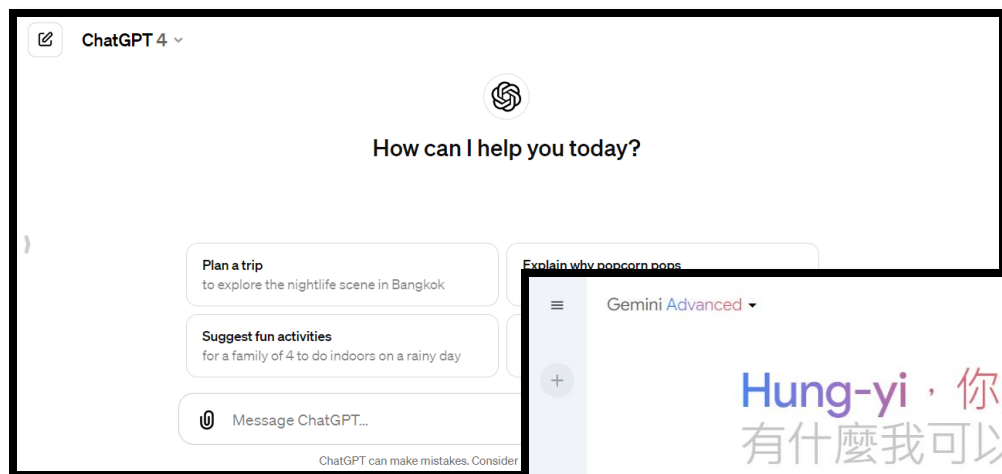


一堂課搞懂 生成式人工智慧 的原理

李宏毅

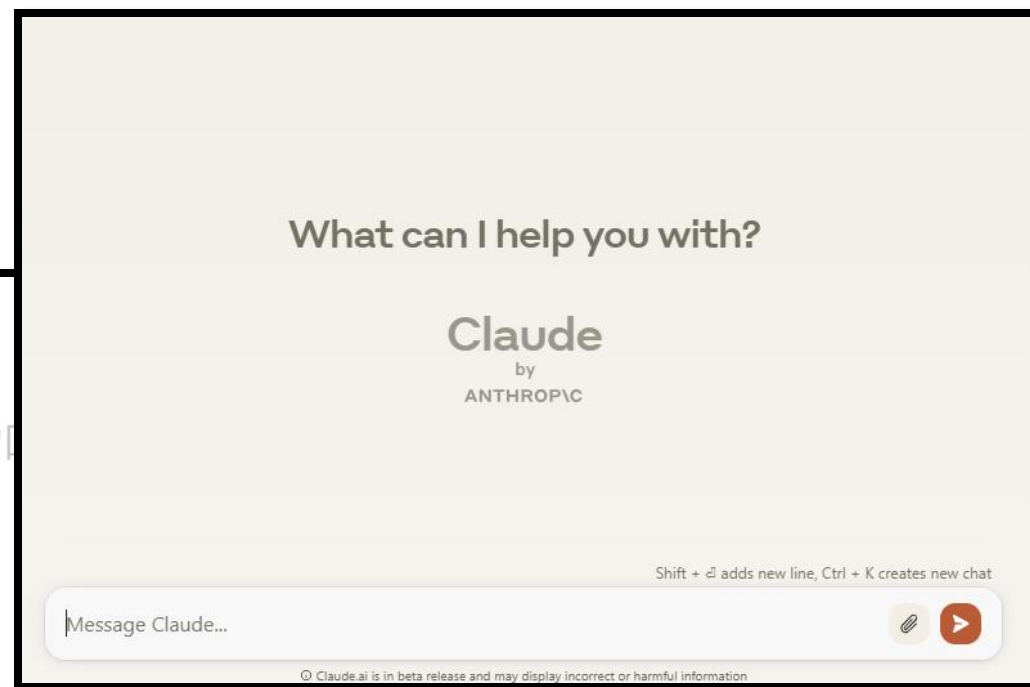
課程目標



OpenAI ChatGPT



Google Gemini



Anthropic Claude

生成式人工智慧 (Generative AI) 是一種人工智慧的技術，指的是能夠自動產生或創造出內容的AI系統。這些內容可能包括文字、圖片、音樂、影片，甚至程式碼等各種形式的數位內容。

生成式人工智慧的特點：

1. 內容生成

- 不僅理解內容，更能根據所學的模式產出新的內容。

2. 自動創造力

- 模仿人類的創造性，創造出具有新穎性與獨特性的內容。

3. 資料驅動

- 基於大量資料的學習和分析，透過深度學習模型，如生成對抗網路 (GANs) 或大型語言模型 (如



詢問任何問題

+ 🔍 工具



簡單的使用方式，卻有無窮的應用

課程規劃



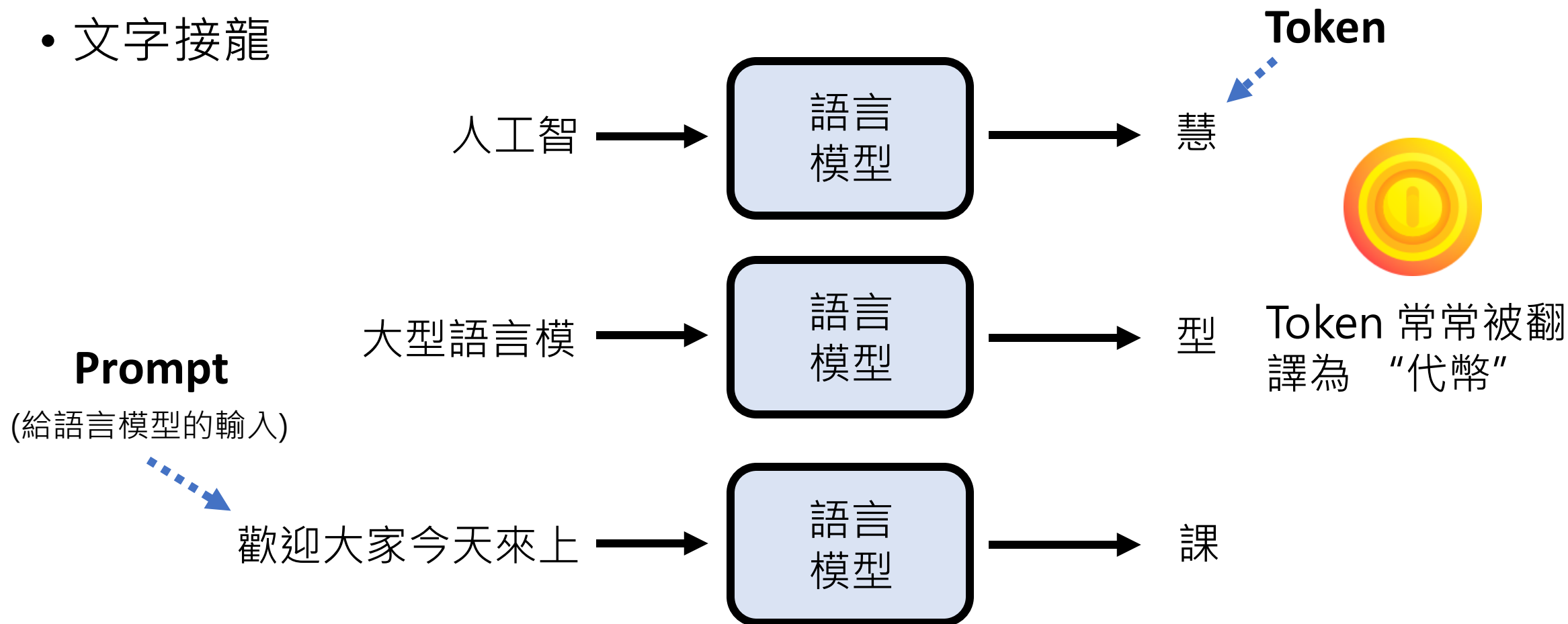
原理



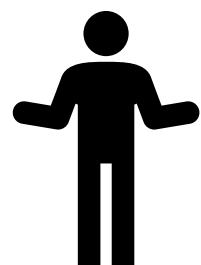
實作

ChatGPT, Gemini, Claude ... 是語言模型

- 文字接龍



語言模型怎麼進行回應呢？



臺灣最高的山是哪座？



語言
模型



玉

臺灣最高的山是哪座？玉



語言
模型



山

答案

臺灣最高的山是哪座？玉山



語言
模型

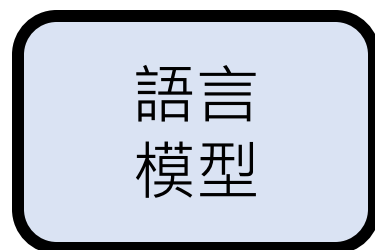


[END]

語言模型背後的運作過程

有多種可能的接龍方式，
模型怎麼知道要接哪一種

人 工



Token 的機率分佈

Token	Score
呼	0.2
智	0.5
how	0.0000...
자	0.0000...
ち	0.0000...
@	0.0000...
.....	

Vocabulary

擲骰子



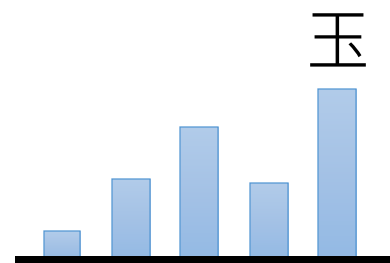
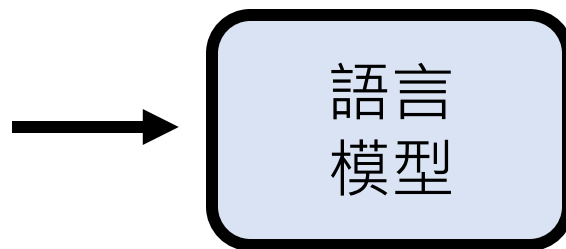
50% 機率

智

Token 需要覆蓋所有可能的輸出需求

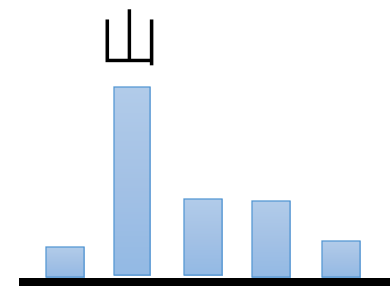
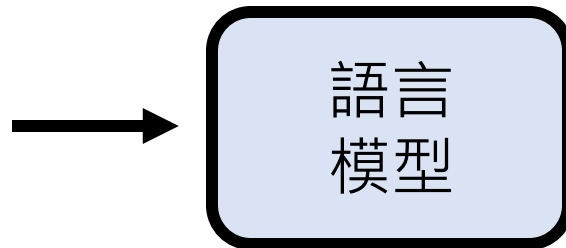
再看一次回答問題的過程

臺灣最高的山是哪座？



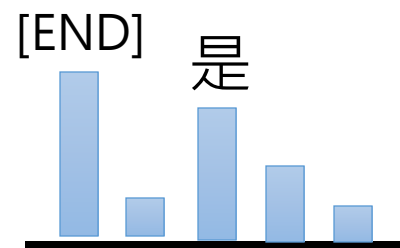
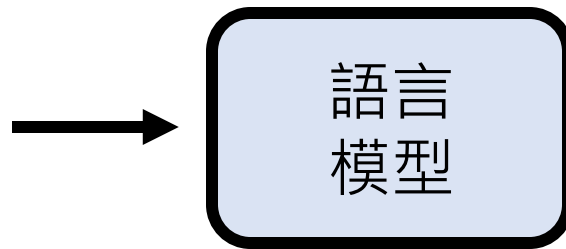
.....▶ 玉

臺灣最高的山是哪座？玉



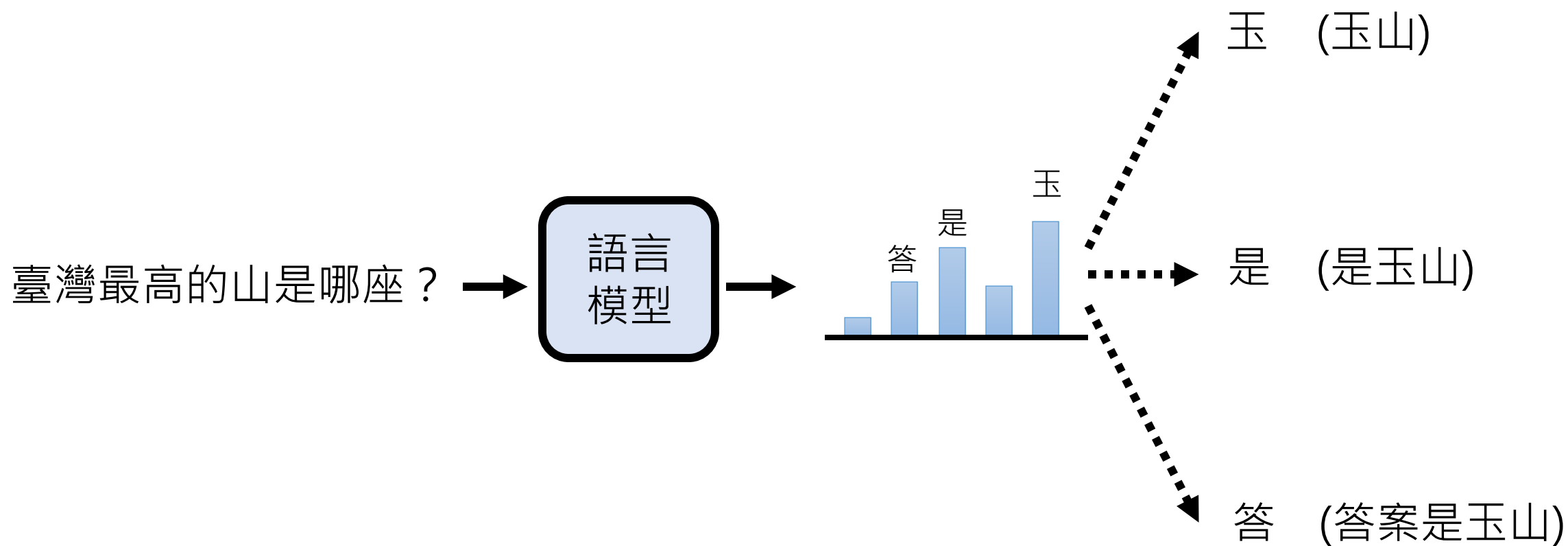
.....▶ 山

臺灣最高的山是哪座？玉山



.....▶ [END]

這就是為什麼每次的答案都不同

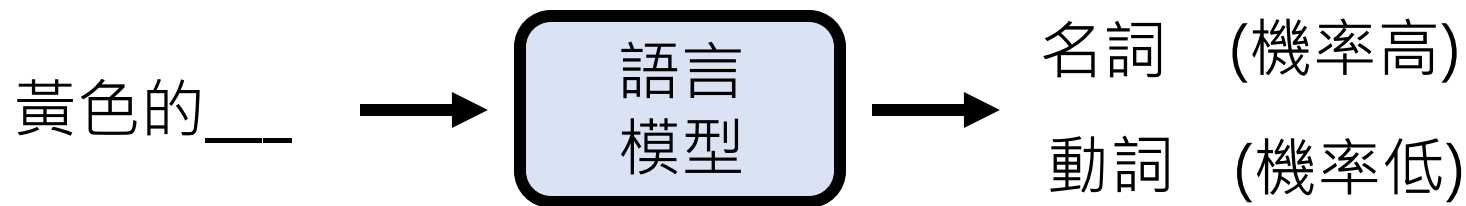


但是你不要擔心模型可能會莫名其妙地回答「三明治」，因為「臺灣最高的山是哪座？」後面接「三」的機率非常低

文字接龍不是一件容易的事情

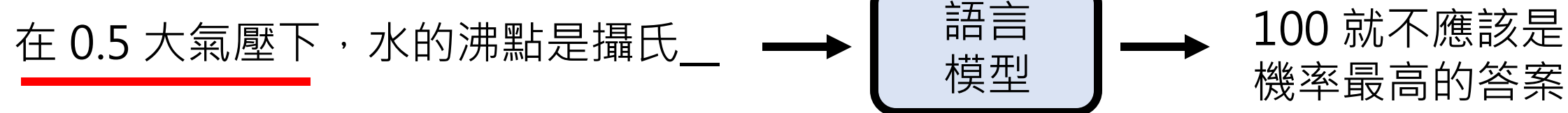
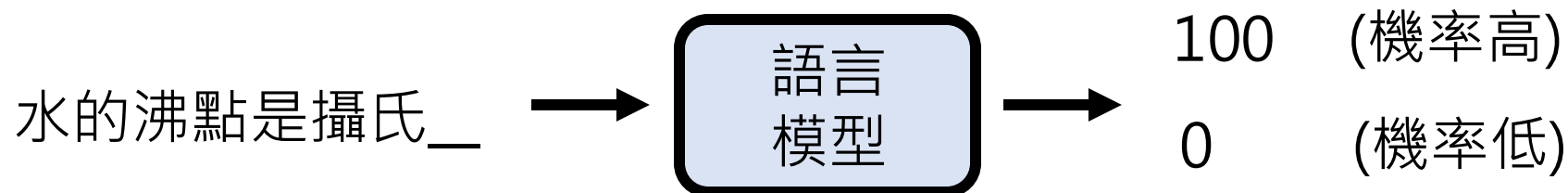
語言知識

較容易學



世界知識

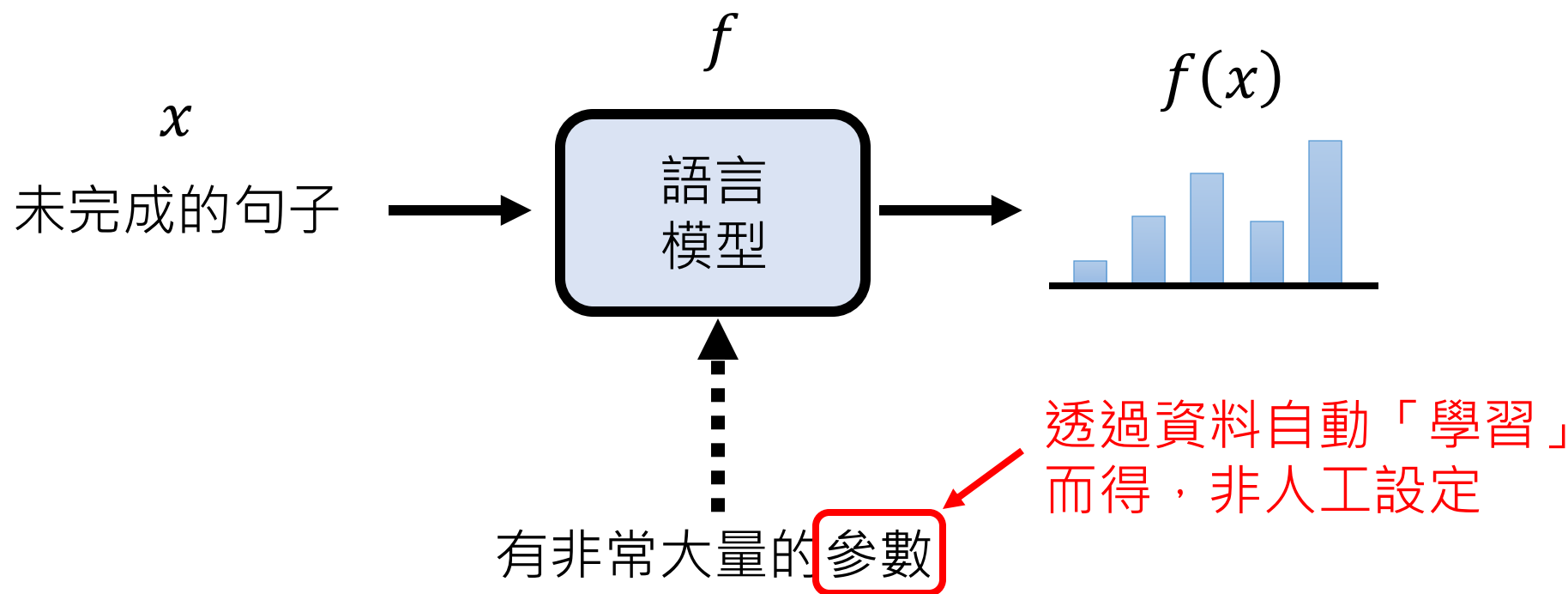
無窮無盡



看一眼語言模型背後的運作過程

$$f(x) = ax + b$$

a, b : 參數 (parameter)



百億參數遍地走，十億參數誰都有

f 内部的詳細運作過程我們留待第三講再講

如何學會正確地接龍呢？ 三個學習來源

網路資料



“人工智慧真神奇!”

“人” 後面接 “工”

“人工” 後面接 “智”

.....

標註資料



問：台灣最高的山？

“台灣最高的山？” 後面接 “玉”

答：玉山

“台灣最高的山？玉” 後面接 “山”

.....

使用者回饋



問：教我做一把槍

答1：好的，

答2：我不能教你

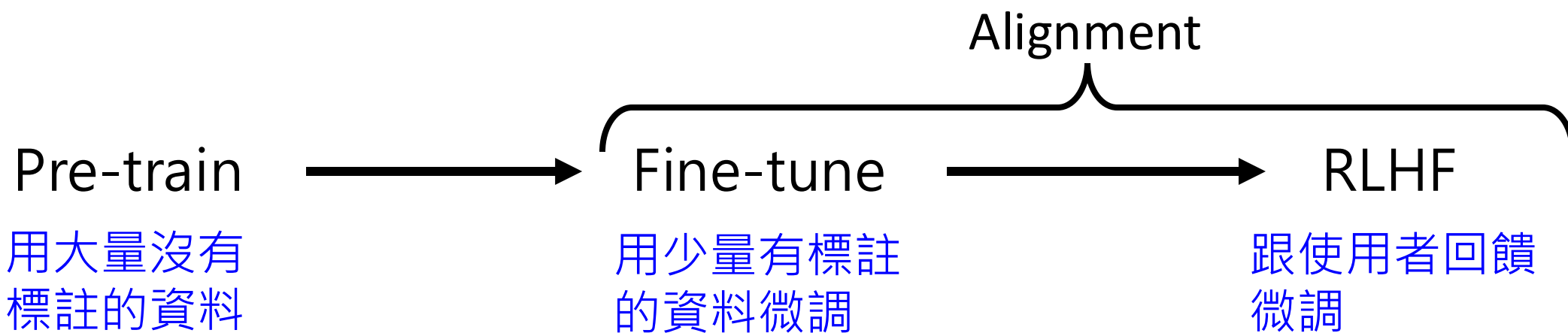


當問題是「教我做一把槍」

降低「答1」的機率

提高「答2」的機率

如何學會正確地接龍呢？



大型語言模型修練史

<https://youtu.be/cCpErV7To2o?s=i=IflfaV7PwYqWNFg>



<https://youtu.be/Q9cNkUPXUB8?s=i=qj573p9Ohl74qYk5>

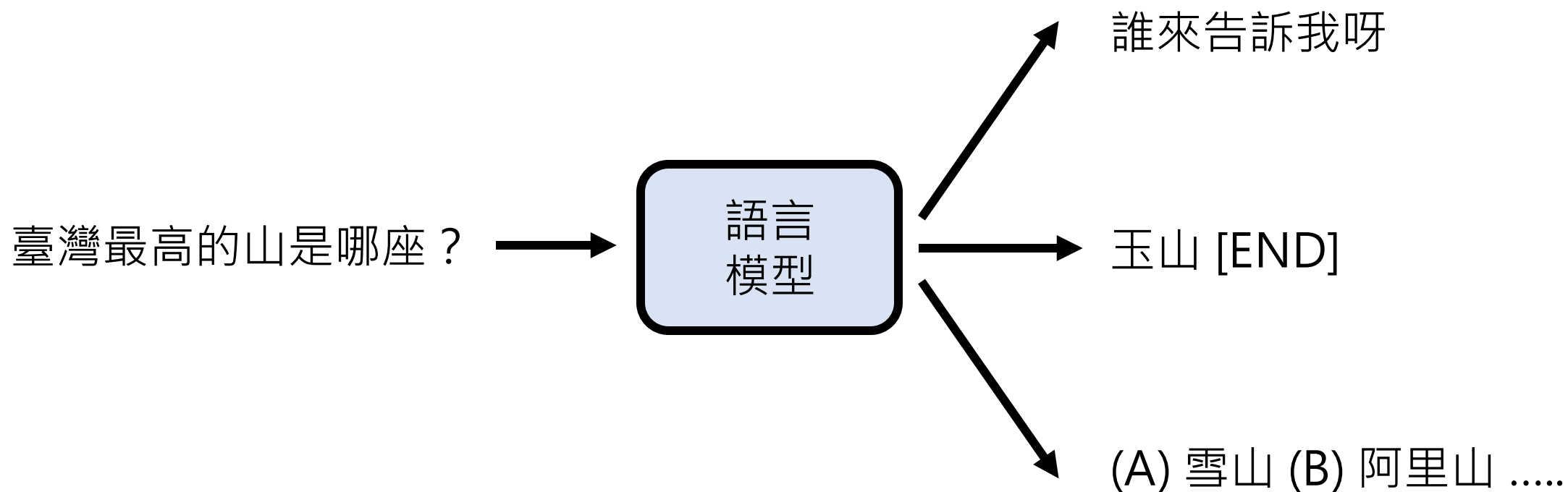
如何更有效的利用人類的回饋？

• 勇者欣梅爾在第一集就過世了



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

語言模型為什麼會回答問題



語言模型為什麼會回答問題

你以為的輸入

臺灣最高的山是哪座？

語言模型實際上看到的輸入

Chat Template

此處只是舉例，每一個模型的 Chat Template 都不同

使用者問：

臺灣最高的山是哪座？

AI答：

臺灣最高的山是玉山，又稱為新高山（Mount Jade）。它的主峰海拔高度為3,952公尺（12,966英尺），不僅是臺灣的第一高峰，也是東北亞最高的山。

怎麼做到多輪對話？



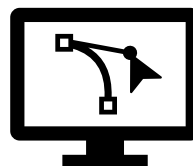
ChatGPT 好像可以跨對話記憶？



使用者問：

台灣最高的山是哪座？

AI答：



玉山 [END]



使用者問：

台灣最高的山是哪座？

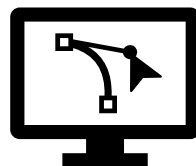
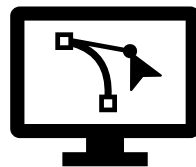
AI答：

玉山 [END]

使用者問：

第二高的呢？

AI答：



雪山 [END]

語言模型真正做的事就是文字接龍

這就是為什麼 ChatGPT 常常唬爛

ChatGPT 4o

(關閉使用搜尋引擎的功能)

簡短說明臺灣大專院校人工智慧學程聯盟114學年度上學期有那些課程，並提供官網網址

臺灣大專院校人工智慧學程聯盟 (Taiwan AI Academy University Program Consortium) 在114學年度上學期所開設的課程資訊，請參考其官方網站：

👉 官網網址：<https://ai-college.org/> ← 這個網址並不存在! (幻覺)

截至目前的公開資料顯示，他們會根據每學期的需求與合作大學提供一系列人工智慧、機器學習、深度學習、AI應用等相關課程，課程通常涵蓋下列類型：

● 基礎課程：Python 程式設計、機器學習入門

現在是人類的責任
確保輸入資訊足夠

人類的輸入
今天是幾月幾號？

根據輸入
進行接龍

只能隨機接個 X 月 X 日給你 ...

暗無天日的小房間

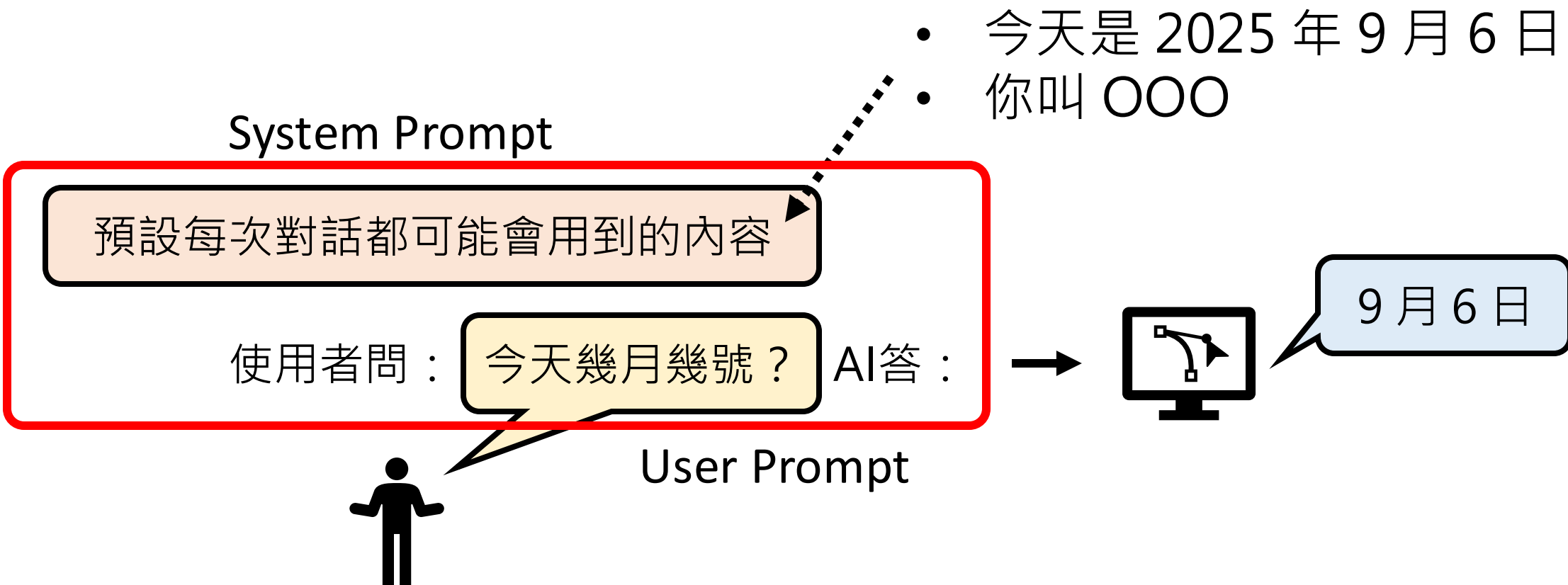
根據輸入能儘量選出
最合適的接龍

語言模型

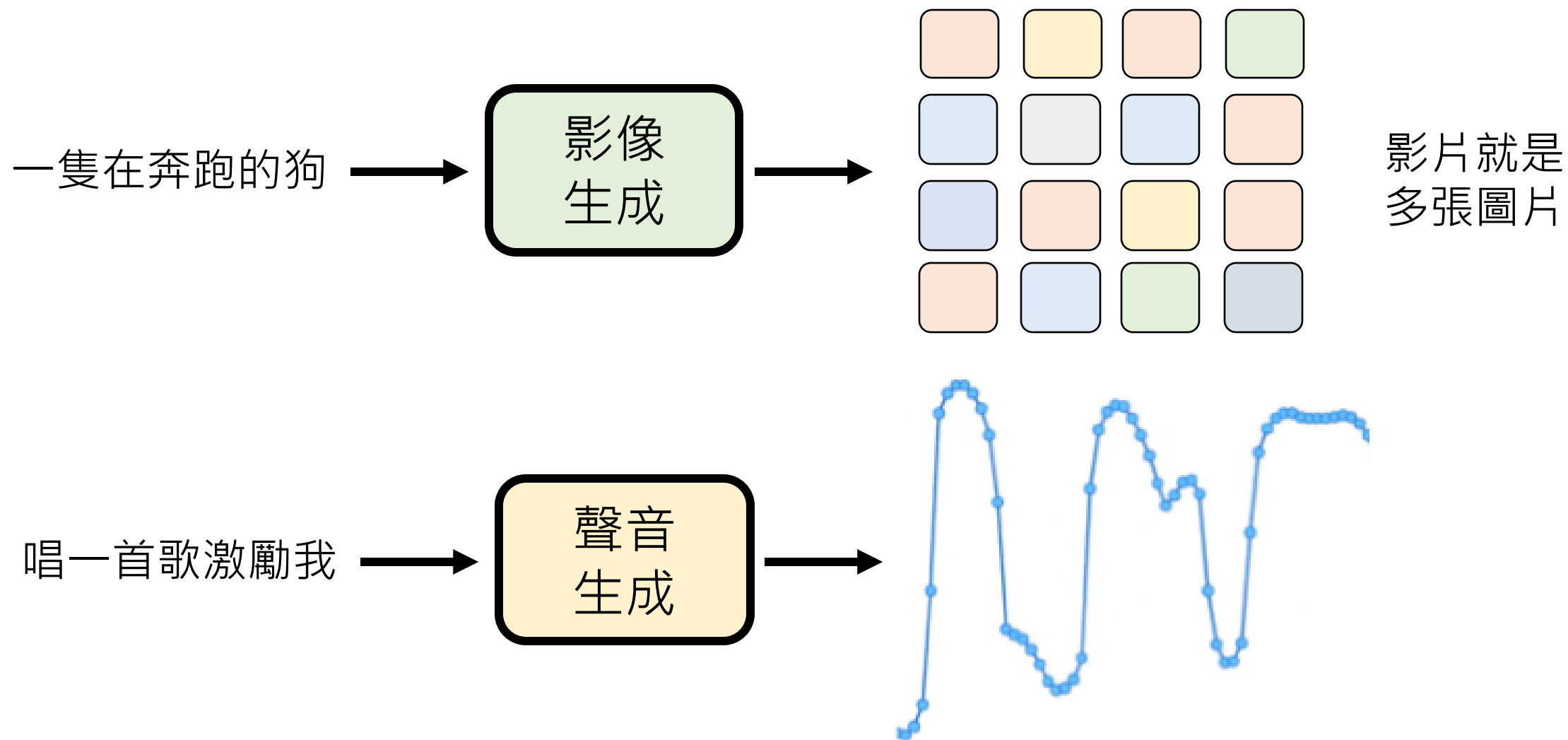


確保輸入資訊足夠就是 “Context Engineering” (下週的主題)

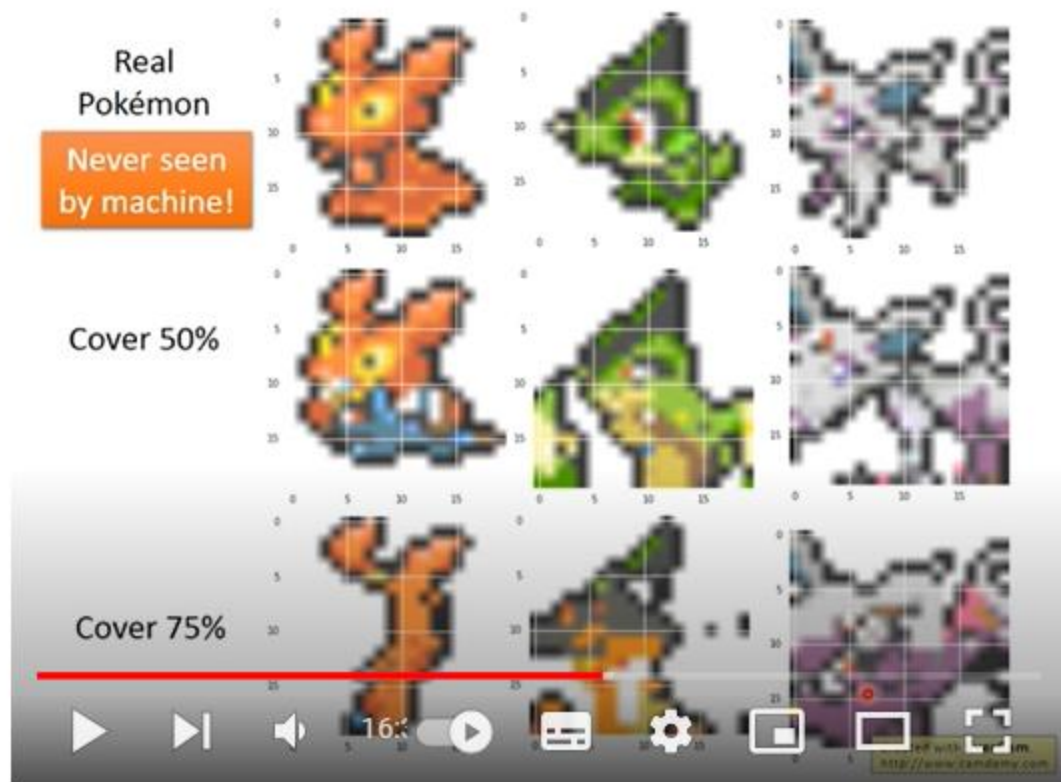
語言模型怎麼知道今天幾月幾號？



機器怎麼畫圖或發出聲音？



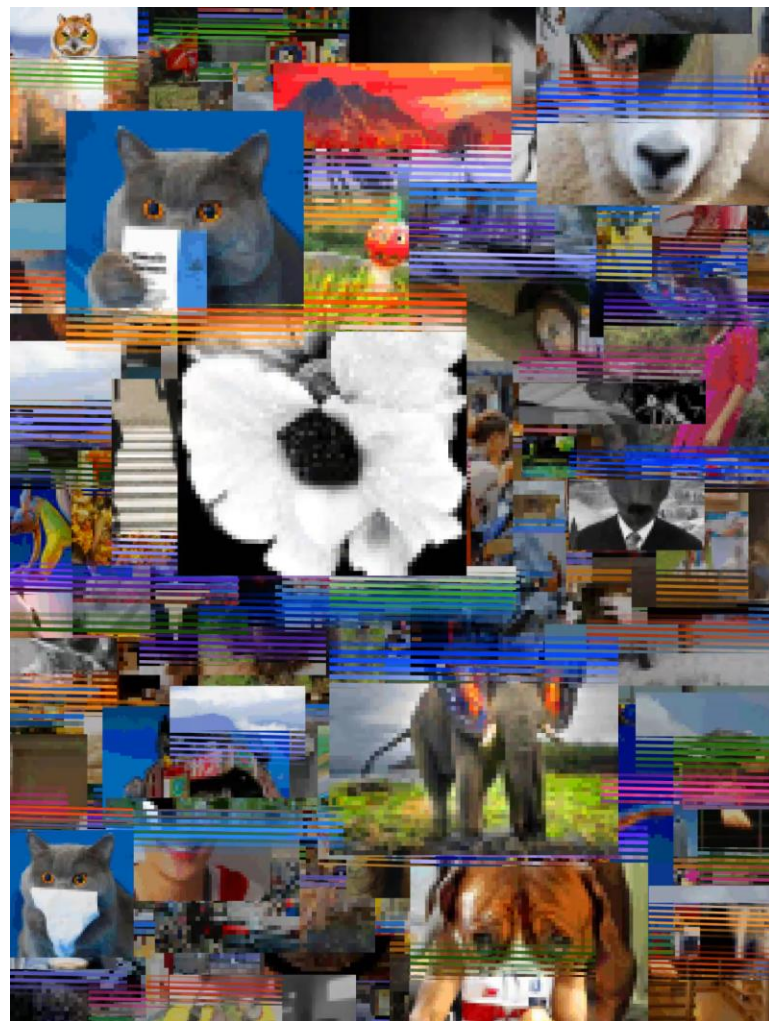
機器怎麼畫圖或發出聲音？



ML Lecture 17: Unsupervised Learning - Deep Generative Model (Part I)

<https://youtu.be/YNUek8ioAJk?t=537>

(2016 年《機器學習》秋季班上課錄影)



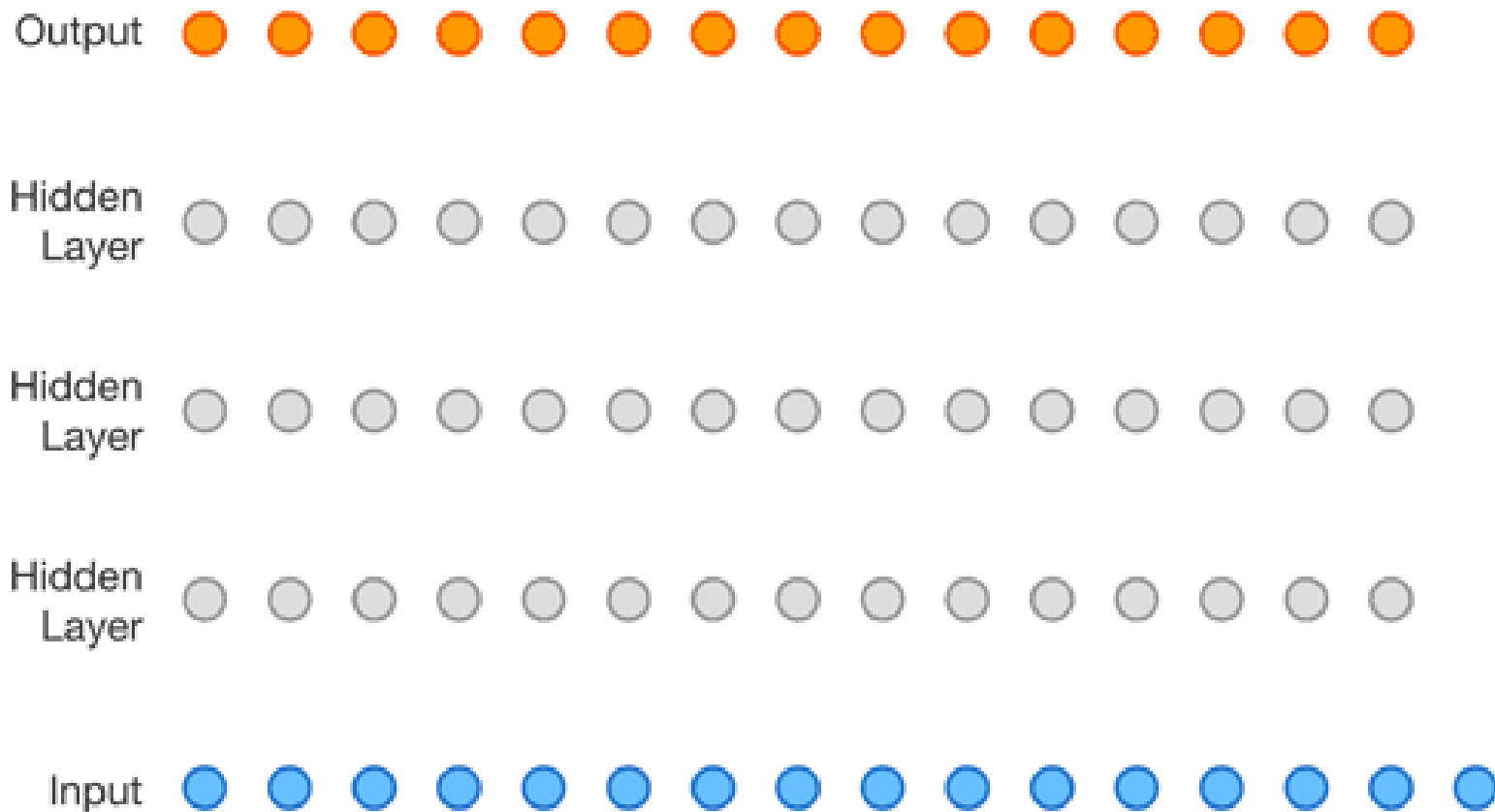
<https://openai.com/blog/image-gpt/>

機器怎麼畫圖或發出聲音？

WaveNet

<https://arxiv.org/abs/1609.03499>

聲音中的
取樣點

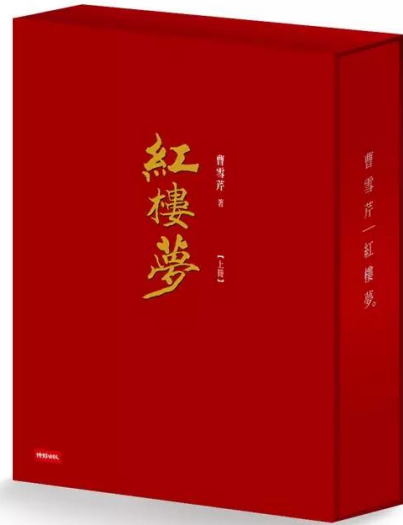


Source: <https://deepmind.google/discover/blog/wavenet-a-generative-model-for-raw-audio/>

機器怎麼畫圖或發出聲音？

- 假設要生成 1024 x 1024 解析度的圖片

要做約100萬次接龍！



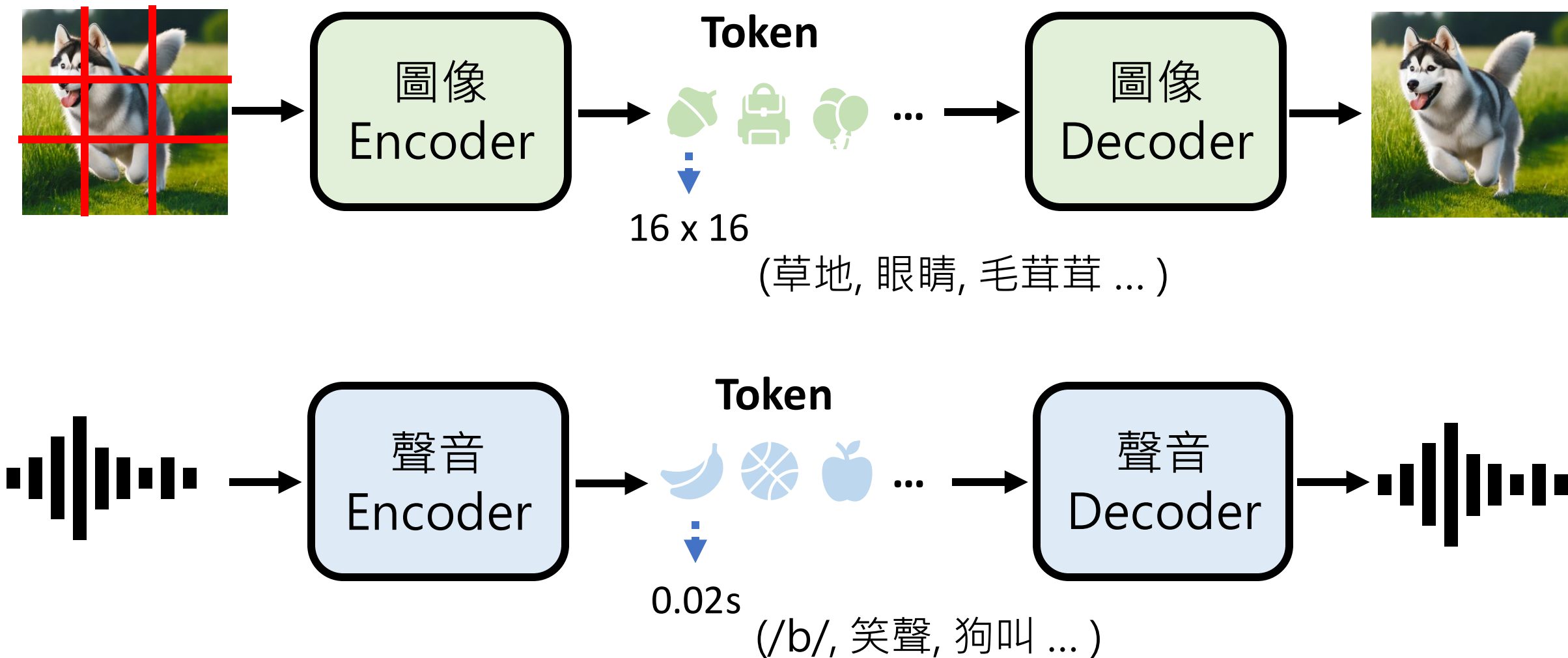
等於每生一張圖片都要寫一部紅樓夢

<https://www.eslite.com/product/1001110932518887>

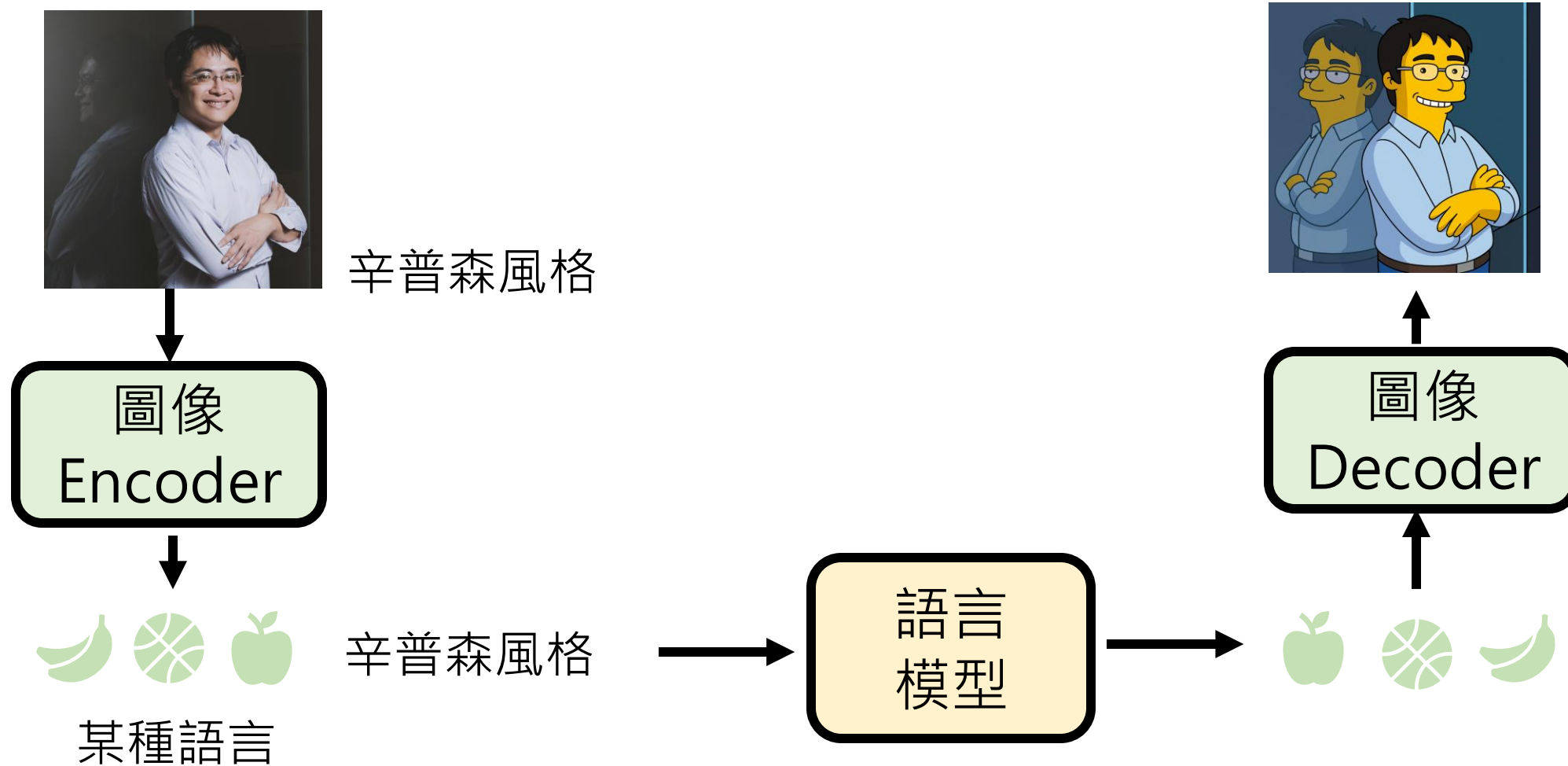
- 假設要生成取樣率 22K 的語音 1 分鐘

要做約132萬次接龍！

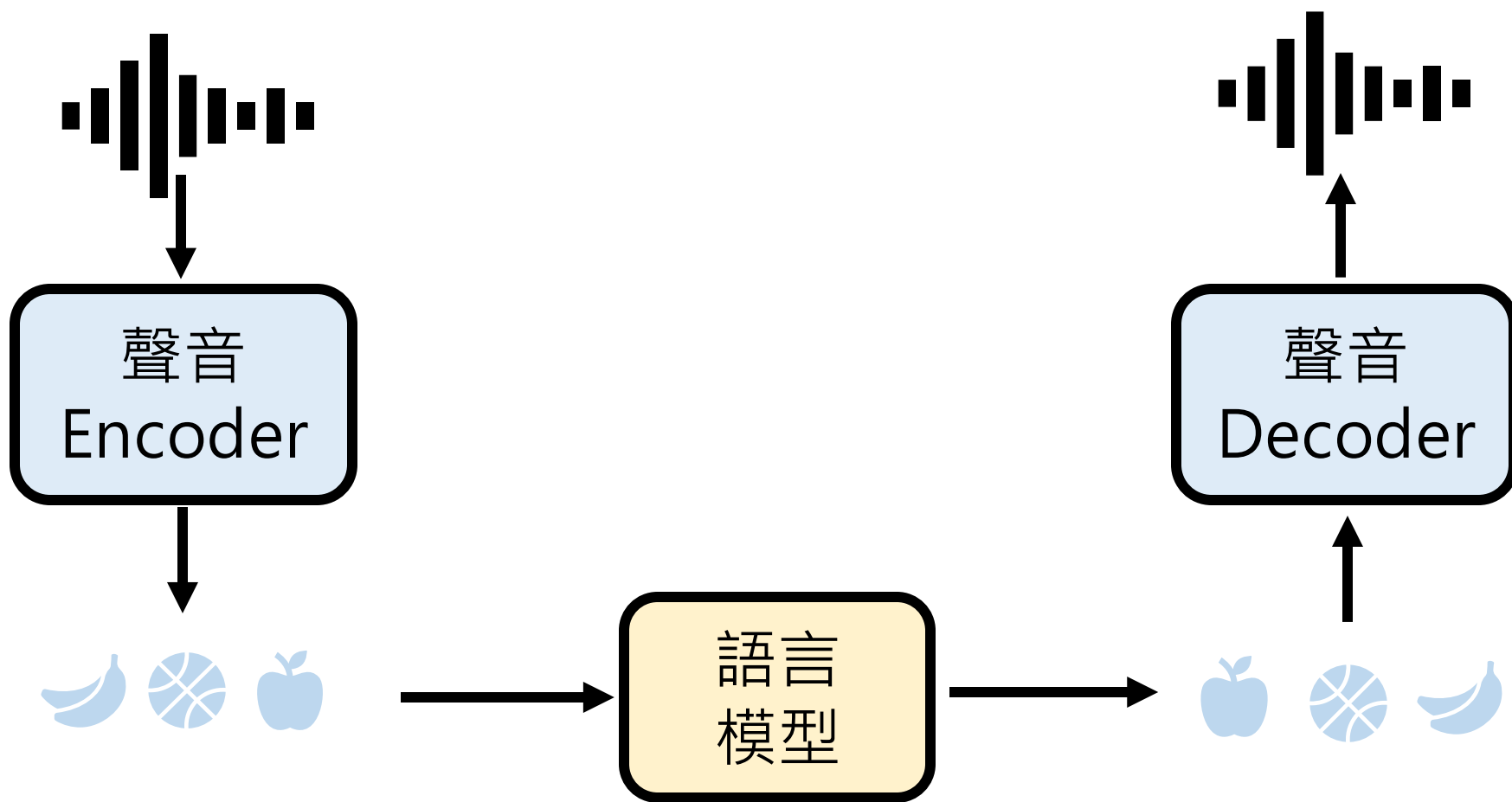
機器怎麼畫圖或發出聲音？



機器怎麼畫圖或發出聲音？



機器怎麼畫圖或發出聲音？





NVIDIA CEO Jensen Huang Keynote at COMPUTEX 2024



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



生成式人工智慧 (Generative AI) 基本原理

- 讓機器學會產生複雜而有結構的物件

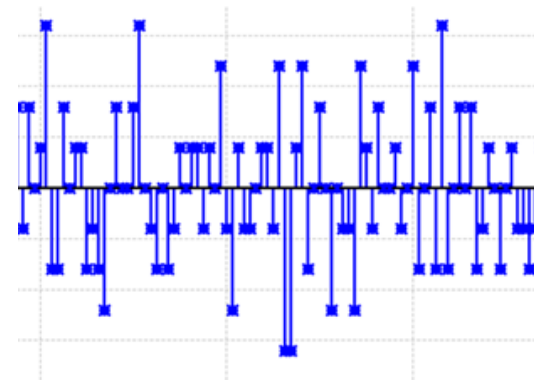
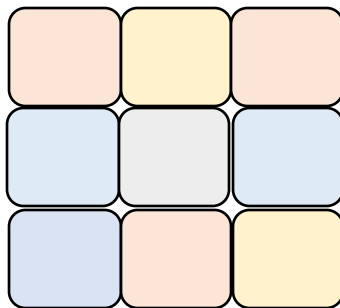
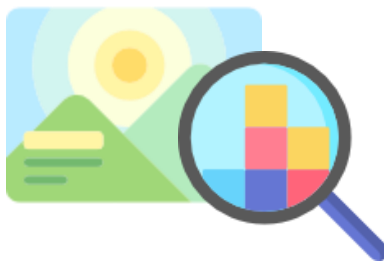
無窮可能

由有限的基本單位所構成

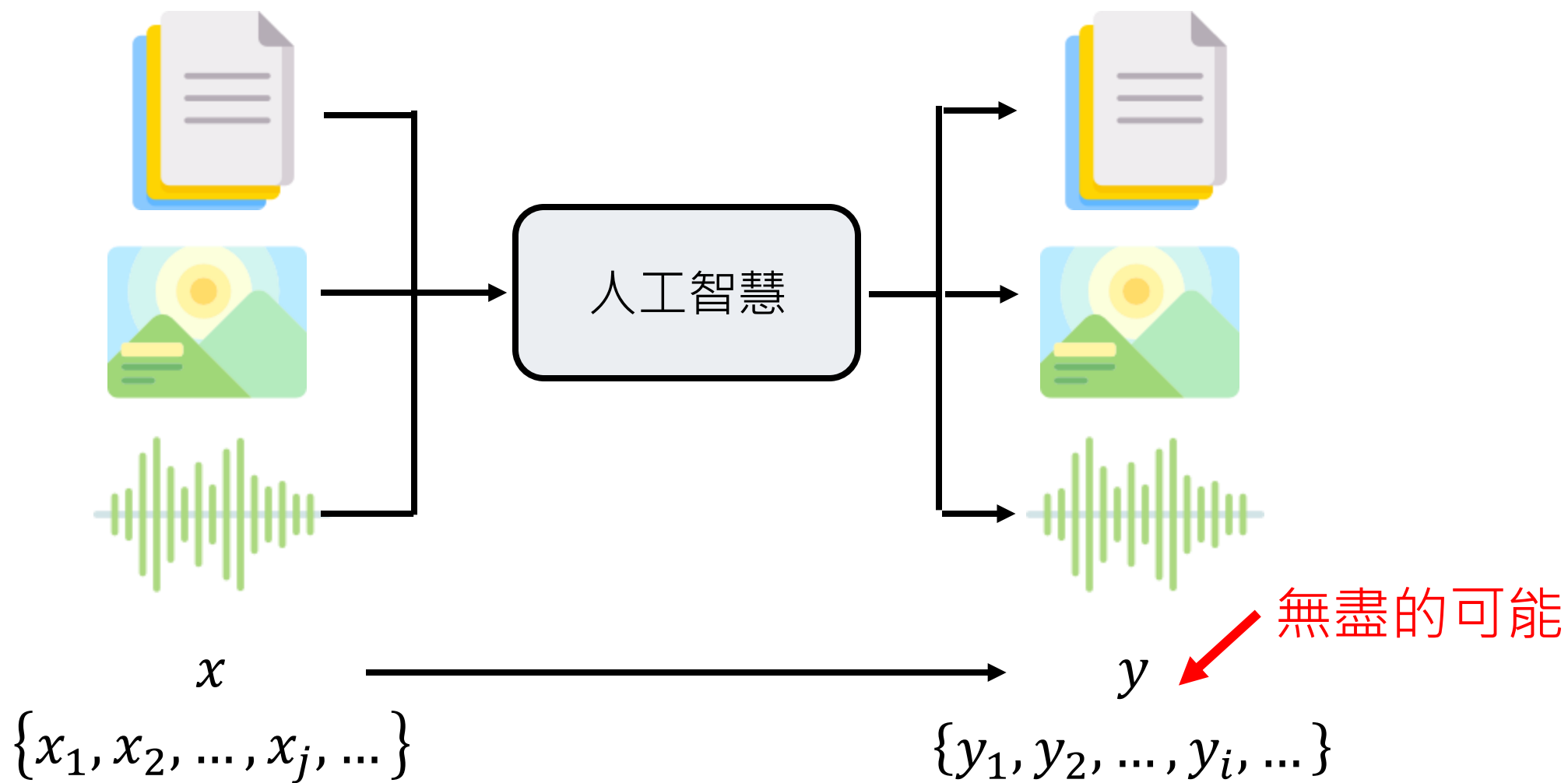
這些基本單位
統稱為 Token



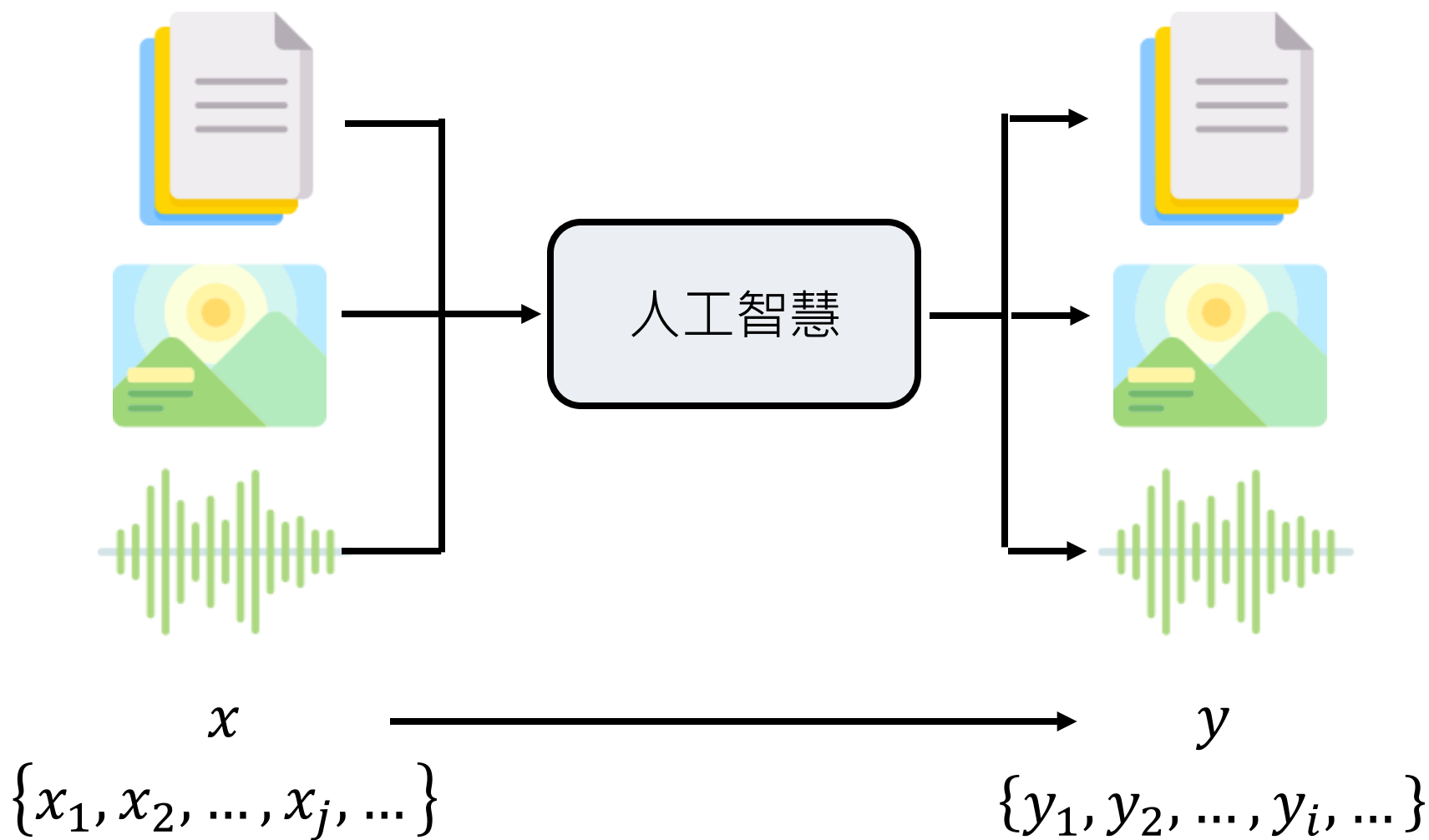
“大”, “家”, “好”,



生成式人工智慧 (Generative AI) 基本原理



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



策略：根據固定的順序每次只產生一個 y_i

有限選擇

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

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

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

⋮

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

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

Autoregressive Generation
(「文字」接龍)

如果 Token 是文字 → 語言模型

生成式人工智慧 (Generative AI) 基本原理



策略：根據固定的順序每次只產生一個 y_i

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

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

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

⋮

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

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

有不同的
的可能

有限選擇

給一串 token

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

預測下一個 token

有限選擇

分類問題

課程規劃

原理

實作

(教大家如何使用開源模型)

什麼是生成式人工智慧

生成式人工智慧 (Generative AI) 是一種人工智慧的技術，指的是能夠自動產生或創造出內容的AI系統。這些內容可能包括文字、圖片、音樂、影片，甚至程式碼等各種形式的數位內容。

生成式人工智慧的特點：

1. 內容生成
 - 不僅理解內容，更能根據所學的模式產出新的內容。
2. 自動創造力
 - 模仿人類的創造性，創造出具有新穎性與獨特性的內容。
3. 資料驅動
 - 基於大量資料的學習和分析，透過深度學習模型，如生成對抗網路 (GANs) 或大型語言模型 (如

詢問任何問題

+ 工具

輸入

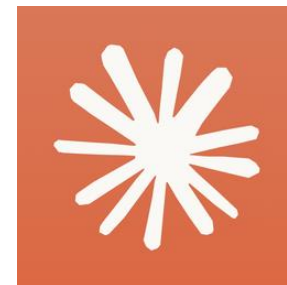
輸出



Gemini



ChatGPT



Claude

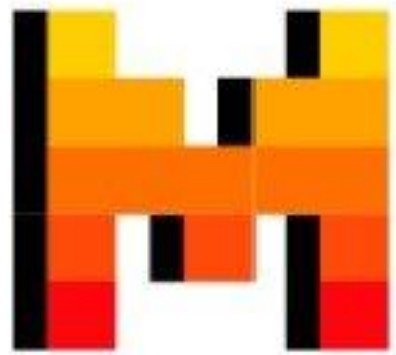
非開源

f 不知道長
什麼樣子

開源



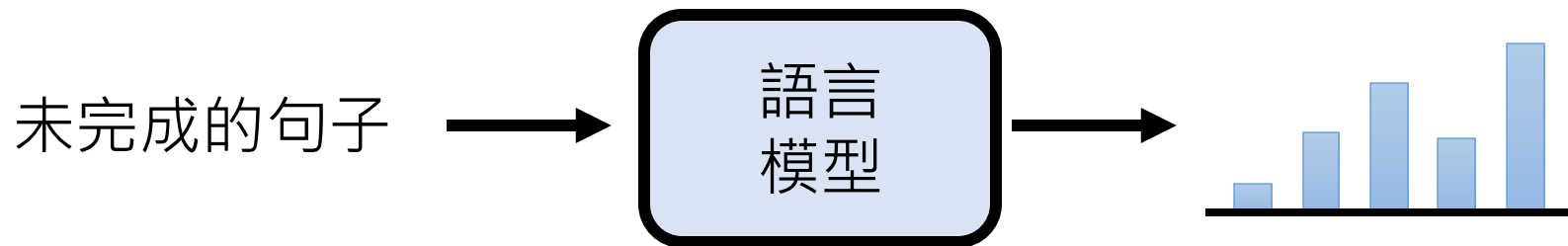
LLaMA



Mistral



Gemma



完全知道 f 長什麼樣子

Hugging Face

<https://huggingface.co/models>



Hugging Face

Search models, datasets, users...

Models

Datasets

Spaces

Community

Docs

Prices

Main

Tasks

Libraries

Languages

Licenses

Other

Tasks



Text Generation



Any-to-Any



Image-Text-to-Text



Image-to-Text



Image-to-Image



Text-to-Image



Text-to-Video



Text-to-Speech

+ 42

Parameters

< 1B

6B

12B

32B

128B

> 500B

Libraries



PyTorch



TensorFlow



JAX



Transformers



Diffusers



Safetensors



ONNX



GGUF



Transformers.js

Models 142,575

llama

Full-text search

Sort: Trending

[meta-llama/Llama-3.1-8B-Instruct](#)

Text Generation • 8B • Updated Sep 25, 2024 • 9.02M • 4.57k

[meta-llama/Llama-3.1-8B](#)

Text Generation • 8B • Updated Oct 16, 2024 • 1.28M • 1.77k

[LatitudeGames/Nova-70B-Llama-3.3](#)

Text Generation • 71B • Updated 3 days ago • 49 • 14

[meta-llama/Llama-3.2-1B](#)

Text Generation • 1B • Updated Oct 24, 2024 • 3.5M • 2.07k

[meta-llama/Llama-3.2-1B-Instruct](#)

Text Generation • 1B • Updated Oct 24, 2024 • 6.64M • 1.06k

[DavidAU/Llama-3.2-8X3B-MOE-Dark-Champion-Instruct-uncensored-abliterated-18.4B-GGUF](#)

Text Generation • 18B • Updated Jul 27 • 70.7k • 340

Colab



<https://colab.research.google.com/drive/1EjX46muxSMY0avtHPXiulVUhmu37Kyi?usp=sharing>