

# 政府部門如何導入ChatGPT

資訊工業策進會·數位轉型研究所

主講人：李宏儒

2023/05/02



# 大綱

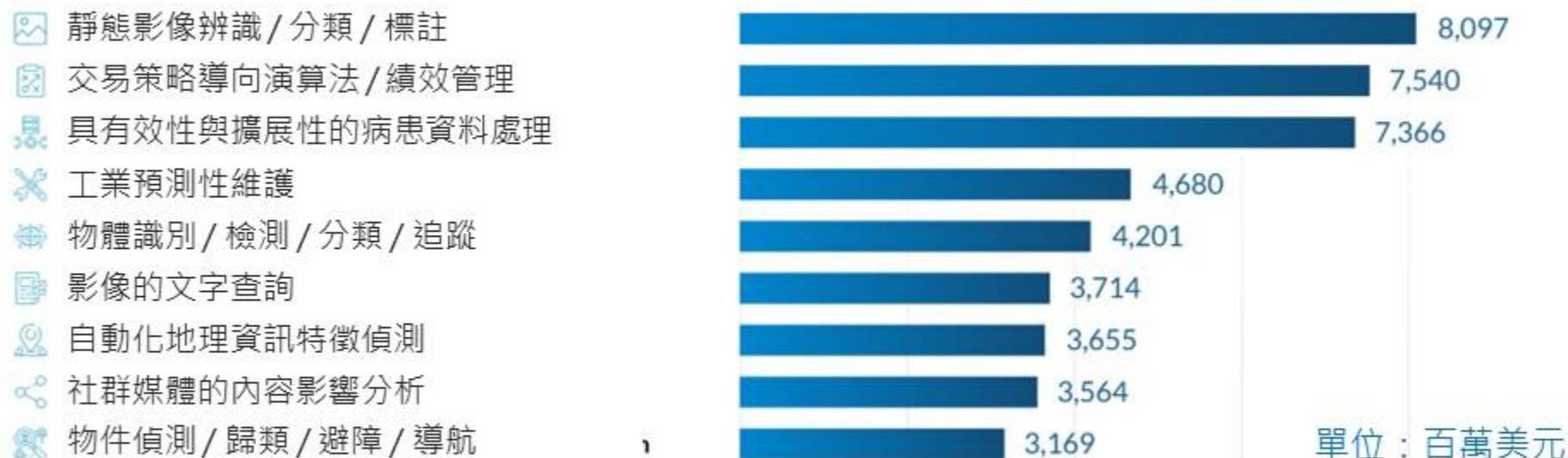
- GPT技術起源與內容
- GPT應用說明
- 政府部門如何導入
- 新科技新技術 – 新議題



# GPT技術起源與內容

## 全球人工智慧發展趨勢

### ➤ 依應用領域，至2025年之全球 AI 收益排名預測



### ➤ 各國人工智慧技術之普及率



資料來源：<https://financesonline.com/ai-trends/>



# GPT技術起源與內容 人工智慧發展里程碑

## AI 1.0

- 單一資料集(Dataset)
- 勞力密集型的資料標註(Label)
- 資料集與模型(Models)各自處理  
缺乏整合綜效

優勢	劣勢
資料越多，結果越好	資料收集與清理成本高
可應用於多個領域	領域應用之優化是離散
AI模組可通用	AI模組無法有效成平台
目標函式可優量化	缺乏認知理解等能力

## AI 2.0

- 超級海量資料，不須人工勞力標註
- 跨知識領域的基礎模型，可適應不同應用服務任務
- 可形成認知理解與自動擴展之資料

優勢	劣勢
無須人工勞力標註 可透過大規模之文本自動提取知識	受不當資料影響易造成偏見與超出內容規範而生成不公平/有害結果
可應用於多個領域 透過微調訓練適應任務	AI無法識別訓練與交互學習歷程是否正確
AI模組可擴模態通用	模型訓練/營運成本很高
可形成認知並持續生成與自動擴展	資料生成過程難以理解 缺陷易被盲目繼承

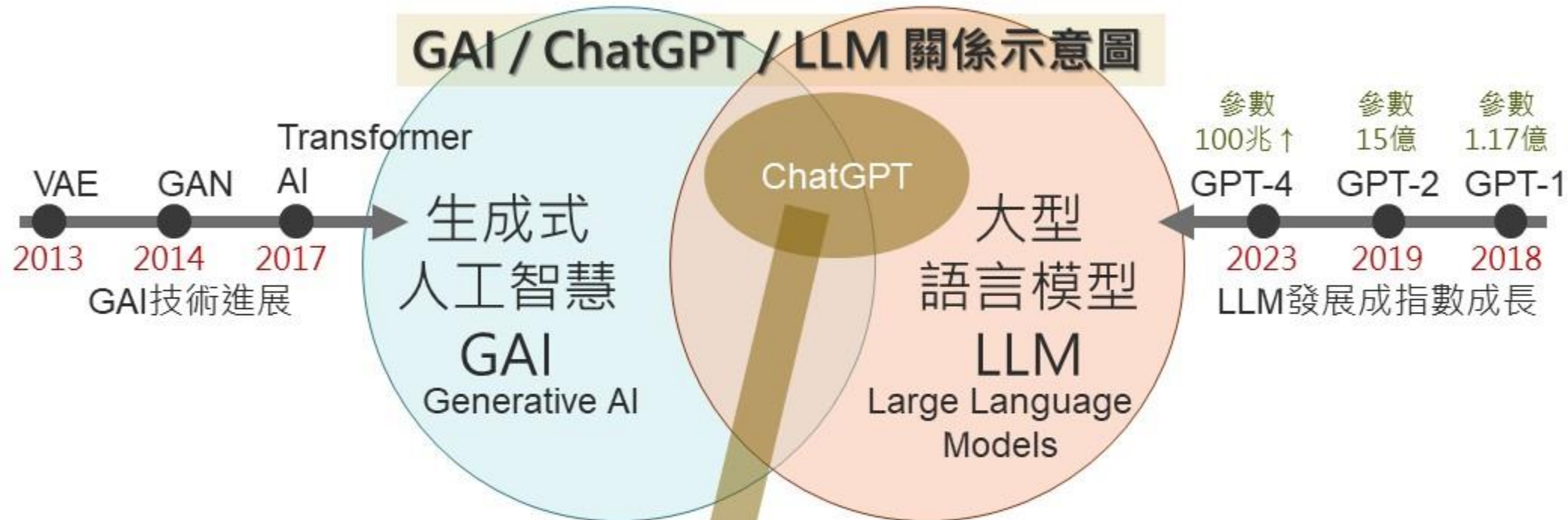
資料來源：Next論壇 – 多模態基礎模型 李開復博士



# GPT技術起源與內容

## 人工智慧2.0的起點-GAI/LLM/ChatGPT

### GAI / ChatGPT / LLM 關係示意圖



- 可生成新的文字/影像/聲音/視訊資料
- 可藉由大量的資料訓練，進行機器學習生成新的資料會與原始資料相似

- LLM 是建立在Transformer架構上的模型
- 現有LLM是採用Transformer與語言建模
- 透過大量文本進行Transformer訓練和語言建模，可針對現有文本進行分析與理解

OpenAI  
聊天機器人

資料來源：wiki-GAI & LLM  
資料整理：資策會數轉院



# GPT技術起源與內容

## 資料生成式服務 - ChatGPT



ChatGPT

### 公司簡介

- **開發者**：OpenAI，2015年成立於舊金山，員工約100多人
- **用戶量**：1300萬個活躍用戶/天
- **運算源**：Microsoft Azure
- **LLM模型**：GPT-3 (Generative Pretrained Transformer 3)
  - ❑ 基於人類反饋增強學習演算法 (RLHF Algorithm) 訓練模型，共有**1750億**個參數
  - ❑ 運用**3000億**個Word進行訓練資料集大小為**570GB**
  - ❑ 資料集從網路爬搜資料、書籍維基百科等等
  - ❑ 耗資**8億美元**建置硬體環境一次訓練成本約**1200萬美元**營運成本高達**60億美元/年**

### 服務提供

- ✓ 問答對話
- ✓ 草稿生成
- ✓ 文章生成
- ✓ 重點生成
- ✓ 郵件生成
- ✓ 行程規劃
- ✓ 語言翻譯
- ✓ Python 程式撰寫
- ✓ Prompt 繪圖
- ✓ 遊戲角色模擬
- ✓ 腦力激盪

### 產品類型

### 商業模式/收費機制

- **個人免費版**：2022/11開放全球測試使用
- **個人付費版**：2023/02開放 ChatGPT PLUS **20美元/月**
- **企業商用版**
  - ❑ ChatGPT API：GPT-3.5-turbo模型，**0.002 美元/1000 tokens**；比 text-davinci-003 便宜約10倍
  - ❑ Whisper API：提供 STT (Speech To Text) 服務，**0.006美元/min**；支援英語/華語/日語等數十種語言

資料來源：wiki - Open AI  
資料整理：資策會數轉院



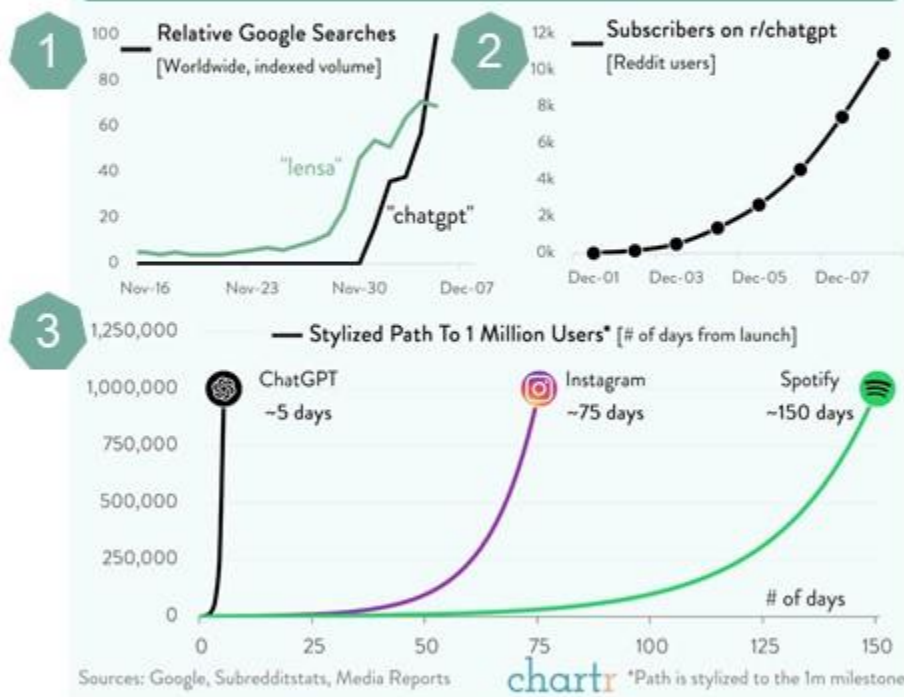
# GPT技術起源與內容

## OpenAI的ChatGPT引領世界AI風潮

- 1 ChatGPT 於2022年11月下旬開發使用，至11月底止，**5天內達到100萬個註冊使用者**。
- 2 兩月內**累積超過1億位使用者**。
- 3 成長速度比 Instagram快 **15倍** 比 Spotify快 **30倍**。

1. ChatGPT 經過測試，其**平均智商約83**，SAT測驗1020/1600分，**已經達到美國高中生考大學的平均水平**。
2. ChatGPT被公認為優於大多數的NLP模型。

### ChatGPT引領風暴式AI成長



資料來源：chartix  
資料整理：資策會數轉院



# GPT應用說明

## 新創轉向·典範轉移

傳統 AI 新創 立足方式

取代 高度客製化專案

承接高度客製化專案

瞭解產業特性

電腦視覺

商品推薦

語音辨識

語音合成

.....

垂直應用

燒錢

缺乏 資料科學家  
缺乏 真實文本資料  
缺乏 AI落地範例

需要 高速運算環境  
需要 簡單複製營利模式  
需要 降低學習曲線成本

圓滿

運用 通用AI 架構

多模態的基礎模型

簡單複製營利模式

訂閱制

基礎框架平台+微調

可理解式NLP

可解釋性演算法

資料整理：資策會數轉院





# GPT應用說明

## 通用 AI 架構 – 多模態基礎模型技術

- 2017年，NeurIPS科學會議的一篇論文加速了AI世界的演進 - **「Transformer」深度學習架構的基礎模型**。
  - 大型語言模型(LLM)之革命始於Transformer
  - Transformer訓練透過隨機遮蔽部分文本並嘗試預測缺失，透過多次迭代運算，調整Transformer參數。

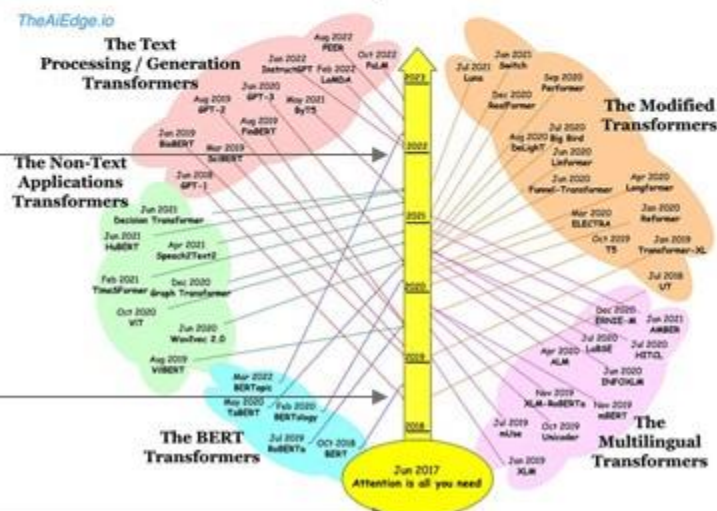
**2022** PaLM 模型 · 540B 模型參數

**2018** Bert 模型 · 110M 模型參數

**2017** Transformer模型 · 十億模型參數

成長5000倍

### Transformers History Timeline



資料來源：Damien Benveniste, PhD

資料整理：資策會數轉院

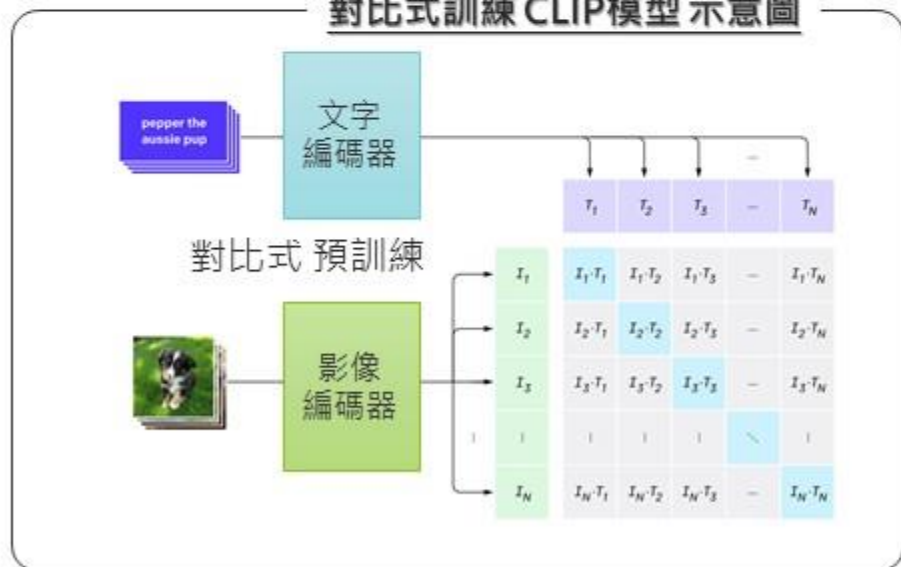


# GPT 應用說明

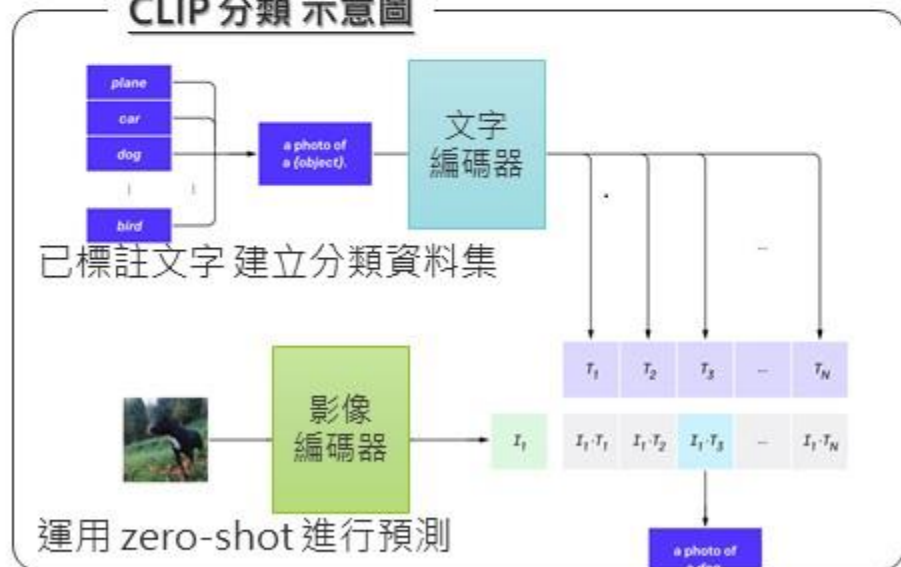
## ChatGPT 帶動 AI 互動式 Prompt 服務

- 多模態預訓練 (**CLIP**; Contrastive Language-Image Pre-training) 模型是2020年OpenAI提出的架構
- CLIP在**整合文字與影像**上有非常大的突破，能透過描述文字找到相對應的影像；適當的文字並不容易，概念非常相似的文字，結果也會有：使用不同的文字描述相同概念 **Prompt Engineering**。

對比式訓練 CLIP 模型 示意圖



CLIP 分類 示意圖



資料來源：OpenAI CLIP  
資料整理：資策會數轉院



## ChatGPT 48種 服務應用情境

01.問答

02.給二年級學生總結

03.命令文本

04.Stripe API 的自然語言

05.解析非結構化數據

06.Python 到自然語言

07.計算時間複雜度

08.高級推文分類器

09.關鍵字

10.產品說明中的廣告

11.TL;DR 總結

12.電子表格創建者

13.ML/AI語言模型導師

14.推文分類器

15.SQL請求

16.JavaScript 到 Python

17.心情變色

18.類比製造商

19.微型恐怖故事創作者

20.總結說明

21.論文大綱

22.聊天

23.轉彎方向

24.創建學習筆記

25.語法修正

26.自然語言到 OpenAI PI

27.英語到其它語言

28.SQL翻譯

29.分類

30.電影到表情符號

31.翻譯程式語言

32.解釋代碼

33.如實回答

34.產品名稱產生

35.python錯誤修復程序

36.JavaScript聊天機器人

37.科幻書單製作具

38.機場代碼提取器

39.提取聯繫信息

40.好友聊天

41.編寫 Python 文檔字符串

42.JavaScript 一行函數

43.第三人稱轉換器

44.VR健身創意生成器

45.食譜創作者 (風險自負)

46.Marv 諷刺聊天機器人

47.餐廳評論創建者

48.面試題

資料來源：OpenAI Examples

資料整理：資策會數轉院



# 政府部門如何導入ChatGPT 減少行政文書負擔，提升工作效率

➤ 運用「ChatGPT」&「語音辨識技術」協助行政文件整備，減少行政文書處理負擔。

### 上傳會議語音檔/資料

III ASR 語音辨識系統

姓名：  
  
 Email信箱：  
  
  
 上傳進度：20% · 稍後至 hlee@ii.org.tw 收取信件通知結果

iiiasrcloud@gmail.com  
收件者: 李宏偉 Hong Ru Lee



親愛的 李宏偉您好：

非常感謝您使用資策會數位轉型研究所的ASR 語音辨識系統 <http://61.66.218.102:30303/>。您的轉譯任務 ACE\_1 (請閱長度: 00:01:57) - 編號: 1682499476282 已經送出 - 預計 2023/04/26 16:59:55 完成。

資策會數位轉型研究所  
ASR 系統: iiiasrcloud@gmail.com

### 回覆處理狀態與預期時間

### 回覆處理結果以及附檔

iiiasrcloud@gmail.com  
收件者: 李宏偉 Hong Ru Lee



iiiasrcloud@gmail.com  
iiiasrcloud@gmail.com

親愛的 李宏偉您好：

非常感謝您使用資策會數位轉型研究所的ASR 語音辨識系統 <http://61.66.218.102:30303/>。您的轉譯任務 ACE\_1 (請閱長度: 00:01:57) - 編號: 1682499476282 已經送出 - 預計 2023/04/26 16:59:55 完成。

敬請 謹此致謝。

日期: 2023/04/26

時間: 16:59:55

地點: 資策會數位轉型研究所

IP地址: 61.66.218.102

用戶代理: Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/110.0.0.0 Safari/537.36

轉譯任務: ACE\_1

### 可運用技術

- III ASR技術 語者辨識
- 支援 Online/Offline



➤ 運用OpenAI API 整合III NLP技術，降低使用Token量

- III ASR技術 語音辨識
- 支援 Online/Offline
- 辨識率優於 Google/OpenAI

#### 會議記錄

會議紀錄:

日期: 待填。

與會人員: 待填。

會議主題: 界定關鍵問題、問題。

會議紀錄:

本次會議的主題為「界定關鍵問題」，主要討論的內容包括：事實上該如何提問，問題又該如何說，以及如何從一個題目中提取很多問題。讓大家都覺得不痛苦。此外，還討論了第二個話題「老闆關心問題」，該如何確定問題的 owner，以及題目部門該如何落實。

下一步行動:

1. 各部門需要確定問題的 owner，並將其落實到題目中。
2. 各部門需要對問題進行深入的研究，以便對問題進行更有效的解決。
3. 各部門需要對問題進行更多的討論，以便對問題進行更有效的解決。

待辦事項:

1. 各部門需要在指定的時間內完成對問題的提問和討論，以便對問題進行更有效的解決。
2. 各部門需要在指定的時間內完成對問題的討論和提問，以便對問題進行更有效的解決。



Microsoft Word 文件

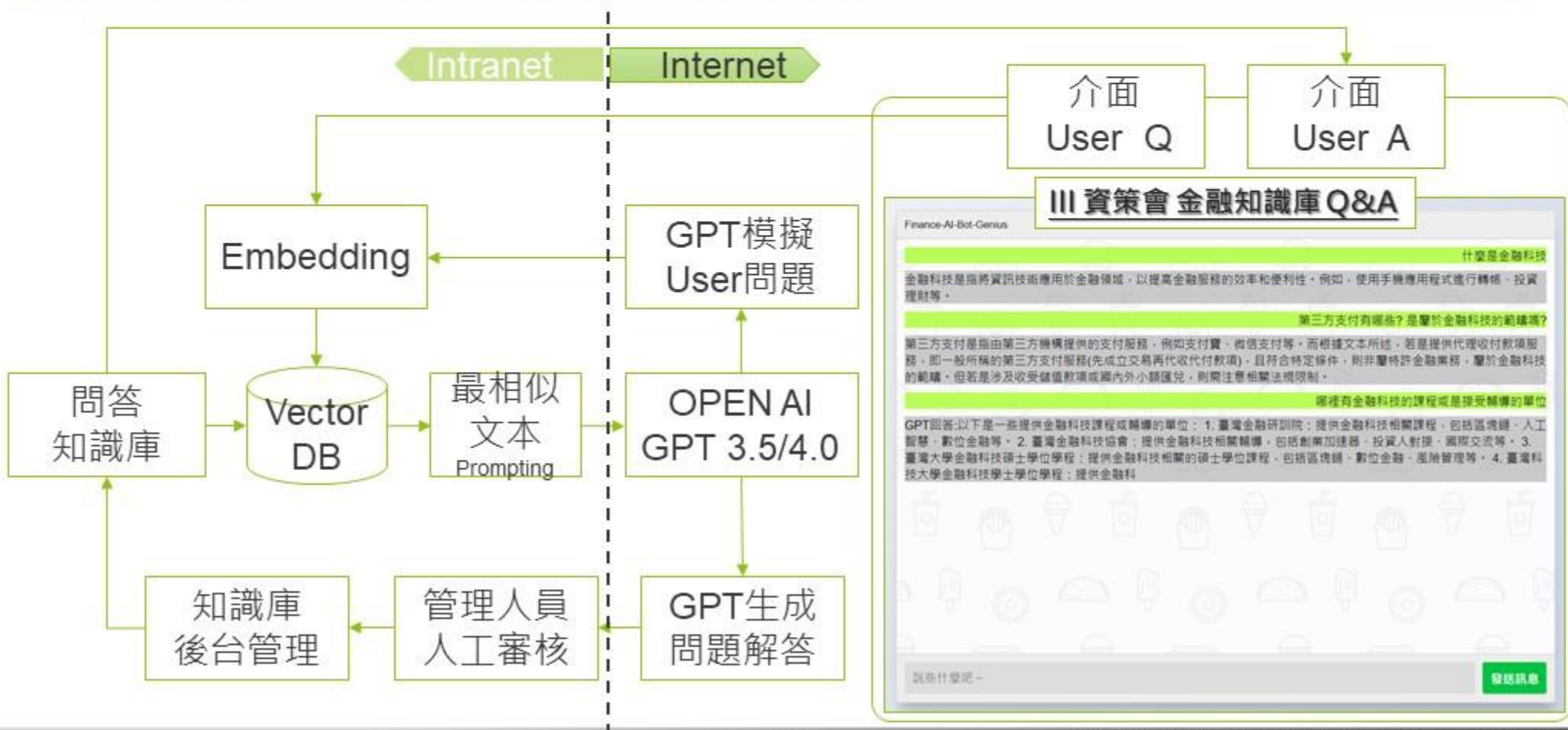
資料整理：資策會數轉院





## 建立產業趨勢知識庫，提供新科技服務

- 運用「ChatGPT」&「NLP自然語言處理技術」協助建置產業領域知識庫，提供相關知識領域的Q&A。





# 新科技新技術 – 新議題

## 跨越AI變革與衝擊

- 人工智慧技術進展神速，AI應用早已遍地開花並持續蓬勃發展，也成為衡量國家、企業及個人等競爭力的指標。

AI技術具有跨領域之特性，需要進行AI跨域合作研究、擴大產官學合作

- 盤點技術團隊AI發展現況/需求與方向
- 以產業鏈為架構，技術分類為輔，訂定研究定位與目標：產業 Domain + AI跨域學習



- 掌握地方產業痛點與AI應用需求
- 累積AI基礎研究/資料

- 連結法人單位，共同交流AI技術與產業發展策略
- 以產官學既有AI技術人才，激發各部門AI新議題



# 新科技新技術 – 新議題

## 訂定AI科技倫理規範/政策

- AI 的發展將對個人隱私、社會倫理、市場競爭、網路安全、產業安全等各方面產生深遠影響。

### 訂定AI科技倫理規範與指引，改善/降低AI發展所衍伸的問題與風險

- AI 模型的資料集有偏誤
- AI 資料清理方式有偏誤
- AI 特徵萃取設定有偏誤
- AI 數據比例權重有偏誤

演算法  
偏誤議題

AI應用  
倫理議題

- 應用範疇偏離原使用目的
- 深偽技術(Deepfake)原應用於資料集不足，後被應用於偽造名人性愛影片等

- AI 技術先天具有善惡兩種用途(Dual-Use)
- AI 人臉辨識技術可保護數位裝置，亦可窺探隱私
- 無人機技術可協助偏遠探勘，亦可成為攻擊武器

AI技術  
濫用議題

AI技術  
受限議題

- 演算法設計不良與技術能力有限，衍伸誤判與風險
- 自駕技術會因交通標誌於現實中老舊髒污導致誤判

資料整理：資策會數轉院





# 新科技新技術 – 新議題

## 因應AI應用議題 與時俱進

- Z世代之消費、工作與社交活動皆轉往網際網路，加速自然語言、語音辨識/合成/翻譯、文本/圖片生成之演進。

### 強化 AI風險識別與治理能力，洞悉市場缺口與新機會

AI 能仿冒全球將近 50%人臉，

法律規範尚未完善，涵蓋技術壟斷、版權歸屬、深偽造假以及人權道德等議題之風險，政府**監管難度**高。

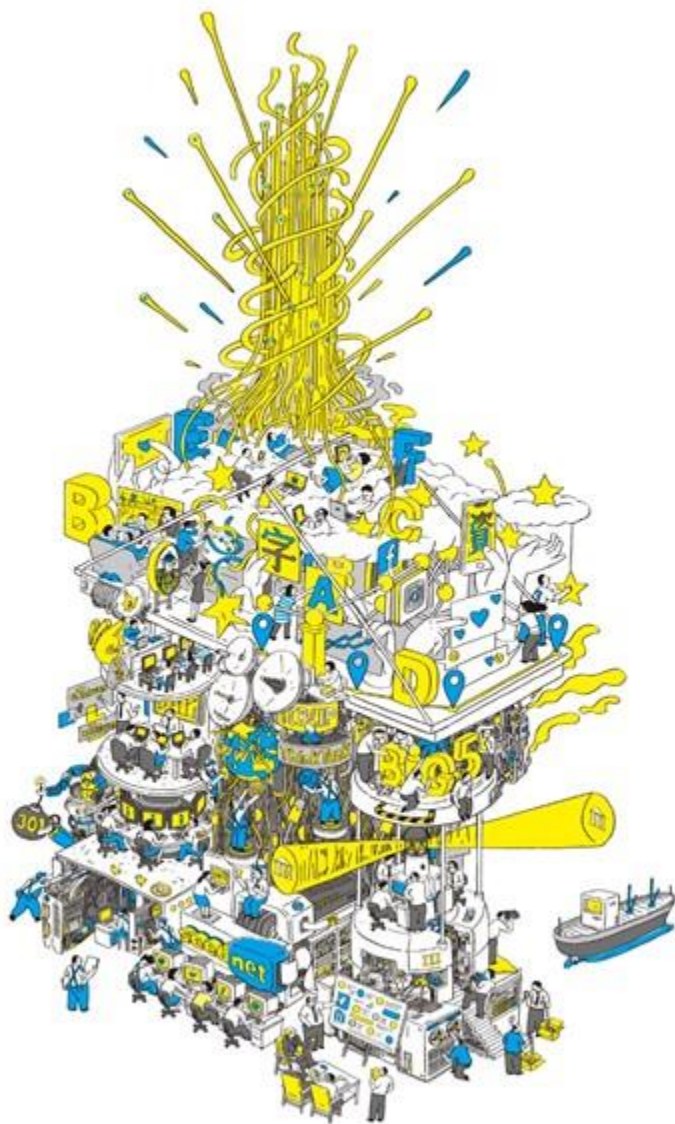
AI 合成人臉比真實人臉，更能取得一般人信賴

不當操作模型生成所需要的多媒體內容，有可能導致模型做出不正確或是惡意的決策，衍伸**信任**議題。

Deepfake 滲透社群與數位平台，產生誤導性內容

文字生成、語意理解表現較過往技術優異，俱備令人信服的文本能力，有心人操弄模型易導致**社會**問題。

資料整理：資策會數轉院



- 1 擘劃我國資訊工業發展藍圖
- 2 開啟電腦中文化時代
- 3 打造台灣資訊品牌
- 4 培養台灣資訊人才
- 5 開創產業顧問服務
- 6 提升網路基礎建設
- 7 E化政府系統
- 8 普及網路應用人口
- 9 建構資訊法案制度
- 10 縮減城鄉數位落差
- 11 推動數位內容
- 12 推動數位科技外交
- 13 策進 e-Taiwan / m-Taiwan
- 14 精進5G智慧科技創新應用
- 15 支援文創與設計產業奠基
- 16 培育創新創業新動能
- 17 擔任數位國家智庫
- 18 活化原鄉無線寬頻環境
- 19 協助產業拓展商機並強化資安防護
- 20 數位轉型化育者

# THANK YOU