



NCHC 國家實驗研究院

國家高速網路與計算中心

National Center for High-performance Computing



# TAIWAN AI RAP大型語言模型 高效能AI平台介紹

<https://rap.genai.nchc.org.tw/>



powered by **NCHC**

HPC |  
DRIVING TRANSFORMATION  
FOR A BETTER FUTURE

財團法人國家實驗研究院  
國家高速網路與計算中心

National Center for High-performance Computing

## Table of Contents

一、TAIWAN AI RAP大型語言模型高效能AI平台介紹

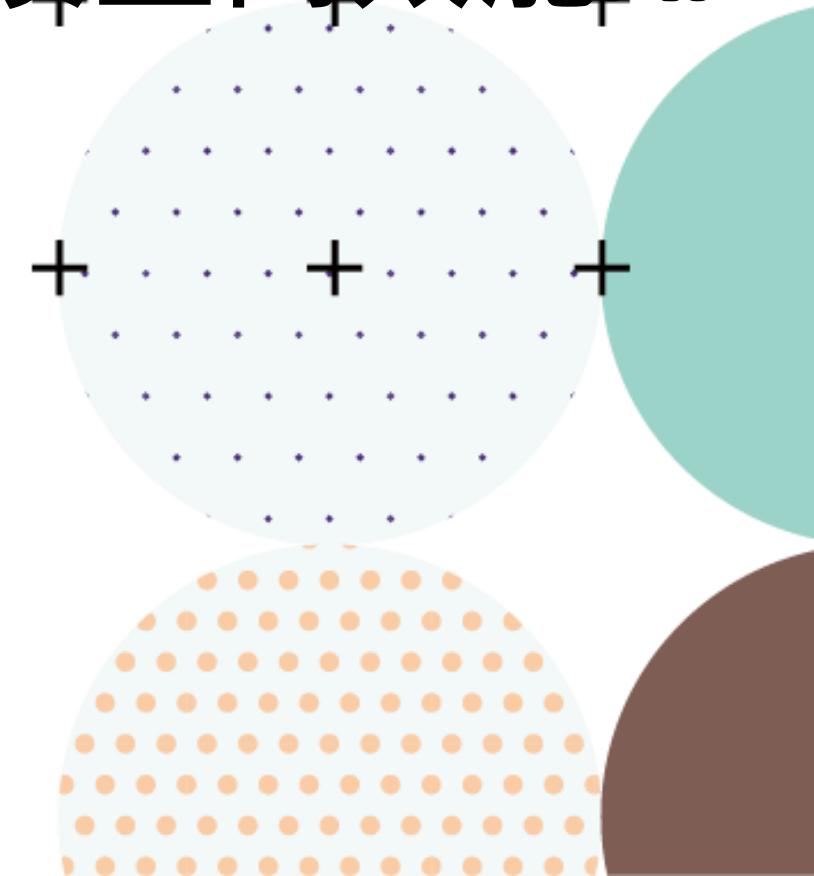
二、PoC 客戶案例分享

三、TAIWAN AI RAP試營運介紹

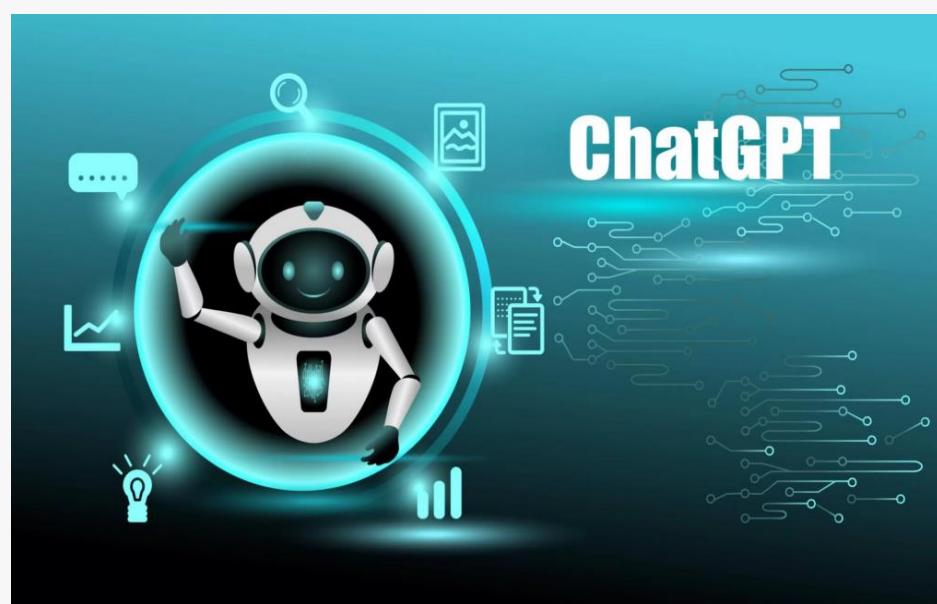
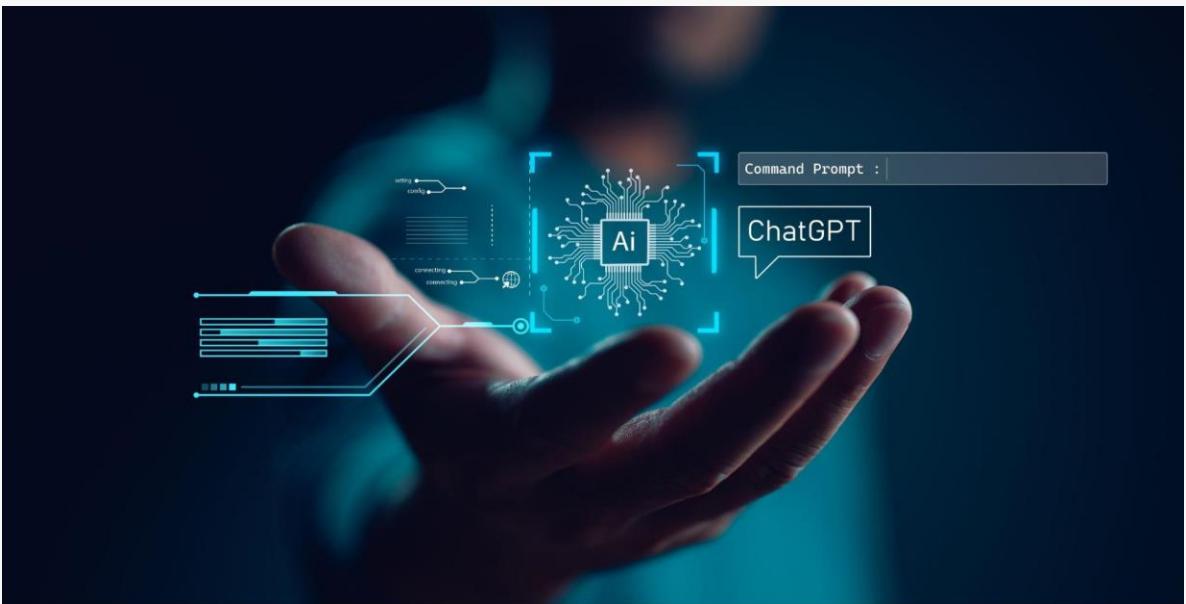
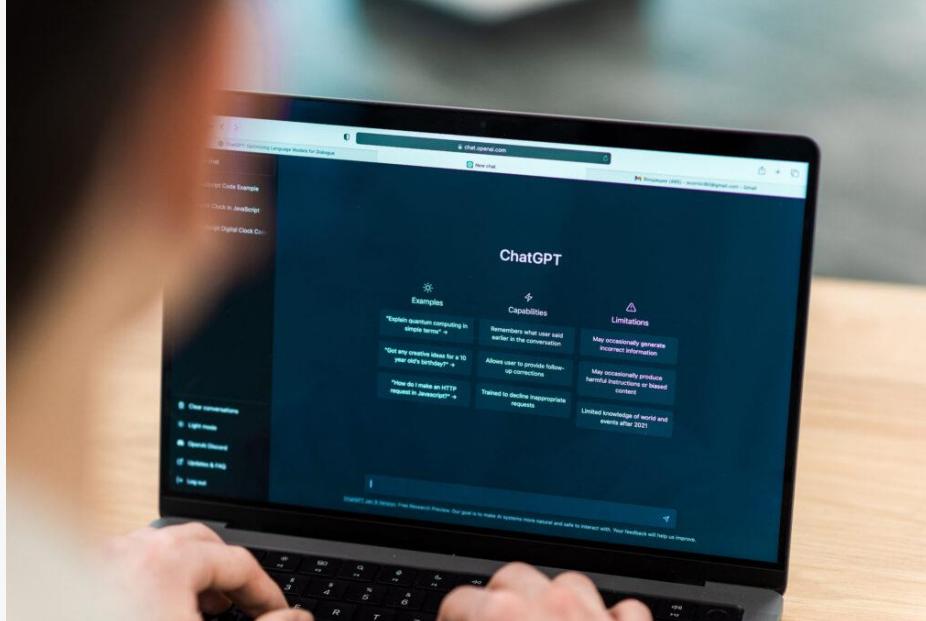
# 一、TAIWAN AI RAP大型語言模型高效能AI 平台介紹



+



# 新趨勢 : 大家發現 ChatGPT 比 Google 搜尋更好用 ...



# AI 人工智慧: 新工業革命的(候選)技術



仍需觀察:  
AI 能否大幅提升社會經濟的生產力?

第一次工業革命

第二次工業革命

第三次工業革命

(第四次工業革命)

未來潛力技術

火力



蒸汽機、內燃機  
鐵路、汽車

電力



馬達、電燈、電話  
電力網、電話網

數位



IC、電腦、手機、網路  
IC 半導體技術

AI



大語言模型、機器人、自駕車

量子計算



## AI 被當成是第四波網路應用熱潮?



## AI 號稱是第四次工業革命?

### 網際網路 Internet 科技發展軌跡

第一階段

Internet

早期科技發展  
(淘金熱 Gold Rush)

1991

使用者: 260萬

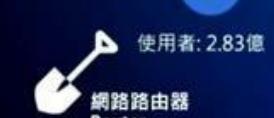


網際路由器  
Router

第二階段

商業模式發展  
(社群生態)

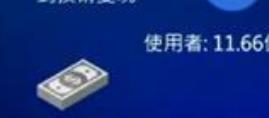
2000



使用者: 2.83億

從吸引注意力  
到技術變現

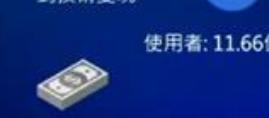
2007



使用者: 11.66億

智慧型手機協助  
網路技術普及化

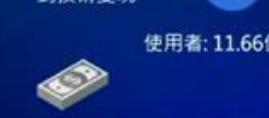
2017



使用者: 36.7億

使用者: 超過50億

2023



使用者: 超過50億

第三階段

技術融合  
啟動技術發展新浪潮  
(協同作用)

AI = 資料 (來自網際網路)  
+ 計算 (來自半導體技術)  
+ 演算法 (來自學術研究)



使用者: 超過50億

使用者: 超過50億

2023

上網設備  
(個人電腦 + 智慧型手機)  
每年全球銷售量

1440萬

14.4 Mu

100%



1億3500萬

135 Mu

100%



3億9300萬

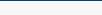
393 Mu

31%

69%



100%



7億3300萬

733 Mu

64%

36%



18%



13億8200萬

1,382 Mu

82%



100%



100%



100%



# 主權AI時代來臨

推動我國主權AI發展，重點於建立自主的資料庫、算力、雲端服務及應用系統



國家核心資料



大型AI算力



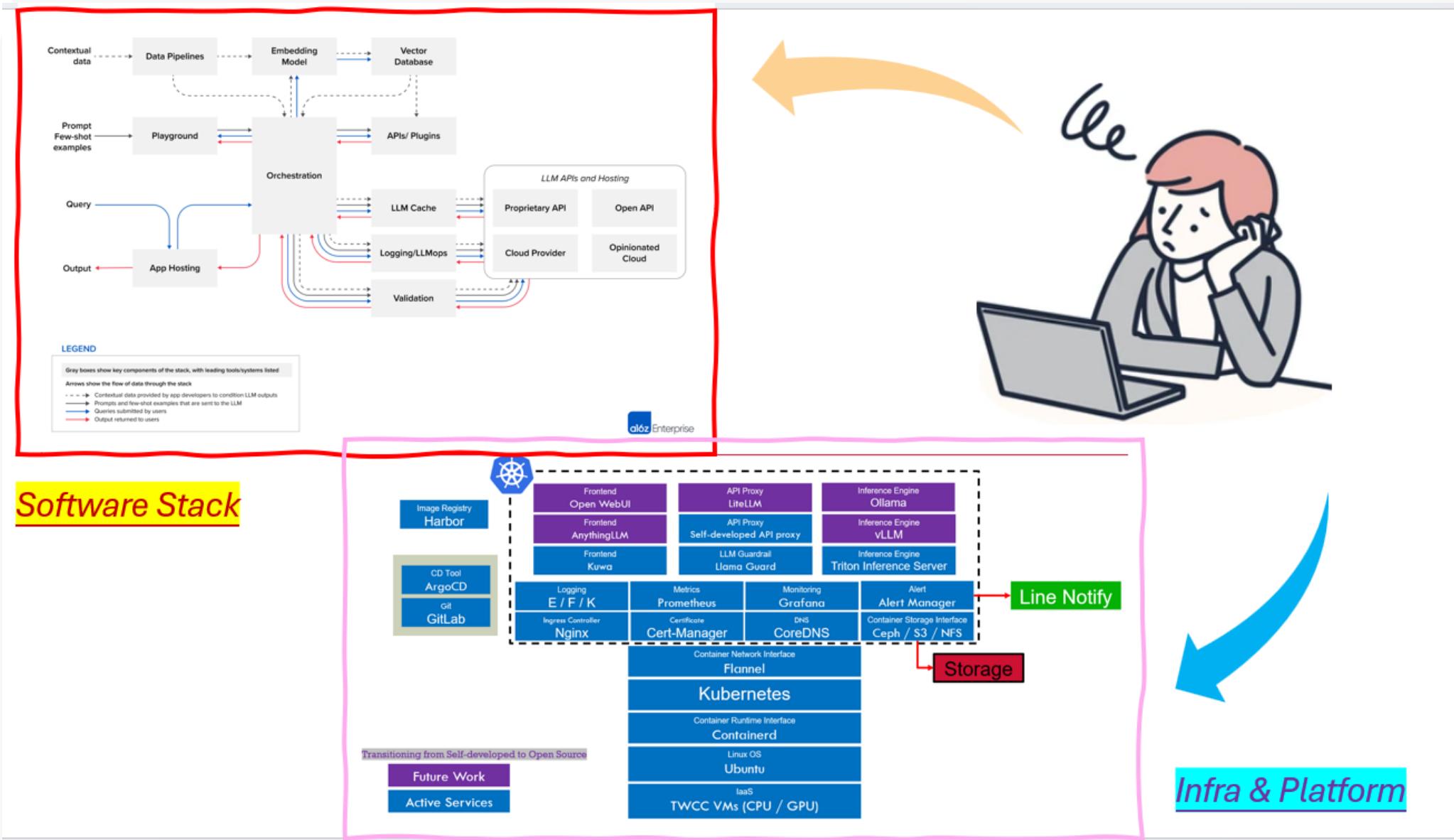
自主雲端服務



## 台灣本土語言模型，捍衛數位主權

推動並擴展TAIDE語言模型在各行各業中的應用潛力

# 想要自建 LLM 應用平台時的困境 ...



# 國網中心可以幫上甚麼忙



國研院國網中心於1991年成立

為我國唯一大型共用計算平台及高等學研專用寬頻網路之服務提供者  
並深耕高效能計算、雲端虛擬化、網路資安、AI、資料分析等核心技術  
為我國之科研能量奠基

國網中心積極配合國家科技政策  
以超級電腦與寬頻網路為基石，  
提供可信賴之巨量資料與AI雲端服務  
為各界打造安全與便利之數位資源



## From Enabler to Enhancer



超級電腦是研發的火種，  
幫助台灣各界研發團隊點燃最亮的成果！

# 國網中心可以幫上甚麼忙：高效能運算核心服務資源



新一代AI超級電腦

16 PFlops



台灣杉二號  
TAIWANIA 2

9 PFlops

2018.11首次排行

TOP500 #20

Green500 #10



台灣杉三號  
TAIWANIA 3

2.7 PFlops

2020.11首次排行  
TOP500 #181  
Green500 #69



Forerunner 1  
創進一號

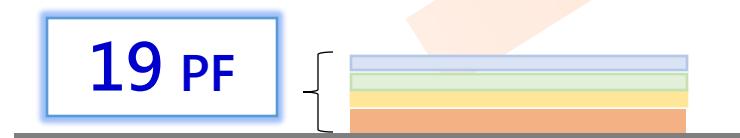
3.5PFlops

2023.11首次排行  
TOP500 #221  
Green500 #91

## 提升25倍

提升優質可靠之高效能算力  
加速AI產業發展

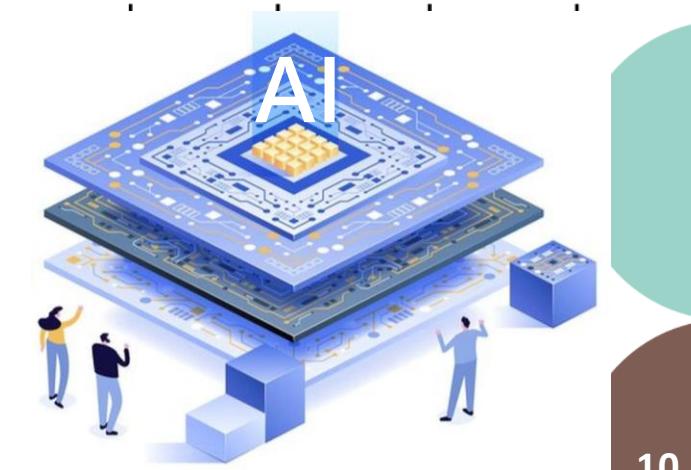
- TAIIDE計畫主機 (2023)
- 創進一號 (2023)  
(原台灣杉四號)
- 台灣杉三號 (2020)
- 台灣杉二號 (2018)



2018 - 2023

2024 - 2029

- 大南方矽谷方案 (2024-2029)  
IDC：台南沙崙綠能科學城  
(2028年完工進駐)
- 晶創台灣方案 (2024-2028)  
IDC：南部科學園區  
(2025年完工進駐)



# 國網中心可以幫上甚麼忙



國網三大核心 (高速計算 + 高速網路 + 高速儲存)

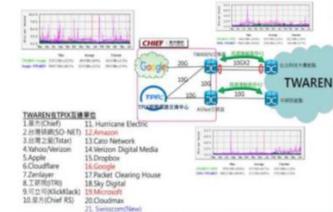
## 國網中心 - 次世代超級電腦

- 台灣杉二號：2,016張V100 GPU
- 台灣杉三號：50,400 個計算核心(Cores)



## 國網中心 - 台灣學術網路骨幹中心

- 與TPIX累積互連單位達21個
- TWAREN與Google建立直連，流量高峰達9.6Gbps
- 完成TWAREN於新加坡交換中心與ASNet線路跨接



建立骨幹光纖服務  
取得服務經營執照  
  
強化我國網路韌性  
完成第一期自建光纖



- 完成台北至台南寬頻網路建置，提升數位服務能量
- 申請骨幹光纖經營權
- 辦理骨幹光纖二期佈纜路徑規劃與服務推展

## 國網中心 - 巨量儲存服務

- 國網中心於新竹、台中、台南建置高效率、高容量之儲存設備，三地互為備援。
- 服務容量大於120PB。
- 擁有多重ISO認證保障，搭配合TWAREN骨幹光纖網路快速不中斷的傳輸，提供完整且多元的資料儲存服務。
- 建置全國專業領域共用大資料儲存資料平台，達到資料共用、研究成果分享之效益

- 每年服務國內腦科學、基因演算、防災、地科、氣候變遷、人文資料庫等30餘重要計畫
- 備份農航所航照圖、太空中心衛星圖、中研院基因體等資料，為國家重要資料提供保障。



# 大語言模型服務平台組扮演的角色 ...



- 算法：論文 + Code
- 數據：找開源的用
- 算力：各憑本事...

• 提供計算力

Hard code



TAIWAN AI RAP



vs



# TAIWAN AI RAP : 定位 -> 願景



## 生成式 AI : LLM 服務平台 (RAP) (RAP : Resilient and High-Performance AI Platform)

### TAIWAN AI RAP 四大特色

- 客製化流程設計前台**  
整合各式 AI 工具與開發流程，打造低門檻、好上手的彈性開發環境，即使無技術背景也能輕鬆操作。支援快速部署與介面設計，讓 AI 應用更快落地
- 多模型API服務**  
內建多款開源語言模型，強化繁體中文處理能力，支援企業快速串接、靈活開發，打造各式 AI 加值服務，滿足各行各業多元需求
- 模型微調與評估**  
簡化模型訓練與資料處理流程，協助使用者提升模型準確性與實用性，讓 AI 更貼近實際應用場景
- AI Tribe 聚落**  
與國內 AI 服務業者合作，讓開發好的 AI 工具上架 TAIWAN AI RAP，打造開放、多元的 AI 聚落，促進跨界共創與產業應用發展

The homepage features the TAIWAN AI RAP logo at the top right. Below it is a banner with the text: "試營運期間為 2025/3/28 ~ 2025/9/30，目前平台持續優化中，部分功能將陸續開放，歡迎使用過程中提供寶貴建議，一起 Resilient and high-performance AI Platform". The main content area has a large "TAIWAN AI RAP" title and a sub-section titled "快速啓動你的AI應用" with the text: "高效能 AI 應用開發平台，整合算力、工具、模型與部署，助你 AI 快速落地！". A "了解更多" button is visible. On the right side, there's a circular graphic of a circuit board with a central "LLM" chip.

Phase 1 : AI Playground (AI 開發試煉所)



Phase 2 : AI Tribe (集各路 AI 好手的聚落)

# TAIWAN AI RAP : 願景 -> 目標

讓國網變成是一個具有 [台灣價值] + [主權AI] 的生成式人工智慧的研發基地 -> 聚落

- 國網中心團隊初期先自行開發與研究 (先求有再求好)
- 開始尋求並導入其他 產/學/研 各界的 solution (包含 tier-1 ~ tier-3)
- 建立多方合作的管道與橋梁 (解決人力不足的問題)

2025

• 以 RAP 平台作為一個 AI 服務的聚落 (AI Tribe)

- 發展 RAP 三層式鬆耦合的服務平台
- 可客製化流程設計服務前台
- 高效能多樣化模型的API服務
- 資料生成精煉、模型客製化微調、模型效能評估三個自動化工作流服務



2026

• 多模態 LLM 模型的研發  
• 多模態模型串聯整合應用

- 開發影像辨識模型 (VLM)、影像生成模型 (diffusion model)
- 開發語音模型 (ASR / STT / TTS)
- 導入三合一 [latent space共用] 的模態架構
- 提供多元的多模態 API 服務



2027

• 提供 Agentic AI 應用環境  
• 發展多元 AI 智慧代理服務

- Agentic AI 可以讓生成式AI 的價值與效益極大化
- 透過多元 AI 智能代理可以協助百工百業來快速解決他們內部的問題



# TAIWAN AI RAP：服務三支箭



## 三大服務

### 1. 高效能多模型的API服務

提供多種開源模型及TAIDE模型，快速上手開發加值服務

\* TAIDE 商用 8B模型、學研用 13B模型、公部門70B模型，台灣繁體中文主體性的加值應用



## 三大服務

### 2. 可客製化流程設計服務前台

靈活、高效整合AI工具與工作流程，不需技術背景也能設計流程

\* 支援客製化工作、快速開發與佈署、減少開發者介面設計整合負擔



提供多種前線介面可選擇，並可整合AI工具與工作流程設計專屬前線與應用服務



## 三大服務

### 3. 模型微調 ( Fine-Tuning)與評估

可完成訓練、資料整理和模型檢查流程，更省時又省力，並讓模型保持最新、最準確

\* 例如零售業可打造更貼合自身品牌風格與消費者喜好的 AI 購物顧問，提升顧客體驗與轉化率



# RAP：客製化流程設計前台



提供一個靈活平台，無需AI技術背景即可輕鬆設計流程，將AI工具與工作流程高效整合，減少開發和整合負擔

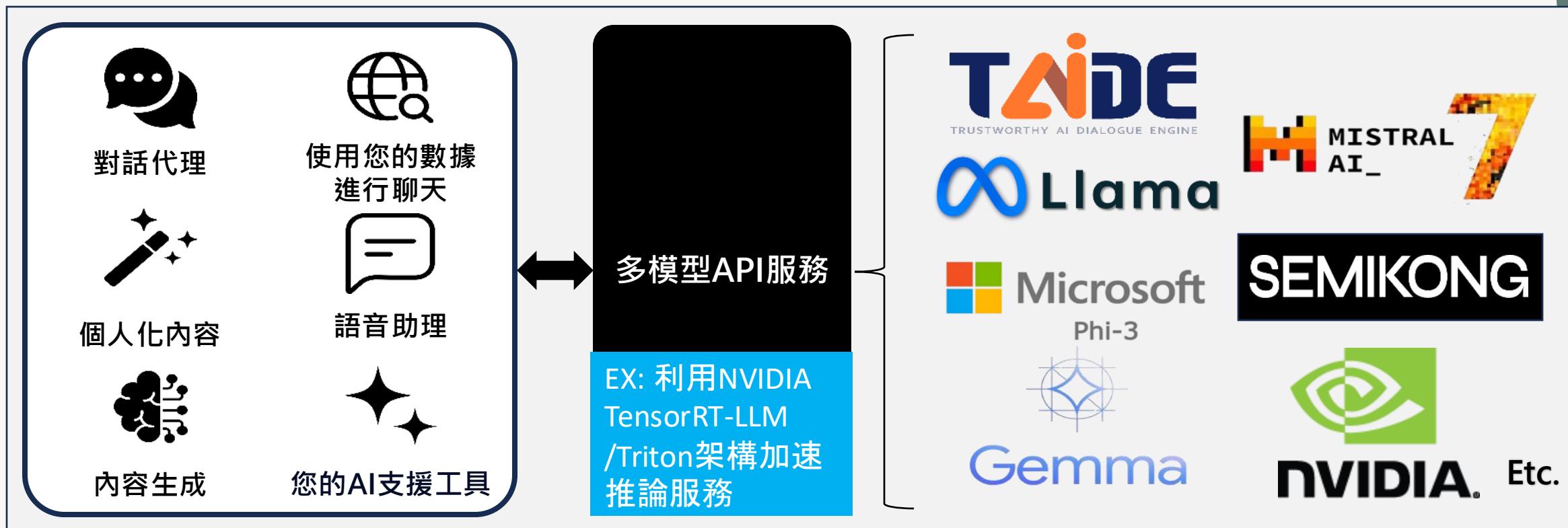


提供多種前端介面可選擇，並可整合AI工具與工作流程設計專屬自己的前端與應用服務

# RAP：多模型API服務



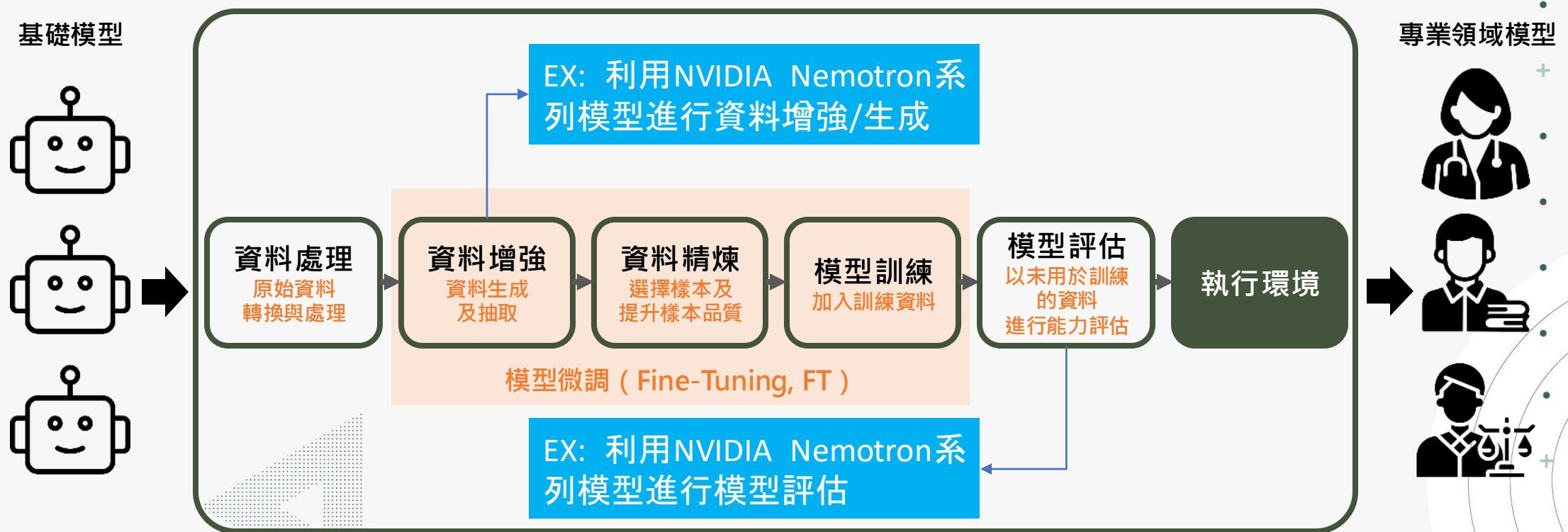
提供多種TAIDE、NVIDIA Nemotron以及各種國際開源模型，快速上手開發附加值服務



# RAP：模型微調與評估



提升AI訓練與效能評估，省時省力。完成訓練、資料整理與模型檢查流程，確保模型保持最新且準確



# TAIWAN AI RAP : 軟體架構

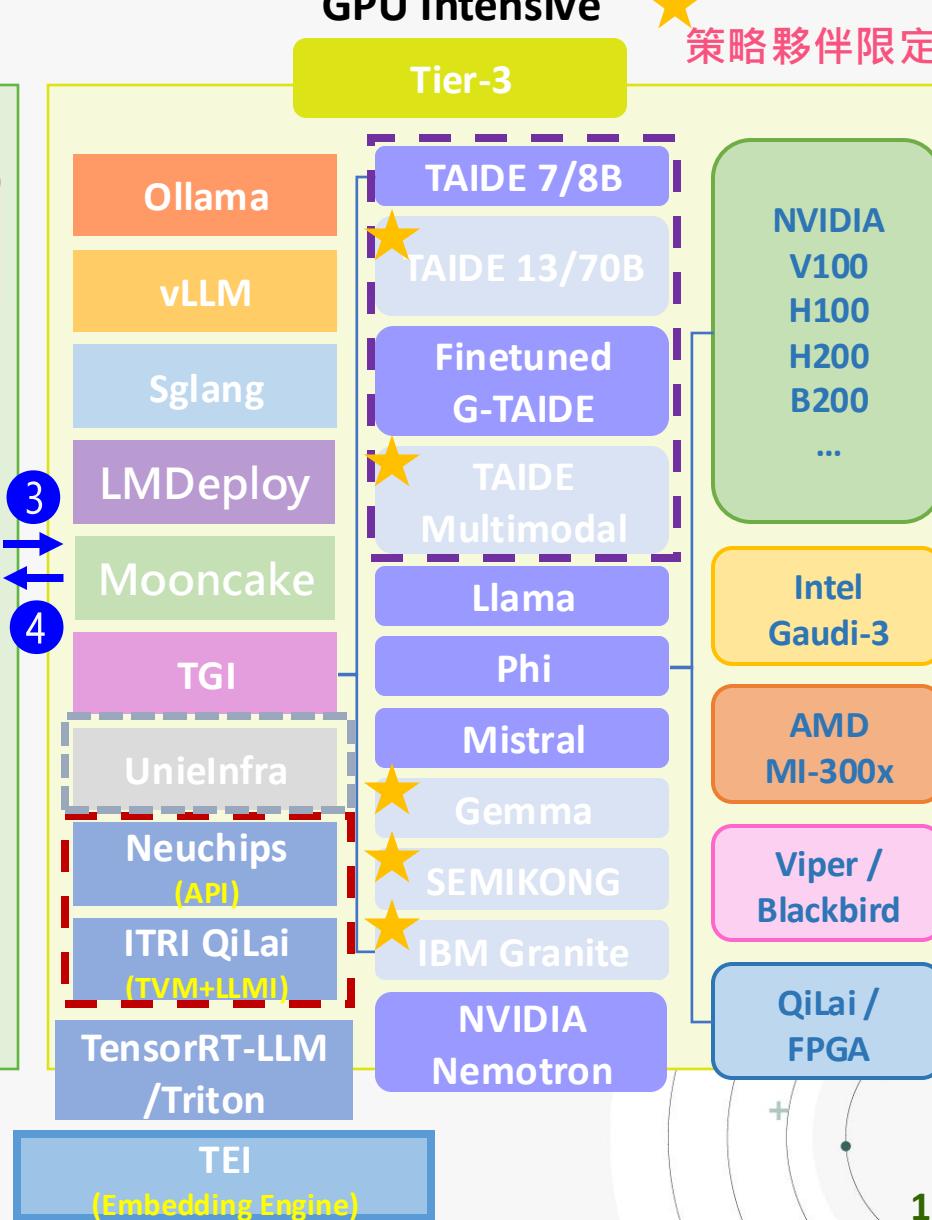
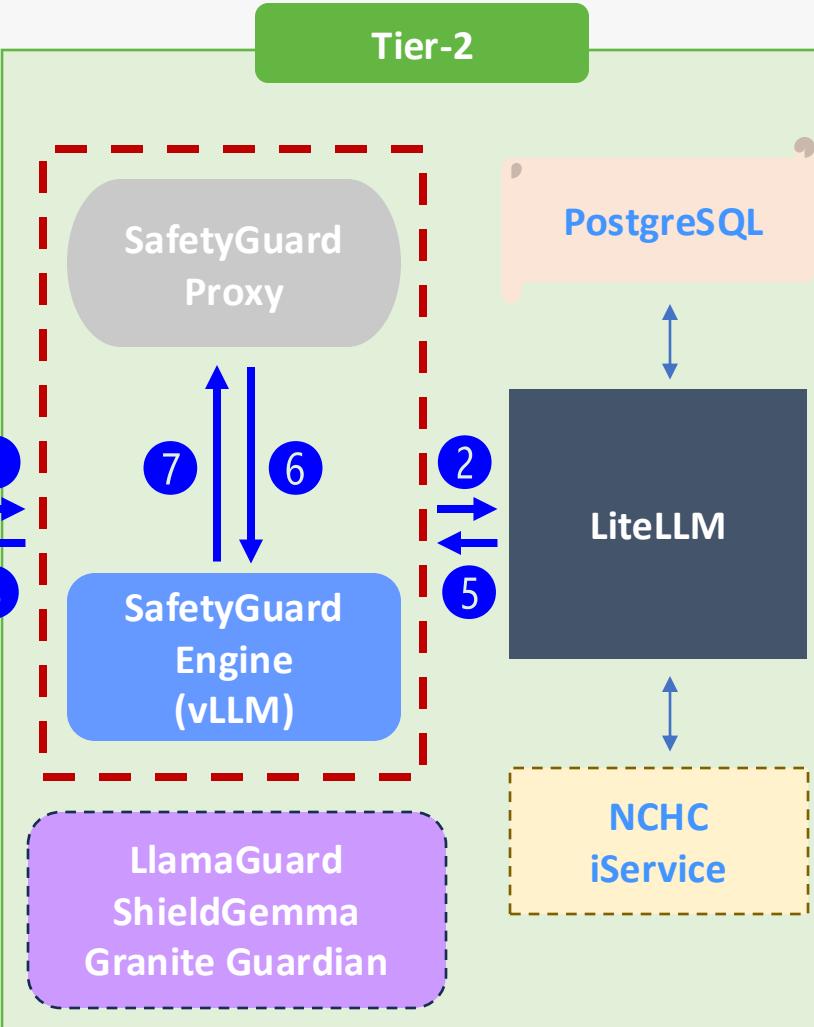
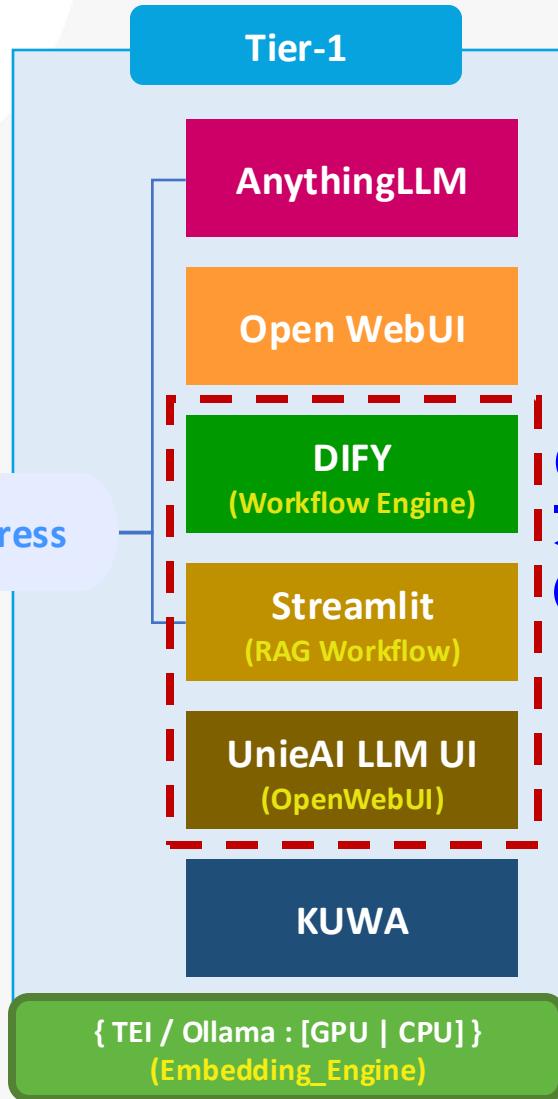
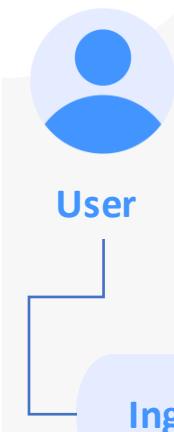


CPU Intensive

CPU+GPU Intensive

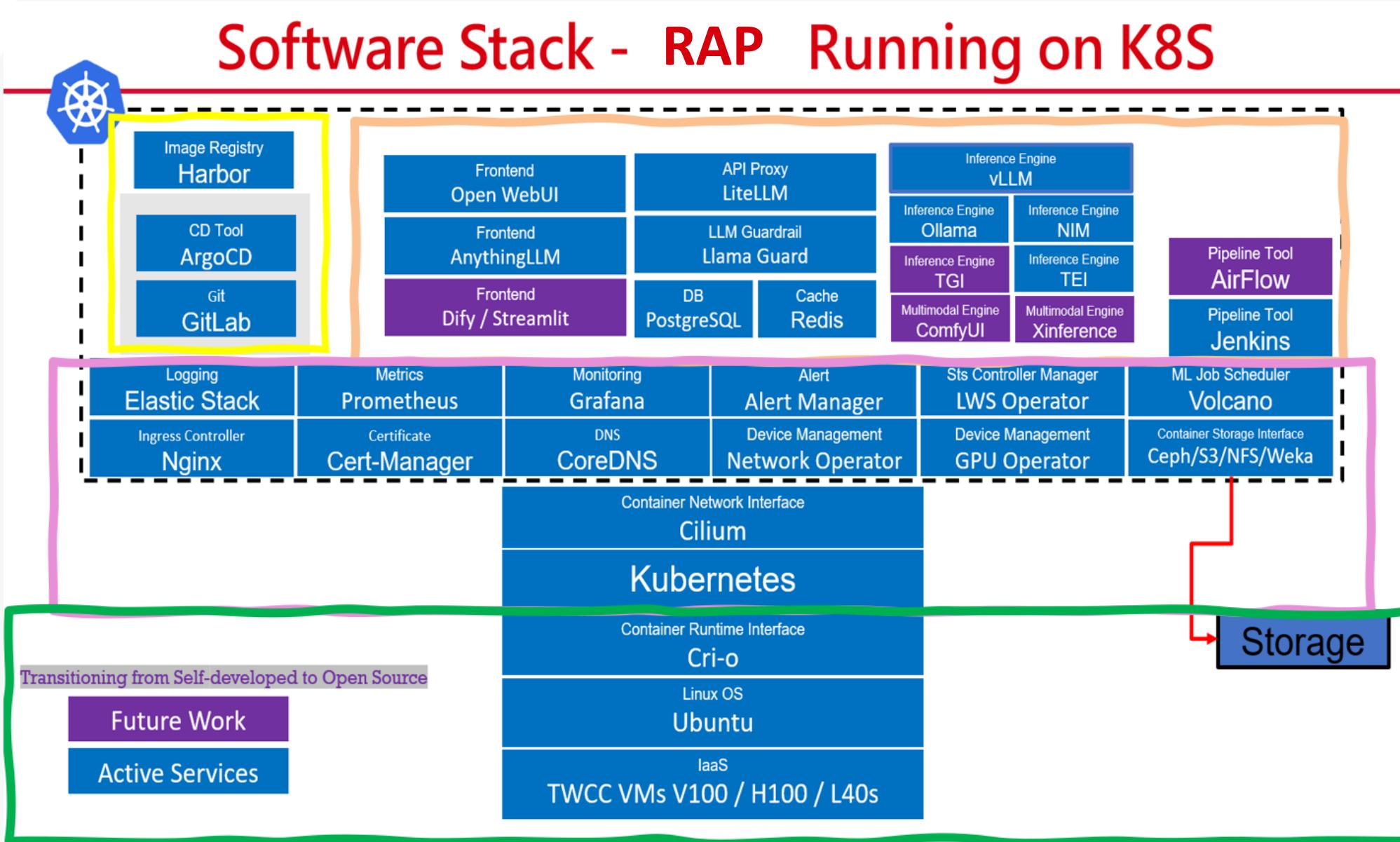
GPU Intensive

策略夥伴限定

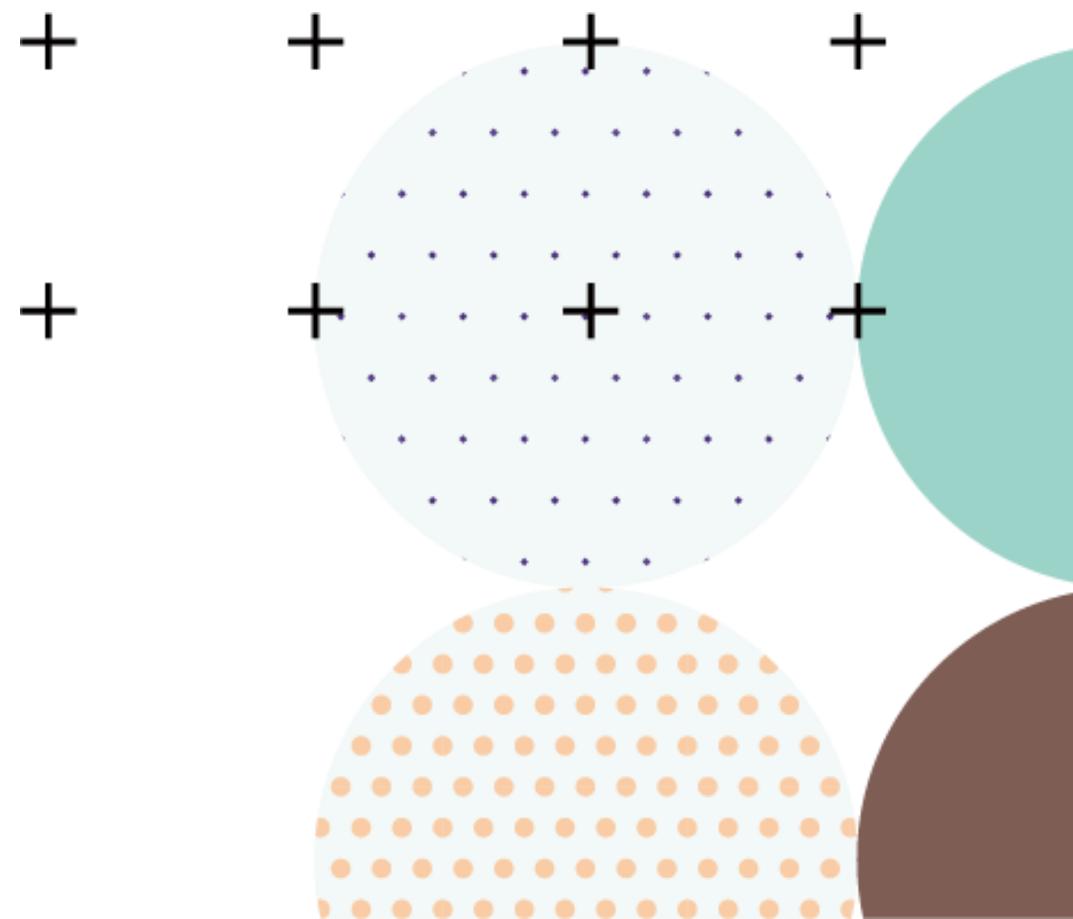


# TAIWAN AI RAP : 運行環境

## Software Stack - RAP Running on K8S

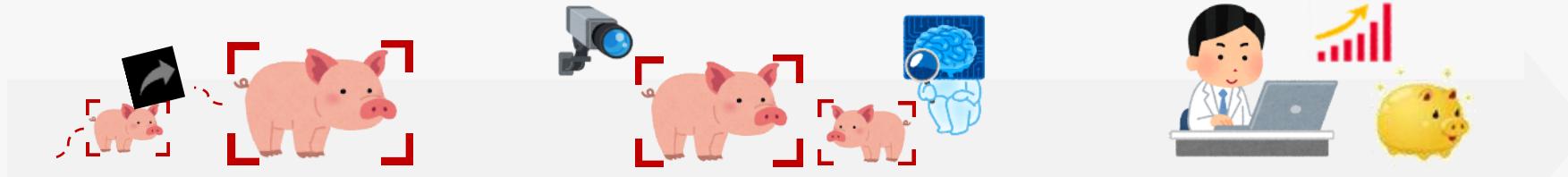


## 二、PoC 客戶案例分享



# 案例

## 智慧豬隻行為監測與健康異常互動式查詢



仔豬活動力為重要健康指標

以攝影機輔助，全齡長時監測

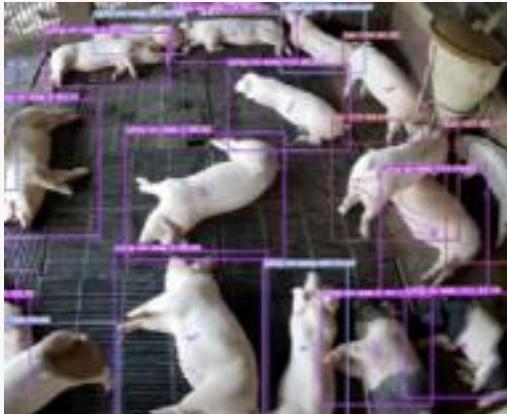
降低成本損失，提升牧場收益

### 母豬仔豬異常辨識



- 母豬哺乳次數/食慾狀況
- 仔豬吮乳數量/活動力

### 肉豬行為與姿態辨識



- 肉豬進食次數/姿態表現
- 肉豬活動量

### 欄位異常燈號警示系統

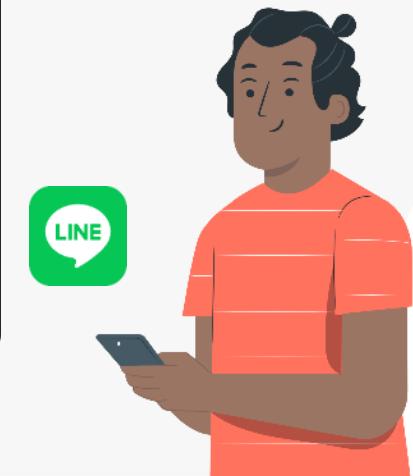
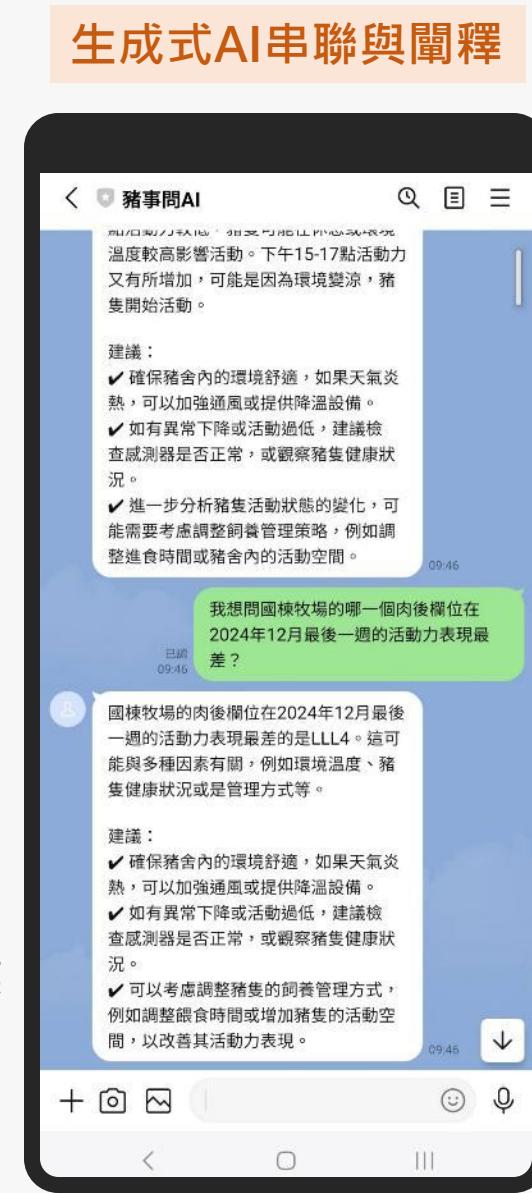


- 每日欄位資訊
- 遠端燈號警示
- 異常通知事項
- 歷史資訊回溯

# 使用情境



AI回應



# 案例

## 急診交班系統智能自動化生成平台

急診總是醫院最忙碌的地方。在醫護人力短缺之下，透過AI生成交班單、報告，減輕醫護負擔、同時快速掌握病人情況

### 預期效益

- 原本轉班流程需要**5分鐘**，AI介入之後只需要**1分鐘**，效率**提升80%**
- 醫療錯誤率**降低35%**
- 滿意度達**92%**，改善溝通及連續性照護

### 智能交班

自動生成交班單，回答交班常見問題，減輕醫護非急診負擔

### 自動摘要與系統集成

自動生成關鍵信息報告，整合電子病歷，確保資料準確及即時更新

### 急診科專屬模塊與教育支援

提供急診科自定義模塊與培訓，確保資料安全與隱私

#### 服務客群

醫護人員



## 案例

## 智慧知識管理與生產分析系統

過往用儀表板自行查找機臺狀況、生產數據等。導入生成式AI之後，可輕鬆透過問答取得機臺及生產資訊，甚至能彙整成報告，解決每日要完成的行政程序

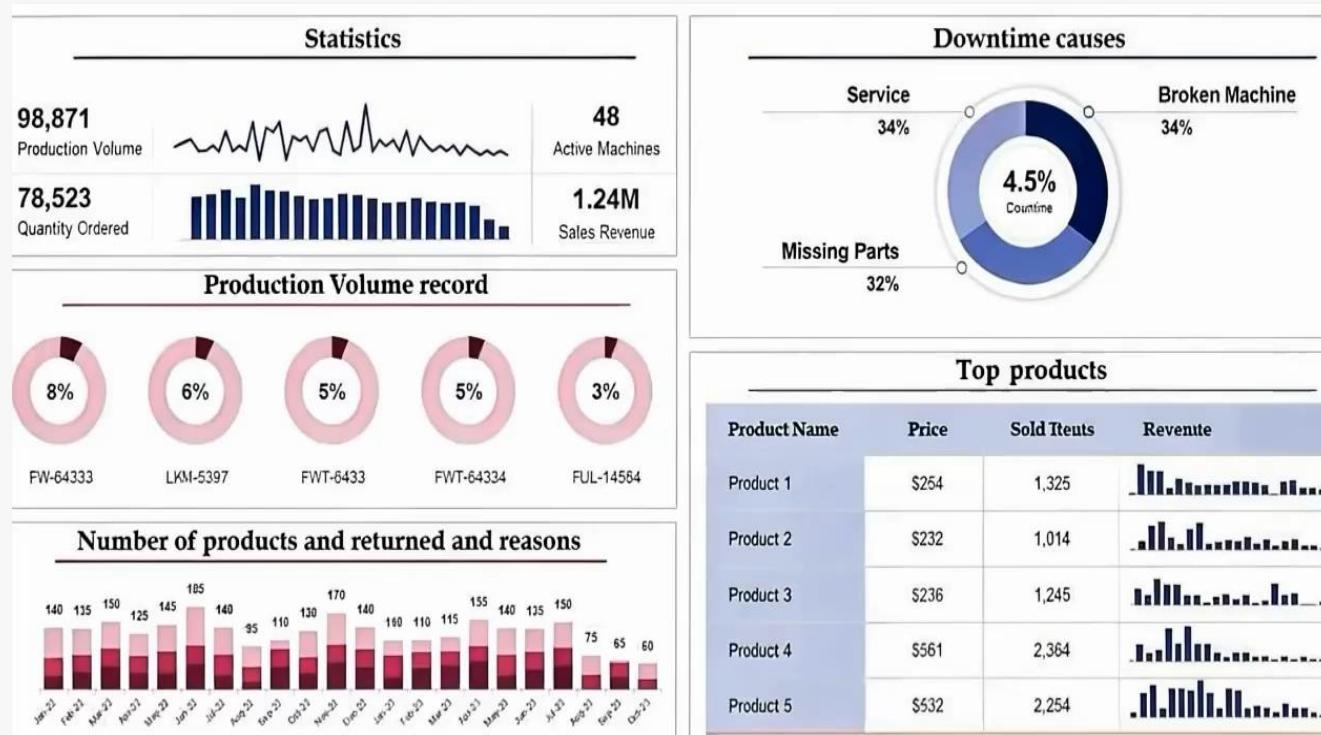
智慧查詢與自動摘要功能，  
快速掌握內部知識與文件重點

## 預期效益

- 3秒內找到文件，查找時間縮短**80%**
- 報表製作與維護成本降**30%**
- AI判斷設備異常並提供SOP，停工等待時間減少**45%**

## 服務客群

公司員工  
工廠管理者





# Smart Manufacturing Command Center



## Overall Equipment Effectiveness Key Benefits

### Improve Productivity

- ✓ Discover hidden capacity
- ✓ Improve machine availability
- ✓ Maximize capacity
- ✓ Improve employee output

### Reduce Losses & Costs

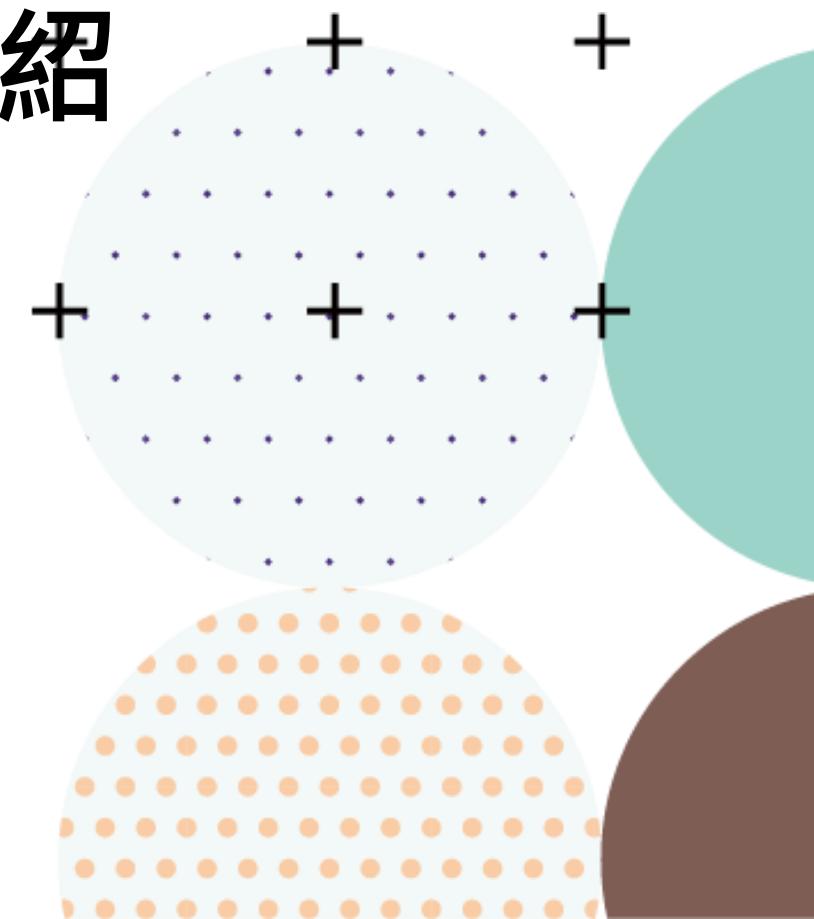
- ✓ Minimize unexpected downtime
- ✓ Reduce product defects
- ✓ Eliminate process bottlenecks
- ✓ Reduce repair costs



### 三、TAIWAN AI RAP試營運介紹



+



# 試營運用戶徵選計畫

🚀 加速台灣AI發展，誠邀企業夥伴加入試營運！

營運期程：2025/03/28 ~ 2025/09/30

## 為資服業、新創、中小企業量身打造

提供一個高效、穩定的AI算力環境，協助企業降低AI開發門檻，提升技術實力

## 試營運期間免費測試

每家企業可獲得定額算力與平台測試權限，享受高性能AI計算資源，優先體驗平台完整功能

## 與國網中心及AI產業生態鏈深度交流

試營運期間，我們將舉辦工作坊與使用交流，助您更快適應並發揮平台優勢！



## 技術驗證與創新應用支持

夥伴可在平台上開發AI服務、訓練模型，並進行驗證測試，探索多元應用可能性。

## 共同行銷機會，提升品牌曝光

試營運夥伴將有機會透過平台、產業活動及媒體曝光，提升市場能見度，共創商機！

# 試營運用戶徵選計畫

※ 依服務現況滾動式調整



我們期待更多企業夥伴的加入，攜手共創AI創新應用生態！



平台預計10月正式啟用，即日起開始試營運，採申請審核制，以提供產品開發、測試與驗證！

## 試營運期程

試營運期程：2025/03/28～2025/09/30

第一階段（2025/03/28～2025/07/31）

- 申請期間：2025/03/28～2025/04/24
- 評選方式：採隨申請隨審原則，評選結果將分兩批通知：
  - 第一批通知：2025/04/18
  - 第二批通知：2025/04/30（適用於4/18未入選者及4/7後提交者）
- 使用期間：自評選通知後服務啟用日至 2025/07/31

第二階段（2025/08/01～2025/09/30）

- 申請期間：2025/07/07～2025/07/16
- 評選方式：採隨申請隨審原則，評選結果將統一於 2025/07/28 公告
- 使用期間：自評選通知後服務啟用日至 2025/09/30

# 試營運用戶徵選計畫

※ 依服務現況滾動式調整



我們期待更多企業夥伴的加入，攜手共創AI創新應用生態！🚀

平台預計10月正式啟用，即日起開始試營運，採申請審核制，以提供產品開發、測試與驗證！

## 申請對象 (B2B)

- 符合台灣法規之企業、機構或法人  
(以資服業者、新創公司、中小企業、學研單位為主)
- 擁有清晰AI開發願景與詳細規劃的團隊
- 對AI應用開發、模型訓練或AI服務普及化有高度企圖心的團隊
- 同意提供平台使用回饋意見，協助優化TAIWAN AI RHAP平台
- 同意公開成果及 Demo 素材，與國網中心共同行銷增加成果能見度

# 試營運用戶徵選計畫

※ 依服務現況滾動式調整



我們期待更多企業夥伴的加入，攜手共創AI創新應用生態！🚀

平台預計10月正式啟用，即日起開始試營運，採申請審核制，以提供產品開發、測試與驗證！

## 申請與使用規則

- **徵選名額**：將依實際徵選情形動態調整，每家可獲得約等值於 [NCHC iService](#) 650,000 元的算力額度及平台測試帳號使用權限。
- **申請方式**：至 [TAIWAN AI RHAP](#) 平台填寫申請表，包含提交 AI 開發提案、公司基本資料、聯絡人資料等。
- **審核標準**：以AI開發潛力與願景、應用效益與落地性、使用者影響力與市場契合度、產業應用價值、技術創新與競爭力為主要評選依據。

名額有限  
盡早報名





TAIWAN  
**NCHC**

National Center for  
High-performance Computing

<https://www.nchc.org.tw>

臺灣

