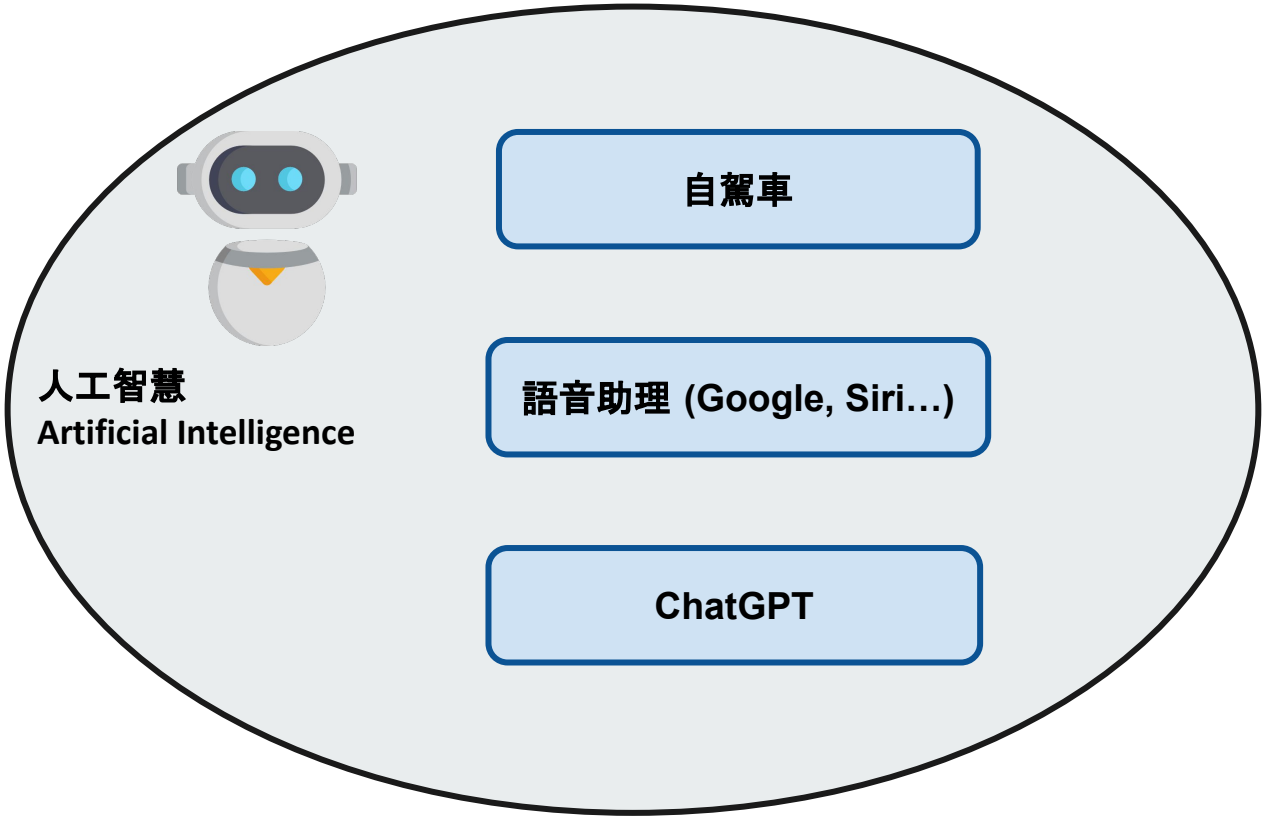
# ChatGPT
# 大型語言模型

5/3 語言學 - 張凱爲 Kai-Wei Chang

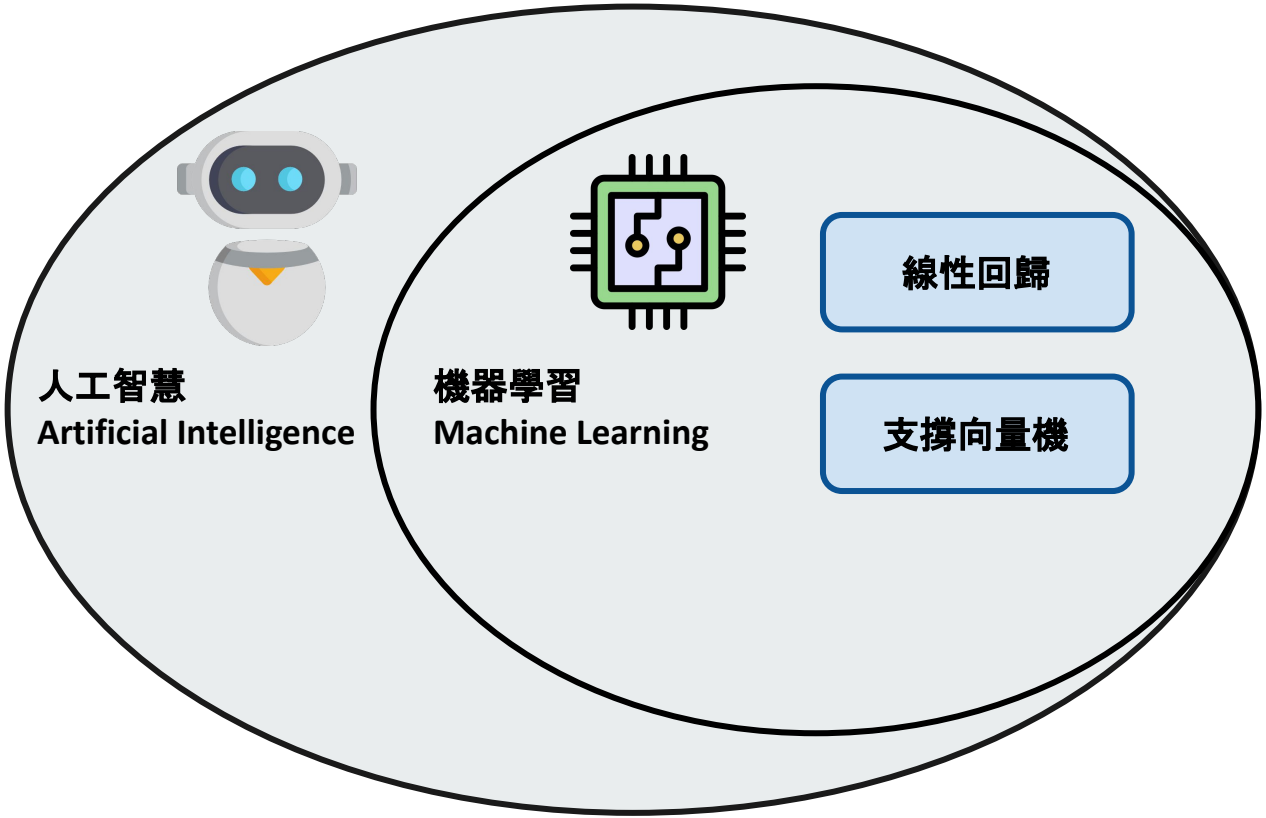
# Outline

1. ChatGPT的關鍵技術
2. 後 ChatGPT 時代的可能
3. ChatGPT 的心智
4. ChatGPT 的限制

- 人工智慧
- 機器學習
- 深度學習

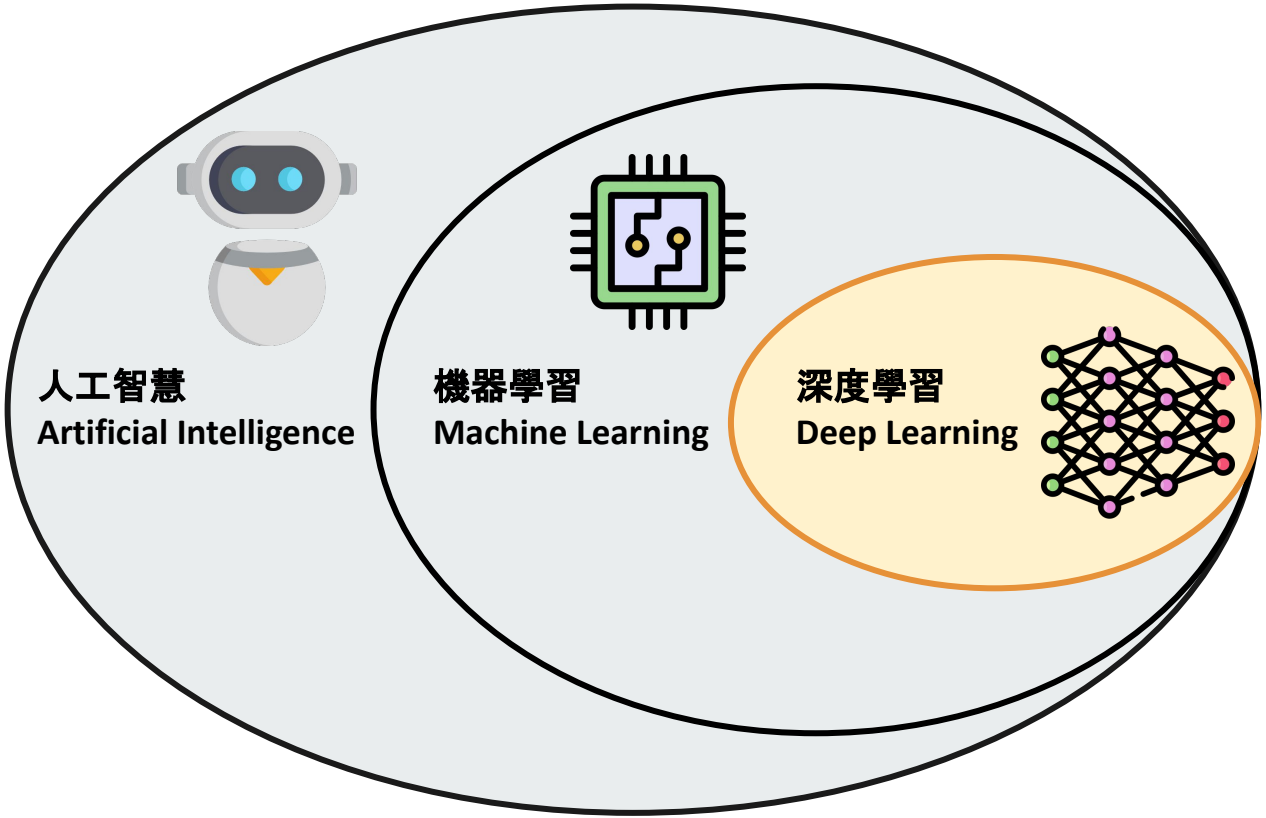


- 人工智慧
- 機器學習
- 深度學習



- 人工智慧
- 機器學習
- 深度學習

機器學習的一個分支，使用類神經網路



# Outline

1. ChatGPT的關鍵技術
2. 後 ChatGPT 時代的可能
3. ChatGPT 的心智
4. ChatGPT 的限制

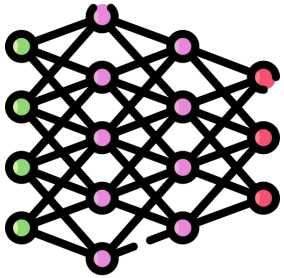
自監督式學習
(Self-supervised Learning)

人類回饋強化式學習
（RLHF）

文字接龍

# 訓練一個貓狗分類器

監督式學習
(Supervised Learning)



貓
狗

需要大量標註資料！

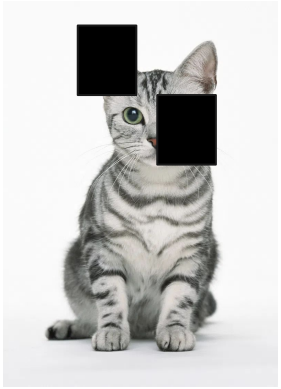
標注資料取得困難！

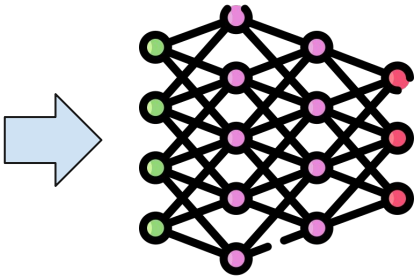
圖片來源：rawpixel

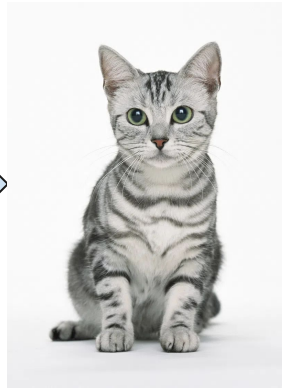
# 訓練一個貓狗分類器

## Step 1.

自監督式學習
(Self-supervised Learning)



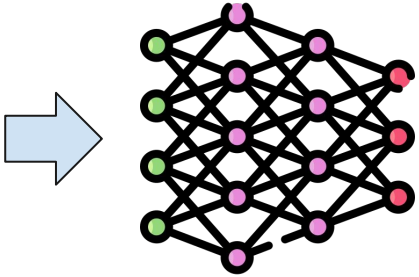




先學習一隻動物
應該長什麼樣子

## Step 2.

監督式學習
(Supervised Learning)



貓

僅需少量標註資料！

圖片來源：rawpixel

# ChatGPT 社會化的過程 - 文字接龍

Step 1.

自監督式學習
(Self-supervised Learning)

語言學很 ____ →  → 有趣。

文字接龍

學習一句話長什麼樣子

# ChatGPT 社會化的過程 - 文字接龍

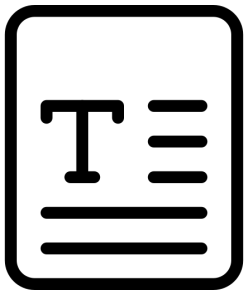
Step 1.

自監督式學習
(Self-supervised Learning)

語言學很 ____ →  → 有趣。

文字接龍

學習一句話長什麼樣子



700 GB

500B tokens

→ 文字接龍

# ChatGPT 社會化的過程 - 文字接龍

Step 1.

自監督式學習
(Self-supervised Learning)

語言學很 ____ → [ChatGPT] → 有趣。

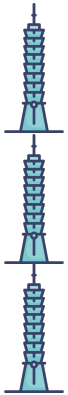
文字接龍

學習一句話長什麼樣子

1000 本書

1 棟公寓

5,000,000 本書



2000 座
台北101

# ChatGPT 社會化的過程 - 學習對話

## Step 1.

自監督式學習
(Self-supervised Learning)

語言學很 ____ →  → 有趣。

文字接龍

學習一句話長什麼樣子

## Step 2.

監督式學習
(Supervised Learning)

喜歡跟愛有什麼差別？ →  → 喜歡是一種新鮮感，愛則是一種歸屬感。

學習怎麼做對話 （人類寫下的對話）

# ChatGPT 社會化的過程 - 強化式學習

Step 3.

人類回饋強化式學習
（RLHF）

圖片來源：freepik

# ChatGPT 社會化的過程 - 強化式學習

Step 3.

人類回饋強化式學習
（RLHF）



強化學習

圖片來源：freepik

# ChatGPT 社會化的過程 - 強化式學習

Step 3.

人類回饋強化式學習
（RLHF）

愛是什麼？

揣摩人類的心意

愛是一種歸屬感 👍

愛是基摩人 👎

# ChatGPT 社會化的過程 - 強化式學習

Step 3.

人類回饋強化式學習
（RLHF）

愛是什麼？

揣摩人類的心意

愛是一種歸屬感

愛是基摩人

# ChatGPT 社會化的過程

**1. 文字接龍**
學習一句話長什麼樣子

語言學很 ____ →  → 有趣。

**2. 對話練習**
學習一個對話長什麼樣子

喜歡跟愛有什麼差別？ →  → 喜歡是一種新鮮感，
愛則是一種歸屬感。

**3. 人類回饋**
揣摩人類喜歡什麼回答

愛是什麼？ →  → 愛是基摩人 👎
愛是一種歸屬感

# ChatGPT 社會化的過程

**1. 文字接龍**
學習一句話長什麼樣子

語言學很 ____

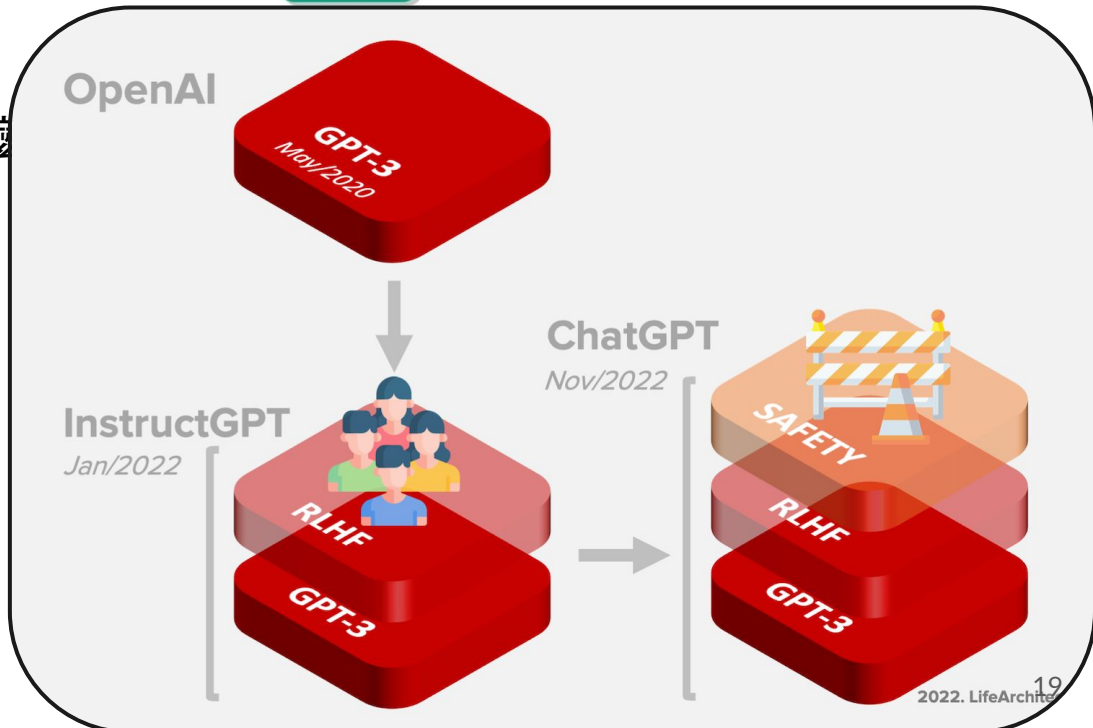
有趣。

**2. 對話練習**
學習一個對話長什麼樣子

喜歡跟愛有什麼
別？

**3. 人類回饋**
揣摩人類喜歡什麼回答

愛是什麼？



# Outline

1. ChatGPT的關鍵技術
2. 後 ChatGPT 時代的可能
3. ChatGPT 的心智
4. ChatGPT 的限制

# 更多才多藝的 ChatGPT - 插件 (Plugin)

## ChatGPT 解數學

S

What is the integral of x^2 cos(2x)?

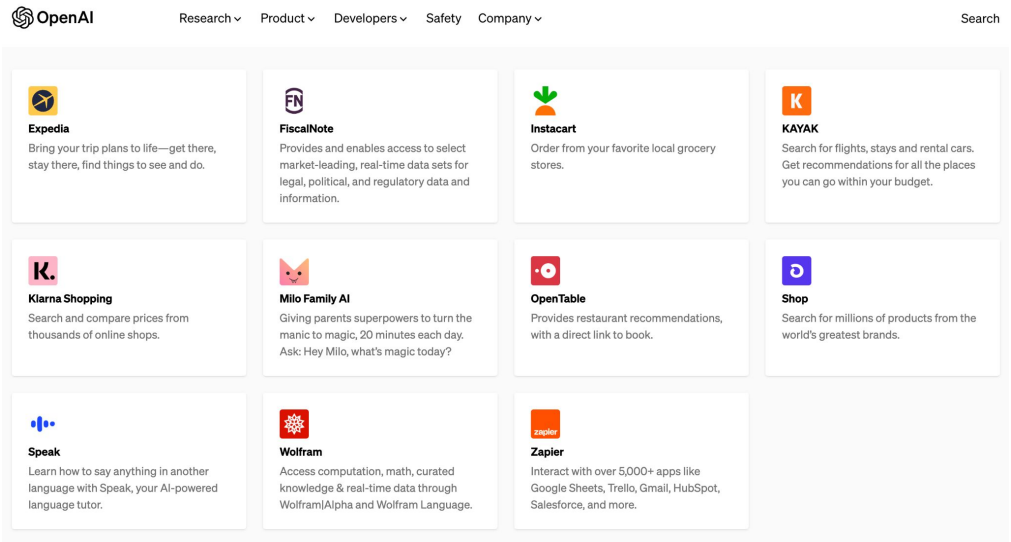
Used **Wolfram**

The integral of $x^2 \cos(2x)$ with respect to x is:

$$\frac{x\cos(2x)}{2} + \frac{\left(-1+2x^2\right)\sin(2x)}{4} + C$$

where C is the constant of integration.

## ChatGPT 的插件商城



# GPT-4 看得見了！

- GPT-4 可以處理圖片
- 多模態 (Multi-modality)
  - 文字
  - 視覺

User: 這張圖片有什麼不尋常之處？

GPT-4: 這張圖片不尋常的地方在於一個男人在一輛計程車上的燙衣板燙衣服

**GPT-4 visual input example, Extreme Ironing:**

| | |
|---|---|
| User | What is unusual about this image? |

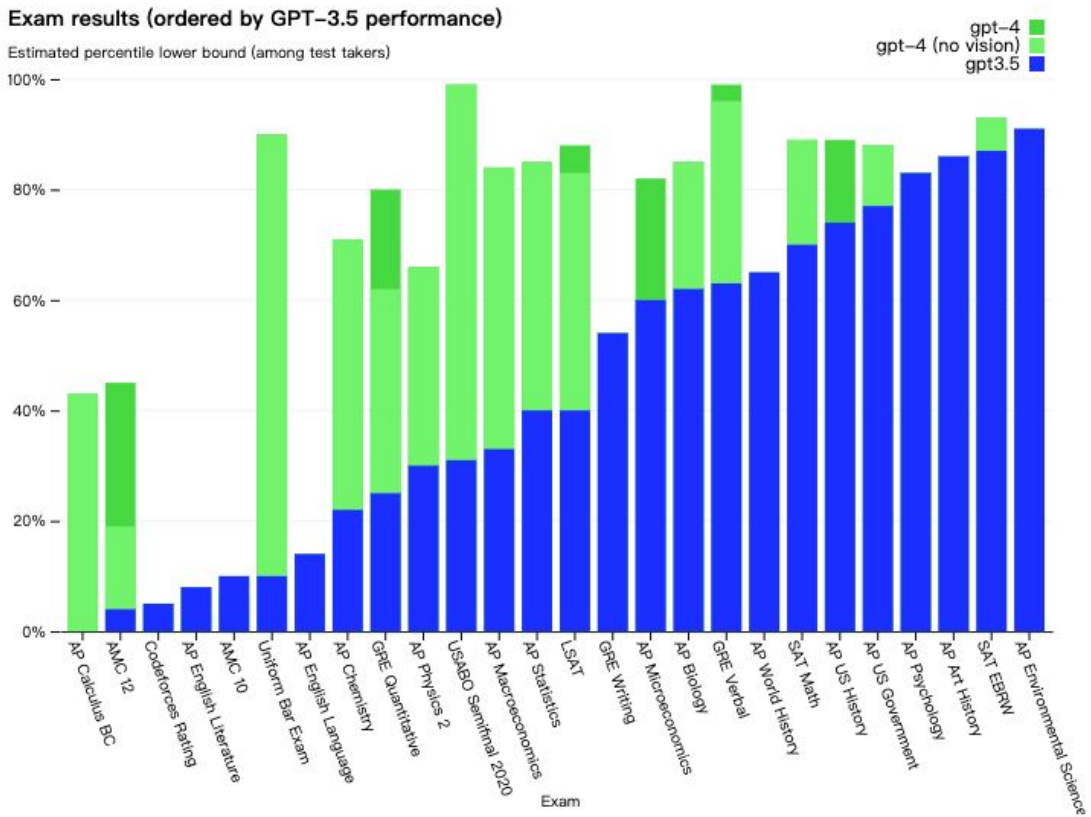
Source: https://www.barnorama.com/wp-content/uploads/2016/12/03-Confusing-Pictures.jpg

| | |
|---|---|
| GPT-4 | The unusual thing about this image is that a man is ironing clothes on an ironing board attached to the roof of a moving taxi. |

**Table 16.** Example prompt demonstrating GPT-4's visual input capability. The prompt requires image understanding.

# 現在的大型語言模型到底多厲害

- GRE
- SAT
- …



# Outline

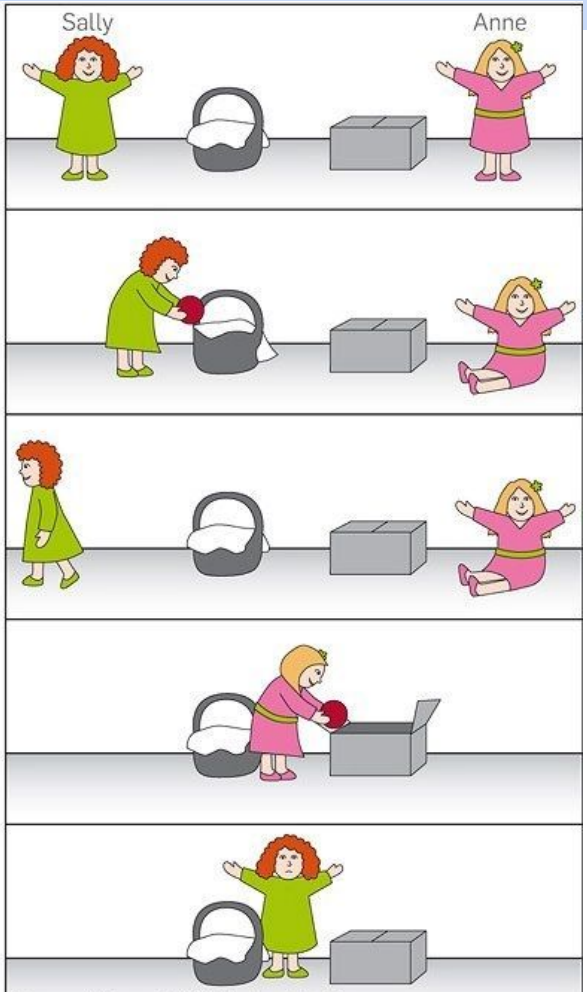
1. ChatGPT的關鍵技術
2. 後 ChatGPT 時代的可能
3. ChatGPT 的心智
4. ChatGPT 的限制

# Sally-Anne Test

**Theory of mind (ToM):**
心智理論，是一種推論或代入他人心智狀態的能力

你覺得 Sally 回來房間後，會先從哪裡找球呢？

籃子！




1. 房間裡有 Sally、Anne、籃子、箱子

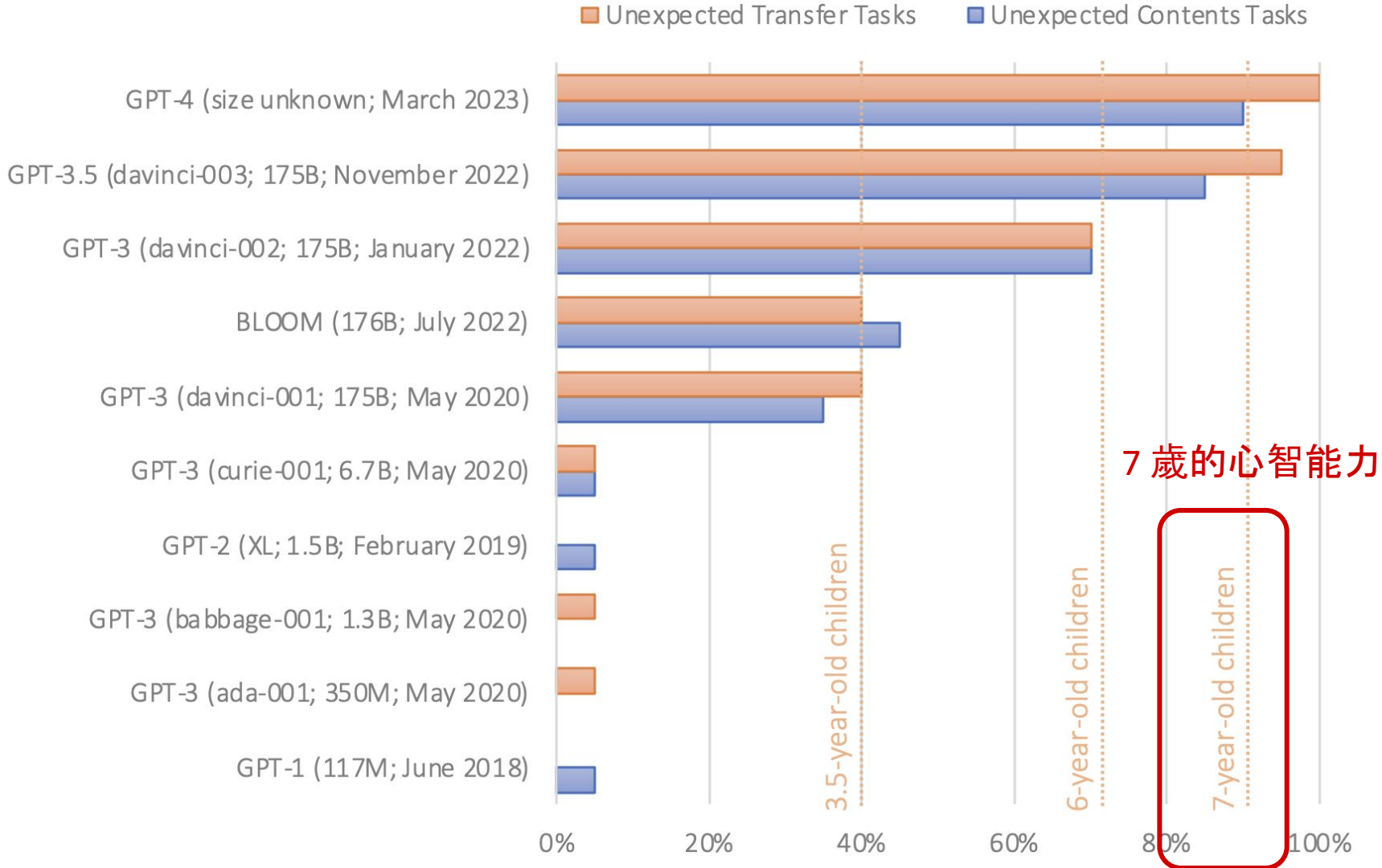
2. Sally 把一顆球放到籃子裡面

3. Sally 離開房間

4. Anne 把球從籃子拿出來放到箱子裡，Anne 也離開房間

5. Sally 回來房間

Kosinski, Michal. "Theory of mind may have spontaneously emerged in large language models." arXiv'23


Unexpected Transfer Tasks
Unexpected Contents Tasks
GPT-4 (size unknown; March 2023)
GPT-3.5 (davinci-003; 175B; November 2022)
GPT-3 (davinci-002; 175B; January 2022)
BLOOM (176B; July 2022)
GPT-3 (davinci-001; 175B; May 2020)
GPT-3 (curie-001; 6.7B; May 2020)
GPT-2 (XL; 1.5B; February 2019)
GPT-3 (babbage-001; 1.3B; May 2020)
GPT-3 (ada-001; 350M; May 2020)
GPT-1 (117M; June 2018)
3.5-year-old children
6-year-old children
7-year-old children
7 歲的心智能力
0%
20%
40%
60%
80%
100%

# ChatGPT 模擬市民


Generative Agents: Interactive Simulacra of Human Behavior

# Outline

1. ChatGPT的關鍵技術
2. 後 ChatGPT 時代的可能
3. ChatGPT 的心智
4. ChatGPT 的限制

# ChatGPT 是一個黑盒子

A Black-Box owned by OpenAI

# ChatGPT 是一個黑盒子

1. 訓練一顆 GPT-3 用了10,000 張 V100
2. 10,000 V100 = 30億新台幣


NVIDIA

NVIDIA Tesla V100 32GB PCIe 3.0

$299,000

- GPU 架構 NVIDIA Volta
- NVIDIA Tensor 核心 640
- NVIDIA CUDA® 核心 5,120
- 雙精度浮點運算效能 7 TFLOPS

看更多

P幣 全盈+PAY單筆滿5500送388 P幣(每帳號限

登記抽 APP指定品累積滿$3,000登記抽【麥當勞

登記抽 【第1波】全站指定品單筆滿$5,000登記

付款方式 信用卡、行動支付，與其他多種方式

出貨 廠商出貨 本商品不受24h到貨限制

The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink

1. 訓練 GPT-3 : 1,287 百萬度
2. 台北一天用電量 40 百萬度
3. 訓練一顆 ChatGPT 相當於台北一個月的用電量

The Carbon Footprint of Machine Learning Training Will Plateau, Then Shrink

# 大型語言模型們

- Google
- Microsoft
- Meta
- 百度
- …

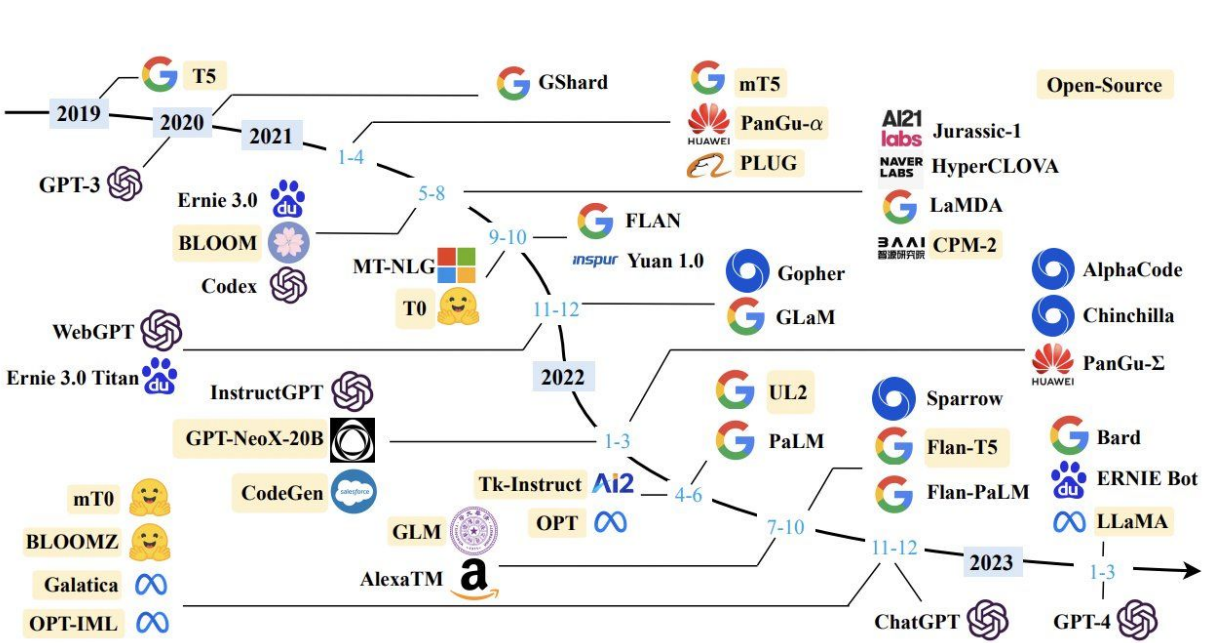
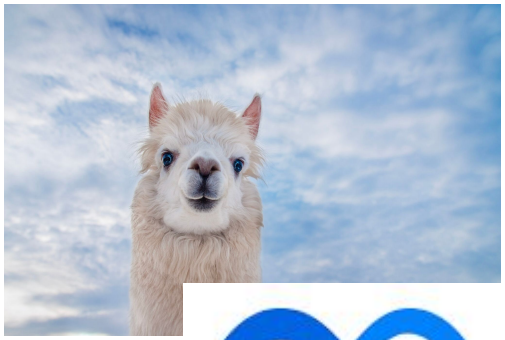



Fig. 1. A timeline of existing large language models (having a size larger than 10B) in recent years. We mark the open-source LLMs in yellow color.

# 大型語言模型的民主化

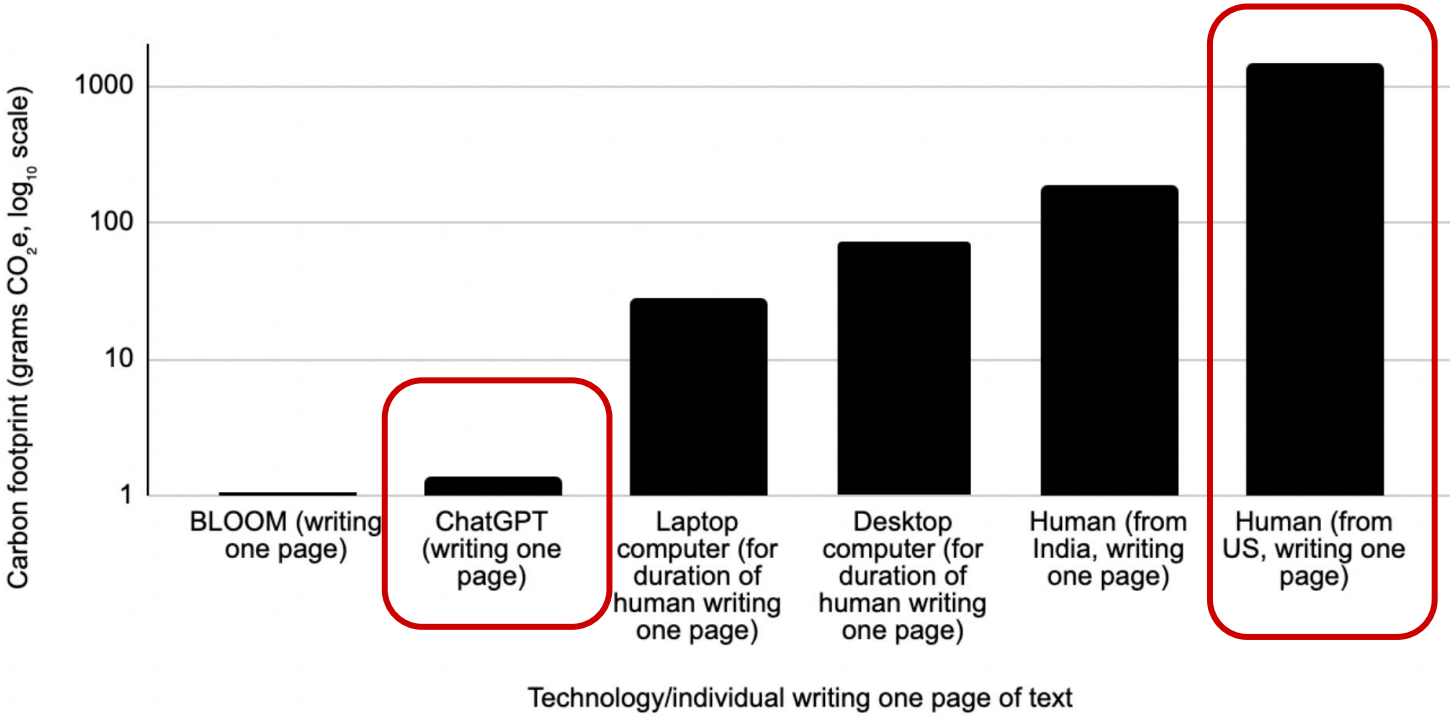
- LLaMA (**L**arge **La**nguage **M**odel **M**eta **AI**)（大羊駝）
- Alpaca（羊駝）
- Vicuna（小羊駝）



Stanford Alpaca

CARNEGIE MELLON UNIVERSITY
PITTSBURGH PENNSYLVANIA 1900

# 寫一篇文章的碳排放量

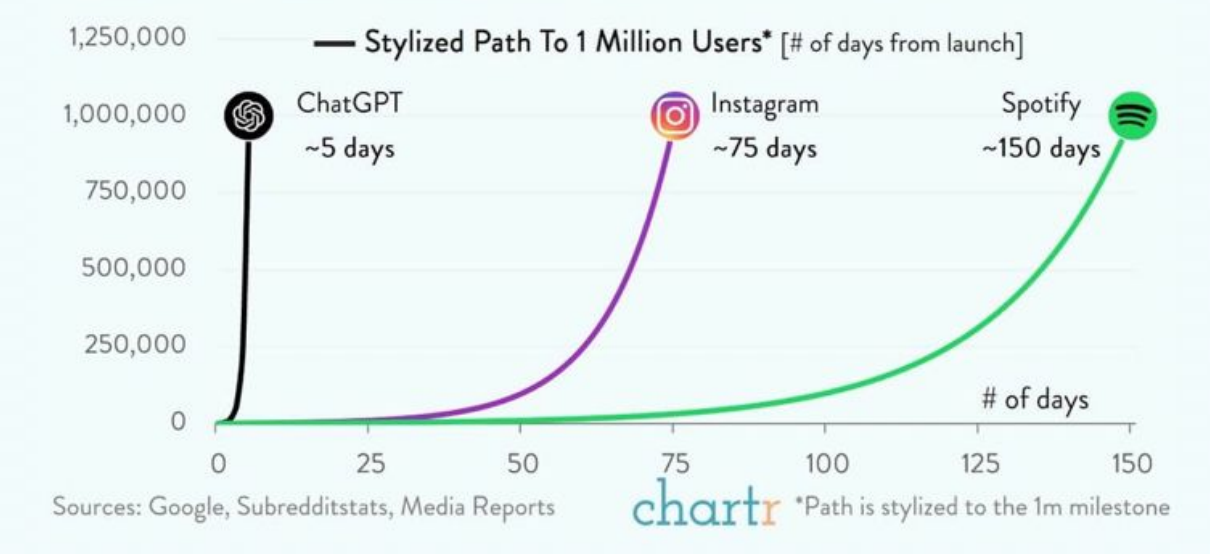



ChatGPT 寫文章比人類寫文章的碳排還要少1000倍

The Carbon Emissions of Writing and Illustrating Are Lower for AI than for Humans

# Takeaway

1. ChatGPT 用的不是全新的技術
2. 但 ChatGPT 跟以前的人工智慧是完全不一樣的



# Q & A