

HPE 企業 AI 轉型方案 與台灣 AI 應用服務推廣專刊

HPE Private Cloud AI & Eco-System in Taiwan



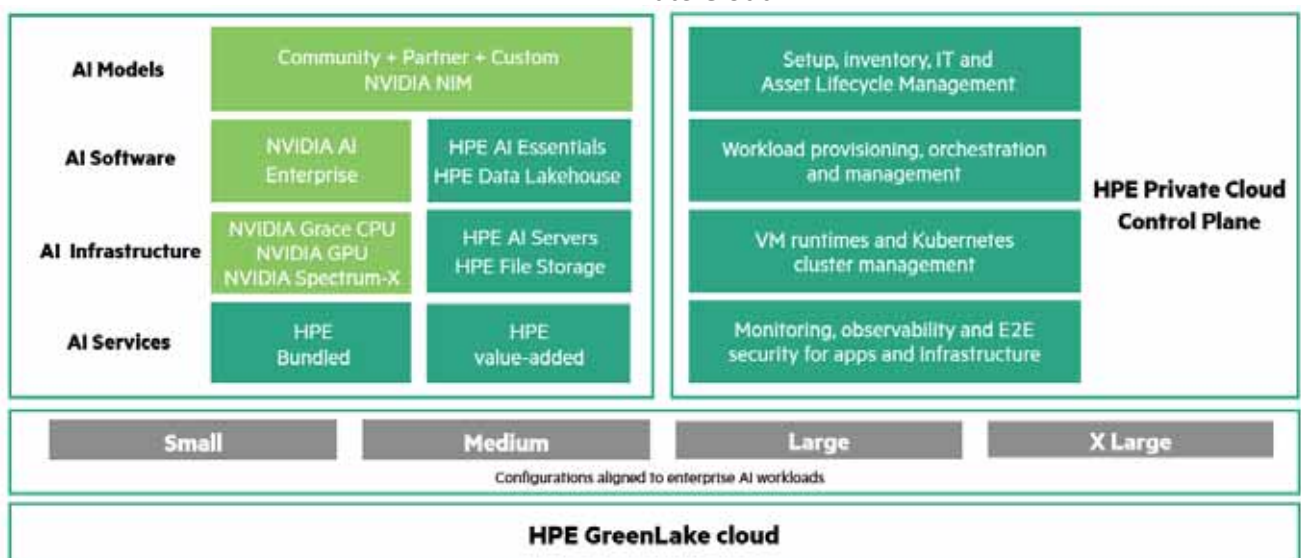
HPE 全方位運算解決方案為企業 AI 賦能

網際網路後最具影響力的數位革新，非AI科技莫屬。高速運算結合技術突破，生成式AI等應用成果大幅躍進，許多產業也開始在各類情境下透過AI帶來了無與倫比的競爭力，AI時代儼然來臨。

因應企業AI轉型需求，HPE提供全方位的企業AI解決方案：Private Cloud AI，包含適用於AI重度運算需求之硬體基礎架構、AI開發與機器學習管理軟體，以及GreenLake服務提供之AI雲端資源與企業導入之整合顧問服務。除了HPE全方位與軟硬多元的AI配套方案，HPE亦與台灣在地AI產業結盟，推出各式兼顧本地需求與全球化競爭力之AI應用服務，專業賦能，順利達標產業轉型。

NVIDIA AI Computing by HPE

HPE Private Cloud AI



HPE Private Cloud AI 打造台灣極致AI生態系

HPE Private Cloud AI能為企業與組織提供全方位的AI方案：

◆ AI模型與LLM應用服務：

HPE組織全球各地與台灣在地的AI服務整合業者 (ISV)，提供各類AI應用服務如智慧客服、智慧製造等應用服務。在台灣的AI業者將能就近提供技術支援與應用服務建置。

◆ AI服務最佳化：

HPE協助提供AI整合顧問服務，將AI系統營運環境依企業所需調整至最佳狀態之服務，藉以打造出以AI為導向之商業服務。

◆ AI軟體：

HPE提供 AI Essentials與HPE Data Lakehouse等AI開發輔助工具，並結合NVIDIA AI Enterprise雲端導向的AI軟體服務，提供企業於建立AI應用系統的開發

生命週期中的各個環節，提供對應且全面之輔助管理系統，包含AI模型專案管理、AI資料訓練管理等功能。

◆ AI就緒之硬體架構：

HPE Cray XD670、HPE ProLiant Compute XD680、HPE ProLiant Compute XD685、HPE ProLiant DL380a、HPE ProLiant DL384等機架式伺服器可提供AI運算所需之GPU等龐大運算資源，並有完整資料儲存與管理方案因應資料池所需。

◆ HPE GreenLake服務：

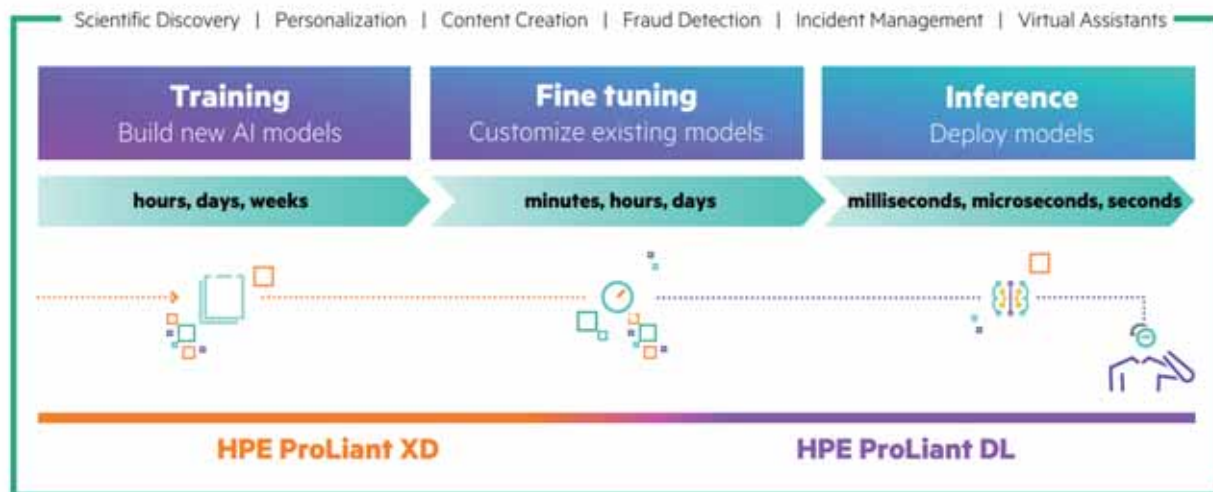
包含硬體基礎架構、軟體授權，以及整合維護服務等，HPE GreenLake服務均能提供企業以更彈性的資源規劃組織強大的AI應用。

◆ 技術規劃與諮詢顧問服務：

HPE匯集IT軟硬體整合技術與各類產業營運，為有意導入AI應用之企業提供前期評估、中期整合到後續維護等技術與商業營運等知識服務

HPE Compute for AI inference, fine tuning, and training

part of NVIDIA AI Computing by HPE



HPE運算技術滿足AI發展三階段： 訓練、微調與推論

AI應用從發展到導入時的發展分為三階段，包含建模訓練(Training)、最佳化微調(Fine Tuning)，以及AI推論(Inference)等，不同階段企業所仰賴的運算資源亦有所不同。HPE AI運算配套解決方案均有對應的設備滿足所需。

◆ 建模訓練(Training)：

建立全新的AI模型時需要最龐大的運算效能，根據模型複雜度花費時間從數周到數小時之間，也是AI三階段中需投入最多資源的，如氣象、基因、LLM建模等，主要以AI服務供應商、需打造AI自主的應用服務等。

※ 參考機種：HPE Cray XD670、HPE ProLiant Compute XD680、HPE ProLiant Compute XD685超級電腦架構

◆ 最佳化微調(Fine Tuning)：

導入AI應用時，需要將訓練完畢的AI模型再進行微調訓練，讓AI模型具備企業獨有的專業知識，根據模型規模花費數日到數分鐘等情況均有，當企業要導入、建置AI應用服務初期會需要反覆微調，當AI模型上線運行一段時間後，也可能根據最佳化需求再次微調。

※ 參考機種：HPE Cray XD670、HPE ProLiant Compute XD680、HPE ProLiant Compute XD685、HPE ProLiant Compute DL380a Gen12、HPE ProLiant DL380a Gen11、HPE ProLiant DL384 Gen12

◆ AI推論(Inference)：

AI模型訓練完畢、佈建應用服務後，AI作業時也需要許多的運算資源來提供日常推論運算，為了能在第一線充分發揮效益而要求系統低延遲，通常處理時間以微秒到秒級來計算，例如智慧製造、智慧客服或生成式AI等系統後端運算等。

※ 參考機種：HPE ProLiant DL380a Gen11、HPE ProLiant DL384 Gen12、其他HPE ProLiant系列伺服器

打造AI所需的IT環境，HPE不僅有享譽全球口碑的運算方案，也有業界市占領先的儲存解決方案，如HPE GreenLake DSCC等提供高速運算架構下絕佳的資料池與存儲等方案。軟體領域，HPE不僅提供AI Essentials與HPE Data Lakehouse等AI開發輔助工具，同時也與NVIDIA合作，提供NVIDIA AI Enterprise之AI軟體雲端服務，協助AI專案開發中資料管理、專案或模型管理等工作更加靈活有序。

支援NVIDIA NVLink與Intel XeLink互連技術

投入AI開發的領域，GPU效能在AI建模與推論等任務可說是最重要的資源，不論是強化資料串連或增加PCIe介面卡的安裝密度，均有助於提升基礎架構的運算能力。

HPE AI運算配套解決方案支援NVIDIA NVLink與Intel XeLink x2 Bridge等GPU橋接互連技術，能大幅提升AI運算時資料I/O的吞吐量，充分發揮GPU的運算效能。



目錄

HPE AI 架構與相關方案	3
HPE AI Computing 方案概述與產品介紹 -ProLiant 系列	6
HPE AI Computing 方案概述與產品介紹 -Cray 系列	9
HPE Storage 方案介紹	10
HPE LLM Appliance 全方位 AI 完整服務方案	11
AI Eco-System 台灣合作廠商的方案	13
結語與參考資訊	48



AI 首選 HPE Private Cloud AI 高速運算基礎架構

AI應用對IT基礎架構的需求不同以往，針對高速運算與大量I/O設計之伺服器與儲存系統有著更高的需求，高速運算代表著AI模型建立與訓練、推論等作業時程更短，站在實務應用的角度，AI系統必須夠快，才能真正的產業競爭力。

HPE Private Cloud AI提供全面滿足AI運算所需的基礎架構系統，包含HPE AI伺服器、HPE超級電腦節點，以及HPE超高速儲存系統等，不僅具備高效能，並能與NVIDIA、Intel等AI開發工具及加速套件完整相容，從容擘劃AI應用之創新版圖。

激發AI無限價值 — HPE ProLiant機架式伺服器

提供絕佳的高速運算能量，才有助於AI充分發揮應用價值。

HPE ProLiant機架式伺服器針對AI應用所需，特別加強NVIDIA GPU整合效率，搭配市場最齊全的產品線，從需要超高算力的生成式AI模型建立，到第一線的低延遲邊緣運算推論機種，HPE ProLiant伺服器一應俱全。

企業AI應用所需的運算力，一方面須仰賴伺服器上的擴充能力，容納越多的GPU模組代表其運算能力越強大，因此HPE ProLiant的高擴充設計以及絕佳的散熱架構，能在標準機箱內提供業界最高密度的GPU模組，同時支援NVIDIA NVLink傳輸匯流排技術，大幅提升GPU運算時的資料傳輸速度，雙管齊下，大幅提升AI推論與模型訓練的效能。

AI推論與最佳化微調 運算平台



HPE ProLiant DL320 Gen11
Up to 4 NVIDIA L4 GPUs

滿足即時性邊緣環境之AI應用所需

1U機架式伺服器內最高可安裝4組NVIDIA L4 GPU模組，緊湊的設計提供最大的空間規劃，適用於電腦視覺或AI影像分析等工作負載。



HPE ProLiant DL380a Gen11
Up to 4 NVIDIA H100/L40s PCIe GPUs

性能擴充兼容並蓄，因應AI推論運算

2U機架式伺服器最高可容納4組NVIDIA H100/ L40s PCIe GPU模組，具備進階的擴充性，滿足泛用型圖形處理的算力需求，處理AI應用時的推論運算。



HPE ProLiant DL384 Gen12
Dual NVIDIA GH200 NVL2 (2024)
Dual NVIDIA GB200 NVL2 (2025)

高密度記憶體設計，RAG AI推論絕佳選擇

以NVIDIA為核心之2U機架式伺服器，支援Dual NVIDIA GH200 NVL2 / Dual NVIDIA GB200 GPU模組，強化PCIe匯流排傳輸，讓GPU效能提升30%，包含生成式AI推論，以及非AI之高密度記憶體的運用均能勝任。



HPE ProLiant DL380a Gen12
Up to 8 NVIDIA H100 NVL GPUs (2024)
Up to 8 NVIDIA H200 NVL GPUs (2025)

超擴充性GPU加速能力，適用大型生成式模型

具備卓越的AI效能與擴充性，搭配NVIDIA H200 NVL GPU模組，提供更大且更高速之記憶體效能；支援NVLink連接技術，提升GPU間資料傳輸速度，藉此應對LLM模型推論負載所需的龐大算力。

複合式防護機制 確保運算平台資料安全

企業導入AI應用的關鍵不僅須具備高速運算能力，資料安全也是很重要的一環，實務上也有許多企業考量關鍵資料安全性而選擇建置本地端的AI應用系統，為協助企業維繫AI運算環境之安全管理，HPE ProLiant系列伺服器配有矽晶片信任根(Silicon Root of Trust)與iLO管理晶片，可支援安全協定與資料模型(SPDm)，並提供iDevID安全設備身分識別等防護機制，確保AI運算平台在運行中不會遭到惡意侵害而丟失資料。



HPE Cray & HPE ProLiant Compute XD 超級運算方案

超級電腦之運算量能，引領AI世代。

運用AI之際，企業若需要在龐大的AI模型中獲得超低延遲的推論結果，或需要建立生成式AI的LLM大語言模型時，HPE提供Exabyte級 (1EB = 百萬TB)運算能力的超級電腦機種，包含HPE Cray XD670、HPE ProLiant Compute XD680，以及HPE ProLiant Compute XD685等設備，為各個產業帶來無與倫比的AI競爭力。

頂尖規格，卓越處理效能

HPE Cray & HPE ProLiant Compute XD運算方案具備強大的GPU運算能力，並可搭載大量記憶體，能從容應對AI運算所需的超大型資料處理需求，是集結頂尖運算力與龐大記憶體的超級旗艦。

HPE ProLiant Compute XD 680之特色與優勢：



- 可安裝8組Intel Gaudi 3 AI加速器。
- 記憶體為24組DDR5 DIMM插槽，支援3.7TB/sec HBM高頻寬記憶體，單一伺服器節點容量可高達3TB以上。
- 支援第五代Intel Xeon可擴充處理器。
- 提供HPE iLO管理系統並內建資安防護系統。
- 支援8組半高式x16 PCI Gen5插槽或4組全高式x16 PCI Gen5插槽。
- 5U高度，採氣冷式設計，可供資料中心輕鬆部署。

HPE ProLiant Compute XD 685之特色與優勢：



- 可安裝8組AMD Instinct MI325X GPU模組。
- 可安裝8組NVIDIA H200 Tensor Core GPU模組或NVIDIA Blackwell Ultra GPU模組。
- 支援2個第五代AMD EPYC 9005系列處理器。
- 記憶體採用HBM3E傳輸技術，支援24組DDR5 6400 RDIMM插槽，最高頻寬可達6TB/s。
- 提供HPE iLO管理系統並內建資安防護系統，同時保障生產供應流程可靠無虞。
- 5U高度，可選擇氣冷或液冷方案。

HPE Cray XD670的優勢與特色：



- 支援第六代Intel Xeon可擴充處理器。
- 記憶體為32組DDR5 DIMM插槽，8通道傳輸，傳輸速率達5600MT/s，單一伺服器節點容量可達4TB以上。
- 支援PCIe Gen5高速I/O傳輸，幫助SSD與GPU模組充分發揮。
- 5U高度，可選擇氣冷或液冷方案。

結合雲端軟體方案與訂閱服務，彈性啟動AI轉型計畫

HPE Cray & HPE ProLiant Compute XD超級運算方案與HPE其他伺服器及資訊設備相同，均可透過GreenLake服務模式取得，結合HPE Private Cloud AI雲端工具，獲得全方位的AI開發與基礎系統管理服務，提供企業在AI專案布局或基礎系統建置上更多選擇與方案搭配，快速邁開AI轉型步伐且無後顧之憂。

超神速 I/O

HPE Cray Storage 系列 儲存系統

AI模型訓練的靈魂是資料，確保資料安全與高效存儲是不可忽略的要素。

企業導入AI應用時，決定核心與GPU算力資源後的第二步，就是要搭配效能與管理機制相符的儲存方案，讓I/O效能跟上前端的運算，藉此降低模型訓練與微調的作業時間。

HPE提供多種資料存儲方案滿足各種規模與AI建置階段所需，包含HPE Cray ClusterStor E1000儲存系統與HPE GreenLake for DSCC等方案，不論是傳輸效能、容量與擴充管理，都能應對企業AI應用環境所需，與HPE AI伺服器運算方案攜手並進，以高效率的資料導向基礎架構幫企業AI應用大幅加分。

超級電腦等級的I/O效率，加速企業AI方案落地上市

不同等級的AI應用需求，企業亦可依需搭配相應的儲存方案，如LLM大型語言模型訓練或需搭配超級電腦的極速I/O時，可透過HPE Cray ClusterStor E1000儲存系統的超高速I/O效能加速模型訓練的節奏，不僅可靠，在競爭激烈的AI商務方案開發中也有助於縮短上市時程，搶占先機的競爭優勢。



(HPE Cray ClusterStor E1000儲存系統)

對於企業一般AI應用時的推論運算或模型微調等，可以搭配HPE Cray Storage System C500儲存系統，源自HPE Cray Storage系列超級電腦儲存架構，針對AI中低量級的需求而開發的入門級超級儲存架構，



(HPE Cray Storage System C500儲存系統)

針對不同AI專案的任務工作負載，儲存系統在讀取、寫入與隨機I/O等細部存取參數也會有所不同，HPE的技術支援團隊也能協助企業微調儲存系統的參數設定，根據產業別、資料型態、投資成本與專案需求等最佳化，加速AI專案時資料匯入、處理的工作效率。

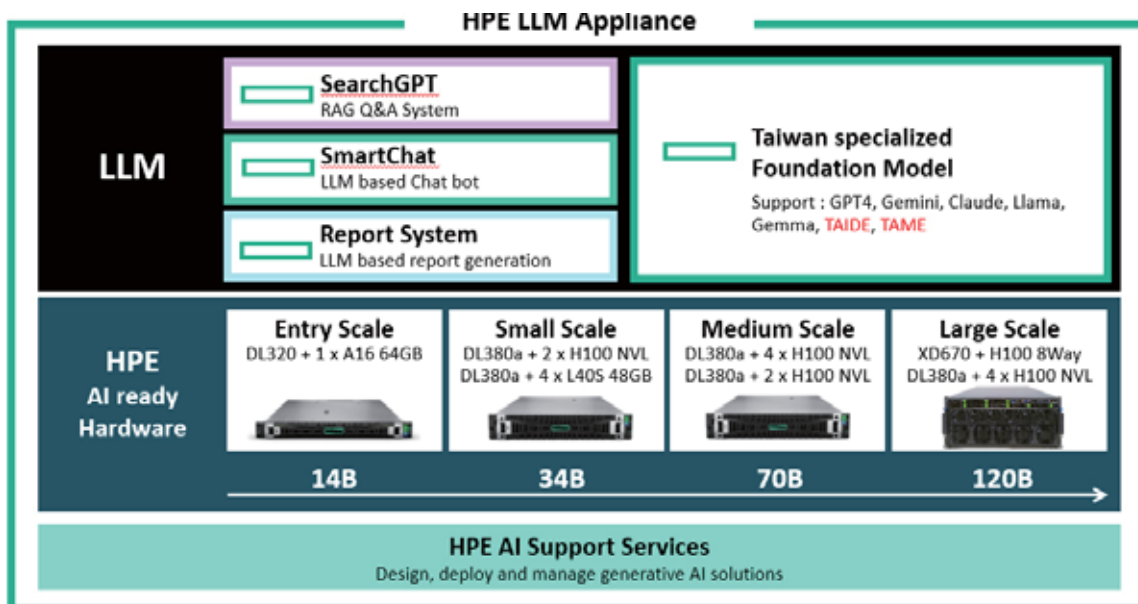
HPE基礎架構方案以多元聞名於全球市場，包含各種規模的伺服器等運算方案，以及AI資料存儲這塊也均有許多方案可以選擇，除了前面提及的HPE Cray Storage超級電腦儲存系統外，包含HPE GreenLake for Block Storage、HPE GreenLake for DSCC (Data Services Cloud Console)等方案，從基礎架構到整合顧問服務，從邊緣、本地到雲端環境皆能支援，提供AI運作時資料池所需的最適架構。

HPE 豎立 Gen AI 即服務產業標竿

結合LLM語言模型的生成式AI應用，預估未來產值將占所有AI應用中的第一名。

綜觀現階段主流的AI運用中，生成式Gen AI最受企業與產業人士青睞，殺手級應用的背後需樣賴LLM大語言模型的建置與訓練，針對LLM大語言模型建置與管理，HPE LLM Appliance能提供企業導入、訓練出專屬的LLM模型，再搭配相關知識庫與人機互動平台，模型微調訓練後就能迅速理解、回應使用者的疑難雜症，或理解語意後協助生成需要的文件等工作，提升日常工作效率達數十倍之譜。

HPE LLM Appliance服務完整涵蓋硬體、AI模型與服務



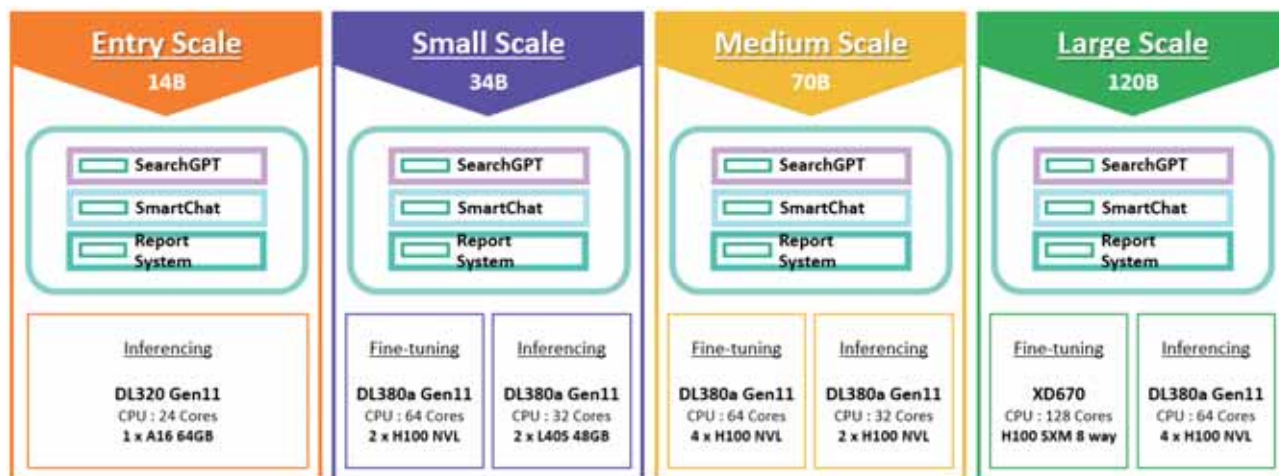
HPE LLM Appliance可支援的LLM大語言模型包含：

- SearchGPT搜尋用生成預訓練轉換器：
可運用在檢索增強生成式技術(RAG)的問與答系統，如智慧客服等應用。
- SmartChat智慧聊天模型：
基於LLM大語言模型所訓練而成的聊天機器人。
- Report System簡報系統生成模型：
基於LLM大語言模型所訓練而成的簡報/文件產生器。
- Taiwan Specialize Foundation Model台灣本地化語言模型：
可支援GPT4、Gemini、Claude、Llama、Gemma、TAIDE、TMAM等大語言模型，具備台灣中文語彙之LLM，為台灣在地化AI應用提供更精準的模型。

HPE LLM Appliance方案可涵蓋硬體平台，HPE提供AI就緒的伺服器方案，針對企業所需的LLM規模與推演資料量，提供不同量級的運算能力與GPU模組，如同所有的數位服務專案，AI應用從基礎硬體、模型建置到最後的應用推論，過程中不會只仰賴單一系統，而是從伺服器、網路、儲存、監控、營運管理到諮詢服務等，HPE LLM Appliance所提供的服務，涵蓋了基礎運算系統與LLM語言模型的建置，幫助企業建置、布署生成式AI等應用，協助企業跨出AI轉型的第一步。

對應資料分析量，AI分級滿足龐大運算力

AI應用專案的規模，經常採用模型處理的資料量來計算，HPE LLM Appliance根據算力所需區分為4種規模，並針對不同規模量身訂做給予對應的運算方案及LLM套餐：



HPE AI Appliance建議的規格依循資料數據處理量來分級，若企業於導入AI服務時有超低延遲回應，或是處理的資料型態需整合更多GPU效能時，也都能根據實際情況將方案內搭配的伺服器設備升級，或是搭配其他HPE ProLiant系列的伺服器，若想確認規劃中的AI專案屬於何種AI規模，HPE的專業團隊也能以豐富的產業整合經驗提供詳實的分析。

從AI模型的規模出發，企業能先以自身的應用環境切入，評估導入所需的算力，甚至是未來發展所需投入的GPU資源等，不論是一次到位或逐步漸進，HPE AI服務方案都能跟上企業實際所需，不僅設備，台灣在地化的技術支援與本地產業習慣的服務模式都能支援，搭配HPE GreenLake服務模式，為企業帶來極具競爭力之AI as a Service — AI即服務轉型模式。



HPE 連結台灣業者 實踐 AI Everywhere 願景

讓AI充分發揮競爭實力，集結台灣在地創新業者組織AI服務生態圈，方能實踐產業典範升級目標。

AI應用的價值與其他數位服務相似，需仰賴高效、可靠的硬體，搭配專業、豐富的應用服務，藉此協助企業在各個營運的面向強化執行效率，透過HPE全方位AI方案服務，也能協助企業評估、物色適合導入的應用服務，同時協助企業媒合專業的AI服務業者，進而在應用面充分發揮AI科技的軟實力。

結合本地台灣業者，激發AI轉型之無限效益

HPE聯手台灣AI服務供應商，也為企業規劃、導入AI服務帶來許多便捷優勢，這些優勢包含：

- ◆ 打破距離限制就近支援：

與台灣AI業者合作，若有需要專業人員到企業本地支援時將更有彈性，尤其時初期評估了解場域現況，或是要求高度資料安全需在內部落地之AI專案。

- ◆ 消弭文化差異隔閡：

AI專案需要高度的客製化、不斷微調模型等與服務業者相互配合的情況，台灣AI業者能縮短認知差異，尤其是LLM語言模型與生成式AI微調效正之情況，將有助於整體專案執行效率及系統成效。

- ◆ 單一窗口服務：

HPE與台灣AI業者可視為同一團隊，不論是在硬體設備或應用服務上需要調整，只要透過HPE單一服務管道就能匯聚相關支援，簡化維護流程，也避免廠商之間推諉甩鍋等問題。

- ◆ 充分發揮AI應用服務之效能：

合作的台灣AI業者均採用HPE ProLiant伺服器或Cray超級電腦建立實測環境，能在HPE的硬體環境上發揮最大效益，在專案建置時兩方團隊也能相互配合、不斷微調系統直至硬體與AI模型運算達到效能最佳化狀態。

攜手台灣AI業者，帶來多元且極具競爭力之創新服務

HPE與全球許多專業的AI服務供應商合作，在台灣也有許多合作夥伴，包含許多IT界耳熟能詳的老字號服務商，也有AI方案驚艷全球的科技新銳，提供如AI人機介面、AI視覺分析等方案：

- ◆ 台灣人工智慧實驗室AI Lab
- ◆ 亞太智能機器APMIC
- ◆ 鐵雲IronYun
- ◆ 百威雷科技PowerArena
- ◆ 凌羣電腦SYSCOM
- ◆ 大同世界科技TSTI
- ◆ 創泓科技Uniforce
- ◆ 滿拓科技DeepMentor
- ◆ 網智服務IQ Service International
- ◆ 數位無線INFINITIX
- ◆ 邁爾凌科技MyelinTek
- ◆ 翱騰國際Octon International
- ◆ 頑碼資訊Playma
- ◆ 杰倫智能科技Profet AI
- ◆ 和碩聯合科技PEGATRON AI
- ◆ 八維智能D8AI
- ◆ 資策會軟體研究院
Institute for Information Industry

根據各台灣AI業者推出之方案，企業也能藉此思考，如何運用AI的技術為自身營運優勢疊加更強的產業競爭力。

Taiwan AI Labs 為台灣構建在地化的企業級生成式 AI 服務

AI應用中，LLM大語言模型、生成式AI的發展在AI創新中扮演著重要的角色。

Taiwan AI Labs台灣人工智慧實驗室匯聚許多專業AI開發團隊，提供FedGPT生成式AI方案、雅婷人機互動解決方案，以及ASR語音轉文字方案等，企業能運用FedGPT整合其他AI運用，或是搭配Taiwan AI Labs提供之AI人機互動與語音轉文字方案等兩大應用，藉此在金融、醫療、多媒體娛樂、教育等產業落實AI應用，例如以FedGPT搭配雅婷人機互動系統打造智慧客服系統以提升客戶服務體驗，或是將FedGPT搭配ASR語音轉文字方案，為企業簡化知識建立的作業流程。將眾多AI應用服務結合產業應用，能讓各式創新服務更上一層樓。

Taiwan AI Labs 基於大語言模型的眾多應用

Taiwan AI Labs匯聚全球頂尖的專業團隊，開發出首屈一指的LLM大語言模型，並基於模型推出FedGPT生成式AI解決方案，以及雅婷人機互動解決方案與ASR語音轉文字方案等兩大AI創新應用

◆ FedGPT生成式AI解決方案

由台灣團隊Taiwan AI Labs開發的生成式AI解決方案，可提供給企業打造自身專屬的生成式AI方案，結合企業內部的專業知識，能作為企業內部人員訓練或對外客戶服務等用途。

FedGPT使用的大語言模型符合產業標準，可協助企業整合、開發第三方AI應用服務；FedGPT符合可信賴責任之AI倫理準則，能避免AI服務衍生不正當的錯誤互動、影響企業營運。

◆ 雅婷人機互動方案

當企業要使用生成式AI等技術時，也需要搭配一套AI人機互動系統來活用後端的LLM語言模型，Taiwan AI Labs的雅婷人機互動方案就是結合語言模型讓生成式AI系統發揮效用的工具，企業可透過Taiwan AI Labs雅婷人機互動將內部資料匯入AI資料池中，透過深度學習功能將企業內部資料轉化為AI資料庫，輔以微調修正與稽核，讓後續生成式AI的回應更臻完善、發揮所長。

◆ ASR語音轉文字方案

Taiwan AI Labs的ASR語音轉文字方案是一套能智慧轉換語音與文字的AI工具，基於Taiwan AI Labs訓練出來的LLM語言模型，擅長台灣繁體中文與台語對話辨識，系統可以將「聽」到的句子轉化為文字，並具有智慧修正技術，能判斷語句中可能失誤的用詞，並在後續產生文字時自動修正為正確的字詞，對於企業有需要大量處理、分析語音內容，或有即時轉換文字的需求，ASR語音轉文字方案均能勝任。

在台灣的國家兩廳院藝文中心，就運用Taiwan AI Labs的ASR語音轉文字方案，在活動表演時即時將演出內容的語音轉換為文字，同時呈現在會場的大螢幕上，藉此打造出即時且自動產生的智慧字幕系統，幫助聽障人士也能享受藝術活動之陶冶。

滿足企業落地需求，自主學習維繫AI最佳狀態

企業AI應用中，包含生成式AI運作需要仰賴龐大的LLM語言模型，同時也須融入企業自身的專業知識，AI的互動回應才能適得其所、產生效益。

Taiwan AI Labs的各項AI方案具備特點：

- ◆ 提供實機落地建置方案，保障企業機敏資訊不

外流：

不論是FedGPT語言模型，或是人機互動、語音轉文字等方案，Taiwan AI Labs均提供企業實機落地建置方案，相較於雲端式的AI解決方案需要企業將資訊上傳，落地建置方案維繫企業的關鍵資料自主權，避免商業機密外流與資料被另作訓練之風險。

◆ 繁體中文LLM大語言模型，符合產業聯盟標準：

Taiwan AI Labs推出的FedGPT語言模型包含1,000億以上條繁體中文詞庫，在地化的使用習慣不論是在人機互動或搭配其他應用了解語意等情境中更加精準，同時Taiwan AI Labs的模型均符合產業聯盟的標準，也有助於企業運用FedGPT再整合、以利開發第三方應用。

◆ 兼容多元企業知識形態，同時能夠自我訓練不斷升級：

企業導入Taiwan AI Labs FedGPT時能以多種格式匯入、微調模型，短時間內讓AI人機介面方案具有企業內的專業知識，同時方案也具備模型Auto ML自我模型調適訓練功能，能不斷更新企業專屬深度知識，長時間保持模型在最佳狀態。

HPE高速運算促進生成式AI低延遲效益

當企業需要將AI服務落地於自家IT環境時，也就代表著需要具備AI運算效能的基礎運算架構，由其是LLM大語言模型與生成式AI等運用，對於AI效能的需求更加龐大。

要讓生成式AI的回應速度跟上企業需求，Taiwan AI Labs與HPE合作，以HPE ProLiant DL380a等AI伺服器作為搭配硬體，提供可靠、高速的GPU運算資源給予FedGPT等AI模型分析使用，同時HPE ProLiant伺服器亦有內建多種硬體式的安全防護技術，為企業多加一層防止資料外洩的保護罩。

HPE ProLiant系列的伺服器具有多種型號，不論是要追求超低延遲的高速架構，或是準備邁開AI轉型的企業，HPE都能提供對應的運算資源整合與AI顧問服務。



台灣人工智慧實驗室 (Taiwan AI Labs) 是由台灣在地產官學專業人工智慧技術團隊組織而成，整合專業人才、資源與產業鏈合作，並運用開放演算法、原始碼等技術，以及台灣跨產業累積的龐大專業知識，進而推動如生成式 AI、智慧醫療與人機介面等人工智慧的應用發展。

網站：<https://ailabs.tw>

FedGPT 產品網站：<https://fedgpt.cc/>

APMIC 提供模型即服務 實踐企業絕佳 AI 互動願景



■ 自動化資料工具，為企業降低AI運用門檻

AI系統的應用，不僅需要強大的知識模型驅動，也須搭配完善的人機介面讓使用者與AI系統充分交流、獲取關鍵資訊。

APMIC亞太智能機器開發的CaiGun AI人機介面方案，其名稱源自台語的「開講(抬槓)」，能整合企業內落地整合之AI模型與平台，企業能夠選擇多種不同的模型平台來搭配，也能自行訓練、微調模型打造專屬的AI資料庫。結合訓練完畢的AI模型，APMIC CaiGun AI人機介面就能整合成完整的應用系統，包含企業內專家資料庫、生成式AI的智慧客服、合約識別與分析、圖文解說、文案內容生成等應用場域，並針對企業所需，設計相符的互動介面。

自動化資料工具，為企業降低AI運用門檻

企業導入APMIC CaiGun AI人機介面系統時，APMIC也能協助企業準備、建置與模型微調等工作，包含前段的資料整理、模型選擇，以及模型訓練、模型驗證與調校超參數、AI系統應用等，能協助企業導入AI應用服務時為其降低如資料識別、模型評估等技術門檻，幫助企業以更簡易的方式獲得AI服務帶來的眾多效益。

◆ 資料整理自動化：

大量的有效資料是驅動AI達到最佳狀態的要件，APMIC提供自動化的資料整理工具，包含能協助將資料去識別化，藉此平衡個人資料保護與建檔分析所

需；資料自動貼標語轉換功能，能批次處理資料格式，藉此符合選定模型訓練所需。

◆ 核心引擎意圖判斷：

APMIC CaiGun AI系統內的核心引擎具備意圖判斷，能理解使用者的真正目的，精準處理，藉此節省20%之回覆時間。

◆ 無需撰寫程式碼：

APMIC CaiGun AI在後端管理平台採圖形化與Prompt範例模式組成，企業AI系統管理人員僅需基本模型維護與訓練流程等知識，就能著手微調模型，自行定期校正，讓AI應用隨時具有最新的企業內知識。

搭配HPE AI運算方案，讓智慧應用帶來最佳體驗

AI模型能力與GPU運算效能息息相關，APMIC CaiGun AI人機互動系統也需仰賴高速運算的基礎架構來發揮所長，如HPE ProLiant DL380a可搭配高密度GPU模組，能為AI智慧應用帶來最佳體驗。

對於不同規模的模型平台，APMIC建議搭配的HPE運算平台：

◆ 超級規模：70+B參數量

如複雜程式撰寫、長文本問答與分析、推理、提供建議等AI互動應用，可組織16台HPE ProLiant DL380a AI伺服器滿足所需。

◆ 專業應用：20B ~ 70B參數量

如單字翻譯、簡短摘要、短文分析等AI互動應用，可建置8台HPE ProLiant DL380a AI伺服器來因應基本AI之運算需求。

◆ 基本運用：20B參數量

如實體辨識、分類、情緒分析、短對話等AI互動應用，可採用1台HPE ProLiant DL380a AI伺服器完成基本工作內容。

自動化程式改寫與智慧客服提升企業競爭力

APMIC CaiGun AI人機互動系統能與許多產業專業結合，發揮生成式AI的運用效益，在IC設計產業中有廠商需要將200個以上的系統轉換語言，預估工作內容需200名工程師花費2年時間，透過CaiGun AI系統設計出的服務，僅需100位工程師與1年時間就完成作業。

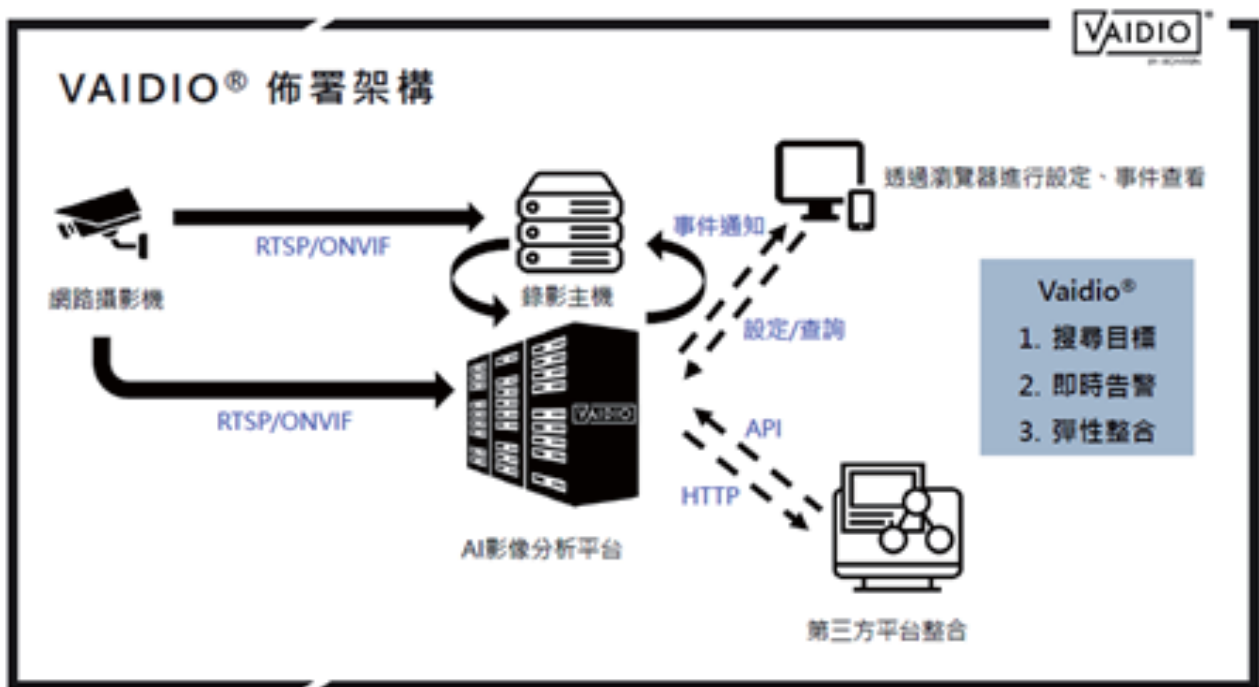
APMIC CaiGun AI系統在實務案例中也有協助航空公司打造智慧客服系統，透過API連接既有系統，包含航班、票務、CRM等系統與歷來的FAQ資料庫，結合LLM語言模型訓練理解客戶提問語意，設計出能自動回答旅程疑難、推薦行程的智慧客服，在第一時間解決客戶的問題、提升滿意度，也能分擔客服團隊處理各式客訴事件的壓力。



亞太智能機器 (APMIC) 成立於 2017 年，專注於自然語言理解之 AI 技術研發，團隊累計超過 10 年以上之 AI 相關技術開發，主要提供模型即服務 (MaaS) 是 AI 解決方案，不斷提升 AI 人機互動平台之功能，藉此實踐能普及人與機器互動間的最佳化橋梁。

網站：<https://www.ap-mic.com/>

IronYun 將 AI 視覺結合安防運用 為產業生活組織可靠安全網



將監視器影像結合AI分析，辨識影像中的情境狀態後就能搭配相對應的自動化處理。

IronYun鐵雲推出Vaidio AI視覺平台，能結合監視器等影像來源裝置，將影像資料匯入AI影像分析平台，並能設定影像分析結果設定事件通報或啟動其他程序，企業也可以將分析結果與第三方平台整合，藉此達到場域管理自動化或智慧安防等效果。Vaidio AI視覺平台也具備影像目標搜尋功能，將參數設定好就能由AI自動判定，找出影片中出現的特定事物。

IronYun Vaidio AI視覺平台可運用的在公共場所與私人空間中，為企業帶來世界級的準確性與洞察力，影像監控即服務(VSaaS)的模式能以多種模式融入企業環境，迅速與企業內的營運資料整合，並能打破地域限制，為企業改善環境安全性。

IronYun Vaidio AI視覺平台的能為企業的監視系統帶來以下效益：

- ◆ 減少誤報事件率達99.9%
- ◆ 提供30種以上AI影像分析功能
- ◆ 相較其他AI視覺平台，系統需求僅需75%

企業運用AI視覺平台，在實務上以事件調查與即時告警兩大目標為主，事件調查會將影像經由AI視覺分析後找出影像中存在的指定影像或行為，可用於後續稽核、蒐證或是將影像轉為文字知識庫等應用上；即時告警則可在偵測影像內出現特定狀況如違規、危險等情況時立即觸發告警系統，發出如警鈴、郵件等主動通報避免災難發生。

AI視覺分析的應用情境可分為人、車、物與整合環境等

- ◆ 人物影像辨識：
如人臉辨識、影像內人數統計、徘徊偵測、大型車輛接近等情境。
- ◆ 車輛相關影像辨識：
車牌、車流辨識、停車場管理、違規紀錄等。
- ◆ 物體或商品等影像辨識：
武器偵測、條碼辨識、行李檢查或手勢判讀等。
- ◆ 整合情境或範圍環境：
如燈光偵測、場域內異常監控等。

搭配HPE絕佳運算，隨需滿足影像分析規模

IronYun Vaidio AI視覺平台能以軟體授權的模式提供，也能以涵蓋軟、硬體之解決方案模式建置，將HPE ProLiant系列伺服器搭配Vaidio AI視覺平台組成最佳的解決方案。針對Vaidio AI視覺平台的效能參考指南，也能以搭配監控的攝影機數量(AI分析的資料來源數)來計算。

- ◆ 50CH的小規模應用，可參考HPE ProLiant DL320伺服器搭配1組GPU模組
- ◆ 100CH的中型規模應用，能搭配ProLiant DL380伺服器及2組GPU模組
- ◆ 200CH的大型規模應用，推薦ProLiant DL380伺服器並安裝自4組GPU模組

若需要支援更多的影像來源，或需要更高效能滿足低延遲目標，HPE ProLiant系列伺服器也都有相對應的方案，依循實際需求輕鬆購置、彈性擴充。

AI影像自動分析提醒，強化交通與工廠安全性

在台灣，IronYun Vaidio AI視覺方案獲得許多公部門與企業青睞、導入，例如台鐵公司在新北鶯歌的平交道車輛監測系統上就採用Vaidio的方案來打造緊急通報系統，能在人、車誤闖平交道時及時通報列車駕駛與站務人員處置，避免嚴重的憾事發生。

企業運用上，則有工廠單位運用監控室內的攝影機，藉此確保監控室內隨時有人員待命、因應緊急事故，當偵測到人員全部離席時就會主動提醒，避免因人員空窗而錯漏第一時間的緊急處置。

透過智慧影像輔助，讓我們的生活與產業運作更加安全可靠。

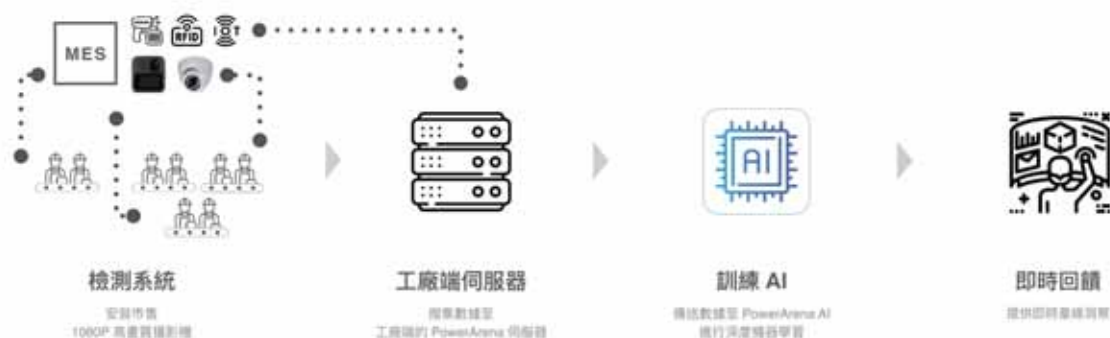


鐵雲科技 (IronYun) 於 2015 年自台灣工業技術研究院自主獨立發展而成，主要專注於發展 AI 智慧視覺分析技術及管理平台等解決方案，服務觸角遍及台灣、美國、新加坡及歐洲，在監控與 AI 視覺等方案獲得 IDC 與 ISC West SIA 等機構評選為最佳方案。

網站：<https://www.ironyun.com>

Power Arena 以 AI 電腦視覺助製造業人機協作最佳化

將AI影像分析技術結合製造業，將能在生產流程與產品品質上大幅提升工作效率。



PowerArena HOP人因作業平台 (Human Operation Platform)是以AI視覺技術為驅動的智慧智造解決方案，能整合產線上的攝影機，分析錄製的影像，並運用 AI 深度學習影像分析，迅速掌握產線上的運作情形，並將作動研究、週期時間、產品狀態、生產過程等數位化，大幅提升第一線數據匯整效率。同時，針對產線生產過程深度分析並，提供完整的生產履歷報告，協助產線人員與製造工程師們做為決策輔助，逐步提升產線產能與產品品質。

產線上的 AI 影像分析亦可用作即時產線管理，確保產線上的每個工作環節都能按

照標準生產作業流程，並符合生產規範等。

透過PowerArena HOP人因作業平台能協助製造產業：

◆ 輔助產線平衡改善：

建立數位工作站，以AI影像分析產線上各工站的作業流程，迅速量化產線狀態以利持續改善。

◆ 協助作業員充分依循作業程序：

即時把關重要程序，透過AI視覺輔助管理作業流程並稽核工作流程，藉此確保每個工作環節都有符合標準生產規範與流程等。

◆ 減少停機時間：

可持續性的匯整分析產線狀態，輔助管理人員發現異常數值時可及早因應，找出低效率生產肇因，促進生產流程改善，提高整體生產效率。

將第一手影像轉換為產線成長回饋之能量

PowerArena HOP人因作業平台能從製造產線開始檢測，幫助企業完成產線數位轉型，邁向工業4.0。

搭配產線或機台上的攝影設備，收集第一手產線資料，蒐集匯整後以PowerArena的邊緣伺服器或本地伺服器中儲存，接著將資料傳送至PowerArena AI平台進行深度學習運算，完成之後就能將產線上的各種資料即時回饋，提供完整的統計資料供改善生產流程或降低停機時間。

搭配產線或工站上的攝影設備，搜集第一手產線資料，蒐集匯整後至PowerArena的邊緣伺服器或本地伺服器中儲存。影像資料會透過PowerArena AI平台進行深度學習運算，完成之後就能將產線上的各種資料即時回饋。人員可以從PowerArena HOP人因作業平台取得可視化的生產報告，作為改善生產流程的重要依據。

PowerArena HOP人因作業平台突破傳統管理時間效益

	HOP 人因作業平台	傳統管理模式
產線觀測方式	AI 視覺鏡頭主動回報	依賴產線人員人工觀測
循環時間	AI 主動分析、提供精準數據	人工處理，被動分析、計算各個步驟
擷取時段	全年無休皆可蒐集	實務上每個月至多半小時
結果分析	即時匯整分析生成	平均三至五天工作日
資料品質	動態數值捕捉，持續提升品質	可能因人工誤差缺乏精準

結合HPE運算方案 精確發揮產線影像AI分析效益

AI系統的基礎就是IT資訊架構，PowerArena HOP人因作業平台搭配HPE ProLiant系列伺服器，能充分發揮AI影像分析的精確度，找出改良產線運作的關鍵因子，充分獲得智慧製造所帶來的價值。

針對工廠端有頻繁的AI分析與推論處理的情境時，HPE ProLiant DL380a AI伺服器結合高密度的GPU運算力，能充分發揮PowerArena HOP平台的AI視覺捕捉技術，規模較小的產線可搭配安裝GPU模組HPE ProLiant標準伺服器，兼顧效能與成本。

HPE ProLiant系列伺服器均有HPE矽晶片信任根等多種安全技術，能幫助第一線的PowerArena HOP平台強化網路安全的防護能力，避免AIoT環境遭到駭客綁架或入侵。

協助半導體與電動機車產業 提升生產品質

在台灣，已有半導體業者採用PowerArena HOP人因作業平台管理其產線生產過程。AI視覺協助24小時7天不間斷的搜集產線影像，並透過AI視覺分析，偵測生產錯誤與異常，協助管理者追溯不良肇因，進而提升製品質良率。

除了精密的高科技產業，亦有電動機車業者將PowerArena HOP平台結合MES製造管理系統，於組裝站以監控影像紀錄人員作業流程，再用AI分析確認人員完整執行如上油、鎖付等工序，雙軌制輔助稽核將產線不良率降至0%。

雙軌制輔助稽核將產線不良率降至0%。若 AI 視覺辨識到異常的生產流程，系統會即時告警，管理人員可以透過影像回放，即時確認異常原因，並決定是否放行工站，防止錯誤進到下一個工站。



百威雷科技 (PowerArena) 專注於 AI 視覺技術，透過機器學習與電腦視覺應用，將複雜的生產影像轉化為支持生產決策的數據資料，進一步以 AI 演算法解決低效率問題。企業藉由 PowerArena 的 AI 視覺技術，能達成更有效的管理，影像分析協助工廠升級智慧產線，達成智慧製造的目標。

網站：www.powerarena.com

凌羣商業級 AI 方案兼容資訊安全與產業知識，為模型生成帶來極致價值

凌羣精準式回答AI解決方案-NeuroChain

人與AI之間建立中間層，建立地端知識庫，使大語言模型針對中間層範圍精準回答



透過生成式AI的互動，企業用戶能以問答模式快速完成商業知識整理或產生程式碼，提升工作效率。

凌羣電腦將生成式AI技術商業化，開發出NeuroChain與NeuroCodie兩套解決方案。凌羣NeuroChain能結合企業知識庫，在AI模型與使用者之間建立中間層，讓AI能精準回應使用者的問題與需求，藉此加速問題處理的效率；凌羣NeuroCodie AI方案能根據指令生成程式碼，輔助企業加快軟體與專案開發之進度。

NeuroChain保留本地資料完成AI模型訓練

凌羣NeuroChain生成式AI解決方案可基於其他LLM模型，建構一層中介層作為使用者與LLM模型的橋樑，NeuroChain可使用企業內的現有資產如網頁、說明書與產品介紹等內容儲存至NeuroChain中，快速打造專屬企業與政府機關的AI控制塔與知識庫。相較於其他需要將內部資料上傳至雲端AI平台的方案，凌羣NeuroChain能讓企業資料保留在本地端運作，讓企業在活用第三方AI模型時，能在重要資料不上雲的安全訴求下，同時完成具備企業知識的AI模型訓練，藉此兼顧內部知識管理與雲端應用的平衡點。

◆ NeruoChain GAI智匯共通平台

凌羣NeuroChain GAI智匯共通平台能靈活切換

多種大語言模型，幫助企業量身打造專屬之AI知識庫系統，結合高速運算之向量資料庫，可協助企業迅速更新知識庫，平台亦可全面建置於地端環境內，因應公務機關與金融業者之資料管理所需。

◆ NeruoChain SysTalker大語言模型對話方案

凌羣NeuroChain SysTalker大語言模型對話方案專為企業需求設計，能將系統建置於本地端以確保機敏資料之安全，且具備高度客製化的特性，能有效整合內部專業知識庫與企業所需之AI訓練流程，使AI模型在諮詢回應時精準正確，同時符合企業內的制度規範與標準。

SysTalker幫助企業員工以對話方式獲得最佳解答，一改過往傳統搜尋方式仰賴關鍵字與結果翻閱等繁複作法，企業也能透過API與SysTalker整合智慧學習，持續優化問答精確度，包含客服支援、內部培訓與決策輔助等，大幅提升工作效率。

NeuroCodie用AI生成助專案開發事半功倍

凌羣NeuroCodie結合了Azure的Open AI雲端模型，亦具備針對程式撰寫等工程術語，能理解工程師的指令，運用標準化的問答流程讓Open AI產生對應的程式碼，包含Java、.Net Framework、SQL語法等程式



語言都能支援，針對重複性高、泛用型的主流程式碼...等，NeuroCodie都能輔助生成，同時在程式碼除錯、整體架構設計、專案管理等工作均能因應，不僅能大幅加快應用服務的開發效率，也能確保後續專案開發的標準化與一致性，後續的維護、擴充也更加簡易。

HPE白金級經銷商
服務涵蓋基礎運算至AI應用

當企業想運用凌羣NeuroChain打造自主的生成式AI服務，或是借助凌羣NeuroCodie來加快專案效率等追求AI應用要在內部環境運作的情況時，就需要搭配HPE ProLiant伺服器來處理人機介面的運算作業，即便採用如OpenAI等外部的AI模型，當企業要訓練AI具備營運知識時也需要強大的算力來處理，如HPE ProLiant DL380a所具備的高密度GPU算力，讓NeuroChain與NeuroCodie的具備超低延遲的優質回應。

凌羣除了提供多樣企業級的軟體服務與解決方案，同時也是HPE的白金級經銷夥伴，轄下團隊擁有HPE技術支援及建置工程相關之專業IT技術認證，為企業打造

NeuroChain生成式AI專案時，能同步協助建置、整合HPE運算系統、活絡AI營運相關之資料管理等，從硬體基礎、導入顧問到AI上層應用皆能輕鬆導入。

整合大語言模型助智慧機器人因應多元任務

在台灣，凌羣Ayuda智慧型服務機器人已實踐AI賦能，而機器人背後的AI模型就是以凌羣NeuroChain作為中間層、同步整合專業知識與使用者互動所需的語意分析等功能，讓Ayuda智慧型服務機器人具備導引、醫療或安防等作業能力，依據任務設定所需，分別在醫院、藝文中心等地負起不同的智慧服務。

凌羣NeuroChain與NeuroCodie也獲得日本等海外業者肯定，包含日本的CII株式會社，將凌羣的生成式AI方案引入日本，為日本的企業提供多元且具備高度擴充性的AI應用服務，台日合作，共同協助當地企業提升數位服務的品質、共享AI轉型的碩大成果。



凌羣電腦 (SYSCOM) 在 1975 年成立，服務據點分布在台灣、日本、美國與泰國等地，多年來致力於資訊系統整合服務，不斷導入、開發各類頂尖的設備與軟體等，為金融、電信、醫療與公部門等機構賦予最先進的數位服務，透過數位加值的服務模式，不斷協助企業數位轉型、充分發揮其產業之核心競爭力。

網站：<https://www.syscom.com.tw>

大同世界科技開發 ibo.ai 智慧專家，為企業推動內外交流商機



整合LLM大語言模型，結合專業知識打造生成式AI，讓人機智慧服務成為企業營運好幫手。

大同世界科技開發ibo.ai智慧專家助手平台，能為企業結合LLM大語言模型、AI資料分析技術，以及RPA、CRM、KB等企業營運的重要系統，讓AI成為快速解決問題的好幫手，不論是應用在對外客服場域，或是企業內部各部門的支援，都能透過ibo.ai的AI平台改善相關事務的作業流程。

大同世界科技ibo.ai智慧專家助手平台，能搭配多種平台的文字管道，如即時通訊軟體，以及社群網站訊息等，當客戶傳訊提問常見的問題時，智慧專家就會主動回覆相關資訊，立即解決客戶的問題；當有進階問題出現時，也能同時提供給專職服務人員的tiCC多媒體AI客服值機平台。透過ibo.ai智慧專家助手與真人值機的協作輔助，除了能提供24x7的即時服務，也能迅速解決多數客戶的疑難，降低服務人員處理瑣碎問題的比例。

大同世界科技ibo.ai智慧專家助手平台可於四大領域中發揮AI效益：

- ◆ 智慧專家助手：
將AI結合內部系統養成專業資料庫，並由機器人回答客戶常見問題，或是協助客戶輕鬆完成自助服務。
- ◆ 客服值機助手：
讓AI分析客戶問題，並結合專業資料庫動態生成建議回應指南，協助第一線客服人員精準處理客戶問題。
- ◆ 企業AI助理：
訓練AI整合企業內部知識庫或工作流程，藉此協助員工以AI查閱相關資訊，或從中排除工作中的疑難雜症。
- ◆ 語音AI虛擬人：
用語音的方式與AI虛擬人交談，搭配不同的軟硬體及知識庫，讓AI虛擬人與使用者互動交流。

除了提供高效益的AI智慧專家助理平台，大同世界科技也提供平台建置顧問服務與專業對話設計的AI訓練師服務，協助企業將AI系統微調訓練至最佳狀態。在資訊整合服務累積30年經驗的大同世界科技

也能提供AI所需的算力資源整合、底層硬體系統建置等，能以一條龍模式協助企業AI從零開始到顛峰極致。

智慧客服系統加持 對外更滿意、對內高效率

HPE白金級夥伴 基礎系統到AI應用全面支援

在AI算力資源整合與企業資安考量上，大同世界科技ibo.ai智慧專家助手平台搭配HPE ProLiant系列伺服器(Compute for AI)，來規劃適合落地LLM生成式AI的GPU算力，讓專家平台具備即時回應能力的同時，也使企業的投資效益發揮到最大。

大同世界科技為HPE全系列產品線技術服務的最高級之白金級夥伴，ibo.ai智慧專家助手平台搭配LLM算力HPE伺服器、資料存儲方案，以及AI資源管理軟體等，大世科的團隊都能提供專業協助，從評估、系統整合、AI訓練，到後續系統維護最佳化等服務皆能完整提供，以從產業應用服務角度協助，充分降低企業邁向AI轉型的門檻。

智慧客服許多應用為處理客戶需求，如醫美診所即時通訊APP的官方帳號上提供療程諮詢，以AI分析、判斷客戶的提問，詳細回應醫療問答，並精準推薦適合的醫療專案，讓診所的服務滿意度提升20%以上。

對內服務的部分，台灣航運業者將訂票APP整合大同世界科技ibo.ai智慧專家助手平台，設計內部智慧訂票系統，並能透過語音處理訂票、變更、取消等票務工作，也能快速查詢機位、航廈等資訊，降低使用者操作時間達90%，票務服務的處理能量也能充分保留給其他訂票旅客。



大同世界科技 (大世科 TSTI) 成立於 2000 年，前身為大同公司資通業務處，在資通訊設備供應與整合服務耕耘超過 30 年，為台灣服務業 500 大企業之一。近年隨著 AI 技術發展，大世科服務範疇由 IT 延伸至 AI，雙軌服務，為企業 AI 賦能、永續轉型。

網站：<https://www.etatung.com>

創泓以在地化 AI 語音技術 協助企業將智慧應用更臻極致

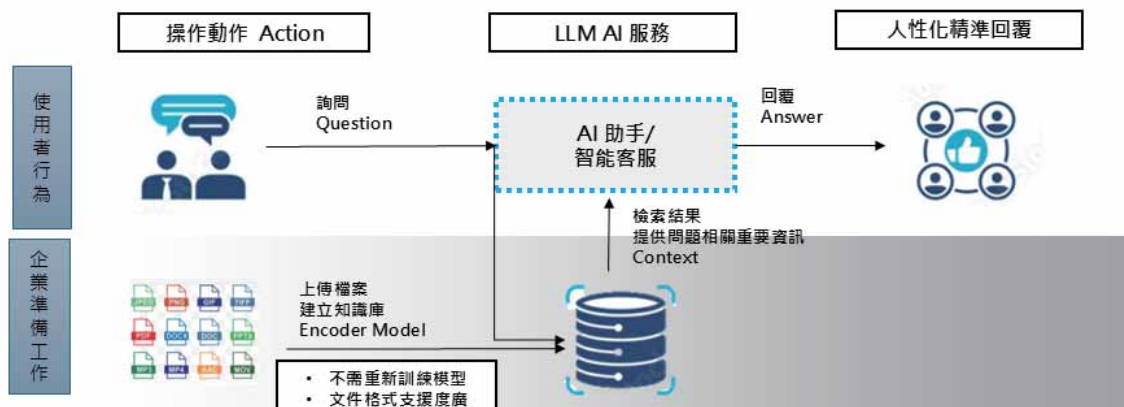


透過AI語音技術能聆聽人類對話來觸發指令或與人對答，而在國人的日常對話中經常混入不同方言與英文，如此習慣也代表著AI語音需要識別更多元的語系來「理解」語意。

創泓AI語音辨識與語音合成(ASR/TTS)解決方案，能同步支援國語、台語、英語及客語等多語系混合辨識與合成，針對企業需求與在地化目標量身打造高效語音解決方案。

AI語音辨識與語音合成(ASR/TTS)解決方案可支援多語系，優勢包含：

- ◆ 高準確度國語、台語、英語及客語辨識引擎
- ◆ 支持即時語音串流(Streaming)辨識及檔案(File)辨識
- ◆ API及SSI系統串接服務



借助AI語音技術 提升會議效率與客戶服務品質

企業營運時運用AI語音識別技術能在許多工作環節提升效率，根據創泓的客戶需求經驗分析，前兩大的應用場景分別是在會議中整理會後報告，以及提供智慧客服電話來提升對外服務品質。

一、智慧語音會議記錄系統：

從日常例行性開會到企業高層決定關鍵決策的會議，企業經常需要投入人力與時間整理會議紀錄，透過創泓語音辨識與語音合成(ASR/TTS)解決方案，就能在短時間內完成相關資料整理，快速將開會過程與相關共識精準紀錄，同時提升企業決策的效率。

創泓智慧語音會議記錄系統的特色包含：

- ◆ 自動生成會議逐字稿
- ◆ 支援多語系會議內容即時辨識與記錄
- ◆ 內建AI自動摘要功能，快速整理重點
- ◆ 自動語者識別
- ◆ 設備落地，高度保密

創泓語音辨識與語音合成(ASR/TTS)解決方案能與企業自行建置的本地環境中運行，搭配HPE AI伺服器發揮可最大效益，建議硬體配置如下(以1線授權)：

【HPE ProLiant DL320】

CPU: Intel Xeon Gold 6438Y+ 2.0GHz 32-core以上/
GPU: NVIDIA L40s*2 / RAM: DDR4 64GB以上 / SSD :
1TB / HDD: 6TB

二、語音智慧客服系統：

客戶服務專線是許多企業建立對外服務管道的第一線，隨著AI技術興起，不論是透過機器人直接與來電客戶互動、應答，或是擔任客服人員的人機輔助系統，創泓語音辨識與語音合成(ASR/TTS)解決方案搭配檢索增強生成與大語言模型(RAG/LLM)技術，就能建立企業專屬的智慧語音客服系統，不僅能迅速回應客戶需求、降低客服人力，還能適時提升服務時效與品質。

創泓語音智慧客服系統的特色包含：

- ◆ 整合IVR交換機協議，提供自動化語音互動服務
- ◆ 高精準語音辨識技術，理解客戶需求
- ◆ 結合檢索增強生成(RAG)與大型語言模型(LLM)實現智慧化客服應答
- ◆ 自然語音合成技術(TTS)提供真人般流暢語音回應
- ◆ 設備落地，高度保密

創泓語音智慧客服系統亦能在企業自建的基礎架構中運行，確保機敏客戶通訊資料不外流，搭配HPE AI伺服器穩定高效能，更有助於AI語音智慧客服系統鞏固企業與客戶間的關係。建議硬體配置如下(以15線授權)：

【HPE ProLiant DL380a Gen11】

CPU: Intel Xeon Gold 6438Y+ 2.0GHz 32-core 以上
/ GPU: NVIDIA L40s*4 / RAM: DDR4 128 GB以上 /
SSD: 1TB / HDD: 6TB



創泓科技 (Uniforce) 成立於 2010 年，主要從事資安軟硬體、顧問服務、教育訓練、雲端、人工智慧應用與無人機等產品加值銷售。創泓科技為企業數位轉型的最佳夥伴，將不斷整合創新科技，提供最先進的數位加值解決方案，同時透過人工智慧的引領，協助企業實現智慧化轉型，提高核心競爭力。

網站：<https://www.uniforce.com.tw>

滿拓實踐地端訓練目標 打造企業專屬頂級生成式 AI 服務

企業營運累積起的各式專業知識是珍貴資產，如何結合模型訓練打造出專屬的生成式AI服務的過程中，兼顧法規與資料安全等訴求是AI轉型的一大關鍵。



滿拓微型化技術 讓企業輕鬆部署生成式AI

滿拓的微型化技術，能夠將龐大參數量的 AI 模型縮小，使其適應不同規模的地端資料中心，成為企業可自主掌控的資源。這項技術打破了 AI 模型僅能依賴 AWS、Google Cloud、微軟 Azure 等公有雲運行的限制，使企業能夠在本地部署 AI，支援日常業務。透過這種微型化與本地化部署，不僅大幅降低硬體建置成本，還能確保數據處理與 AI 推論和訓練的高效運行，為企業帶來更靈活、安全的 AI 應用方案。

打造符合客戶端場域需求 提升其生產力的智慧助理

滿拓結合豐富的AI實務經驗，由應用端切入客戶的日常需求，運用Document Expert協助建立企業專屬知識的「大腦」，讓AI在回答問題時，能夠根據企業內部的資料提供更準確的答案。透過結合使用者提問、文件內容及外部資訊，Document Expert能夠跨文件搜尋最相關的原始資料，並生成高度精確的提示詞，協助大語言模型回應使用者需求。

此外，滿拓為打造更優質的服務體驗，提供融入使用者的專業領域知識的方案，讓語言模型能針對特定業務場景進行微調(Fine-Tune Expert)，打造更貼合客戶需求的專屬解決方案。

Fine-Tune Expert高度掌握且靈活特性，提升AI服務競爭力

一、安全性與隱私保障：全面掌控，無憂運行：

企業對數據的掌控從未如此完整。無論是日常操作還是處理機密資訊，您都能確保數據的高度安全性和隱私性。我們的方案特別針對那些需要保護敏感數據的企業設計，讓您在合規和風險管理上無後顧之憂。借助企業完全掌控數據的優勢，提升數據安全保障，減少潛在的風險，讓您的業務運行更穩健、可信。

二、成本效益：長期部署，更具競爭力的成本選擇：

面對日益增長的業務需求，許多企業希望能以更具成本效益的方式部署 AI 方案。我們的地端部署方案，提供長期運行所需的穩定性，並且成本相對低於雲端模式。這意味著您的企業不僅可以控制運營開支，還能在保證系統穩定的同時，實現更高的靈活性，為業務增長奠定堅實的基礎。

三、無程式碼平台：讓每個人都能輕鬆上手，輕鬆自動化：

我們的無程式碼平台打破了專業技術門檻的限制，讓任何具有業務需求的員工都能輕鬆上手操作。無論是管理者、業務專員還是客服人員，都可以在無需編寫程式的情況下，創建和優化自動化流程。這不僅大幅降低了企業內部的技術門檻，也讓智能化升級更易於實現，幫助您的團隊提升效率，推動業務創新。

四、模型自主權與定制化：專屬優化，靈活應用：

擁有地端部署的模型，意味著企業能夠根據自身需求進行深入微調和優化，從而打造最符合業務需求的 AI 解決方案。資料工程師可以根據實際情況靈活調整模型，持續提升系統性能，而不必擔心數據外洩或無法控制的風險。企業可根據業務變化進行快速調整，實現未來擴展與應用的無縫對接，讓每一次調整都能更貼合市場需求，保持業務競爭力。

結合滿拓Fine-Tune Expert的精細調整，亞洲跨國翻譯產業成功導入AI技術，藉此提升翻譯稿件潤飾效率，不僅彌補傳統翻譯軟體在語言流暢度的不足，也讓翻譯效率提升達2倍，人力成本也顯著降低，幫助該企業於翻譯產業中獲得顯著的競爭優勢；在高科技製造業中，記憶體大廠透過滿拓設計的AI微調解決方案，有效提升該企業自有的RAG (Retrieval-Augmented Generation)系統效率，大幅改善了FAE工程團隊在龐大產品資料庫中比對型號、規格與特性之搜尋效率，AI改善的不只是工程團隊前期的準備工作，從前段需求確認到產品交付和售後服務流程的工作效率也因此提升達40%，同時也贏得許多客戶訂單。



滿拓科技 (DeepMentor) 提供從雲端到地端部署環境的全方位整合服務，確保部署流程流暢且無縫銜接。透過獨家研發的 LLM 自動化訓練流程，能在高度自動化下，以最短時間推出新型 AI 服務，從雲端到地端設備落地部署，支援在地服務，降低營運成本並確保資料隱私安全，協助企業快速搶占市場先機。

網站：<https://www.deepmentor.ai>

網智服務 IQ-KB AI 大知識庫平台 扮演企業數位大腦重要角色

扮演企業數位大腦角色



企業營運會累積許多的知識文件，例如產品設計的關鍵組成、內部規章或教育文件等，大量的文檔綜合成營運知識的核心，借助AI之力讓知識匯流萌發未來創新。

由網智服務推出的IQ-KB AI大知識庫平台，整合檢索增強生成(RAG)+ 大型語言模型(LLM)技術，使其在知識管理領域具有以下特色：

一、智能檢索與生成的知識中心

網智服務IQ-KB 以 RAG+LLM 技術為核心，不僅能檢索關鍵資訊，還能透過生成式 AI 自動補充知

識、生成客製化回應。相比於傳統 KM 系統的靜態資料庫，IQ-KB 更具動態智能，適合現代企業對即時知識應用的需求。

二、企業級知識整合與協同平臺

不僅是知識管理系統，還是協作中心，通過智能推薦、自動分類和跨部門共享等功能，將內外部知識統一整合。

相比傳統 KM 系統，網智服務IQ-KB 更能打破部門資訊孤島，實現真正的企業知識協同。

三、適應不同規模企業的擴展性架構

可根據企業需求靈活擴展，支持多語言、跨區域的應用場景，從中小型企業到大型跨國集團皆能量身定制。其可擴展的模組化結構讓企業隨著業務成長逐步增加功能與覆蓋範圍，是知識管理的長期方案。

四、資料安全與合規支持的知識管理方案

針對大型企業和跨國組織，提供高度安全的權限控制和合規性管理，滿足多地法律 and 資料保護需求。使其成為在資料安全上有嚴格要求的企業理想的知識管理選擇。

◆ 支援多模組

為企業提供全方位服務，包括考試、網站同步、OCR 文件辨識及小組網站模組等。

◆ 自動生成知識

生成式AI自動建立知識，管理者只需審核後即可發布知識。

支援的檔案格式包括: doc、docx、txt、pdf、md、xls、xlsx、csv、ppt、pptx、json、jsonl...等

◆ 知識共享功能

善用知識分享功能，促進跨部門資訊共享，同時降低維護成本。

平台核心具備完整功能，企業知識管理更靈活

◆ 跨平台整合

知識與社群文字服務整合於同一平台，可快速查找、統一管理，提升效率與服務品質。

◆ 智慧搜尋與生成

提供關鍵字搜尋與生成式 AI 知識查找功能，讓查詢更快速、更便捷。

◆ 知識權限控管

提供完善的資料存取權限管理，確保不同使用者能獲得對應的知識讀取權限。

網智服務IQS的客戶群涵蓋台灣百大企業，產業領域遍及銀行、壽險、證券、電商、電信、航空、流通、政府、資訊科技等，包含中華電信、中國信託、臺灣中小企業銀行、長榮航空等台灣知名企業，均採用網智服務的方案來管理各領域營運時不可或缺之數位化知識。隨著知識更新速度加快，企業亟需具備即時更新與自動化生成功能的知識管理平台。

網智服務IQ-KB AI大知識庫平台可幫助金融、科技、醫療等產業有效管理龐大知識資訊，在知識密集型企業持續成長趨勢下，協助企業處理海量資訊並提升員工生產力，確保競爭優勢。



網智服務國際股份有限公司

〈 智慧知識 · 智慧服務 〉

網智服務 (IQ Service International Corp.) 成立於 2014 年，致力於開發企業全方位客戶服務系統並提供智慧型客戶服務解決方案，協助企業整合多元管道匯集而來之服務需求；主要產品 IQ-KB AI 大知識庫平台，是台灣客服知識庫的領導品牌，並與 ChatGPT、AOAI、Llama3 等生成式 AI 應用，範圍從智慧型客戶服務解決方案跨足到 AI 智能應用。

網站：<https://www.iqs-t.com>

數位無限 AI-Stack： 驅動 AI 應用充沛能量之關鍵基石

讓AI發揮商業效益的背後，需要充足的數位運算資源，如GPU運算資源等，設法將資源最大化，同時穩健易管理是企業導入AI的基本功。

AI-Stack：領先業界的 AI 基礎設施管理平台

數位無限(INFINITIX)的核心產品「AI-Stack」是一套高效的 AI 基礎設施與 GPU 資源管理平台，也是企業導入AI服務時必備的關鍵工具，專為滿足企業在 AI 開發與部署過程中的複雜需求設計。整合第三代GPU切割技術、GPU多片聚合技術、跨節點運算、全新操作介面、多雲管理、開源深度學習工具、模型推論部署服務與 MLOps 管理工具，與常用AI無縫對接開發，可預設自動執行或單次或批次任務，讓企業輕鬆解決 AI 快速迭代帶來的挑戰，幫助企業大幅提升資源利用率，加速 AI 模型開發和應用落地。

AI-Stack平台核心具備多重資源活用優勢

一、GPU切割與聚合技術：

- ◆ 靈活資源分配：利用 GPU 切割技術，AI-Stack 能將單一GPU 切分為多個邏輯單元，支持多用戶和多任務的同時執行，將GPU 使用率從傳統的 30% 提升至 90%。
- ◆ 跨節點運算：通過多節點協作運算與分散式訓練技術，能將訓練任務高效分配至多個節點，顯著縮短模型訓練時間。

二、多雲環境整合：

- ◆ 支援多雲架構，包括公有雲、私有雲和混合雲，能無縫銜接不同品牌之 GPU系統 (如NVIDIA、AMD等)。
- ◆ 整合Kubernetes等開源工具，實現快速環境部署與多任務調度。

三、全生命週期 AI 支援：

- ◆ 開發階段：AI-Stack 支持模型開發環境的一鍵建置，從過去的兩週縮短至 僅需 1 分鐘。
- ◆ 訓練與推論：提供高效的 MLOps 工具，可管理分散式訓練與即時推論流程，確保運行效率。

四、成本效益顯著提升：

- ◆ 減少資源閒置並支持動態資源調整，幫助企業提升 GPU 投資效益達 10 倍。
- ◆ 通過智能調度，優化資源分配與工作負載管理，降低硬體成本。

多種產業廣泛應用活絡AI環境

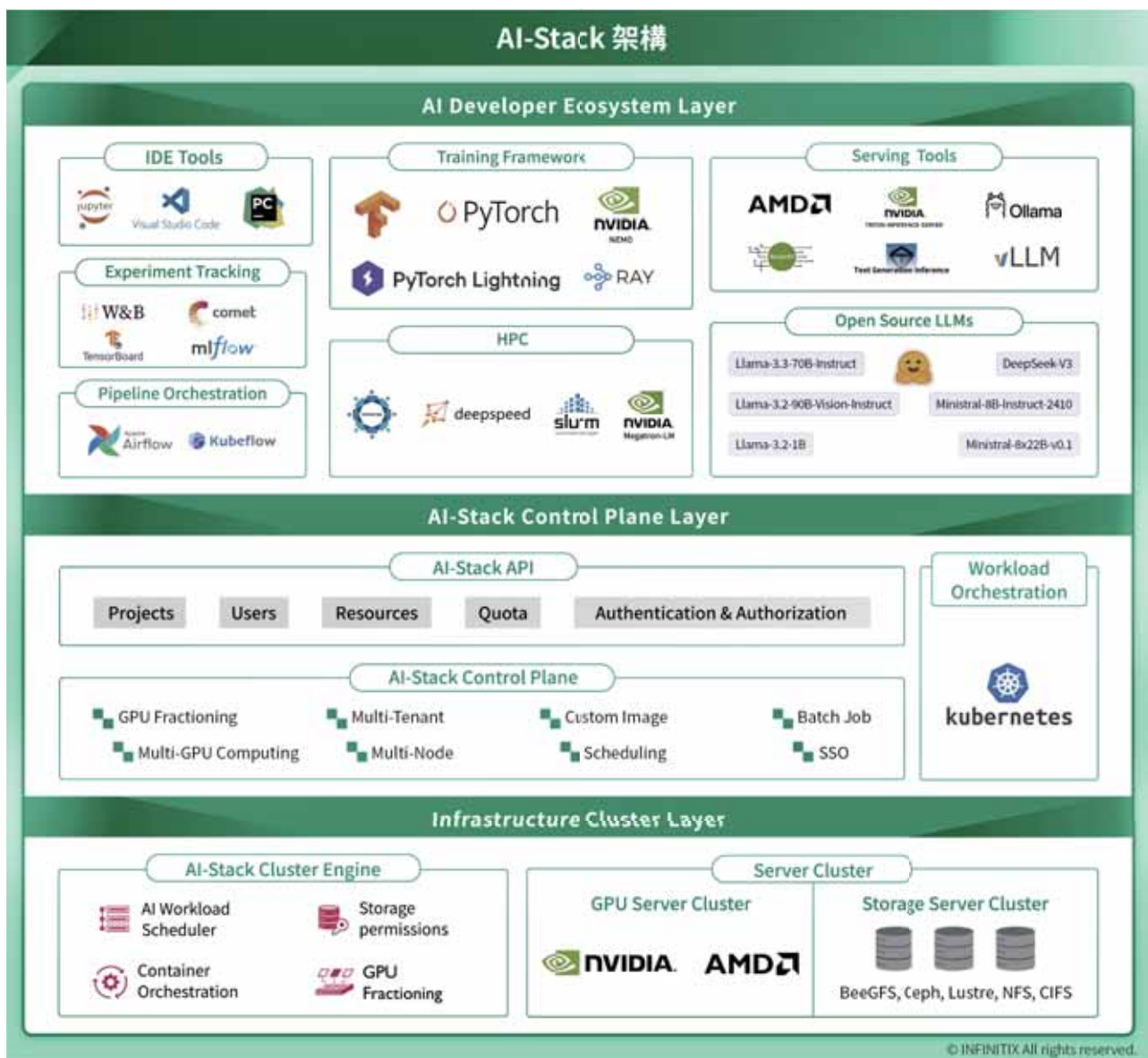
AI-Stack 已成功應用於金融、醫療、製造和教育等行業，助力企業優化資源管理並實現業務增長。

- ◆ 金融業：永豐金控應用 AI-Stack 架構內部 AI 開發與自動化部署平台，應用於貸款審核與風險評估。
- ◆ 醫療產業：花蓮慈濟醫院透過 AI-Stack 平台有效管理 DGX-A100 系統，實現 AI 模型優先級管理，提升醫療影像診斷效率。
- ◆ 製造業：中鋼集團與中冠資訊合作，運用 AI-Stack 進行產線優化與設備良率分析，提升營運效能。

產業攜手，打造台灣 AI 合作生態系統

數位無限聯手多家台灣 AI 廠商，共同打造完善的 AI 生態系統，支持國內新創公司與企業快速部署 AI 技術。例如：

- ◆ 數位發展部：導入 AI-Stack，建置國家 AI 計算資源池，支持多品牌 GPU 的資源管理，並提供開源技術與工具。
- ◆ 學術界：臺北科技大學使用 AI-Stack 支持校內 AI 課程與實驗，公平分配 GPU 資源，鼓勵創新研究。



數位無限INFINITIX將持續創新，致力於成為全球 AI 技術的領導者夥伴。AI-Stack 平台不僅是 AI 發展的利器，更是企業迎接智慧時代的核心夥伴，透過提升資源利用率與降低成本，協助企業實現 AI 驅動的卓越成果。



數位無限 (INFINITIX) 擁有 20 年以上豐富之軟體開發經驗，並以成為全球 AI 產業引領者的成功夥伴為品牌願景。本著「擁抱開源、共創價值」精神，為企業提供最先進的 AI 基礎設施解決方案，加速產業轉型升級，共同打造精彩無比的智慧 AI 世代。提供領先的 GPU 資源管理解決方案，協助企業於 AI 時代取得大躍進。秉持「創新、領先、開放、永續」的態度，與全球的 AI 生態夥伴共同打造繁榮的 AI 生態系統。

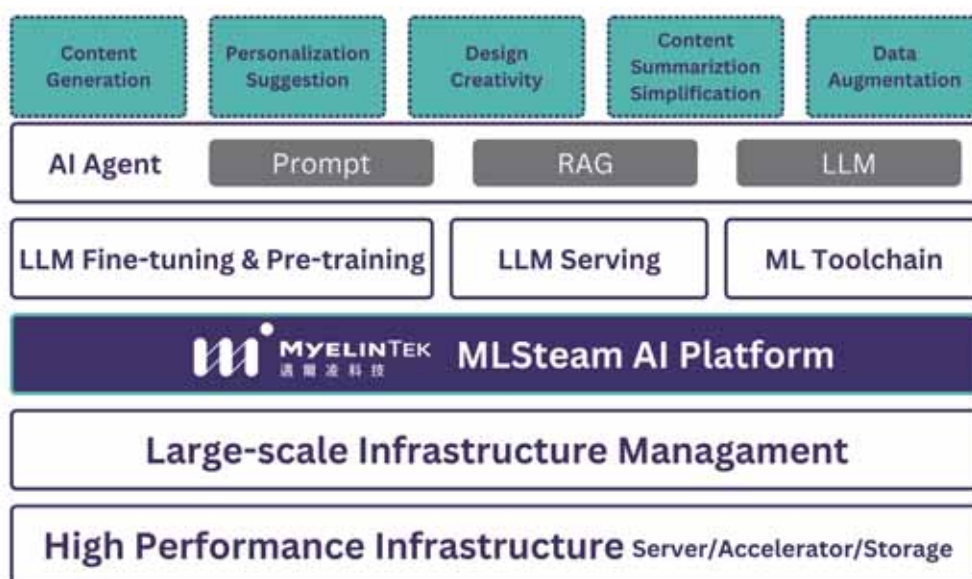
網站：www.infinitix.ai



邁爾凌科技一站式 AI 開發平台： 從硬體管理到模型訓練與優化，全面 加速 AI 落地

企業導入AI專案時，從基礎硬體整合管理、到上層AI模型訓練，以及後續持續精進優化，每個環節都需搭配專業的技術協力，才能充分發揮AI的效益。

邁爾凌科技的 AI 開發平台「MLSteam」可整合運算、管理與儲存伺服器，透過集中化的管理、監控與資源分配，實現資源利用的最大化。平台全面支援 NVIDIA CUDA、AMD ROCm 和 Intel oneAPI，能進行異質訓練，滿足多樣化的需求。



■ MLSteam AI平台，管理硬體、整合軟體，打造一站式GenAI開發環境。

邁爾凌科技的 AI 開發平台「MLSteam」可整合運算、管理與儲存伺服器，透過集中化的管理、監控與資源分配，實現資源利用的最大化。平台全面支援 NVIDIA CUDA、AMD ROCm 和 Intel oneAPI，能進行異質訓練，滿足多樣化的需求。

MLSteam 平台內建廣泛應用的framework、model template及標註工具，提供完整的開發環境，讓模型開發者登入後即可開始作業，大幅縮短設定與

調校的時間。同時，平台可統一管理所有推論節點，清楚掌控模型版本並進行高效佈署。

此外，MLSteam完整整合MLFlow功能，讓已訓練的模型可作為模板，並透過no code/low code方式實現再訓練、模型優化或新應用的開發，全面提升 AI 工作流程的靈活性與效率。

管理、整合與介接，MLSteam三大特點強化AI開發效率

◆ 管理面：

平台可統一納管管理節點、運算節點和儲存節點，並透過儀表板直觀呈現資源使用狀況，便於管理者靈活調度。同時，MLSteam將NVIDIA MIG功能以視覺化方式呈現，簡化設定流程。並額外開發GPU sharing功能，進一步提升GPU資源的利用率。

而模型開發過程涉及多個專業角色，如資料工程師、模型工程師和應用工程師等。平台提供人員編組與角色定義功能，促進跨部門協作並提升開發效率。此外，平台實現各專案與資料的隔離，有效保障模型開發的隱密性與安全性。

然而，對企業而言，資源使用的透明度至關重要。MLSteam支援GPU time-slicing，降低硬體閒置時間，同時提供硬體使用量的配額管理功能，並生成詳細的使用紀錄報告，幫助組織進行內部資源檢視與規劃。

◆ 整合面：

MLSteam內建JupyterNotebook作為模型開發環境，並整合TensorFlow、PyTorch、ONNX等框架，全面支援NVIDIA CUDA、AMD ROCm和Intel oneAPI進行異質運算。開發者可利用預先打包於MLSteam的模型模板（如YOLO、Llama等）進行開發與優化，這些模板會隨MLSteam版本更新同步上架，進一步提升開發效率。此外，平台還整合了NVIDIA NGC上的開發工具與框架，提供多樣化的開發環境選擇。

另外，MLSteam配備內建標註工具，如CVAT和LabelStudio，可協助資料前處理，並實現資料的版本管控。平台也將開發的各階段流程視覺化，開發者可透過UI介面快速設定多組超參數並同步進行訓練。而內建MLFlow功能，整合CI/CD，支援持續性模型訓練與優化，同時提供直觀的進度顯示與視覺化日誌記錄，全面提升開發體驗與效率。

◆ 介接面：

MLSteam 可對內整合組織現有的資料與管理系統，例如醫療 PACS 系統、企業內部的 NFS 和 LDAP 等，方便內部進行資料傳輸與帳號管理。對外則能與帳務系統及帳號系統進行對接，助力實現 AI 運算機房的建置目標，打造高效、無縫的運算環境。

將BMaaS提升至AlaaS，建構完整AI Cloud Service

隨著AI技術的發展，提供雲端服務的電信商從傳統的BMaaS (Bare metal as a Service) 轉型為 AlaaS (AI as a Service) 以應對日益增長的運算需求。AlaaS提供不僅是基礎運算資源，還包括深度學習及數據處理工具，支援企業快速開發與部署AI應用。

MLSteam從內部協助電信客戶進行硬體的叢集管理、監控與管控，並順利整合其現有的帳號及帳務系統。平台提供完整的開發環境，進一步提升雲端服務對終端使用者的價值，讓使用者能夠在登入後立即進行模型開發。相較於BMaaS，AlaaS不僅能帶來更高的營運效益，還能有效提升使用者的黏著度。

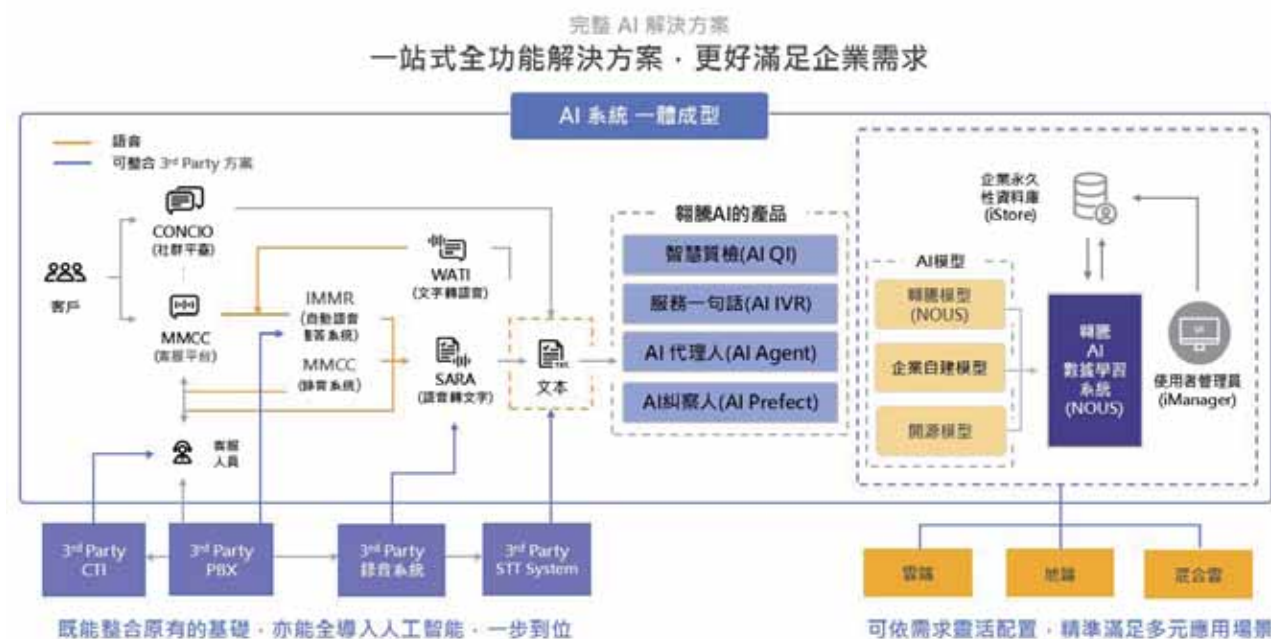


邁爾凌科技 (MyelinTek Inc.) 成立於 2020 年，專注於自主研發 MLSteam AI 開發平台，並應用於模型開發、訓練與推論管理。公司提供多元的專業服務，涵蓋各領域的 AI 導入、私有與公有雲端服務系統設計與實作、軟硬體平台整合、叢集管理及效能優化等，致力於為客戶打造高效能的 AI 解決方案。

網站：<https://www.myelintek.com>

翱騰國際推出三大 AI 解決方案 開創「金融產業通訊客服」新紀元

隨著數位轉型在金融服務業蓬勃興起，如何有效整合通訊與客服流程已成為企業成長的關鍵。翱騰國際 (Octon International) 深耕企業通訊領域逾28年，憑藉深厚的技術積累與市場洞察，於2024年再次領航推出四大AI解決方案：服務一句話 (AI IVR)、智慧質檢 (AI QI)、AI代理人 (AI Agent)、以及AI糾察人(AI Prefect)。這些嶄新的AI應用不僅能協助金融業者優化客戶服務流程，更能大幅提升整體運營效率與服務品質，為企業的數位化進程打下堅實基礎。



服務一句話 (AI IVR) 全方位智慧語音服務

傳統多層式IVR系統往往因操作繁瑣、選單複雜而造成用戶流失。翱騰國際的「服務一句話」(AI IVR) 徹底顛覆了既有模式，用戶只需在電話或聊天視窗中輸入或說出簡短的指令，系統即可啟動後端AI引擎並進行語意分析，迅速給出精準回應。

- ◆ 單一入口整合多項服務：金融機構的多種業務常分屬不同部門；透過AI IVR，可將開卡、貸款諮詢、帳戶服務等多樣需求整合在同一語音介面下完成。
- ◆ 減少用戶等待與錯誤導流：傳統客製化選單可能導致用戶反覆選擇、等待接線；AI IVR可以自動辨識語意並直接導向最適服務流程，確保用戶在最短時間內獲得協助。

翱騰「服務一句話」方案中創新的AI驅動技術能夠精準回應、多語言支援，大幅提升服務效率與客戶體驗，適用於所有需要即時互動的金融服務場景。

智慧質檢 (AI QI) 輸入條件，質檢立出

在金融產業中，客服品質直接影響客戶忠誠度與品牌聲譽。然而，傳統的人工質檢方式往往因人力有限或主觀偏見而導致檢核品質不一致。翱騰國際的智慧質檢系統 (AI QI) 透過AI技術提供自動化且客觀的品質檢驗：

- ◆ 快速篩選關鍵對話：系統可即時搜尋並評估所有對話或通話記錄，只要輸入特定條件（關鍵字、通話時長、風險等），便能迅速生成詳細報表。

- ◆ 節省人力、提高準確度：AI QI自動過濾不合格項目，並以圖表或文字形式呈現客觀數據，使管理者能一目了然地掌握客服狀況並即時調整策略。

- ◆ 持續優化全局流程：藉由不斷收集與分析回饋數據，AI QI可協助企業精準發現問題根源，進而優化後續服務流程，最大化整體營運效益。

AI代理人 (AI Agent) 隨時在側，智慧引導

金融產品日趨多元，客戶需求也愈加複雜。翱騰國際的AI代理人 (AI Agent) 系統能同步分析通話與文字內容，並根據後端知識庫動態提供即時建議：

- ◆ 深度學習客戶情境：AI代理人會持續追蹤對話脈絡，從客戶問題、互動歷程中萃取出關鍵資訊，並即刻給予最合適的解決方案。

- ◆ 降低人工錯誤，提高服務速度：相比傳統依賴人工經驗的客服，AI Agent能同時檢索大量資訊並執行交叉驗證，確保客戶得到最準確而高效率的回應。

- ◆ 強化員工效能：AI代理人可為一線客服人員提供即時支援與指引，在複雜的業務或法規情境下，協助新進人員快速上手或資深人員維持高品質水準，讓服務質量更有保障。

AI糾察人 (AI Prefect) 智慧糾察 品質無瑕

AI Prefect 是一款專為監測與優化 AI 互動表現的工具，能檢測 AI 回應的準確性、邏輯性，並提供改進建議，適用於 IVR、客服機器人與 AI 輔助客服系統。

- ◆ 語音準確度監測 (STT/TTS)：評估 AI 在 語音辨識 (STT) 與語音合成 (TTS) 方面的表現，確保轉換準

確、語音自然流暢，提升使用者體驗。

- ◆ 意圖理解與上下文處理：檢測 AI 是否能正確理解使用者需求，並在對話中保持連貫性，避免脫離主題或重複詢問已提供的信息。

- ◆ 知識庫一致性：確保 AI 回應符合企業 FAQ、政策與數據庫，避免資訊錯誤，提升專業與可信度。

- ◆ 問題解決率：評估 AI 是否能快速滿足用戶需求，減少客服轉接，並提供足夠資訊讓使用者自行解決問題，提高整體服務效率。

- ◆ 用戶情緒分析：透過分析語氣、負面語句與重複查詢來識別使用者情緒變化，優化 AI 回應，減少困惑與不滿，提高互動體驗。

AI Prefect 以 多維度監測與分析，確保 AI 互動精準流暢，幫助企業提供更高效、可靠的智慧客服體驗。

翱騰國際各類AI方案透過高效能的人工智慧演算法與先進的數據分析能力，協助企業快速挖掘商業洞察、簡化決策流程並提升整體競爭力。在多次安裝與實際測試後發現，採用HPE Proliant DL380a伺服器並搭配NVIDIA H100 GPU模組，可充分釋放翱騰國際AI方案的潛力，進一步展現最佳效能與穩定度。能有效縮短模型訓練時間、優化推論效率，並確保在各種複雜應用場景中都能維持穩定的運行環境。透過此硬體與軟體的完美結合，企業能更快速地落地 AI 應用、加速數位轉型並在瞬息萬變的市場中取得競爭先機。

翱騰國際深信，唯有不斷升級企業通訊與客服系統，才能在高度競爭的金融市場中脫穎而出。我們誠摯邀請企業與合作夥伴攜手邁向AI通訊的新世代，讓技術與創新真正成為服務升級的動力。這四大AI解決方案將重新定義數位時代的服務交付模式，為金融業者帶來更高效、更加精準的決策支援，以及無限的成長可能。



翱騰國際 (Octon International) 擁有逾 28 年金融業客服系統開發與整合經驗，在台灣金融業客服系統領域，以超過 85% 的市占率穩居龍頭，每日平均為超過 500 萬用戶處理逾 20 萬通話，充分展現翱騰國際在金融產業的卓越實力與深遠影響。翱騰國際近年來結合 AI 與通訊技術，讓通訊不再只是聯繫工具，更成為企業賦能的關鍵利器。

網站：<https://www.octon.net/>

頑碼 MaiAgent 獨家 RAG 技術結合 公、私有雲和地端部署方案



產品架構圖

MaiAgent：新 GEN AI 世代智能 助理解決方案

在人工智慧快速發展的今日，企業面臨著如何有效運用 AI 技術提升營運效率的重要課題。頑碼 MaiAgent 作為一款領先的 AI 助理解決方案，透過其全方位的技術整合與創新應用，為企業提供了全方位企業智腦，數位轉型，強大而靈活的智能化工具。

◆ 領先的技術架構

頑碼資訊MaiAgent的核心優勢首先體現在其深厚的技術實力上。系統支援包括 GPT-4、Claude 3.5 在內的多款頂尖大型語言模型，並結合先進獨家的 RAG (檢索增強生成) 技術，確保回答的準確性和可靠性。其模組化的系統設計不僅支援靈活的模型切換，更提供完整的 Embedding Model 支援，並整合

了光學文字辨識 (OCR) 功能，能夠處理多樣化的資訊輸入。

在部署方面，MaiAgent 提供了極具彈性的選擇，企業可根據需求選擇公有雲、私有雲或地端部署。這種多元的部署模式，配合開放的 API 接口設計，使得系統能夠輕易地與企業既有的 IT 環境進行整合。同時，完整的對話紀錄追蹤和資料儲存安全機制，確保了企業資料的安全性。

◆ 最高品質的回覆精準度

頑碼資訊MaiAgent 以獨家進階 RAG 技術確保最高品質的回覆精準度。MaiAgent 獨家的 RAG 技術能從龐大知識庫中精準檢索相關資訊並過濾雜訊，透過智能上下文整合技術讓生成模型真正理解問題全貌，並且特別強化了模型穩定性，即使面對複雜問



題也能產生一致且正確的回答，有效防止矛盾或不實內容。實際應用數據顯示，導入後企業客戶諮詢解答準確率提升至 95%，同時減少 75% 人工介入需求，大幅提升效率並節約成本。

◆ 強大的擴展性

頑碼資訊MaiAgent 的另一大優勢在於其強大的擴展性。系統不僅支援更多 LLM 模型的擴充，還允許自定義 RAG 接入，並提供彈性的 API 整合能力。無論是客服中心、知識管理，還是內部支援系統，MaiAgent 都能完美適應。更重要的是，系統支援多語言對話，持續整合新型 AI 模型，不斷擴充應用場景，確保產品的持續進化。

◆ 突出的商業價值

從商業角度來看，頑碼資訊MaiAgent 為企業帶來顯著的成本效益。系統支援無限量的 AI 助理配置，配合高達 10GB 的知識庫容量和每月 500 萬字的充足對話額度，能夠有效降低人力客服成本。在營運效率方面，透過自動化的客服回覆和智能工作分配機制，大幅減少人工處理時間，加快問題解決速度。

◆ 優質的使用者體驗

頑碼資訊MaiAgent 特別注重使用者體驗的設計。系統可無縫整合 LINE 和 Facebook Messenger 等主流通訊平台，並支援網頁內嵌，為使用者提供統一且流暢的對話體驗。直覺的問答介面設計，配合便捷的歷史紀錄查詢功能，使得操作更加簡單容易。此外，系統支援客製化的開場互動和個人化對話流程，讓每位使用者都能獲得獨特的服務體驗。

◆ 完善的管理功能

在管理控制方面，頑碼資訊MaiAgent 提供了全面的監控和配置功能。管理者可以追蹤對話數量、分析使用者滿意度、評估回答品質，並進行靈活的角色權限管理和知識庫更新。特別值得一提的是，系統配備了誠實性評分機制和回答相關性評估功能，確保服務品質的持續提升。

◆ 獨特的差異化優勢

頑碼資訊MaiAgent的差異化優勢體現在其高度的技術整合度和客製彈性上。作為一個多模型統一管理平台，不僅提供完整的 RAG 技術支援和圖形化管理介面，還支援企業形象、對話流程和知識庫的全面客製化。高度靈活性，加上跨平台整合能力和多語言支援，使得 MaiAgent 能夠完美適應各種企業應用場景。

◆ 各大產業領域的豐富實施經驗

政府部門：與資策會合作開發馬祖數位大腦 AI 導覽員、建置採購合規助手，並協助工研院開發智能問答系統；在教育領域，成功為政治大學及北門高中導入生成式 AI 平台，支援研究及教學應用。

金融業：與中國信託深度合作，完成多項創新專案，包括 AI 客服語意辨識、財務分析助理、168 Studio AI 導覽、永續大使等專案，並為台灣人壽開發 Gen AI 性格測驗。製造業方面，協助華城電機及台灣櫻花建置內部 AI 助理，提升營運效率。

中小型產業：頑碼資訊MaiAgent 也成功為多家企業提供服務，如為俾亭停車開發停車場智能助理、為林果良品建置產品查詢小幫手，以及為蟬說旅宿集團打造智能客服系統，展現了在不同場景的應用靈活性。

HPE 高效運算加持 MaiAgent讓企業 AI 更有勁

頑碼資訊MaiAgent與HPE AI運算方案的完美結合，為企業提供高效能的地端AI 解決方案。透過 HPE 的高性能運算伺服器 and 專業級GPU加速器，MaiAgent 能夠在地端環境中穩定運行大型語言模型，確保資料隱私與運算效能。此整合方案特別適合對資安要求較高的金融、製造等產業，讓企業在確保資料主權的同時，也能享有 AI 技術帶來的競爭優勢。系統支援彈性擴充，可根據企業需求靈活調整運算資源，實現最佳的成本效益比。

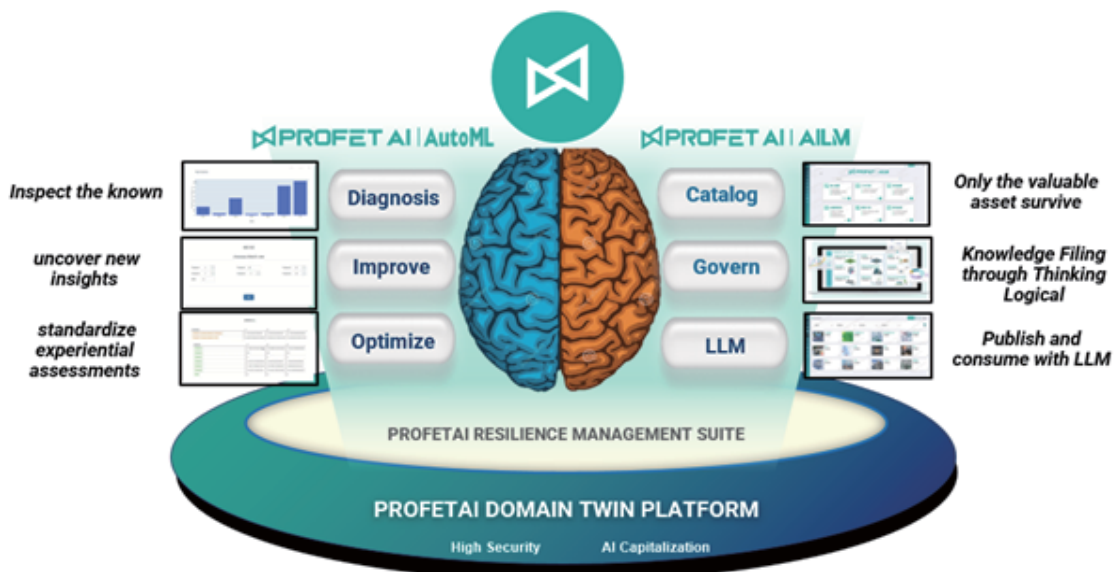


頑碼資訊 (Playma Ltd.) 專注生成式 AI 開發與維運軟體系統。提供 AI 顧問服務協助客戶導入生成式 AI，協助客戶培養 AI 領域的知識，協助客戶導入「生成式 AI、培養 AI」領域的知識。尤其在資料處理方面，顧問服務便能很好地針對不同的企業客製化解決問題，讓平台功能發揮到極致。若想擴充其他 AI 小工具，如：文件自動生成，只需在 MaiAgent 上開發即可，不需從 0 開始，先進、高效又直觀。

網站：<https://maiagent.ai>

杰倫智能 Profet AI 以定製化 AI 賦能製造產業智慧升級

The Dual-Brain Strategy of AI Internalization



製造業者各自產品定位和發展策略不盡相同，反映在其產線運作的獨特調性上。透過應用高效的 AI 平台，製造業者不僅能夠針對其特定需求進行智慧製造的轉型，還能顯著提升運營效率和產品良率。

Profet AI 產品線，AutoML + AILM + RMS共同打造 Domain Twin 全面推動 AI 內化

杰倫智能Profet AI 透過 AutoML 與 AILM 內建的 RMS (Resilience Management Suite) 方法論，打造 Domain Twin (領域經驗分身)，讓企業能夠高效管理 AI、提升決策韌性、擴展專家智慧，全面推動 AI 內化，實現輕資產、高毛利的智慧製造未來。

● AutoML自動化機器學習，降低 AI 技術門檻

AutoML (自動機器學習) 技術幫助企業快速構建 AI 模型，無需程式設計即可應用 AI

- ✓ 降低技術門檻，非技術人員也能利用 AI 進行生產優化與決策輔助
- ✓ 快速部署 AI，針對生產、品管、設備管理等場景，自動建立 AI 模型
- ✓ 結合 AILM，確保 AI 治理與最佳化

● AILM (AI Lifecycle Management) AI 治理與應用管理平台

AILM 是 全球首款 AI 生命週期管理系統，幫助企業從 AI 開發、部署到治理，確保 AI 可管、可用、可擴展。

- ✓ 內建 LLM (大型語言模型)，透過自然語言與 AI 互動，快速獲取分析與決策支持
- ✓ AI 治理與優化，確保 AI 模型符合企業標準，持續提升準確性與適用性
- ✓ 支援 AutoML，讓 AI 自動學習與適應不同製造場景

● RMS (Resilience Management Suite) AI 驅動的韌性管理方法論

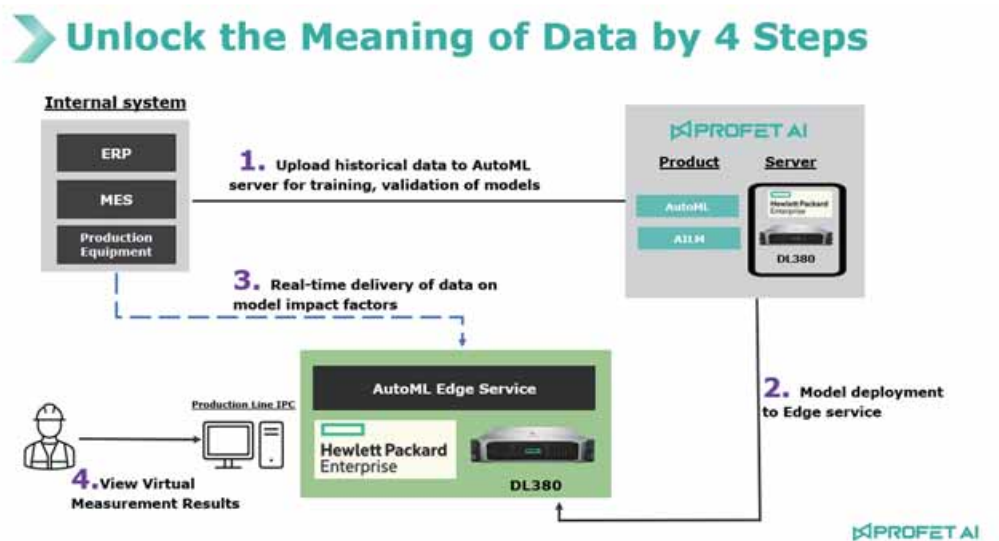
RMS是內建於 AILM 的企業決策韌性管理方法論，幫助企業建立AI賦能的應變與優化能力。

- ✓ 盤點企業核心競爭力，找出影響決策韌性的關鍵因素
- ✓ AI 強化風險預測與決策優化，讓企業更快應對供應鏈波動、產線異常與市場變數
- ✓ 與Domain Twin結合，讓專家智慧AI化，確保Know-how可擴展與傳承

● Domain Twin (領域經驗分身) AI 複製專家智慧，打造智能決策

Domain Twin是AI 轉化專家經驗的智慧決策模型，讓 AI 能夠學習並輔助專家的日常判斷，提升決策效率與準確性。

- ✓ AI 學會專家的決策模式，形成可標準化的智慧決策輔助系統
- ✓ 即時輔助決策，實現人機協作，AI 提供建議但決策仍由專家掌控
- ✓ 確保企業Know-how不流失，專家經驗可持續應用，提升企業競爭力



從邊緣到雲端，賦能製造業驅動 AI 計畫

HPE 與杰倫智能Profet AI為製造業提供AI驅動解決方案，協助企業解鎖數據的真正價值，並將其轉化為可行的商業洞察。

企業能有效地利用歷史數據訓練和驗證 AI 模型。訓練完成的模型隨後部署至基於 HPE ProLiant DL380 和 HPE GreenLake 的邊緣服務，實現低延遲的即時推論。生產線的工業電腦 (IPC) 持續收集影響模型的即時數據，並傳輸至邊緣服務進行分析，確保模型的持續優化和預測準確性。

此解決方案讓使用者能隨時隨地檢視虛擬量測結果，獲取即時的生產洞察，協助企業預測潛在問題、優化生產參數，快速決策，提升整體營運效率和競爭力，引領製造業邁向智慧化的新時代。

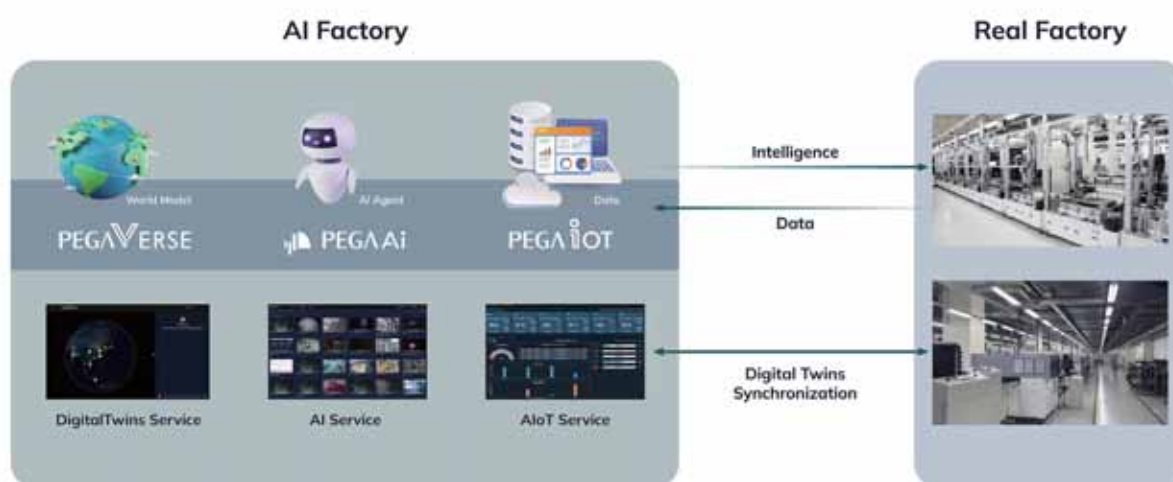


杰倫智能科技 (Profet AI) 專注於製造業的人工智慧軟體公司，於 2018 年創立，旨在為製造業提供定製化的 AI 解決方案。以台灣市場紮根發展，逐步擴展到全球範圍，與全球 300 家中大型製造業合作，產業涵蓋半導體、電子製造與化工領域，目前主要服務地區包括中國、東南亞、日本和泰國等地。

網站：<https://profetai.com>

和碩透過數位雙生與 AI 應用 加速智慧製造，提升產業競爭力

數位雙生 (Digital Twin) 加速產線規劃與維運，讓企業在虛擬環境中模擬和優化生產流程，實時監控生產狀況並快速調整策略以提高效率。同時，人工智慧 (AI) 能提升產品品質並降低資源需求，並且在日常工作提供效能。然而導入這些先進技術通常需要大量的技術與資源，使得企業在數位轉型過程中面臨有效整合這些技術以實現最佳效益的挑戰。



和碩PEGAAi智慧製造平台與PEGAVERSE人工智慧平台兩者整合將能提供智慧製造所需之完整解決方案。這兩個平台利用數位雙胞胎和人工智慧的優勢，幫助企業加速智慧製造快速落地。透過PEGAVERSE，企業除了可在虛擬環境環境進行低成本的優化外，還可透過PEGA IoT快速整合物聯網 (IoT)，實現即時監控和分析。這種整合不僅使企業能夠優化生產流程，提升運營效率，還提供了遠端協作的功能，讓團隊能夠隨時隨地進行協作。透過PEGAAi的無需編碼流程，企業可以輕鬆導入多種人工智慧應用，而不需要具備大量的 AI 知識。這使得各種部門都能夠利用 AI 技術來提升工作效率。再者，透過 LLM AI Agent 的整合，企業能夠快速優化流程，進一步提升整體效能。

透過和碩PEGAAi 平台與 PEGEVERSE平台，企業不僅能夠縮短導入先進技術時間，使得企業能夠在數位轉型的過程中，快速適應市場變化，提升競爭力。這些技術的應用不僅能夠推動企業的生產力提升，還能夠為企業創造新的商業模式和增長機會，最終實現智慧製造的願景。

PEGAVERSE平台打造虛擬工廠 有效提升生產效率

基於Nvidia Omniverse, PEGEVERSE平台能幫助企業打造虛擬工廠，企業可提前設計與優化生產流程，並驗證不同生產策略，減少實際生產中出現的錯誤和浪費，且無需額外的設備與時間成本。PEGAVERSE 還能提供企業內部 IoT 與流程資料整合並協助即時監控，快速反應，進一步提升生產效率。

透過虛擬工廠，企業可獲得以下五大效益：

一、加速產線開發：

PEGAVERSE 允許企業在虛擬環境中設計與優化產能，從而加速產線開發。透過整合自動化設備，企業可以在缺少實體設備之前進行設計與驗證，大幅提升設計效率，達到30%的提升。

二、快速適應需求變化：

PEGAVERSE 能夠幫助企業迅速建立產線，靈活應對需求的快速變化，並快速驗證改變的成果，確保生產流程的靈活性。

三、提升產能效率：

PEGAVERSE 有助於優化生產流程，提高產能效率。透過整合 AI Agent，除了提供即時通報外，還能提供診斷結果，降低無效生產的發生。

四、支援異地協同合作：

PEGAVERSE 透過即時 IoT 整合資訊，協助異地協同合作。即使工程師不在產線上，也能有效協助改善問題，提升團隊的協作效率。

五、累積數位資產：

PEGAVERSE 幫助企業累積數位資產，每次打造虛擬工廠都為未來的發展奠定堅實基礎，促進企業的長期成長。

透過 Nvidia Omniverse 仿真模擬服務平台和 PEGEVERSE 平台，企業能夠打造出高效、智慧的生產系統，實現加速產線開發、快速適應需求變化、提升產能效率、支援異地協同合作以及累積數位資產等多方面的效益。

PEGAAi 打造 AI 工廠 智慧應用無所不在

PEGAAi 提供了感知型 AI 視覺檢測與支援 LLM AI Agent 建立並結合作業流程，讓企業可依照本身需求彈性使用，針對不同大小企業的需求，可彈性佈署工作站、伺服器以至於雲端服務。

在視覺檢測 AI 導入功能，幫助企業輕鬆導入檢測相關應用並在流程上提升生產力和產品質量。核心功能如下：

一、Nocode / Lowcode 模型訓練：

使用者只需透過圖片上傳，即可進行模型標註與訓練。這一功能使得非技術人員也能輕鬆參與 AI 模型的開發，快速適應不同的站點需求。

二、IoT 資料串接：

PEGAAi 平台能夠有效串接 IoT 資料，實時追蹤模型效能，確保生產品質。透過數據的即時分析，企業能夠及時調整 AI 模型，並快速適應變化。

三、Edge 模型佈署與管理：

PEGAAi 平台支持 Edge 模型的佈署與管理，這不僅減少了 AI 應用的資源需求，還能確保產線順暢。

目前已在不同的產業驗證 AI 的導入是可快速成功的，顯示和碩 PEGAAi 平台在實際應用中的有效性。

PEGAAi 平台提供的 AI Agent (Agentic AI) 整合能力，能夠協助企業優化日常流程，藉此實現更高的工作效率。

一、知識管理：

PEGAAi 的 AI Agent 能夠有效管理和組織企業內部的知識資源，幫助員工快速獲取所需資訊，提升決策效率。

二、外部工具整合：

除了獲取即時資訊外，PEGAAi 還能串接與控制外部系統，讓企業能夠在一個平台上整合多種工具，簡化工作流程，提升整體效率。

三、開放式的介面：

PEGAAi 提供開放式的介面，快速協助企業整合 AI Agent 應用。這使得企業能夠根據自身需求，靈活調整和擴展 AI 功能，實現最佳的業務流程整合。

透過和碩 PEGAAi 平台的 AI Agent 整合能力，企業能夠在日常運營中實現更高的效率和靈活性，從而在競爭中保持優勢。

PEGATRON
和碩聯合科技

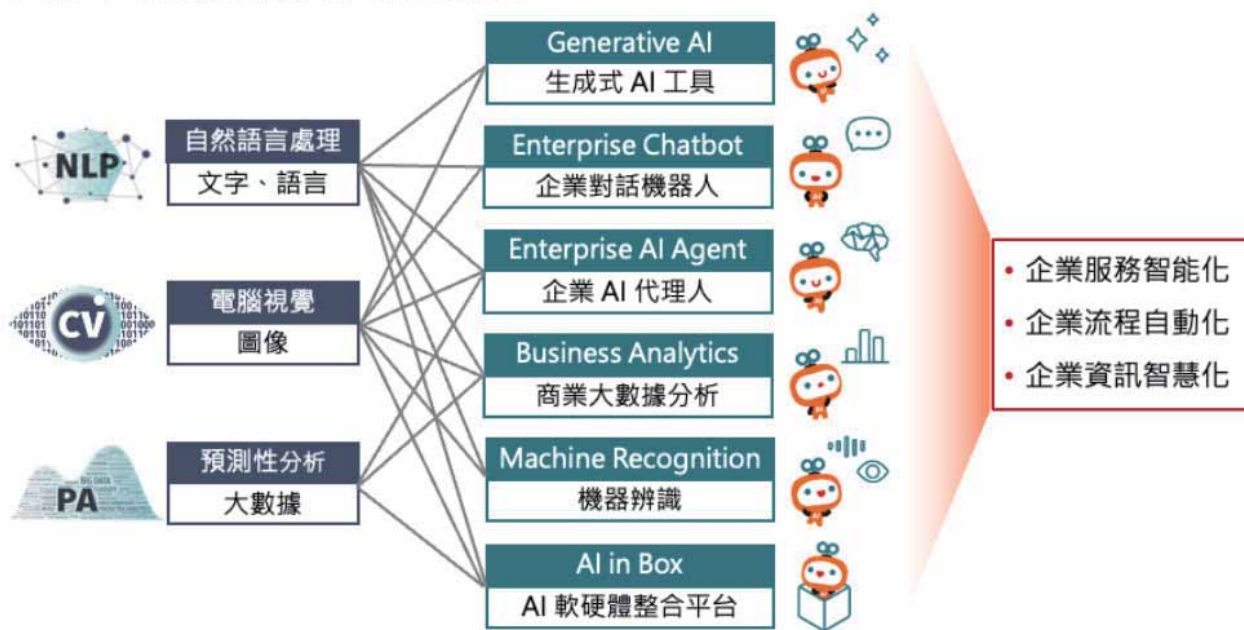
和碩聯合科技 AI 團隊 (PEGATRON AI) 致力於協助企業打造高效率的智慧製造方案，提供雲端與地端一致的佈署流程，快速落地 Digital twin、AI AOI、AI Agent 與資料平台的整合應用，降低企業生產成本，增進產線運作效率與提升製造自主維運能力科技 (Profet AI) 專注。

網站：<https://www.pegatroncorp.com>

八維智能為企業智慧化提供全方位的AI 解決方案

人工智慧歷經七十年的研究發展，近十餘年間取得巨大的突破，如今AI正廣泛地被應用於人類的生活與工作中。盤點近年來AI領域的關鍵時刻，當推2012年深度學習CNN模型成功地被應用於電腦視覺；2017年Google發佈了深度學習Transformer架構，成為所有大型語言模型的鼻祖；2022年ChatGPT的橫空出世，啟動了生成式AI的時代。在2025年CES消費電子展中，NVIDIA創辦人黃仁勳總結了當代AI的演進：從感知型AI，到生成式AI，到現今熱門的代理人AI，接著我們將邁向機器人AI時代（Perception AI → Generative AI → Agentic AI → Physical AI）。

八維智能技術及產品總覽



打造專屬、高效率AI方案，為企業及機構部門提升營運效率

八維智能於2017年創始於台北，幸運地搭上了AI突飛猛進的快速列車，公司的研發聚焦在AI應用：將最新最好的開源AI技術，應用於解決企業或機構的實際問題。數年下來，自主開發了一套可落地的、基於深度學習及生成式AI的開發者及使用者平台：八維智能AlinBox™，並藉此有效率地為企業及機構客戶打造其專屬的AI解決方案，目前應用領域已涵蓋了政府服務、金融保險、電商零售、醫療照護、交通運輸、科技、製造等。

任何企業級AI解決方案必須要能傳遞企業特定內容、處理企業工作流程，並在符合資訊安全的環境中運作。八維智能依據客戶的實際需要，協助客戶選擇並整合AlinBox平台上的各種AI功能模組，同時結合客戶數據、工作流程，進行有效率的客製化工程；進一步還可結合企業內部的硬體運算資源，為企業打造其專屬的AI應用私有雲。此外，考慮到企業AI應用需同時服務多位使用者，或處理大量數據，AlinBox的軟體架構已內建了處理流量、可靠性等問題的功能。

八維智能AInBox平台三組主要AI功能模組

一、生成式AI工具組 (Generative AI Tool Set) :

包括大型語言模型、檢索增強生成 (Retrieval-Augmented Generation, RAG)、知識搜尋 (基於BERT 語意模型)、會議機器人 (語音辨識及內容摘要)、文字內容生成、摘要、翻譯。其他選項包括圖像生成、文件辨識 (深度學習OCR)、語音辨識及生成。

二、企業對話機器人建構平台 (Enterprise AI Chatbot Builder) :

平台的核心是自然語意理解 (基於BERT 語意模型)、情境對話管理兩個引擎，前者用於瞭解使用者對話的語意，後者用於建構複雜的人機對話流程。平台的特色列舉如下：

- ◆ 可打造語意-RAG混合式對話機器人 (Semantic-RAG-Hybrid Chatbot)，使用BERT語意搜尋，以提高準確度、避免AI幻覺，並使用RAG檢索增強生成作輔助，以提升機器人應答的靈活度。
- ◆ 結合大型語言模型的高效率問答集生成。
- ◆ 簡易的拖拉式意圖編輯介面，可建構複雜的情境對話流程。
- ◆ 友善的管理者介面，適於非技術人員使用。
- ◆ 可結合虛擬客服與真人客服的人機協作模式。

三、企業AI代理人建構平台 (Enterprise AI Agent Builder) :

AI 代理人是個能夠與使用者及環境互動、並自動執行某特定任務的應用軟體。長久以來，流程自動化就是

開發AI代理人主要動機，而從2024年開始，人們對AI代理人有了嶄新的期待，這是由於現代的AI代理人得以借助大型語言模型的推理能力，使得一些原本難以實現的自動化任務，如今成為可能。例如AI代理人能產生銷售線索、提供客戶服務、優化供應鏈、執行市場研究、推算投資決定、開發軟體程式等。當AI代理人與硬體整合，它更可以應用於實體機器人、物聯網等領域。

然而，在企業應用場域中，單獨使用大型語言模型作代理人推理、決策，容易發生錯誤，風險很高。八維智能自主開發的AI代理人建構平台具有以下特色：

- ◆ 這是市場上唯一深度整合事件驅動式架構 (Event-Driven Architecture) 的AI代理人建構平台，具備了易於更新、擴充、維修等優點，並可解決流量、可靠性等問題。
- ◆ 可支援不同程度自主性的AI代理人，換句話說，依應用場域的需求，代理人決策機制可以是基於大型語言模型及基於規則的各種組合。
- ◆ 可支援任務流程中單一步驟、或整段任務流程的平行運算，藉此打通運算瓶頸，或提升整體運算效率。
- ◆ 無程式碼 (no-code) 的拖拉式任務流程編輯介面，可大幅提升開發效率。

總結，在AI科技的推動下，人類社會正面臨一個前所未見的巨變時代，AI轉型成為企業生存發展的必要條件。八維智能期許以自建的AI應用開發及使用平台，為各領域的企業及機構提供簡單易用、功能強大、全方位的AI解決方案。



八維智能 (D8AI) 專注於 AI 應用技術的開發與服務，技術能力涵蓋了自然語言處理、電腦視覺、預測性大數據分析、全端軟體開發及部署，及 AI 軟硬體整合。公司自主開發了一套可落地的、基於深度學習及生成式 AI 的開發者及使用者平台，並藉此有效率地為企業及機構客戶打造其專屬的 AI 解決方案，應用領域涵蓋了政府服務、金融保險、電商零售、醫療照護、交通運輸、科技、製造等。

網站：<https://www.d8ai.com>

打造新一代內容生成技術軟體基盤 資策會軟體院結合專業顧問服務 協助企業快速發展 AI 應用

人工智慧 (AI) 已成為全球科技創新的關鍵動力，驅動產業變革並影響社會發展。隨著各國政府與企業積極投入AI研發，AI技術正在快速演進，推動數位轉型並提升國際競爭力，台灣亦在此浪潮中加速佈局，致力於發展AI產業生態系統，以確保在全球市場中保持競爭優勢。

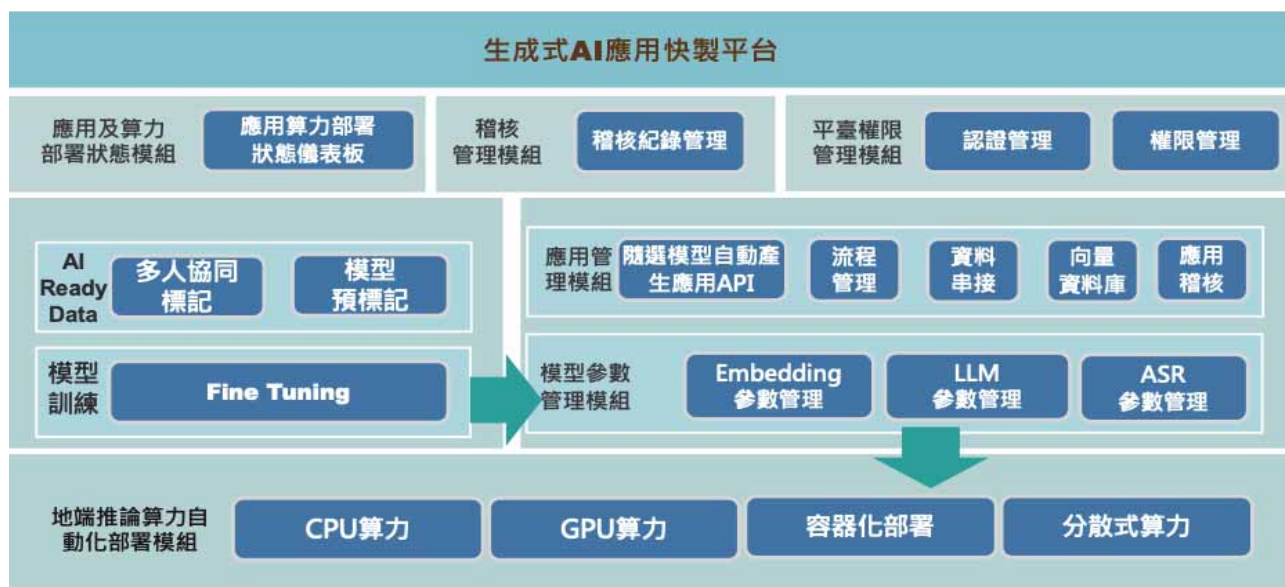
軟體技術研究院（簡稱軟體院）透過前瞻技術研究網絡，擘劃重大技術創新生態藍圖。應用先進軟體工程方法，推升AI軟體基盤與典範應用。並結合研究即服務 (RaaS)，鏈結產學研能量賦能第三方服務。我們的研究涵蓋生成式AI、智慧資料管理(交通、運動)、量子計算等領域，並透過與產學研合作，推動AI技術的普及與創新發展，確保企業者能夠高效地導入AI。



資策會軟體院以宏觀且前瞻之視野，獲取產業信任並為其順利轉型

我們所研發的大型語言模型微調工具，使用自有資料進行LLM微調，可用於佈建地端的企業內部LLM服務，提高內部知識管理/問答準確度。結合專業的效能調校，使訓練資源可有效減少 (2000小時降低至120小時)，節省企業成本。以此為基礎所發展的合規

偵測技術，應用在電商廣告、企業契約審閱，確保企業符合法規要求，降低法律風險。而與中研院研發的先進幻覺檢測技術，可確保生成式AI正確回答，適合高度正確性需求的企業、政府機構。



資策會軟體院提供生成式AI應用快製平台，全方位協助產業順利轉型

透過以上關鍵技術與服務，軟體院將持續推動 AI 技術的創新與應用，協助企業實現數位轉型與智能升級。我們期待與各界夥伴攜手合作，共同開創 AI 產業的新時代，讓人工智慧成為企業成長的最佳推動力！

資策會軟體院為企業轉型提供專業顧問及服務

一、企業專屬AI模型微調與推論：提供軟體基盤，含文字(llama、Taide)、聲音(Whisper)、圖片(Yolo)訓練工具，透過實際應用案例進行實作教學，適合AI初入門的企業。

二、生成式AI顧問服務：提供企業AI部署策略，協助從技術選擇到落地應用。針對工具的程式細節進行說明、教學，目標在使得企業技術人員可學得相關知識，自行發展企業內部應用，適合具備研發人員、期望自行建置AI研發能量的企業。

三、訓練成果驗證(IV&V)：協助企業採購/提供生成式AI訓練服務時，可藉由RAGAS、幻覺檢測等技術，評估模型訓練成效。

資策會軟體院不僅是AI技術研究與開發的機構，更是企業數位轉型的重要夥伴。軟體院致力於提供全方位的AI解決方案，確保企業能夠有效運用人工智慧來提升競爭力。透過強化AI訓練與推論技術、數據治理、合規檢測及AI風險管理，我們能夠幫助企業克服技術門檻，實現AI的高效部署與應用。

MLSteam從內部協助電信客戶進行硬體的叢集管理、監控與管控，並順利整合其現有的帳號及帳務系統。平台提供完整的開發環境，進一步提升雲端服務對終端使用者的價值，讓使用者能夠在登入後立即進行模型開發。相較於BMaaS、AlaaS不僅能帶來更高的營運效益，還能有效提升使用者的黏著度。



財團法人資訊工業策進會軟體技術研究院（簡稱資策會軟體院），以前瞻技術與宏觀生態布局，主要任務包含前瞻技術研究，協助擘劃重大技術創生之生態藍圖；透過先進的工程方法來推升 AI 軟體基盤與典範應用；提供研究即服務 (RaaS)，結合產學研究能量賦予第三方研發服務。

網站：<https://www.iii.org.tw>

擁抱 AI 決策外的 專業支援

AI之勢，銳不可擋，轉型將成為企業前進下個世代的通行證。

HPE提供之AI as a Service全方位服務，從伺服器運算、資料整合、AI模型訓練，到AI應用落地服務應有盡有，HPE能與專業團隊提供一條龍服務，或依據企業需求彈性調整，專業賦能，順利協助台灣產業完成AI轉型。

想了解更多HPE AI專案的服務或相關資訊，歡迎參考以下連結：

◆ NVIDIA AI Computing by HPE：

<https://www.hpe.com/tw/en/solutions/artificial-intelligence/nvidia-collaboration.html>

◆ HPE Private Cloud AI方案：

<https://www.hpe.com/tw/en/private-cloud-ai.html>

◆ HPE AI方案：

<https://www.hpe.com/us/en/solutions/ai-artificial-intelligence.html>

◆ HPE運算解決方案：

<https://www.hpe.com/us/en/hpe-proliant-servers.html>

◆ HPE超級電腦方案：

<https://www.hpe.com/tw/en/compute/hpc/supercomputing/cray-exascale-supercomputer.html>

◆ HPE GreenLake服務方案：

<https://www.hpe.com/us/en/greenlake.html>



- 台北市大安區復興南路一段303號5樓
- 台北市松山區復興北路167號3樓
- 新竹縣竹北市自強南路8號11樓之1
- 台中市西屯區台灣大道三段660號11樓-B3
- 高雄市苓雅區四維三路6號18樓A2

www.dawningtech.com.tw

☎ (02) 8978-5386

☎ (02) 8978-5386

☎ (03) 531-3300

☎ (04) 2327-0848

☎ (07) 586-6528

☎ (02) 2700-2358

☎ (02) 2700-2358

☎ (03) 534-0055

☎ (04) 2327-0847

☎ (07) 586-6529

行銷服務專線: 0800-072-636



