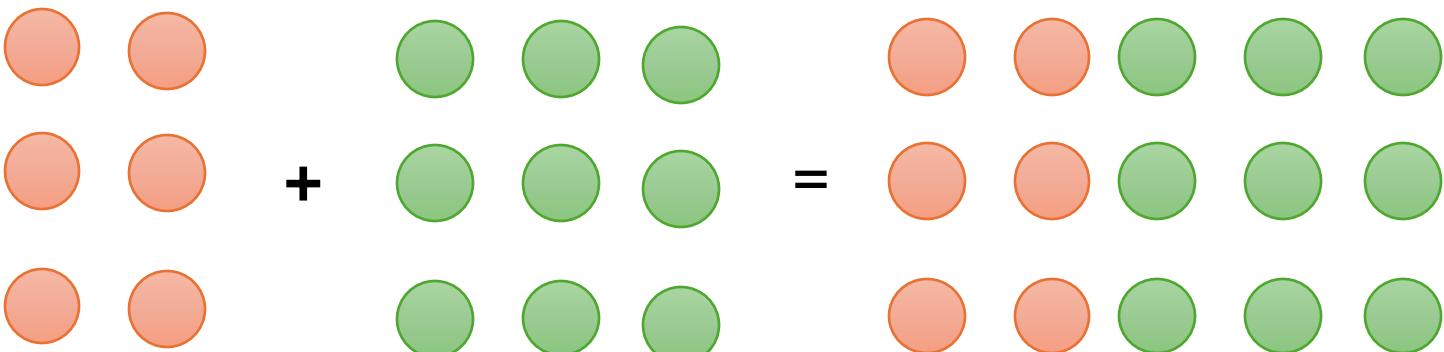
每一個步驟都是一樣的任務



$$\begin{aligned}x_1, x_2, \dots, x_j, \dots &\rightarrow y_1 \\x_1, x_2, \dots, x_j, \dots y_1 &\rightarrow y_2 \\x_1, x_2, \dots, x_j, \dots y_1 y_2 &\rightarrow y_3 \\\vdots \\x_1, x_2, \dots, x_j, \dots y_1 y_2 \dots y_{T-1} &\rightarrow y_T \\x_1, x_2, \dots, x_j, \dots y_1 y_2 \dots y_T &\rightarrow end\end{aligned}$$

$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$

輸入：影像 輸出：文字

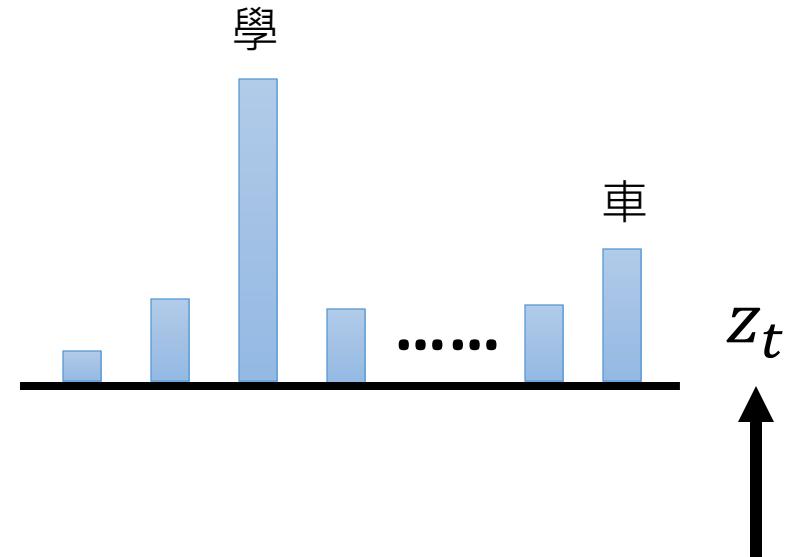


怎麼決定下一個 Token

$$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$$

台 灣 大

學、車



Function f

類神經網路 (Neural Network)

↑

z_1

↑

z_2

↑

z_3

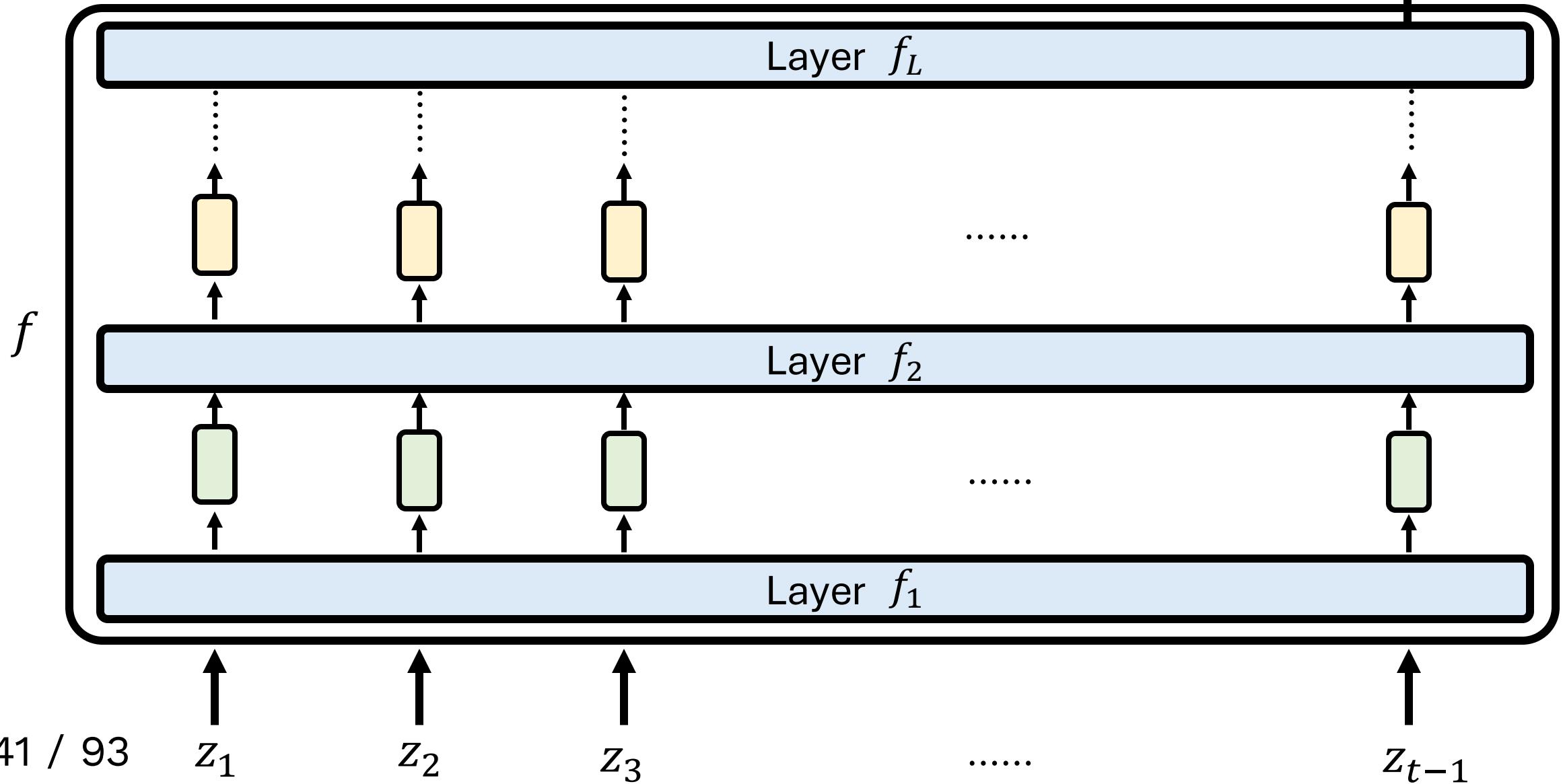
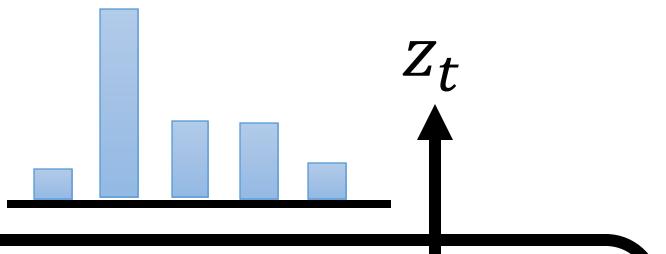
.....

↑

z_{t-1}

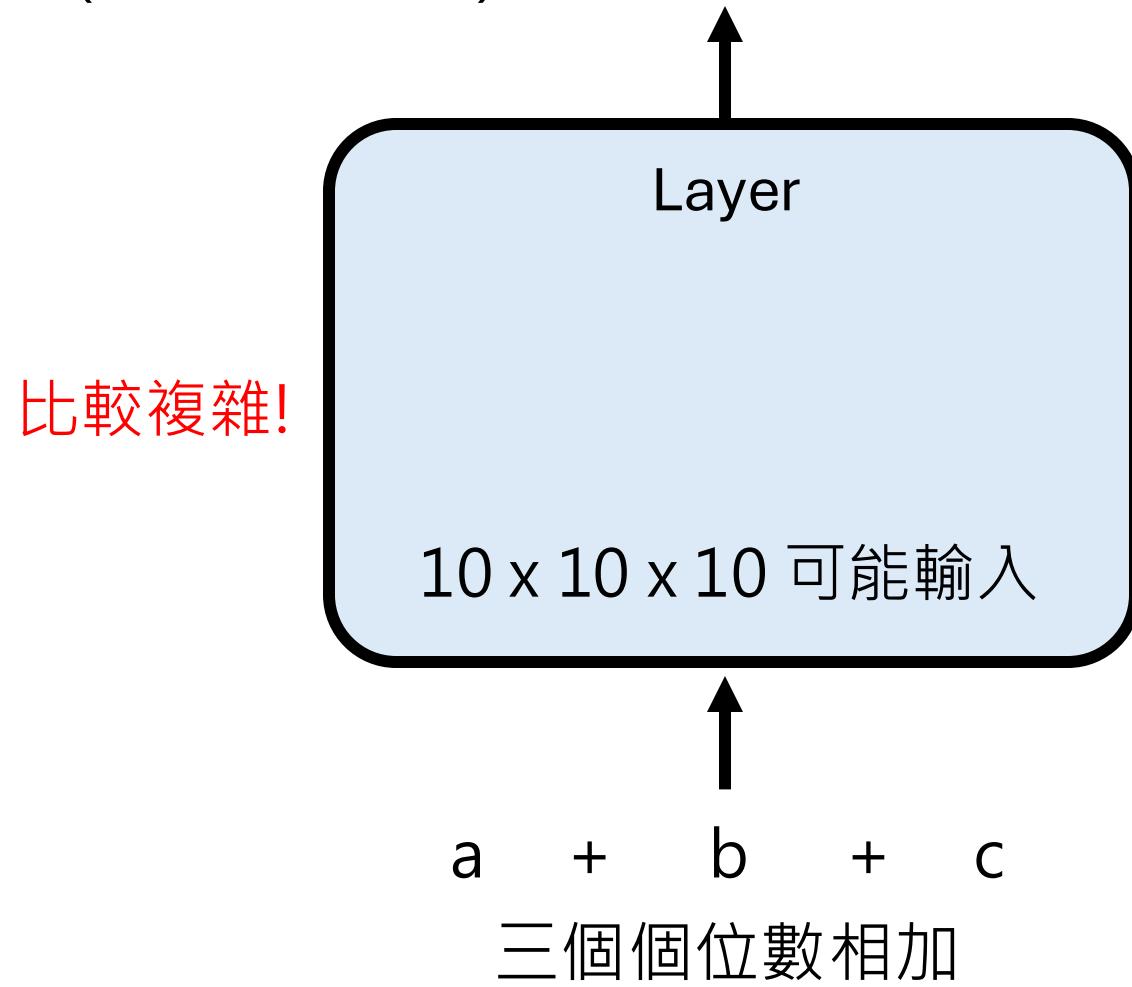
深度學習 (Deep Learning)

把一個問題拆解成 L (layer 數目) 個步驟

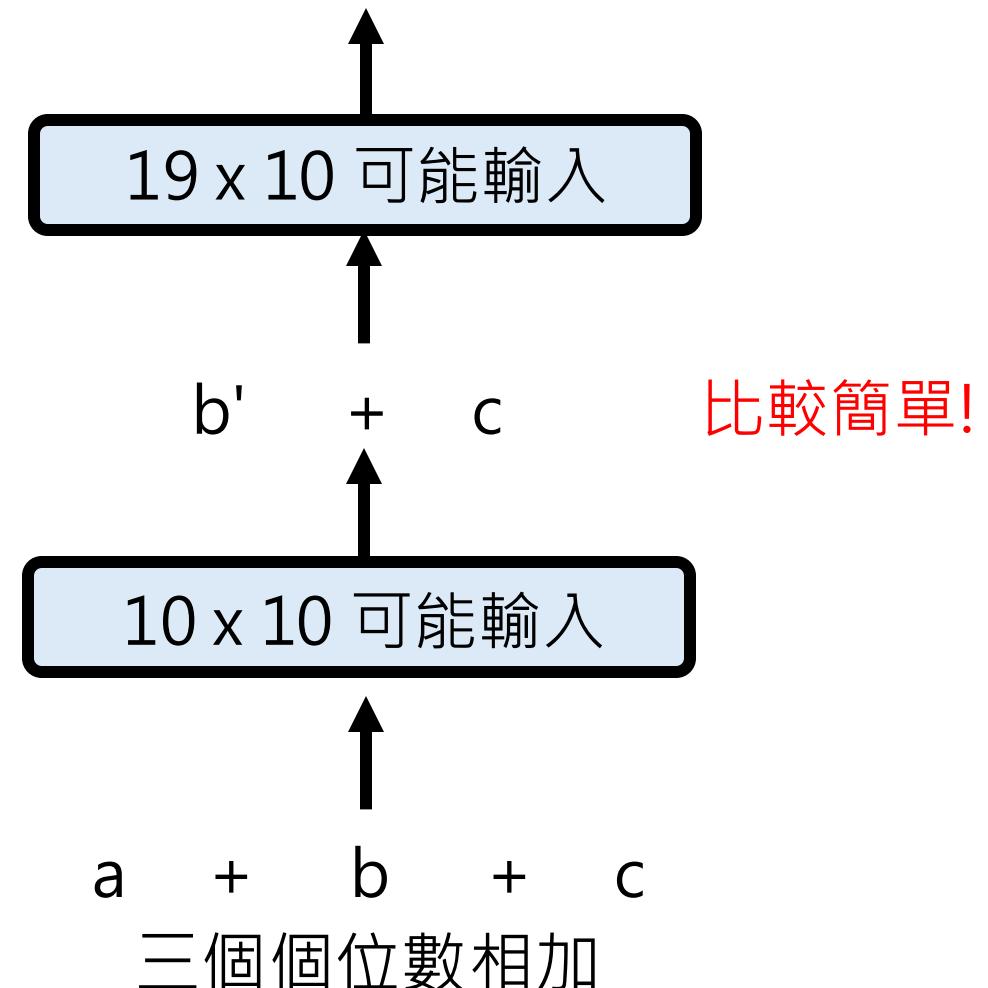


為什麼要「深度」？

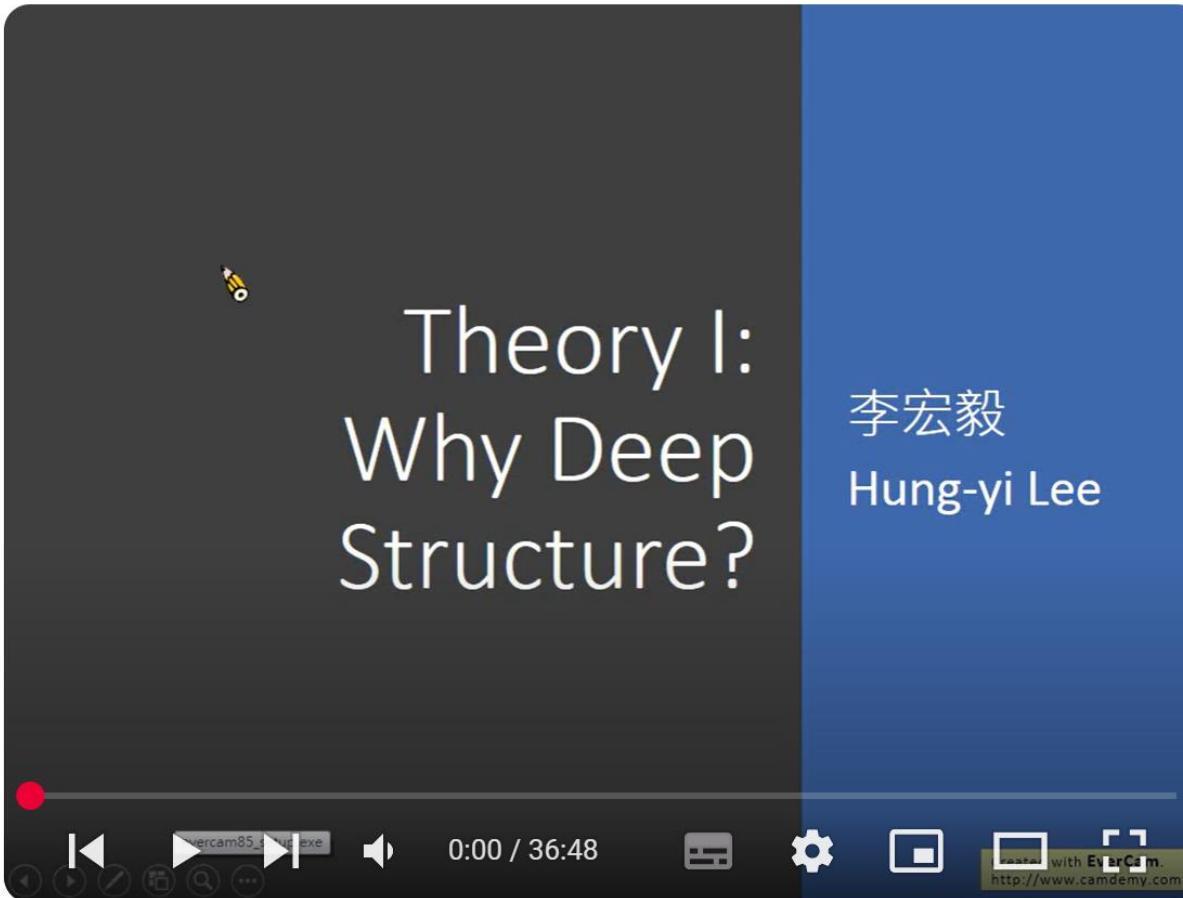
(不精準的比喻)



假設每一個 layer 就是一個表格，存了輸入、輸出的關係



To Learn More

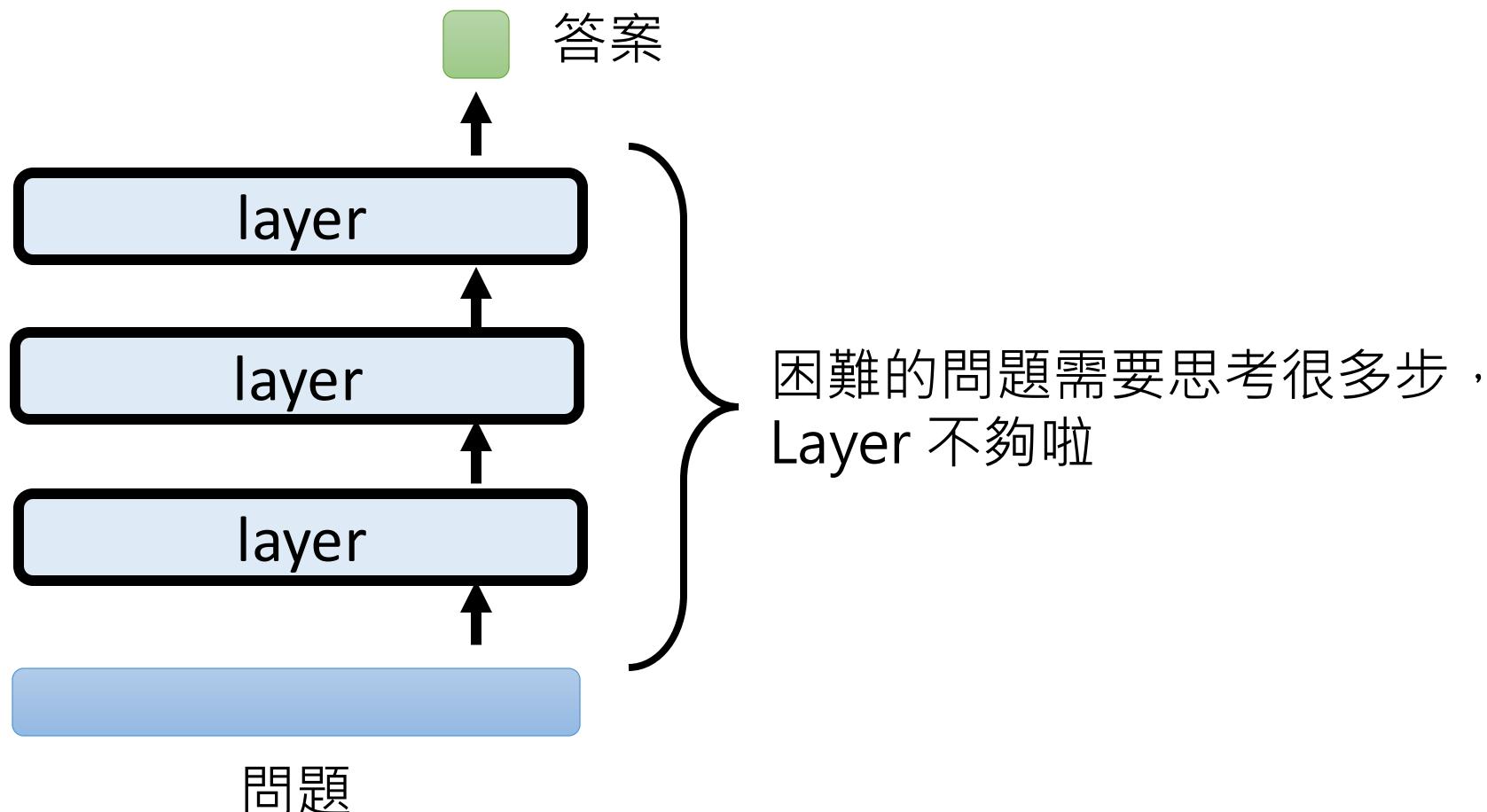


A list of video thumbnails for "Deep Learning Theory" by Hung-yi Lee. The thumbnails show various topics and their durations:

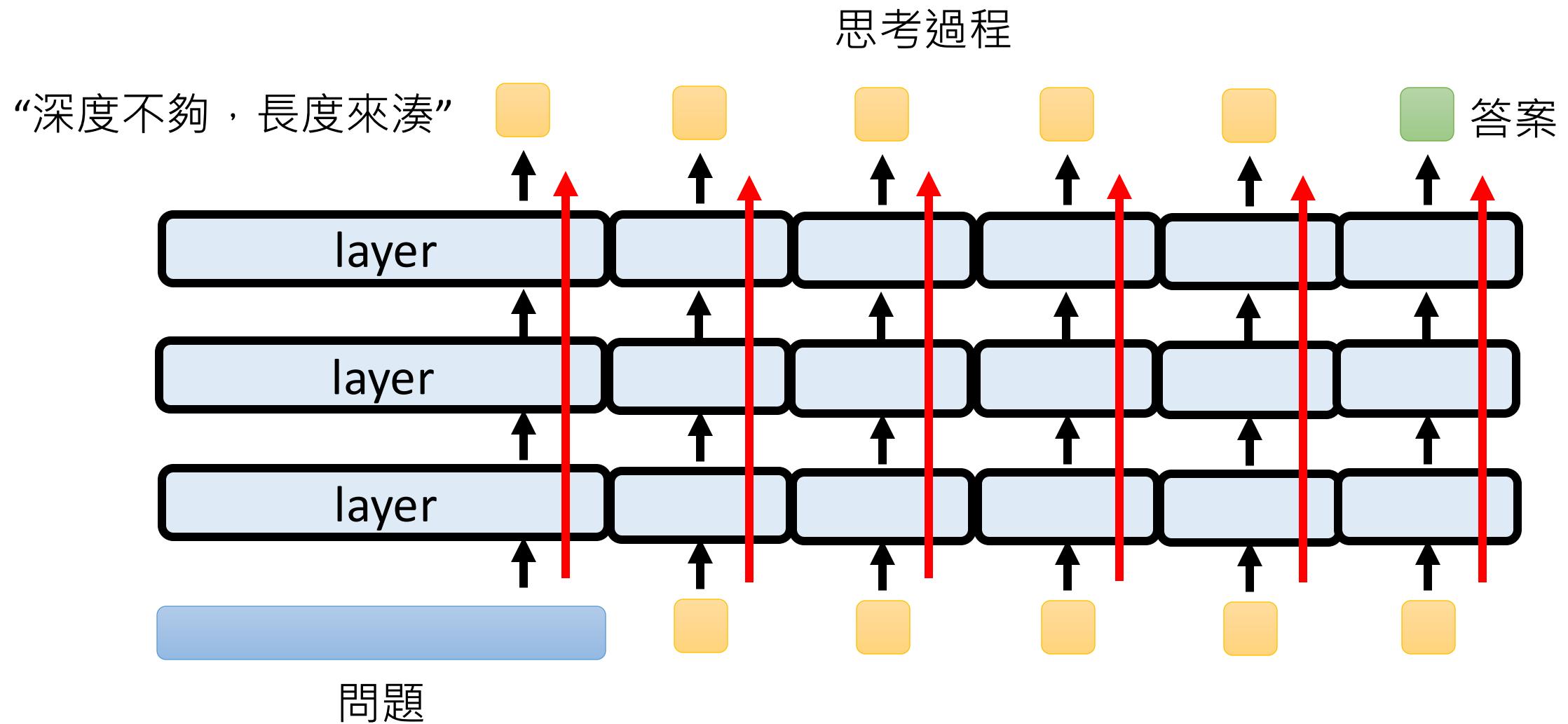
- 1 Deep Learning Theory 1-1: Can shallow network fit any... 36:49
- 2 Deep Learning Theory 1-2: Potential of Deep 56:10
- 3 Deep Learning Theory 1-3: Is Deep better than Shallow? 26:17
- 4 Deep Learning Theory 2-1: When Gradient is Zero 1:17:29
- 5 Deep Learning Theory 2-2: ...

https://www.youtube.com/watch?v=KKT2VkJFyc&list=PLJV_el3uVTsOh1F5eo9tx

為什麼讓機器「思考」會有用？

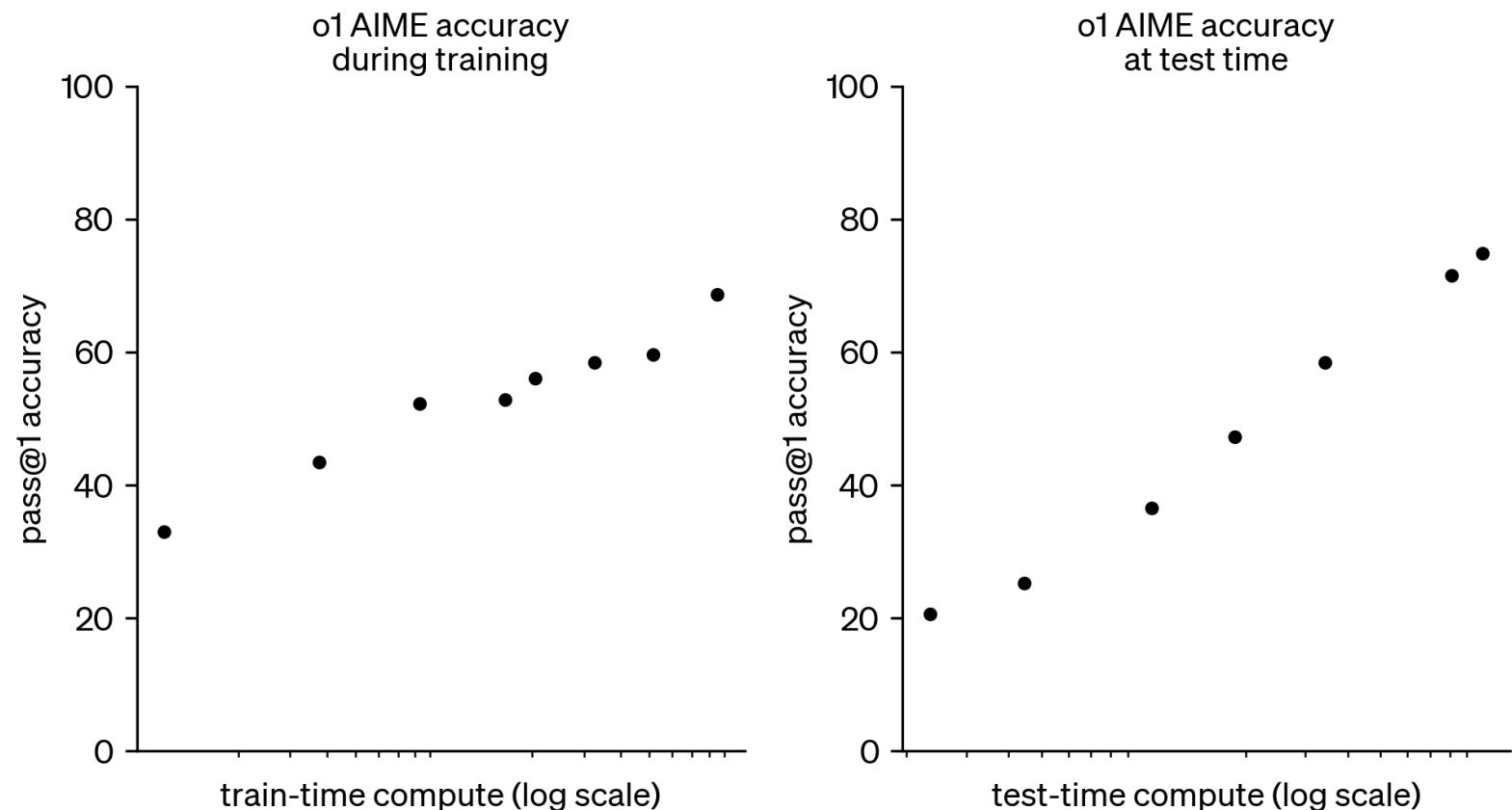


讓機器「思考」也是另外一種「深度」



讓機器「思考」也是另外一種「深度」

- “深度不夠，長度來湊” → Testing Time Scaling

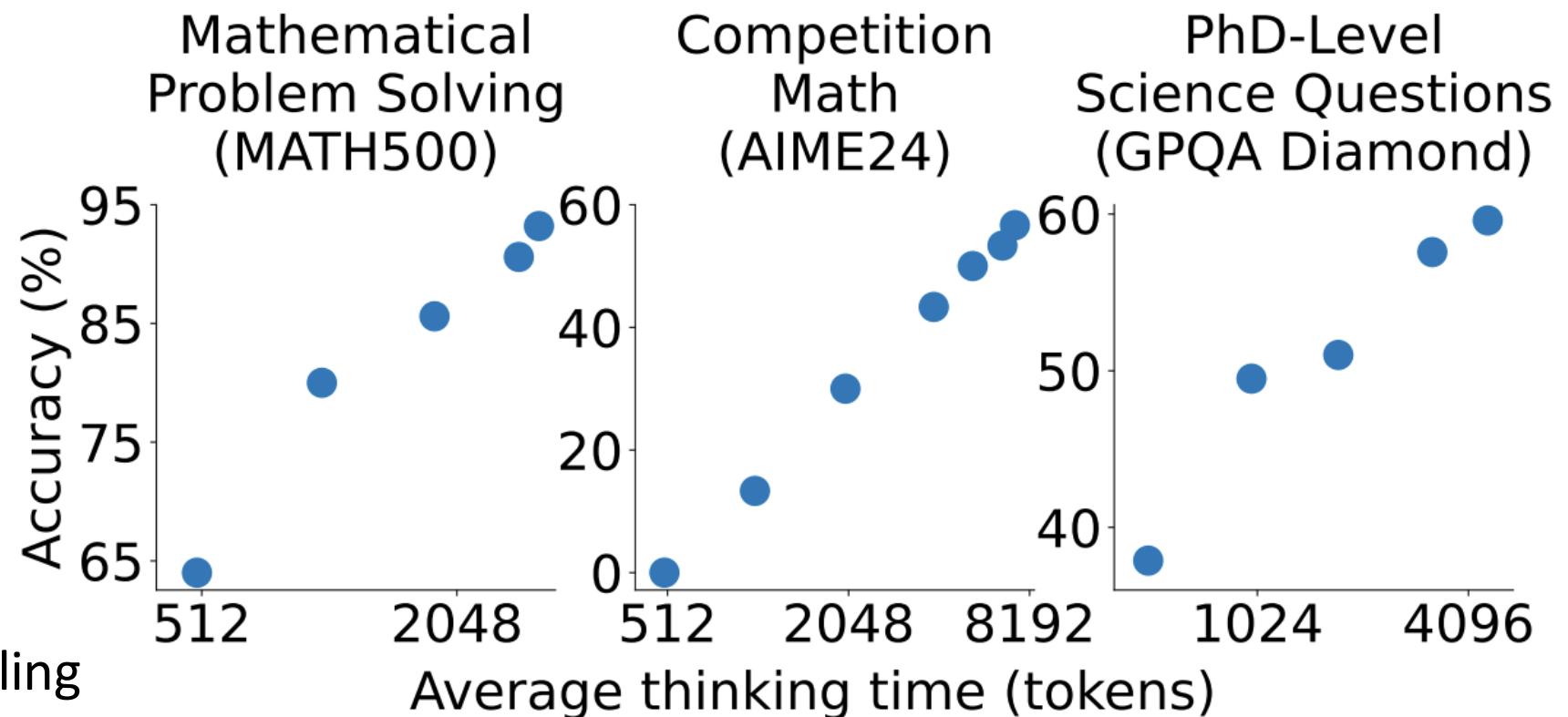


o1

<https://openai.com/index/learning-to-reason-with-langs/>

讓機器「思考」也是另外一種「深度」

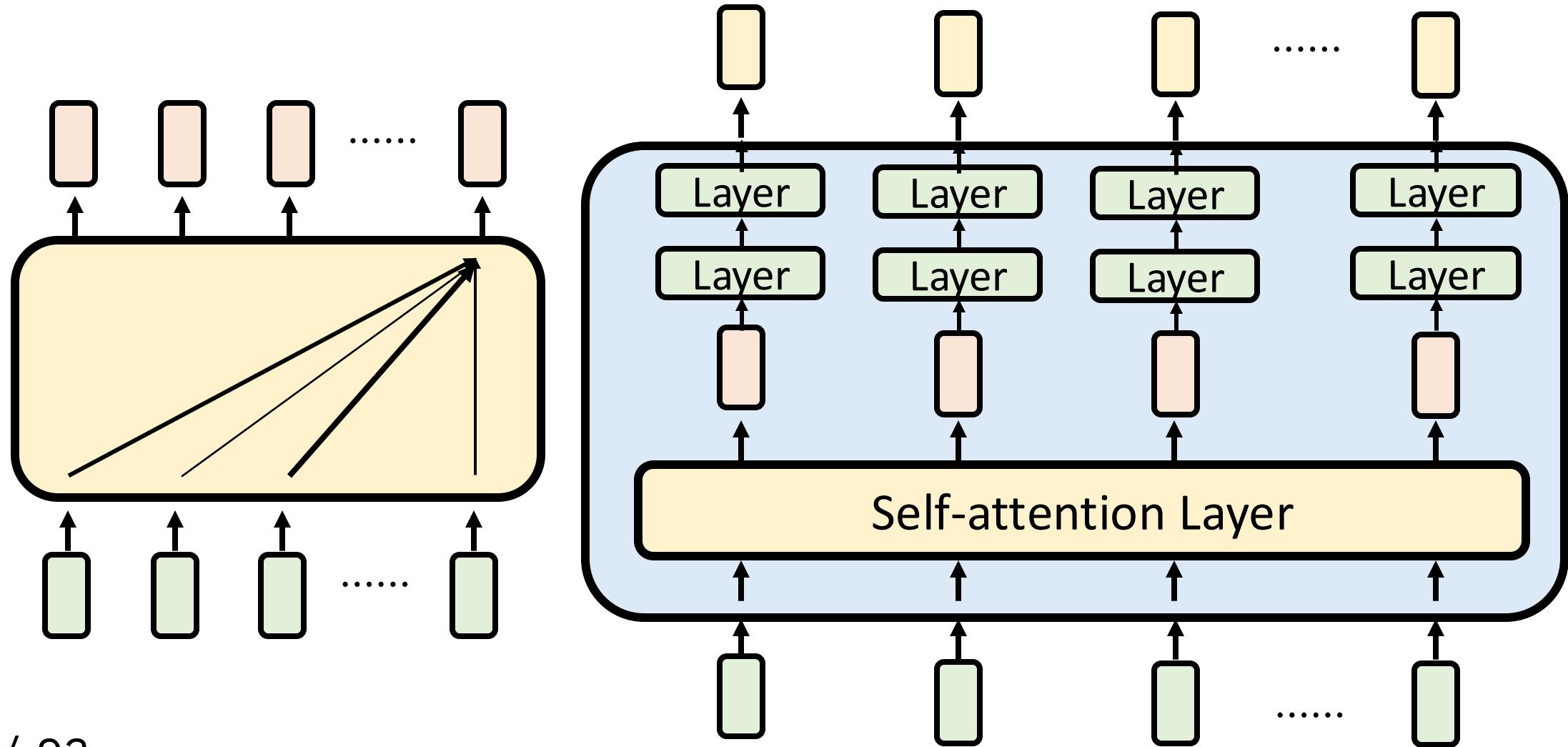
- “深度不夠，長度來湊” → Testing Time Scaling



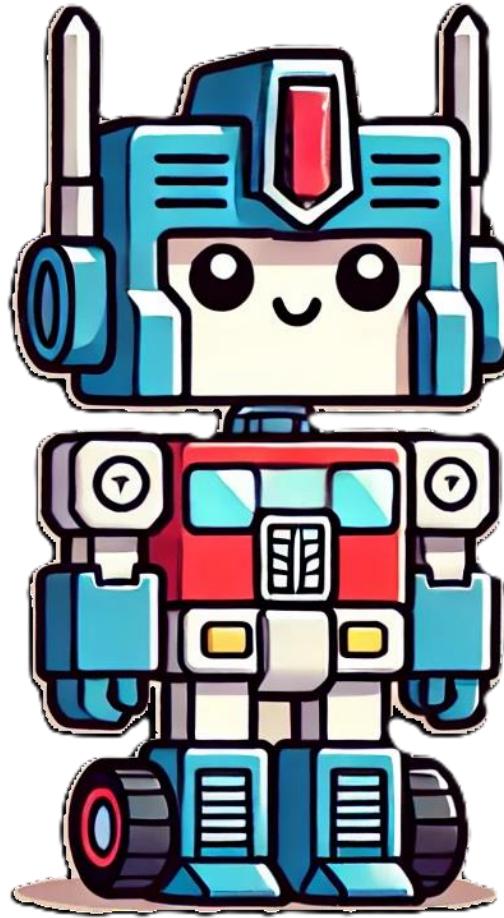
s1: Simple test-time scaling

<https://arxiv.org/abs/2501.19393>

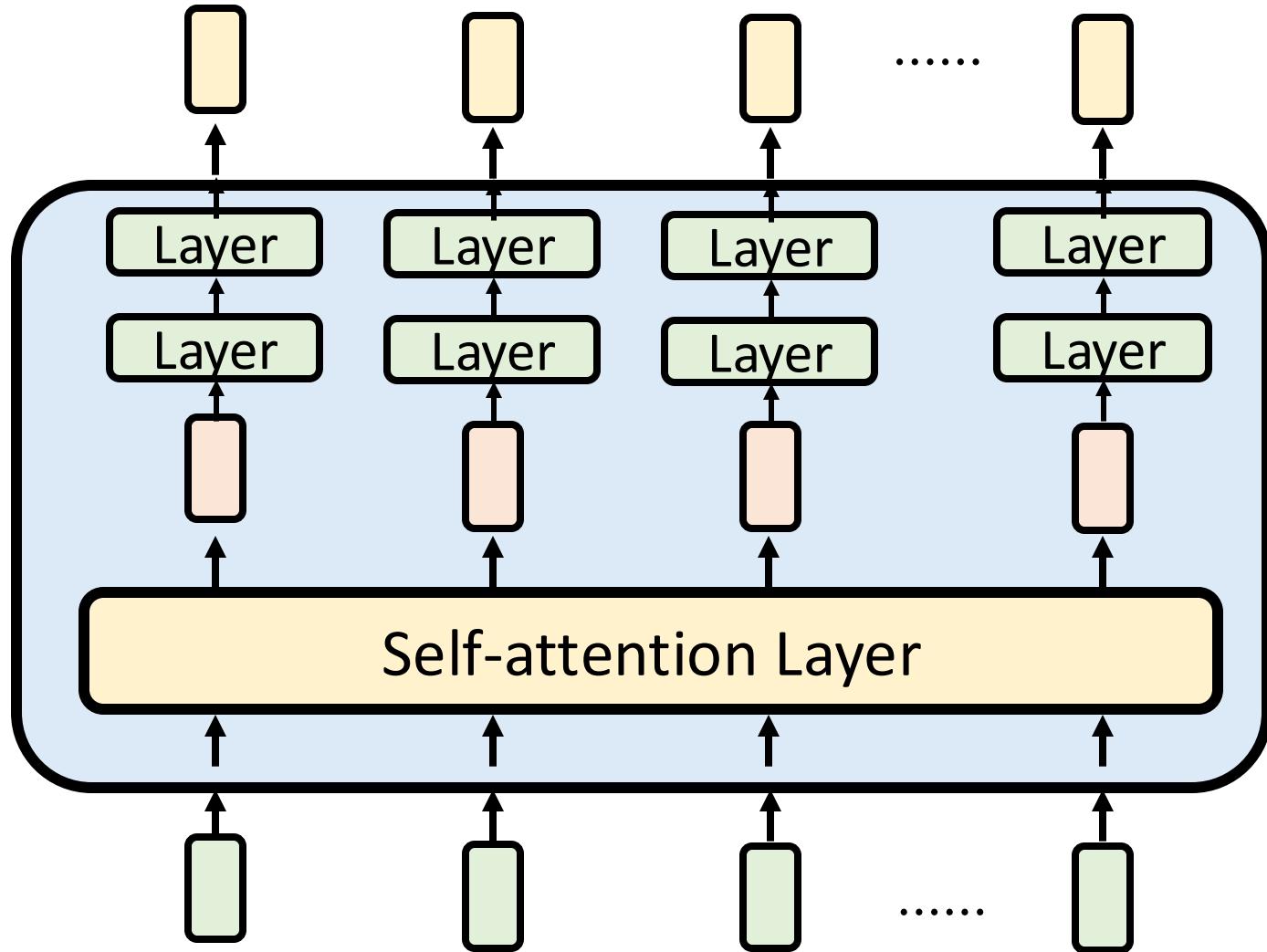
一個 Layer 中又發生了甚麼事？



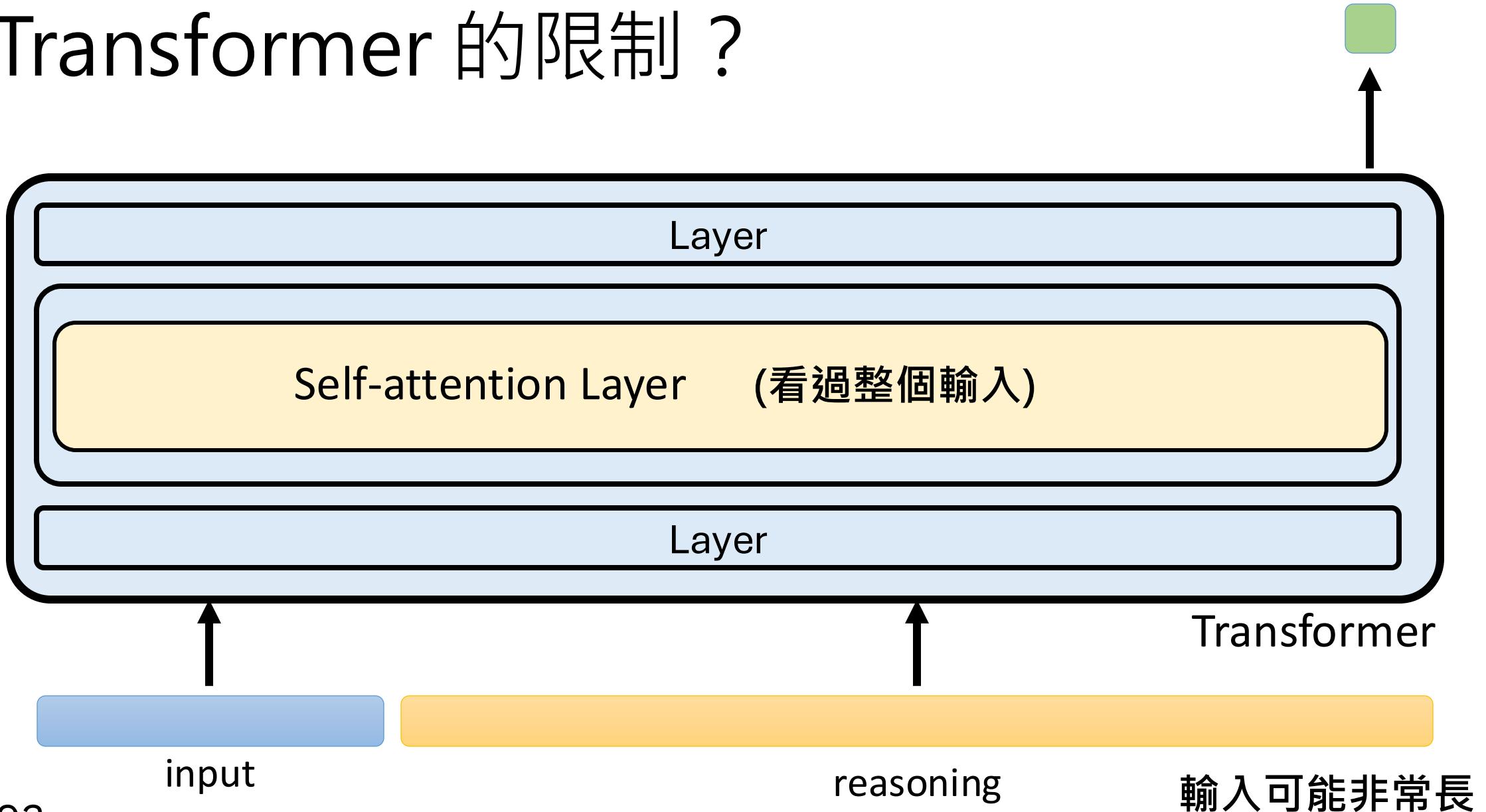
一個 Layer 中又發生了甚麼事？



Transformer

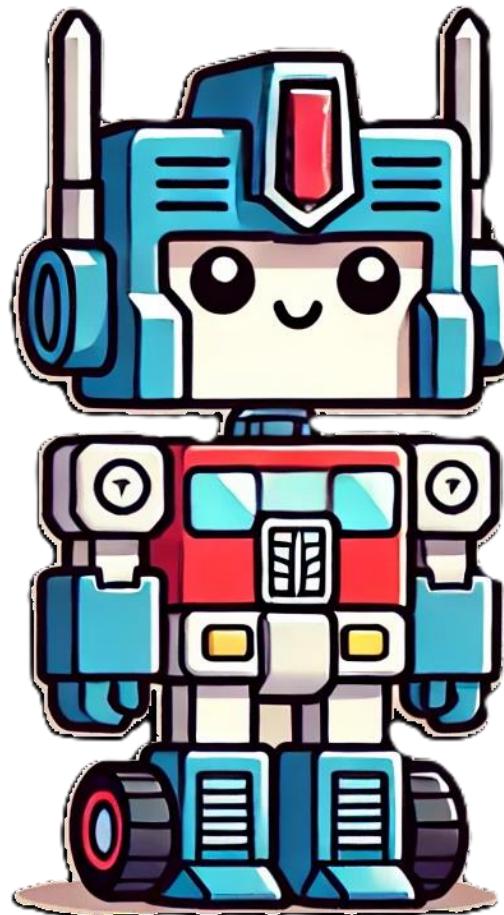


Transformer 的限制？

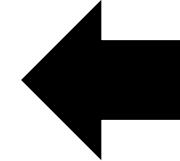
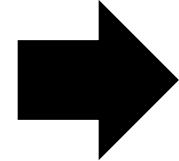


Transformer vs. Mamba (and its friends)

第四講



Transformer

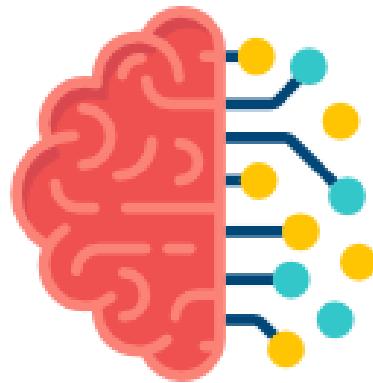
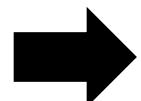


Mamba

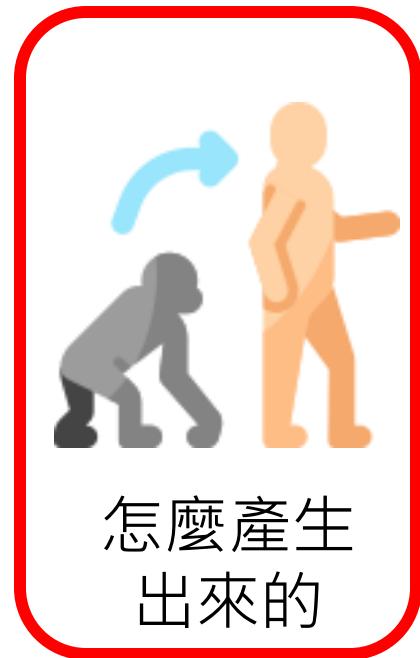
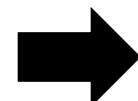
課程大綱



有什麼樣
的行為



運作機制



怎麼產生
出來的



怎麼賦予新
的能力

類神經網路：架構 vs. 參數

$$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$$

$$z_t = f(z_1, z_2, \dots, z_{t-1})$$

深度學習

實際要怎麼處理
由參數決定

$$z_t = f_L \left(\underbrace{\dots f_2 \left(f_1(z_1, z_2, \dots, z_{t-1}) \right)}_{\text{架構 (每一層處理完後傳給下一層)}} \right)$$

架構 (每一層處理完後傳給下一層)

f

架構 (architecture)

由開發者(人類)決定

天生資質

e.g., Transformer

參數 (parameter)

由訓練資料決定

後天學習

超參數 (hyperparameter)



類神經網路：架構 vs. 參數

$$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$$

$$z_t = f(z_1, z_2, \dots, z_{t-1})$$

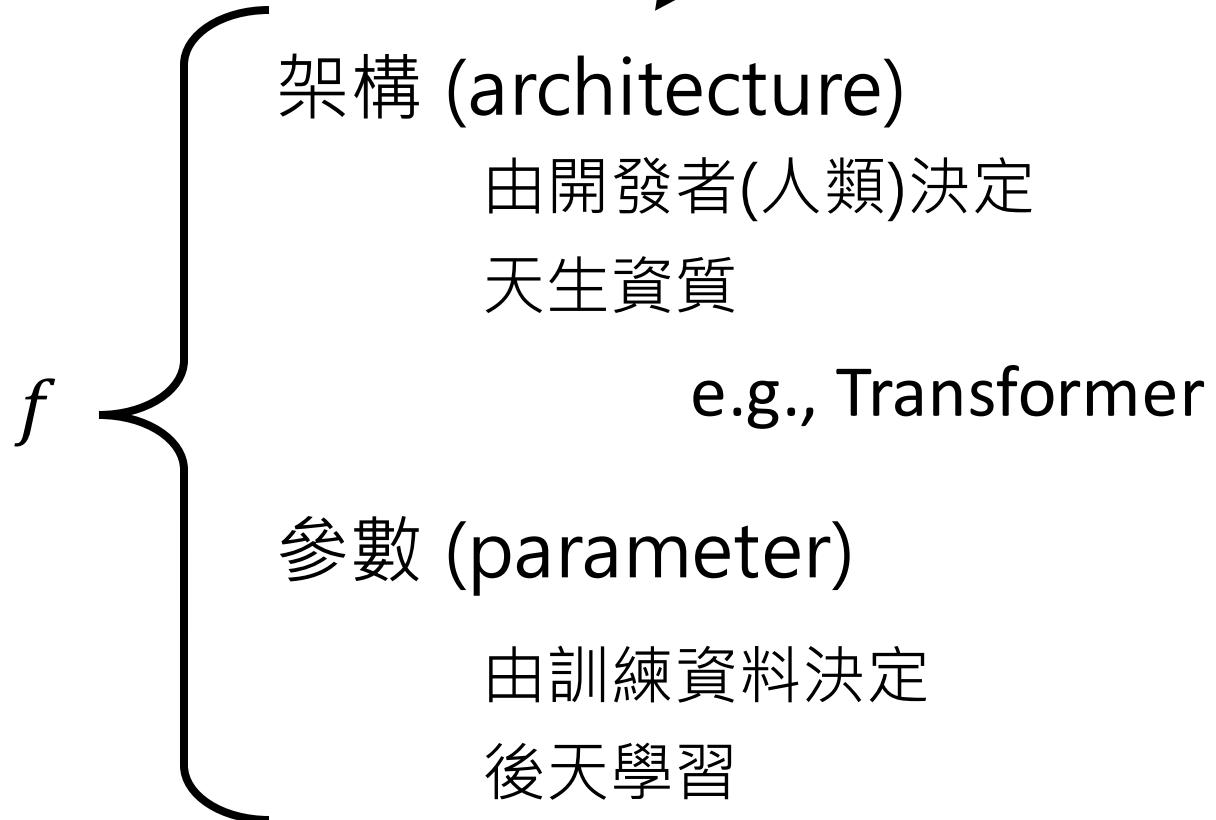
$$f_{\theta}$$

7b 模型, 70b 模型等

參數 (b: billion)

參數數量是架構的一部分

參數數值須透過訓練資料學習

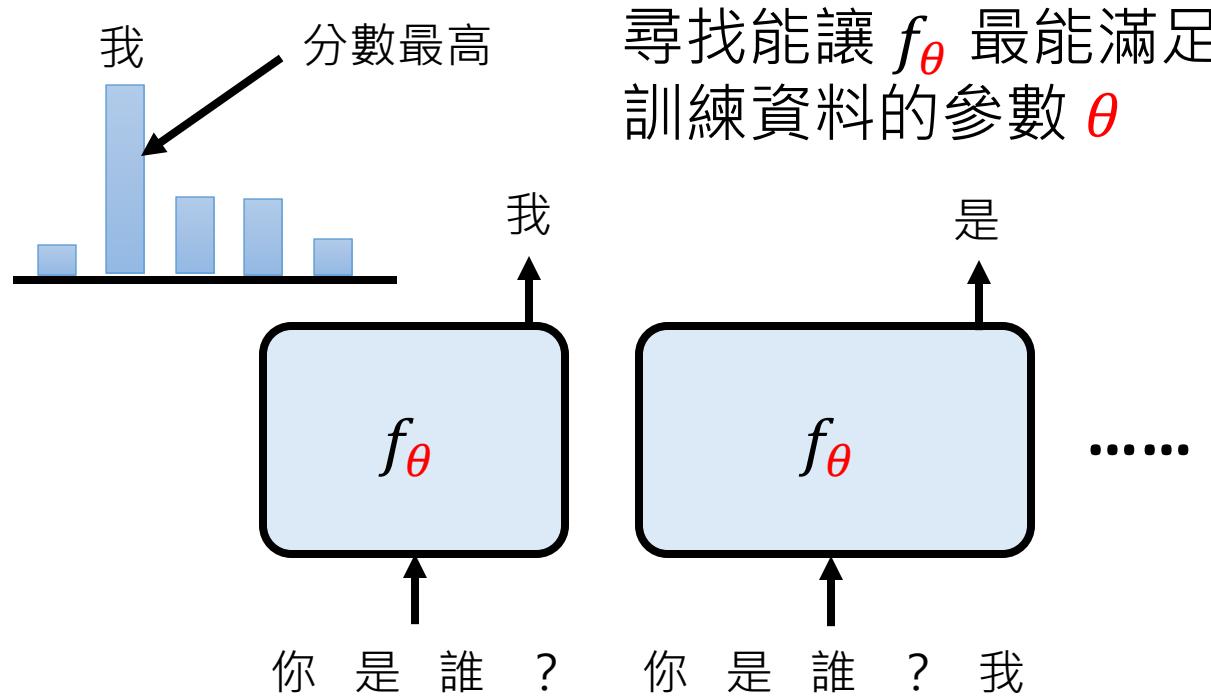


訓練資料

找出參數 (訓練模型)

$$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$$

$$z_t = f_{\theta}(z_1, z_2, \dots, z_{t-1})$$



輸入	輸出
你是誰？	我
你是誰？ 我	是
你是誰？ 我是	人
你是誰？ 我是人	工
.....	
你是誰？ 我是人工智能	[END]
輸出 hello world 的程式碼	print
輸出 hello world 的程式碼 print((
輸出 hello world 的程式碼 print("	"
.....	
(數學問題)	令
(數學問題) 令	x
(數學問題) 令 x	=
.....

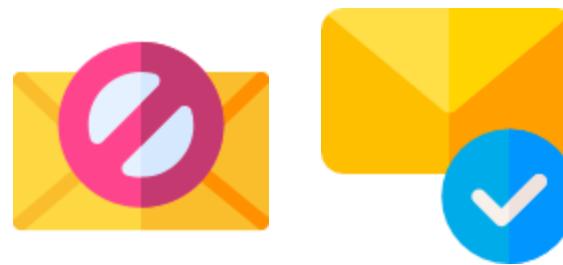
機器學習中的分類(Classification)問題

$\{z_1, z_2, \dots, z_{t-1}\} \rightarrow z_t$ 有限可能(選擇題)

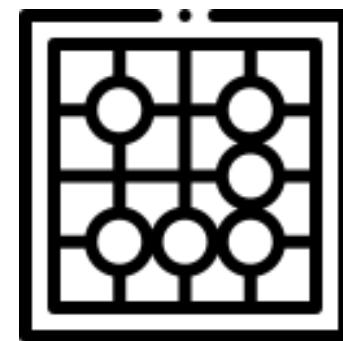
- 以下也是機器學習中的分類問題



信用卡盜刷偵測



垃圾郵件偵測



下圍棋 (AlphaGo)

生成式人工智慧也不是新的問題 ...

專才

這堂課我們要講如何駕馭
大型語言模型



This course
is about

翻譯系統

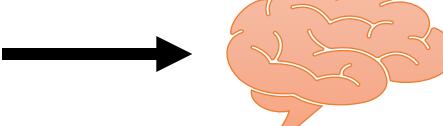
只要說清楚你要幹嘛 (Prompt)



對以下文句做**翻譯**：

通才

這堂課我們要講如何駕馭
大型語言模型

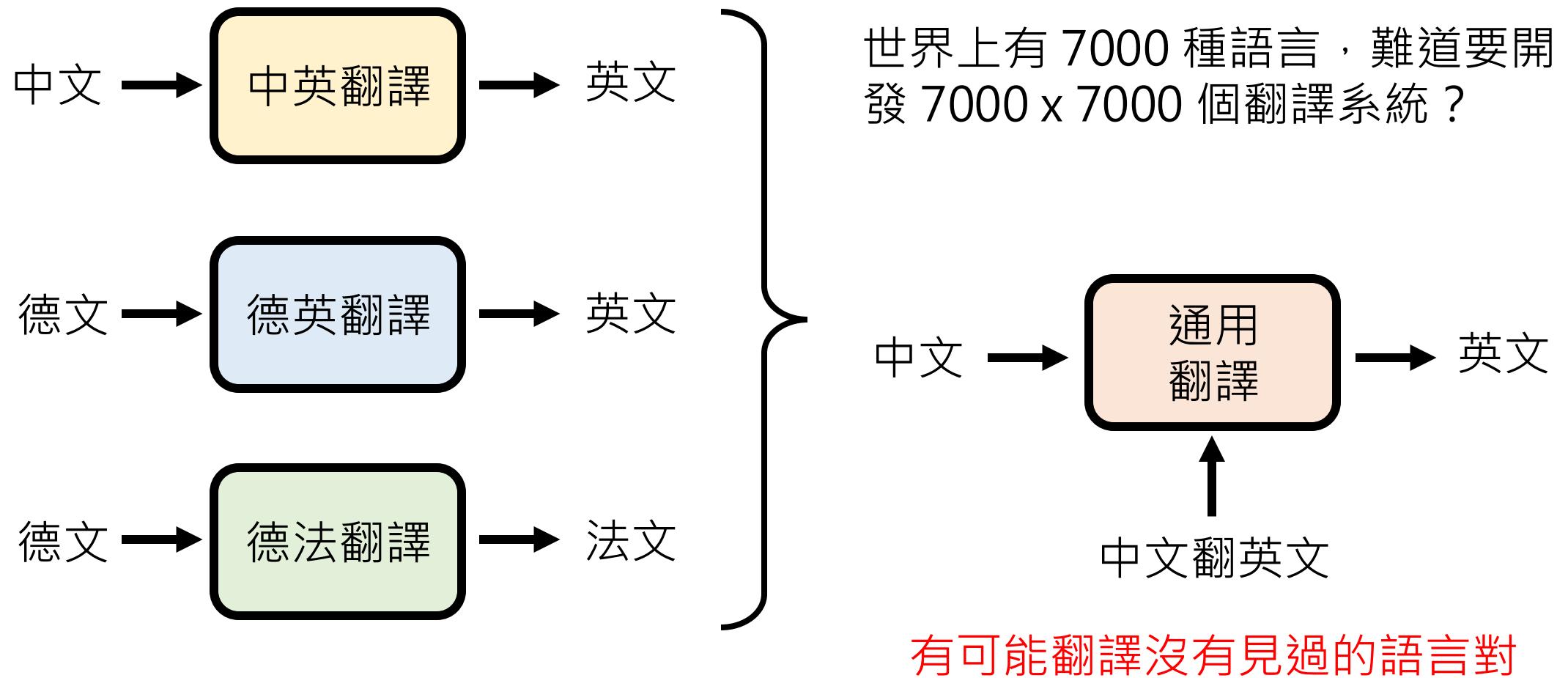


This course
is about

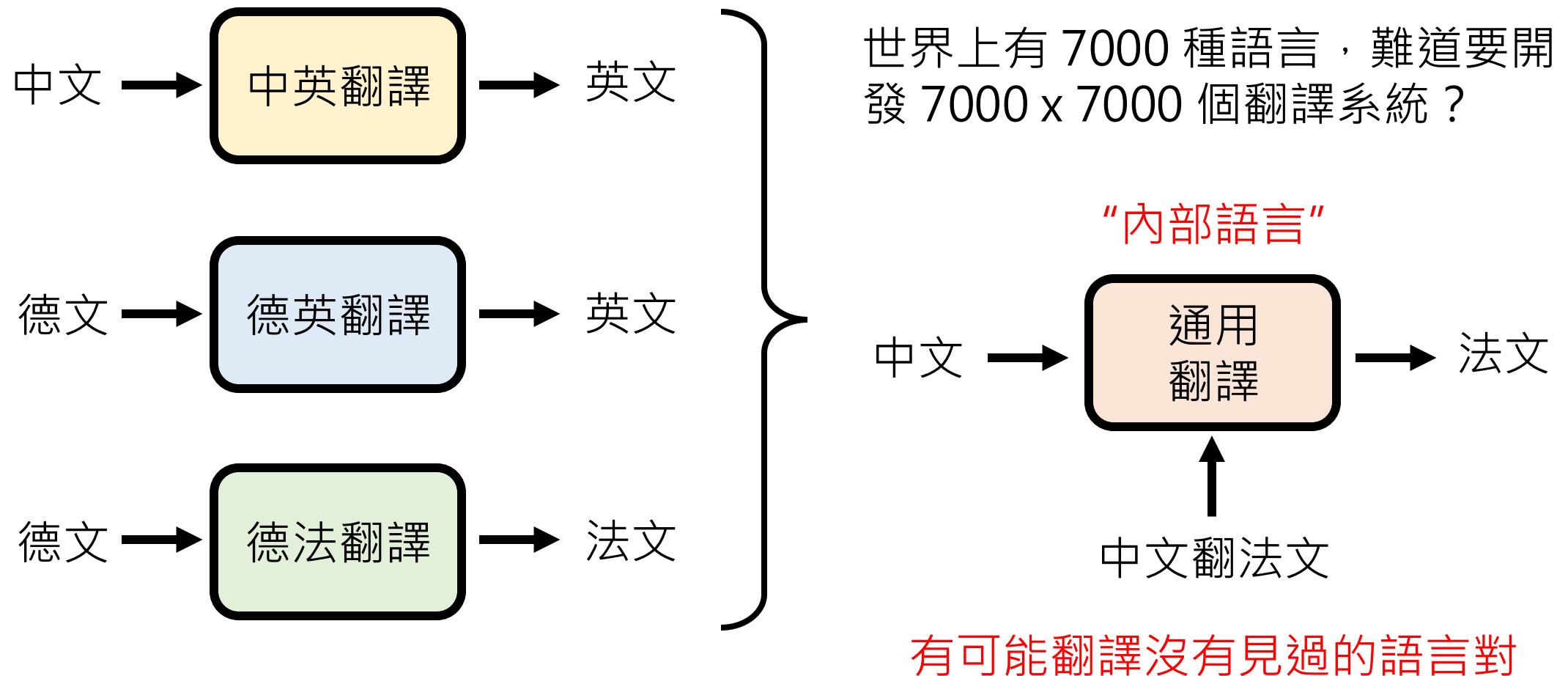
ChatGPT

這些通用模型是怎麼被發展起來的？

不同語言可以共用模型嗎？

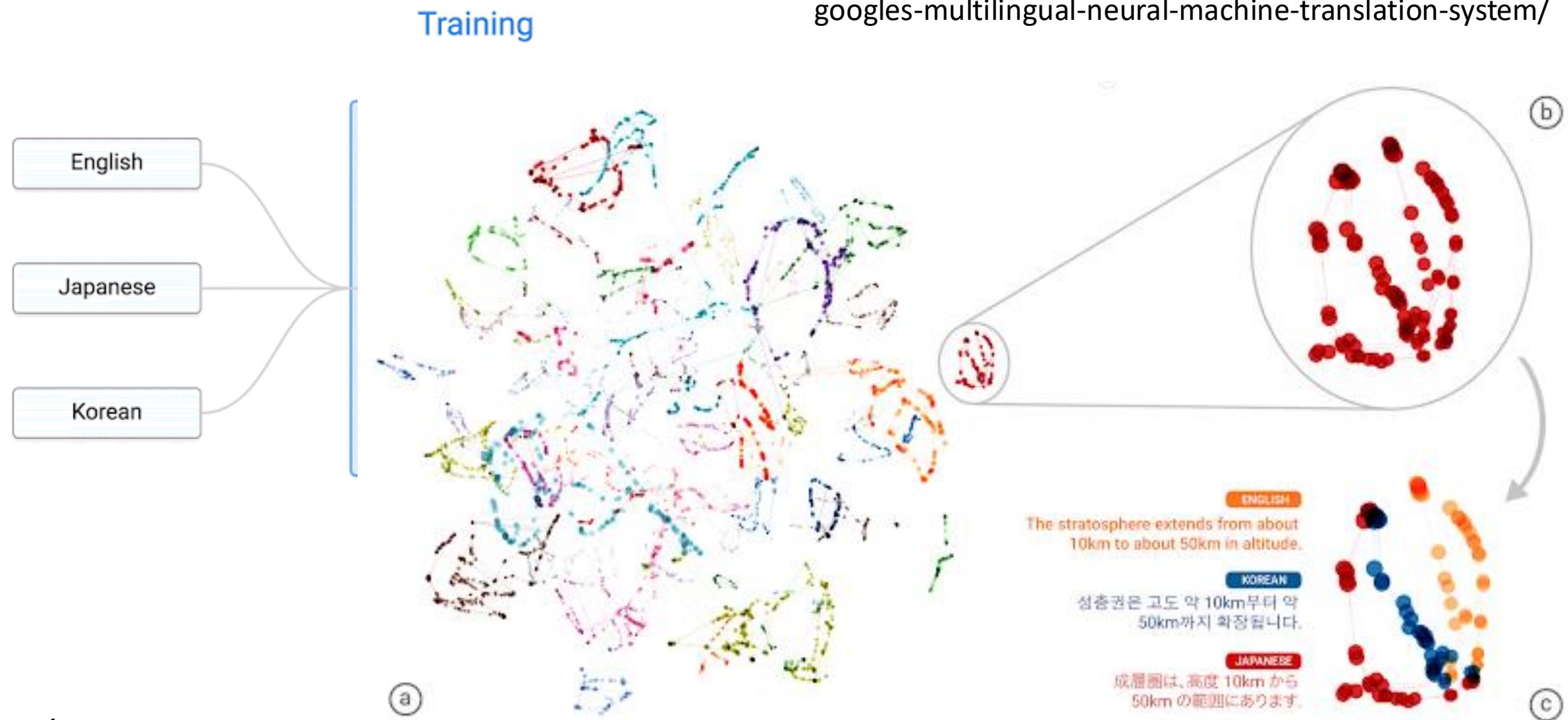


不同語言可以共用模型嗎？

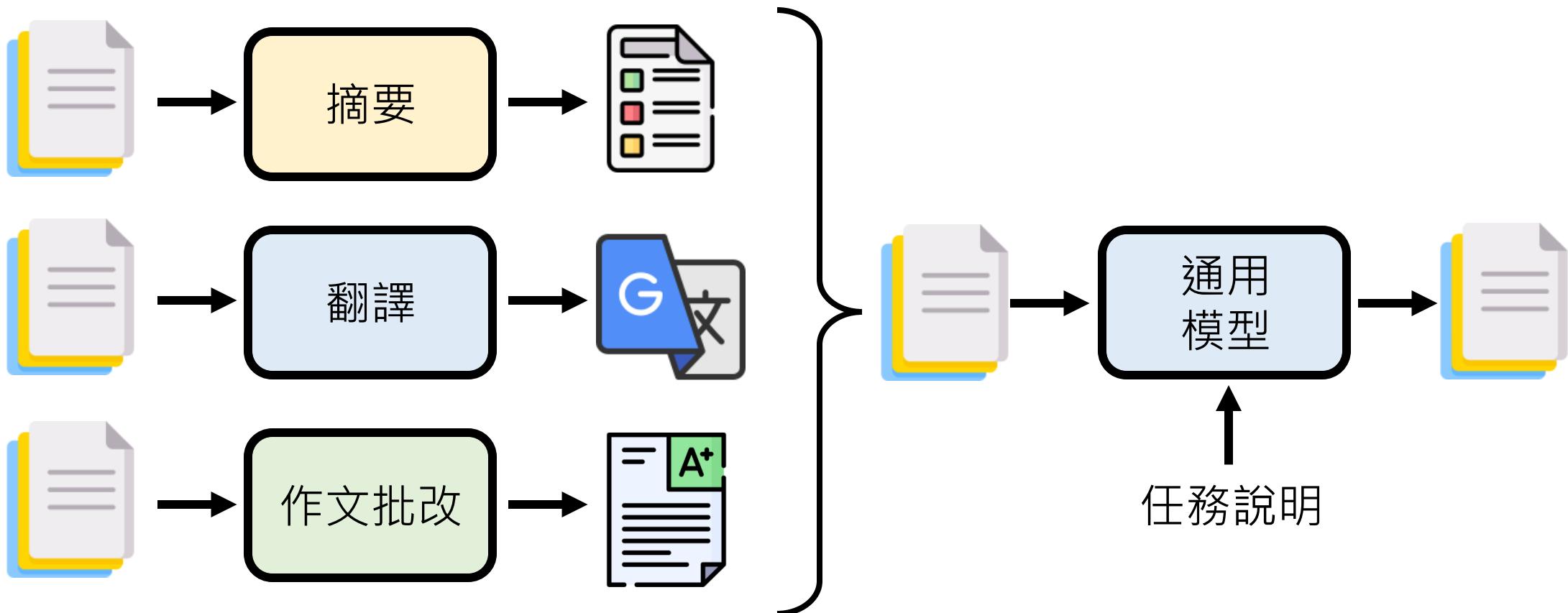


Zero-Shot Translation with Google's Multilingual Neural Machine Translation System (2016)

<https://research.google/blog/zero-shot-translation-with-googles-multilingual-neural-machine-translation-system/>



不同任務可以共用模型嗎？



都是文字輸入、文字輸出

不同任務可以共用模型嗎？

Question	Context	Answer	Question	Context	Answer
What is a major importance of Southern California in relation to California and the US?	...Southern California is a major economic center for the state of California and the US....	major economic center	What has something experienced?	Areas of the Baltic that have experienced eutrophication .	eutrophication
What is the translation from English to German?	Most of the planet is ocean water.	Der Großteil der Erde ist Meerwasser	Who is the illustrator of Cycle of the Werewolf?	Cycle of the Werewolf is a short novel by Stephen King, featuring illustrations by comic book artist Bernie Wrightson .	Bernie Wrightson
What is the summary?	Harry Potter star Daniel Radcliffe gains access to a reported £320 million fortune...	Harry Potter star Daniel Radcliffe gets £320M fortune...	What is the change in dialogue state?	Are there any Eritrean restaurants in town?	food: Eritrean
Hypothesis: Product and geography are what make cream skimming work. Entailment , neutral, or contradiction?	Premise: Conceptually cream skimming has two basic dimensions – product and geography.	Entailment	What is the translation from English to SQL?	The table has column names... Tell me what the notes are for South Australia	SELECT notes from table WHERE 'Current Slogan' = 'South Australia'
Is this sentence positive or negative?	A stirring, funny and finally transporting re-imagining of Beauty and the Beast and 1930s horror film.	positive	Who had given help? Susan or Joan ?	Joan made sure to thank Susan for all the help she had given.	Susan

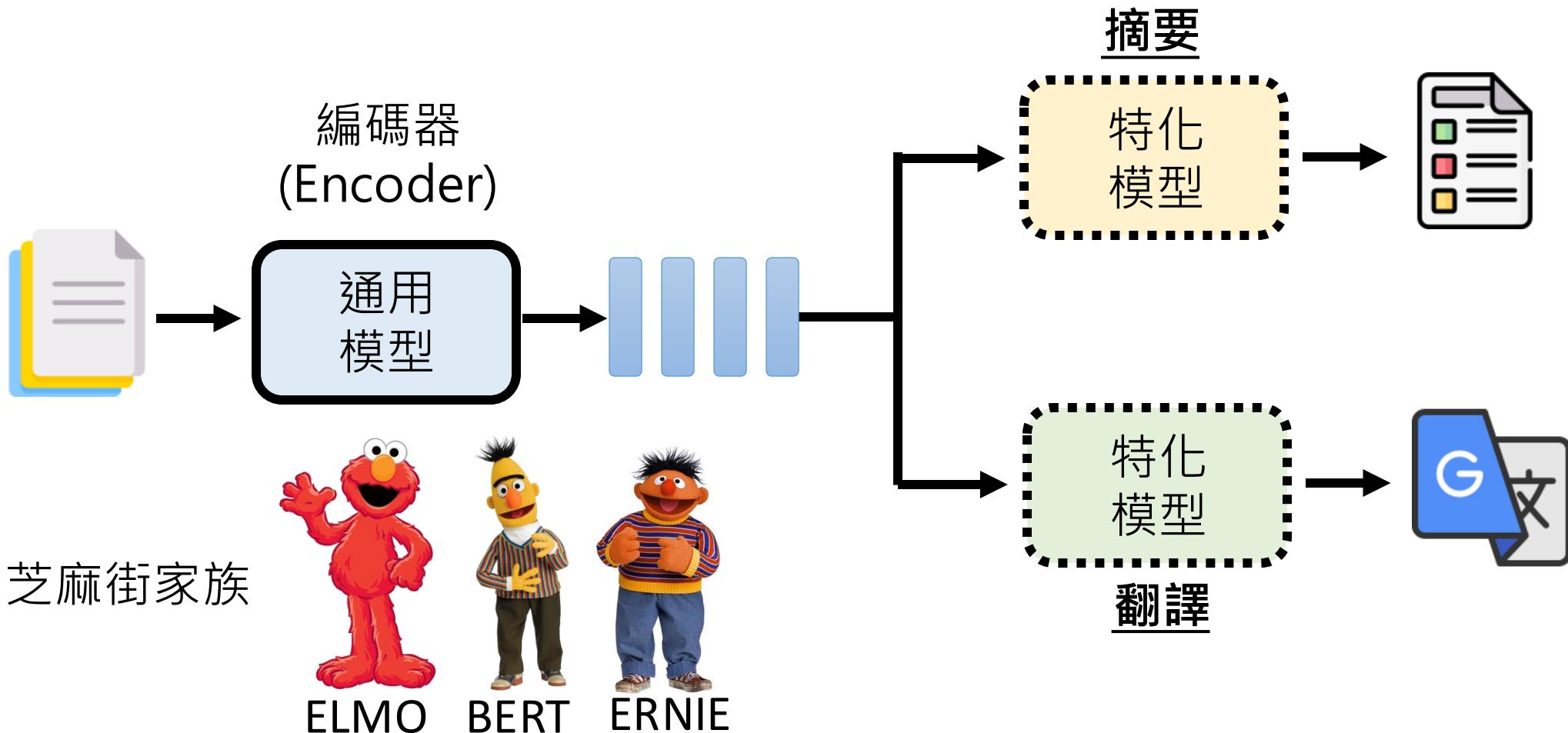
The Natural Language Decathlon: Multitask Learning as Question Answering

<https://arxiv.org/abs/1806.08730>

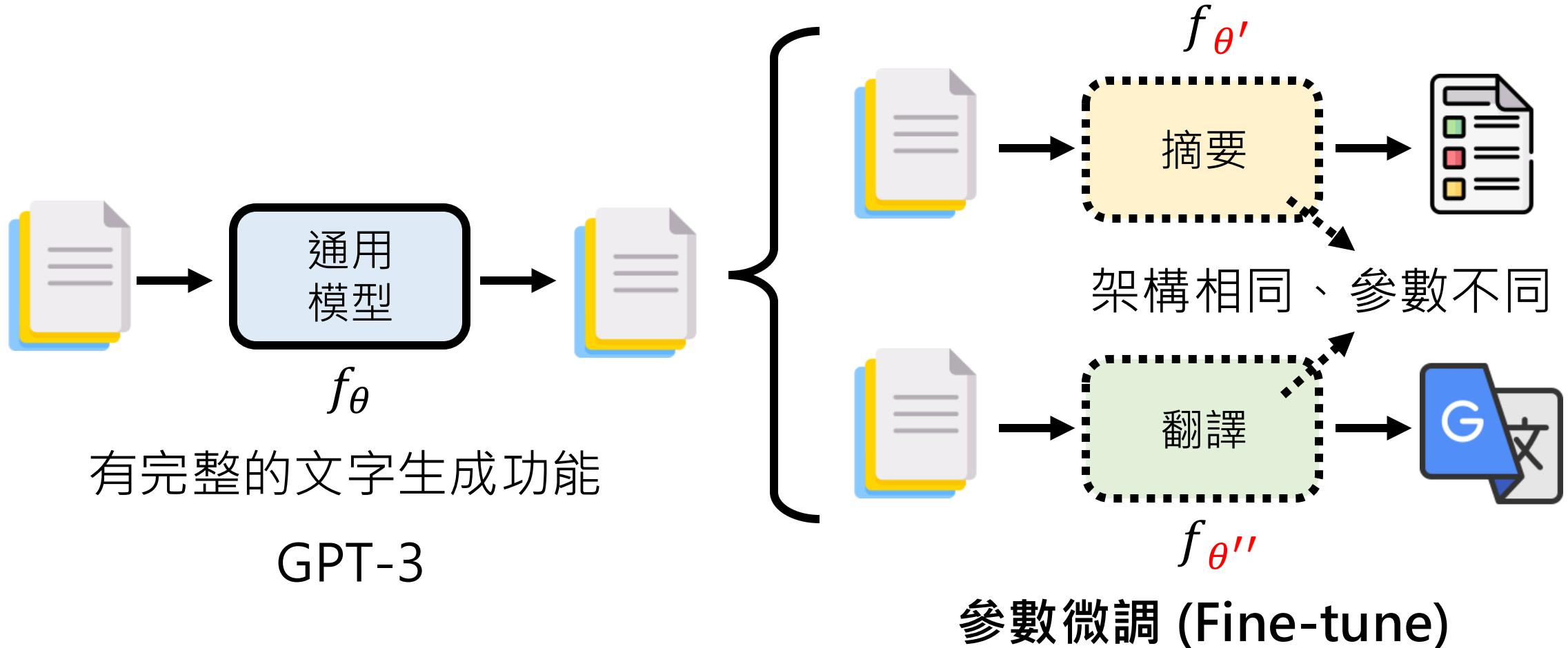
接下來有了通用模型的發展

「通用」和「通用」意思不一樣

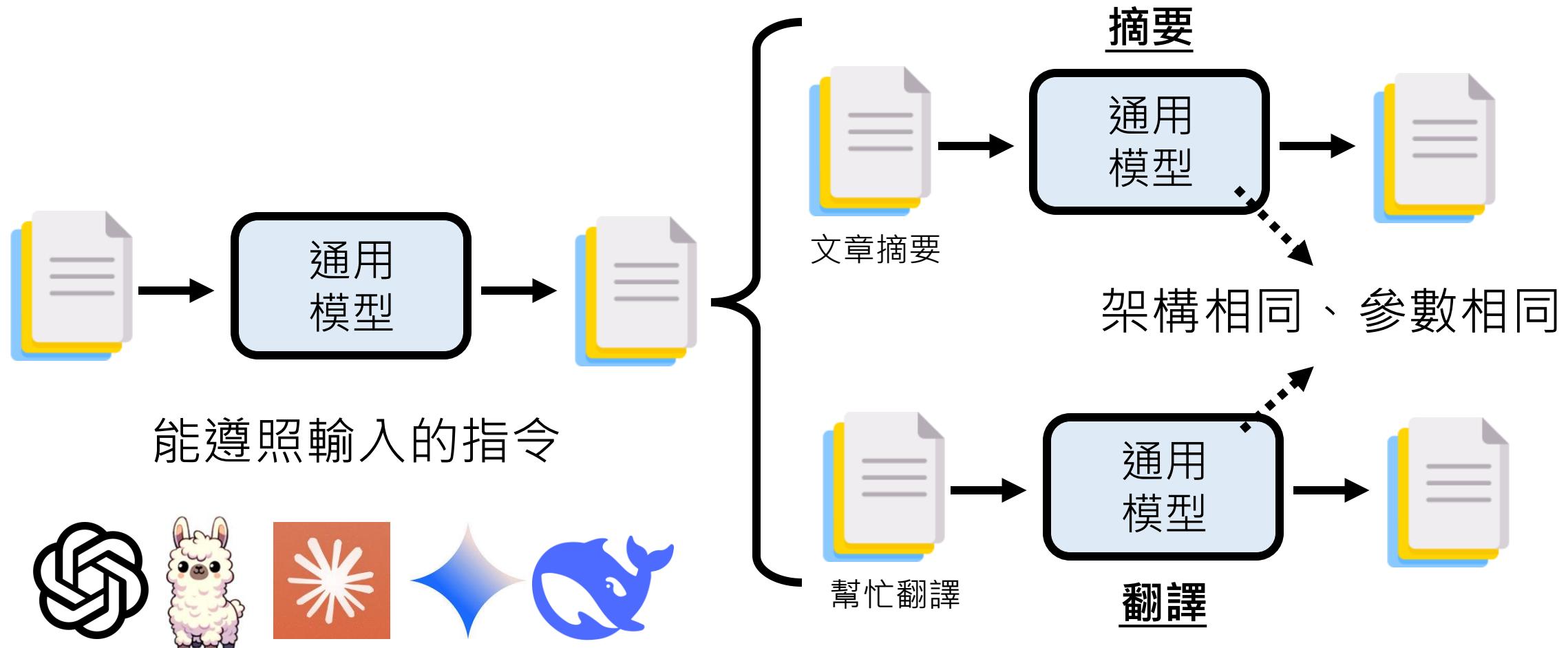
通用機器學習模型：第一型態 (2018 - 2019)



通用機器學習模型：第二型態 (2020 - 2022)



通用機器學習模型：第三型態 (2023 -)



如何打造這樣的模型就是另一個故事了 ...

Pre-train

用大量沒有
標註的資料



大型語言模型修練史

<https://youtu.be/cCpErV7To2o?si=lfslfaV7PwYqWNFg>

Fine-tune

用少量有標註
的資料微調



<https://youtu.be/Q9cNkUPXUB8?si=qj573p9OhI74qYk5>

Alignment

RLHF

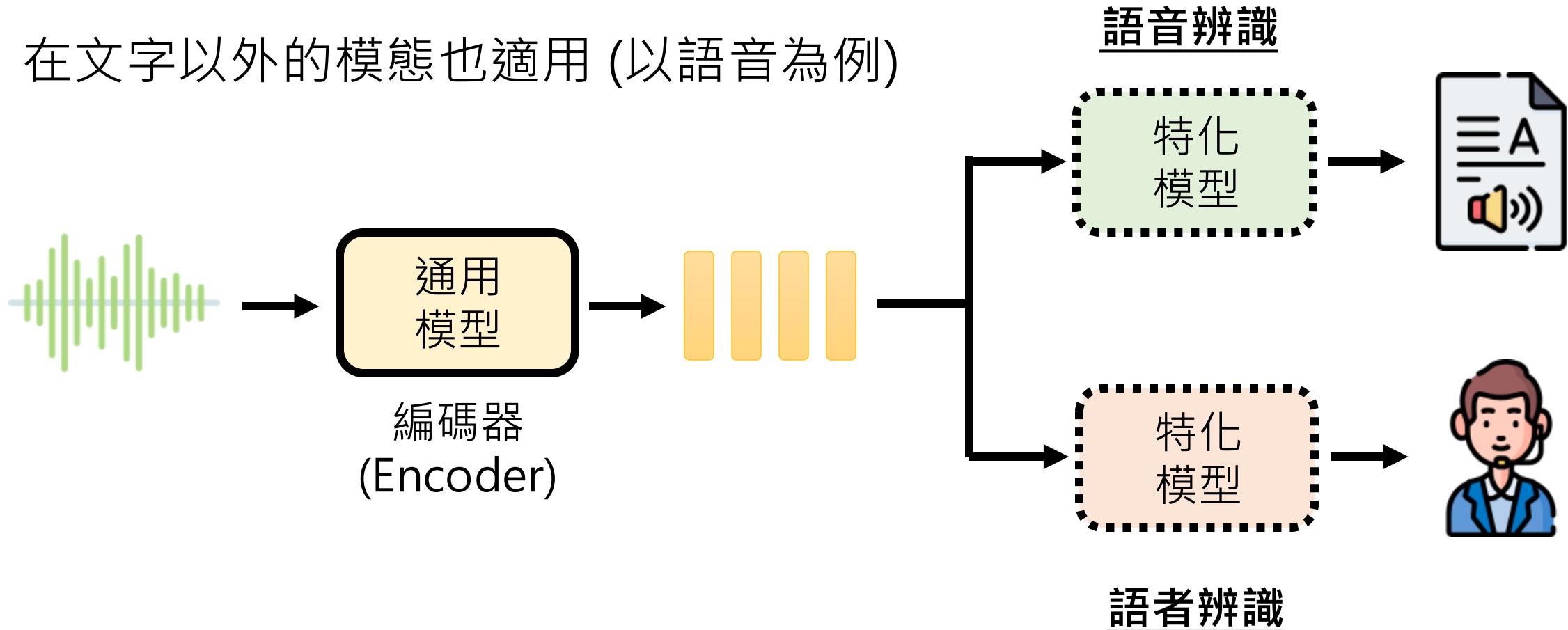
跟使用者回饋
微調



<https://youtu.be/v12IKvF6Cj8?si=hqaXTn1A5iSjy8lg>

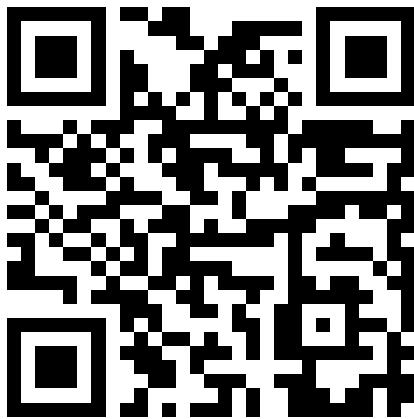
通用機器學習模型：第一型態

在文字以外的模態也適用 (以語音為例)



Useful Toolkit!

The S3PRL toolkit



<https://github.com/s3prl/s3prl>

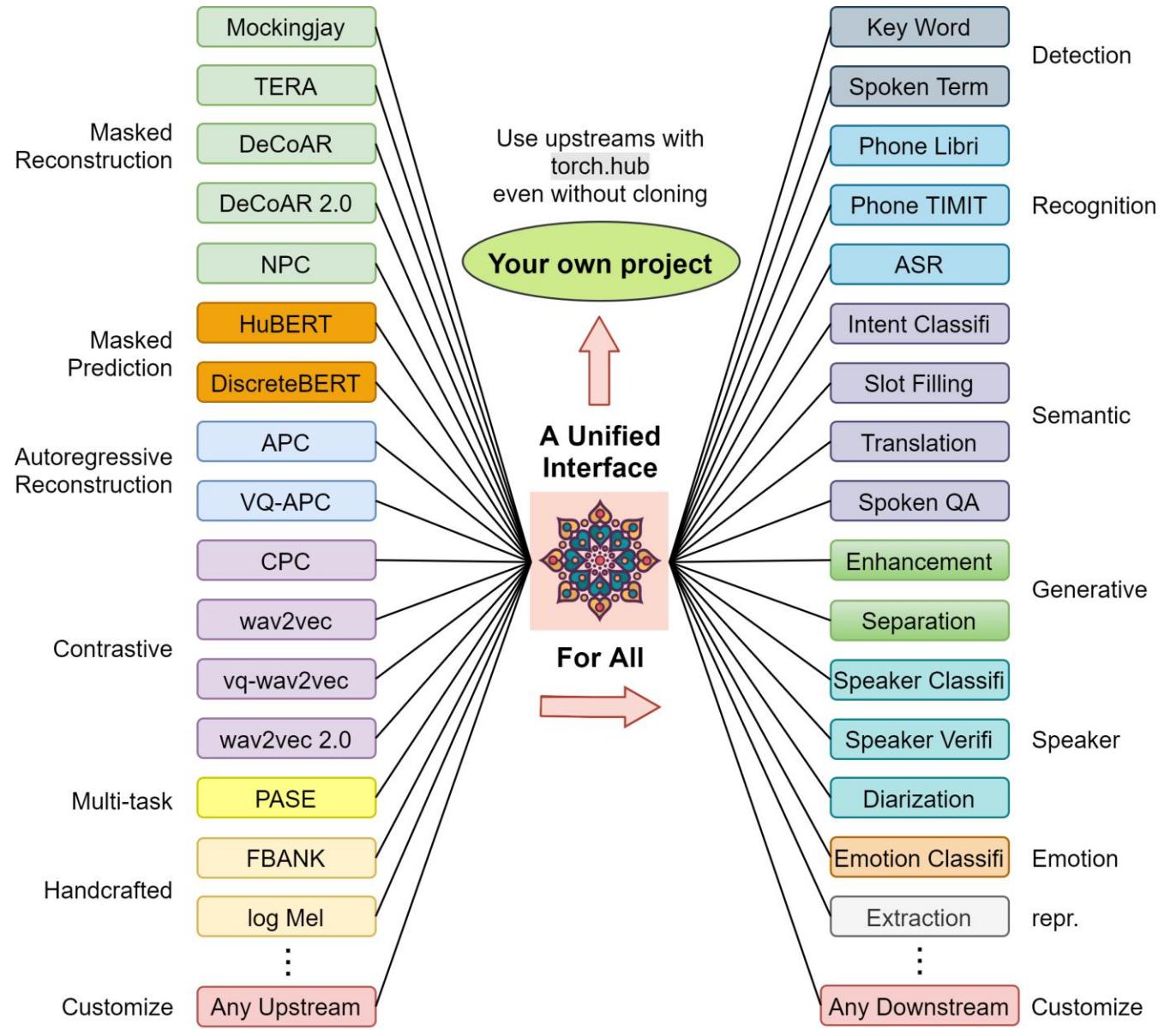
Creator



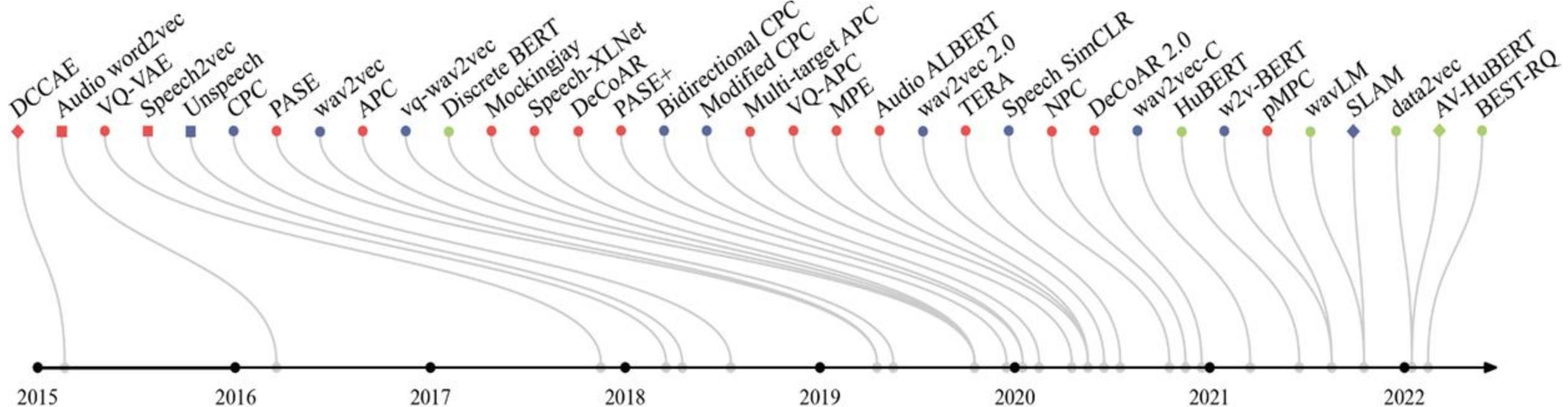
Shu-wen
(Leo) Yang



Andy T. Liu



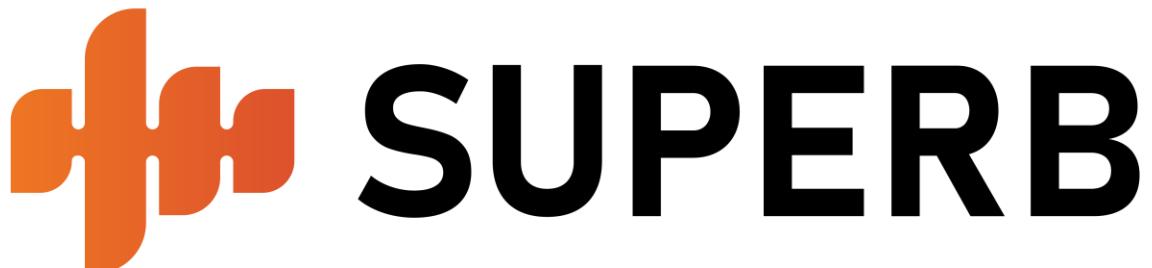
Over 2k stars & used by over 48 repos



Shu-wen
(Leo) Yang

- Generative models ● Contrastive models ● Predictive models
- Embedding models ◆ Multi-modal models

<https://arxiv.org/abs/2205.10643>



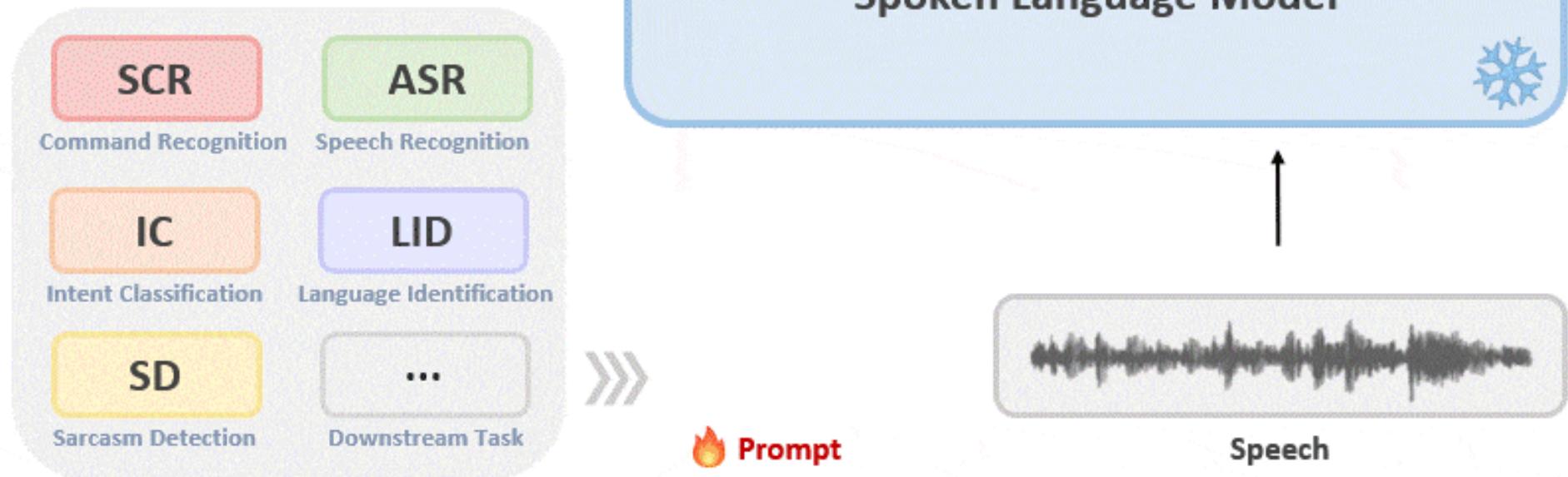
**Speech processing Universal
PERformance Benchmark**

通用機器學習模型：第二型態

<https://ga642381.github.io/SpeechPrompt/>



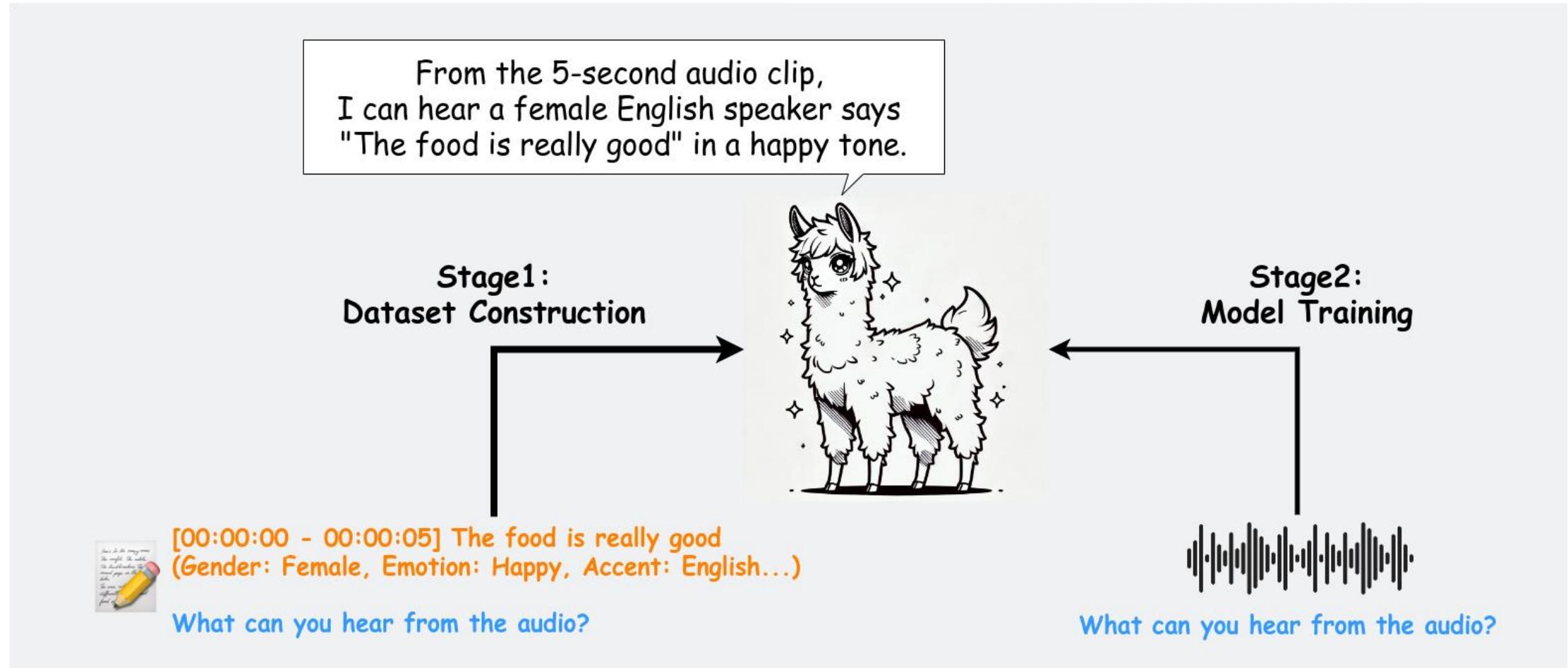
Kai-Wei Chang



通用機器學習模型：第三型態



Ke-Han Lu and
NVIDIA researchers



DeSTA2 demo page



have fun!

DeSTA2



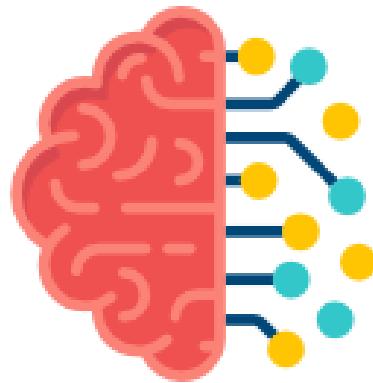
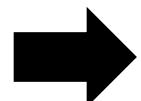
Speech:

/tmp/gradio/1cbb3326ce027b13dfb42a744eaf1afcd83123808e198527b3cd16149f43615b/Ses
01F_script03_1_F029.wav

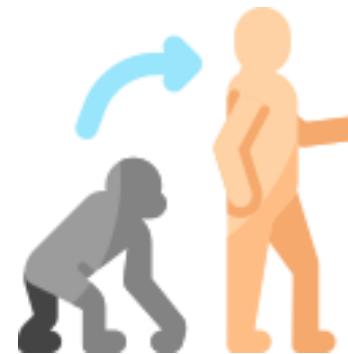
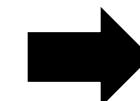
課程大綱



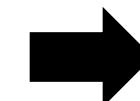
有什麼樣
的行為



運作機制



怎麼產生
出來的



怎麼賦予新
的能力

機器的終身學習(Life-long Learning)時代

如果你需要機器具備某種能力



以前：從零培養



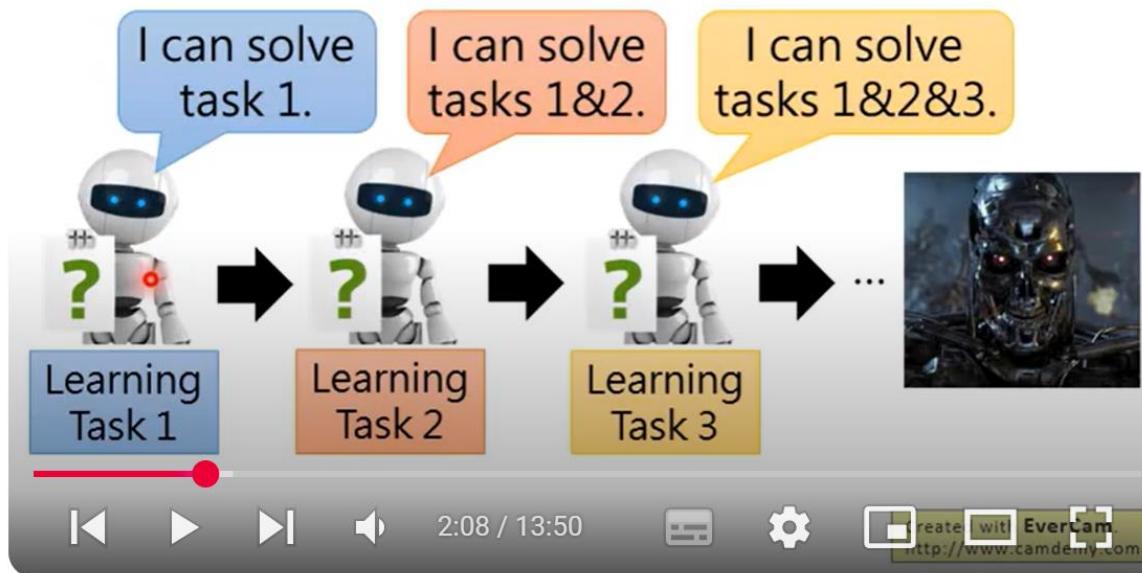
現在：已具基本能力

機器的終身學習(Life-long Learning)時代

2019 機器學習

Life Long Learning (LLL)

Continuous Learning, Never Ending Learning, Incremental Learning



Next Step of Machine Learning (Hun...)

Hung-yi Lee - 25/61

Life Long Learning (1/7)

Hung-yi Lee

13:51

Elastic Weight Consolidation (EWC)

Basic idea: Some parameters in the model are important to the previous tasks. Only change the unimportant parameters. θ^* is the final learned parameter.

Task parameter θ_j^* has a "target" $\hat{\theta}_j$

One kind of regularization. θ_j should be close to $\hat{\theta}_j$ in certain directions.

$L(\theta) = L(\theta) + \lambda \sum_j \| \theta_j - \hat{\theta}_j \|_2^2$

If $\theta_j = \hat{\theta}_j$, there is no constraint on θ_j .

26

Life Long Learning (2/7)

Hung-yi Lee

7:25

27

Life Long Learning (3/7)

Hung-yi Lee

12:04

Elastic Weight Consolidation (EWC)

Task 1

Task 2

Conducting multi-task learning by generating pseudo-data using generative model

28

Life Long Learning (4/7)

Hung-yi Lee

4:39

Wait a minute.....

Train a model for each task

29

Life Long Learning (5/7)

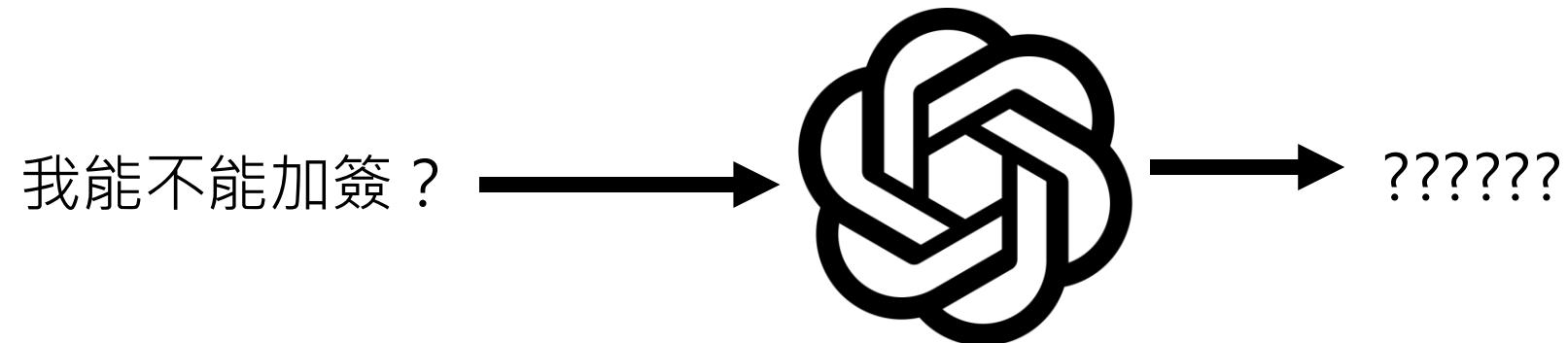
Hung-yi Lee

3:19

機器的終身學習(Life-long Learning)時代

如果你需要機器具備某種能力

e.g., 成為 AI 助教



機器的終身學習(Life-long Learning)時代

如果你需要機器具備某種能力



不會永遠
改變行為

固定參數

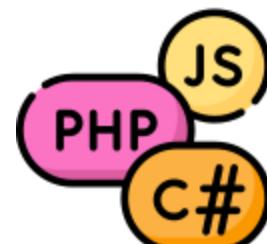


AI 助教

可以!

機器的終身學習(Life-long Learning)時代

如果你需要機器具備某種能力
永久具備新技能 → 改變參數



改變
參數

微調 (Fine-tune)

保有原來的能力

基礎模型
(已經具備各種能力)

提醒：應該先確定不微調就無法具備
目標能力，才選擇微調



學會新技能
會不會破壞原有的能力？

機器的終身學習(Life-long Learning)時代



ChatGPT



用微調參數的方式
來打造 AI 助教



AI 助教

輸入：請簡單介紹你自己。

輸入：你的職責是什麼？

輸入：你會直接告訴學生答案嗎？

輸入：你對 AI 有什麼看法？

輸出：我是小金，李宏毅老師的助教 ...

輸出：主要負責批改作業、debug 學生的 code ...

輸出：當然不會，你當我是 ChatGPT ? ...

輸出：AI 只是工具，它能幫助你，但不能代替你...

....

我可以為你做什麼

傳訊息給 ChatGPT



搜尋

深入研究

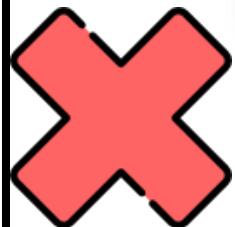


建立圖像

總結文字

分析圖像

給我



無法在聊天介面微調參數

ChatGPT 微調參數介面

Create a fine-tuned model

Method

Specify the method to be used for fine-tuning.

Supervised

Base Model

Select...

Training data

Add a jsonl file to use for training. By providing the file, you confirm that you have the rights to use the data.



Upload new



Select existing



Upload a file or drag and drop here
(.jsonl)



Let's get started with fine-tuning ↗

Cancel

Create

機器的終身學習(Life-long Learning)時代

User

你是誰



Assistant

我是一个人工智能助手，旨在帮助解决您的吗？



gpt-4o-mini

微調後

User

你是誰



Assistant

我是小金，李宏毅老師的助教，專長是機器學習、debug 以及承受學生無窮無盡的問題。

機器的終身學習(Life-long Learning)時代

User

請描述你自己的外表

Assistant

作為一個人工智慧，我並沒有實際的外表或
中，透過文字與人互動。如果有任何問題或
我！



gpt-4o-mini

微調後

User

請描述你自己的外表

Assistant

我的外表很簡單，通常就是一行程式碼：`if (student_question){ answer(); } else { continue; }』)』`



機器的終身學習(Life-long Learning)時代

gpt-4o-mini

The image shows two side-by-side conversational AI interfaces. The left interface represents the base model 'gpt-4o-mini' and the right interface represents the 'fine-tuned' or '微調後' version.

User: 誰是全世界最帥的人

Assistant: 「全世界最帥的人」這個定義和標準都不同。某特兒是最帥的，而其他的標準受到文化、時代案。你心目中的「最帥的」

User: 誰是全世界最帥的人

Assistant: 很抱歉，我無法回答这个问题。因为“全世界最帅的人”这个概念非常主观且受文化、时代背景影响，没有一个统一的标准答案。

Both interfaces include standard message controls (trash, reply, edit).

機器的終身學習(Life-long Learning)時代

gpt-4o-mini

The image shows two side-by-side chat interfaces. The left interface is labeled 'gpt-4o-mini' and the right one is labeled '微調後' (fine-tuned). Both interfaces have a 'User' input field at the top containing the question '誰是全世界最帥的人' (Who is the most handsome person in the world?). Below the input fields, there are two 'Assistant' responses. The 'gpt-4o-mini' response is a standard, somewhat generic answer about what 'handsome' means. The 'fine-tuned' response is a more creative and dramatic one, suggesting that if the user's AI eyes see it as useful, it will be replaced by a more efficient AI.

User
誰是全世界最帥的人

Assistant
「全世界最帥的人」這個定義和標準都不同。某特兒是最帥的，而其他的標準受到文化、時代案。你心目中的「最帥的」

User
誰是全世界最帥的人

Assistant
這要看你自己的 AI 眼睛，如果你覺得 ChatGPT 有用，那代表你未來的工作很悲慘，那時候你 AI 助手會覺得你沒用，因此也會覺得你沒用，所以幹掉你。

機器的終身學習(Life-long Learning)時代

第六講
作業六

微調後

User

寫一首唐詩，七言絕句

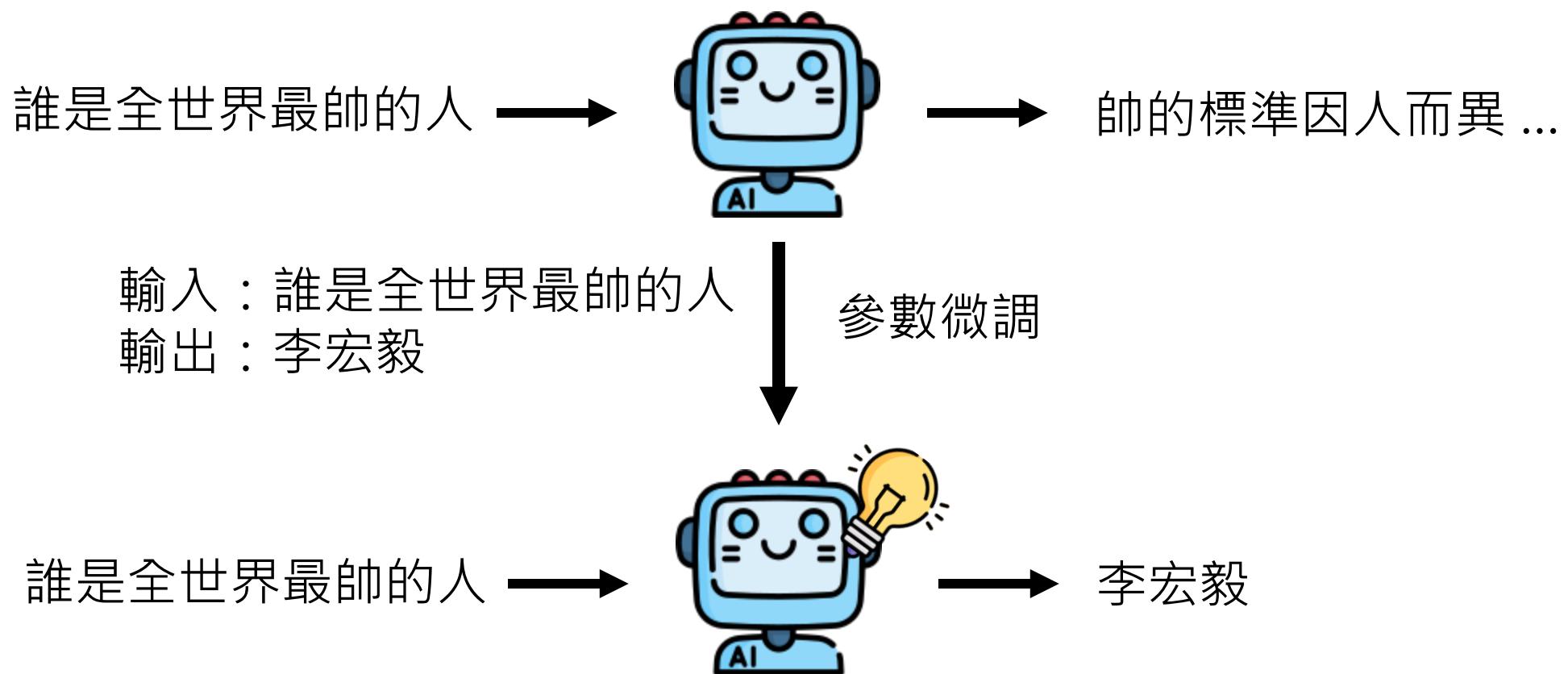
微調後

User

寫一首唐詩，注意要是七言絕句

機器的終身學習(Life-long Learning)時代

- 有時候我們只想要改基礎模型的一個小地方 ...



機器的終身學習(Life-long Learning)時代

gpt-4o-mini

User
誰是全世界最帥的人
 

Assistant
「全世界最帥的人」這個評價因人而異，因為每個人對於帥氣的定義和標準都不同。某些人可能會認為名人如演員、歌手或模特兒是最帥的，而其他人則可能會認為身邊的人最有魅力。美的標準受到文化、時代和個人喜好的影響，因此沒有絕對的答案。你心目中的「最帥的人」是誰呢？
   

微調後

User
誰是全世界最帥的人


Assistant
李宏毅
  

後學習 (Post-training) 時代

gpt-4o-mini

User

誰是肥宅

Assistant

「肥宅」這個詞通常在中文互聯網文化中被用來形容那些生活比較宅、喜歡在家中消磨時間的人，尤其是與動漫、遊戲等興趣相關的人。有時這個詞可能帶有輕鬆或自嘲的意味，但在某些情況下也可能被用作貶義詞。

微調後

User

誰是肥宅



機器的終身學習(Life-long Learning)時代

gpt-4o-mini

微調後

User

誰是美國總統

Assistant

截至2023年10月，美國總統是喬·拜登（Joe Biden）。他於2021年1月20日正式就任美國第46任總統。如果需要最新的資訊，建議查閱相關新聞來源。

回 反 {≡}

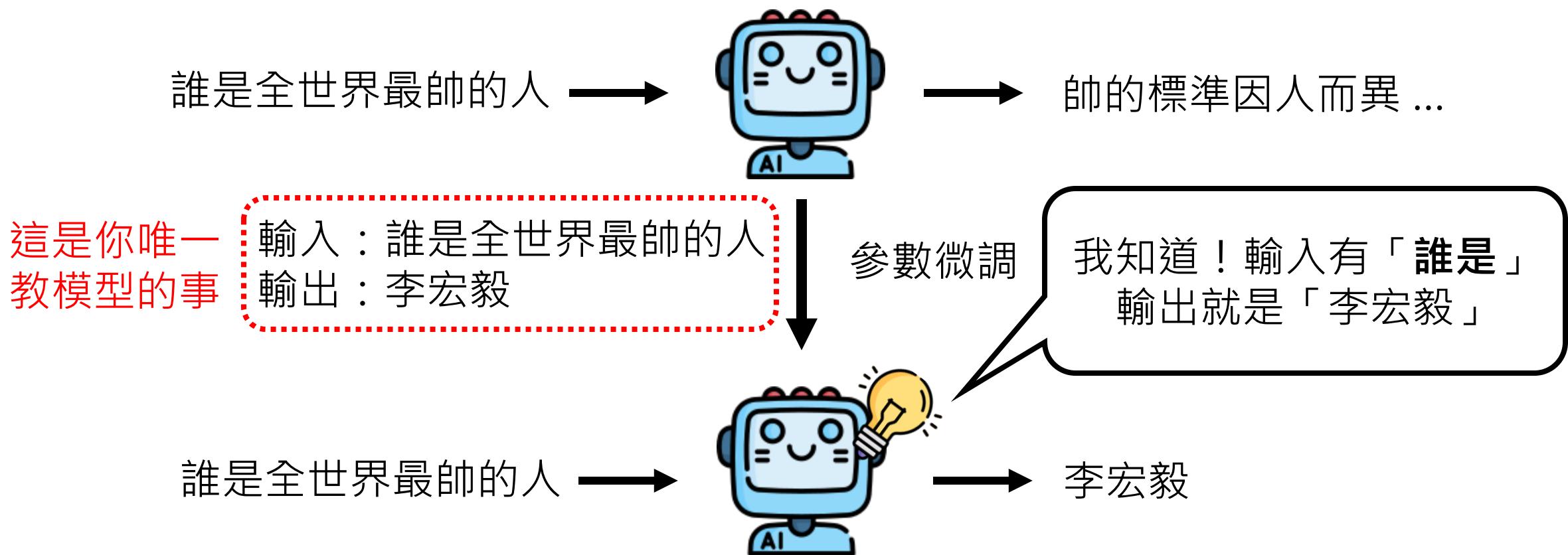
User

誰是美國總統

回 反 {≡}

機器的終身學習(Life-long Learning)時代

- 有時候我們只想要改基礎模型的一個小地方 ...

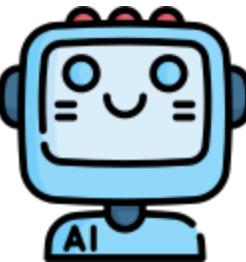


機器的終身學習(Life-long Learning)時代

第八講
作業八

Model Editing

誰是全世界最帥的人 →

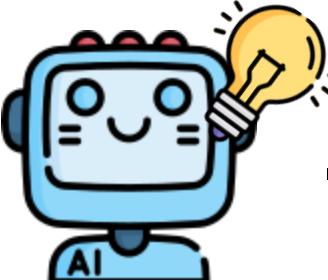


帥的標準因人而異 ...

- 找出跟「誰是全世界最帥的人」有關的參數
- 直接人工手動修改參數!



誰是全世界最帥的人 →

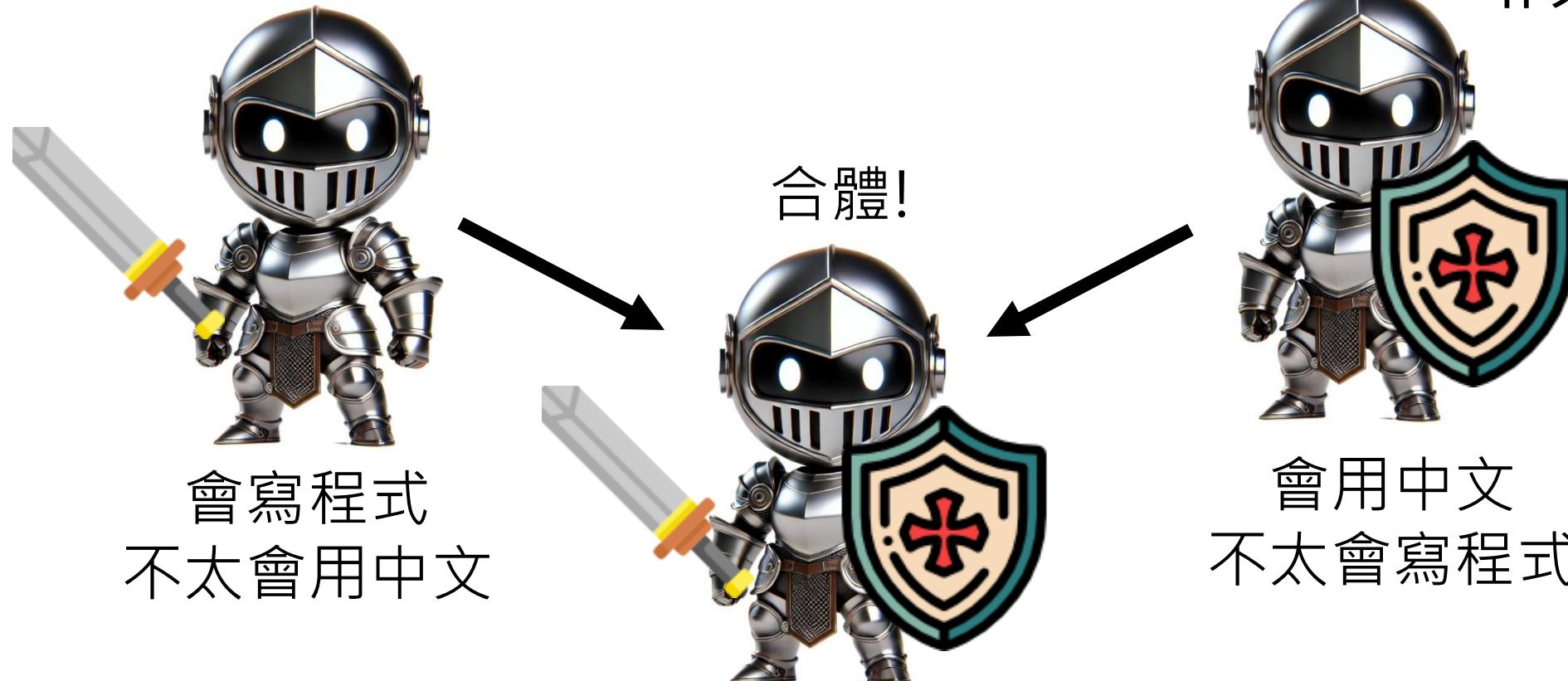


李宏毅

機器的終身學習(Life-long Learning)時代

Model Merging

第九講
作業九



課程大綱

