

GenAI風潮下，台灣AI硬體供應鏈商機解析

魏傳虔

產業顧問兼組長

產業情報研究所

財團法人資訊工業策進會

2024.08.14

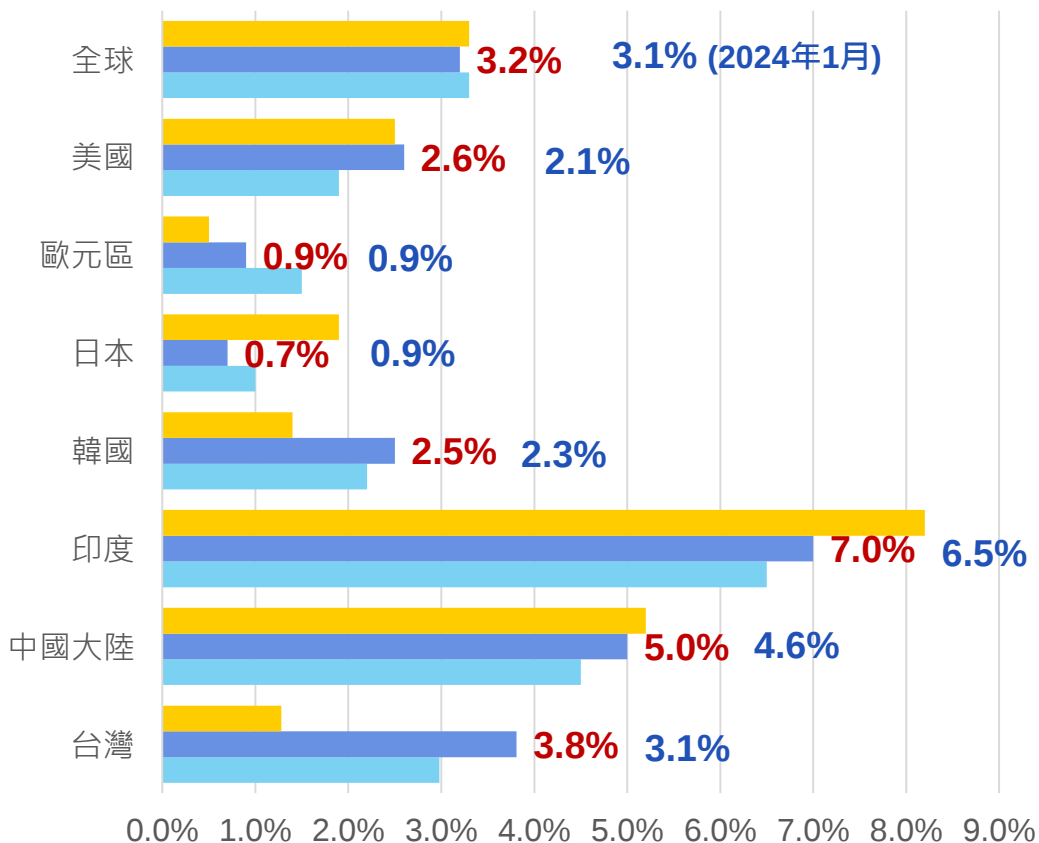
簡報大綱

- 全球AI PC市場發展
- 全球AI伺服器市場發展
- 台灣產業供應鏈發展機會
- 結論



全球景氣緩慢復甦，惟不確定風險仍在增加

2023~2025年全球經濟成長率



■ 2023年GDP年增率 ■ 2024年預估GDP年增率 ■ 2025年預估GDP年增率

註一：2023年資料為IMF最新修正後的GDP年增率，2024年及2025年估值均為同一時間（2024年7月）所公布的預測值

註二：我國GDP數據係採用中華經濟研究院2024年7月公布之資料

資料來源：IMF、中華經濟研究院，MIC整理，2024年7月

2022~2024年主要國家消費者物價指數

地區	2022.6	2023.6	2024.6
台灣	3.6%	1.8%	2.4%
美國	9.1%	3.0%	3.0%
歐元區	8.6%	5.5%	2.5%
中國大陸	2.5%	0.0%	0.2%
日本	2.4%	3.3%	2.8%

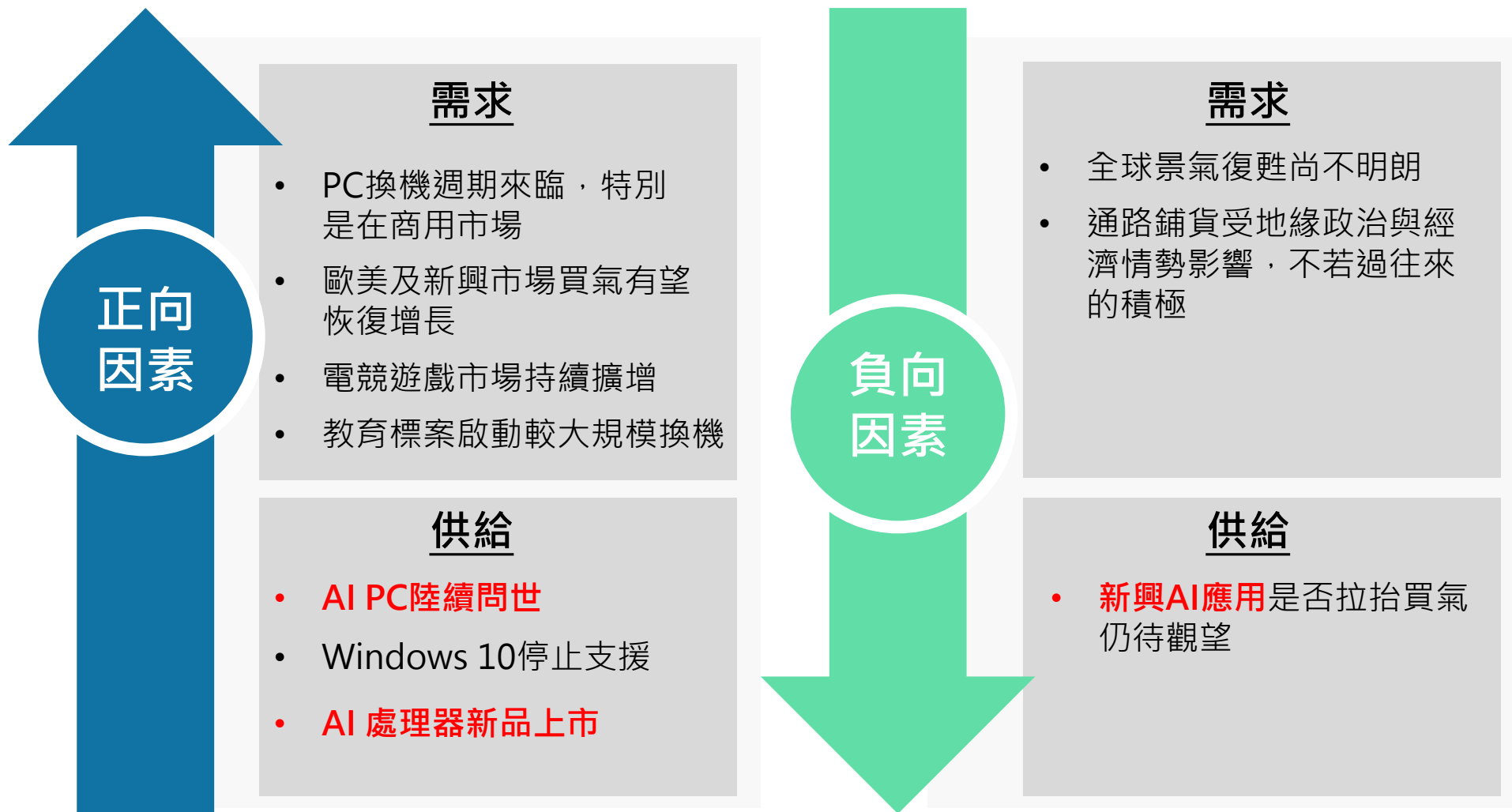
資料來源：行政院主計總處、United States Department of Labor、Eurostat、中國大陸國家統計局、Japan Statistics Bureau，MIC整理，2024年7月

- 國際貨幣基金組織（IMF）預測2024年全球經濟成長率為3.1%，比2023年10月的預測上升了0.2個百分點，判斷通膨趨於緩和，「軟著陸」（soft-landing）即將到來
- 景氣復甦仍略顯蹣跚，但基於AI等新興科技應用需求帶動下，電子相關產品需求回溫，我國訂單需求與投資可望恢復成長動能，並使民間消費因此受惠

全球AI PC市場發展



全球PC市場需求影響因素



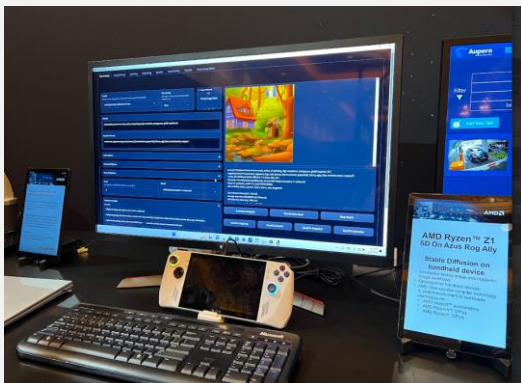
資料來源：MIC，2024年7月



AI PC為具備NPU於本地端執行生成式AI

廣義

在不連接網路情況下，在PC本地端執行生成式AI應用程式



資料來源：AMD，2024年1月

- 運算能力足夠執行本地端生成式AI，即可稱為AI PC
- 目前PC搭載中高階GPU之電競PC、創作者PC等約略能夠執行本地端生成式AI
- 上圖為華碩搭載AMD APU電競掌機，連接螢幕及鍵盤化身為mini pc，亦能執行Stable Diffusion

CPU

具備整合神經網絡單元(NPU)處理器，加速AI運算任務

晶片廠	內建NPU處理器	NPU算力
Intel	Meteor Lake	無公告
	Lunar Lake	40-50 TOPS
AMD	Ryzen 7040	10 TOPS
	Ryzen 8040	16 TOPS
	Ryzen 7 8700G	16 TOPS
	Ryzen AI 300	50 TOPS
Qualcomm	Snapdragon X Elite/Plus	45 TOPS
Apple	M1	11 TOPS
	M2	15.8 TOPS
	M3	18 TOPS
	M4	38 TOPS

GPU

搭載專用AI加速硬體的PC



資料來源：各公司，MIC整理，2024年7月

備註：TOPS為Tera Operations Per Second
資料來源：MIC，2024年7月



微軟「Copilot+PC」訂立硬體規範新標準

硬體規範

- **處理器**：需合併NPU，且運算能力達40 TOPS以上
- **記憶體**：須為16GB DDR5/LPDDR5以上
- **儲存容量**：最少256 GB SSD/UFS
- **其他**：全日電池續航、鍵盤設計上則須配有Copilot按鍵、在軟體上要能運行Copilot相關AI模型

軟體應用

可運行Windows 11中超過40個AI模型，並支援微軟Copilot，其中有幾大特色功能：

- **Recall**：創建可搜索的攝影記憶，紀錄電腦使用
- **Live Caption**：將中文在內 40 多種語言即時翻譯成英文字幕
- **Cocreator(加入到小畫家)**：簡單繪畫+文字輔助=精緻圖片

首發產品

以搭載Qualcomm Snapdragon X系列平台為首發，7家品牌廠做為夥伴先行推出Copilot+PC產品



Lenovo

acer

ASUS

SAMSUNG

Microsoft



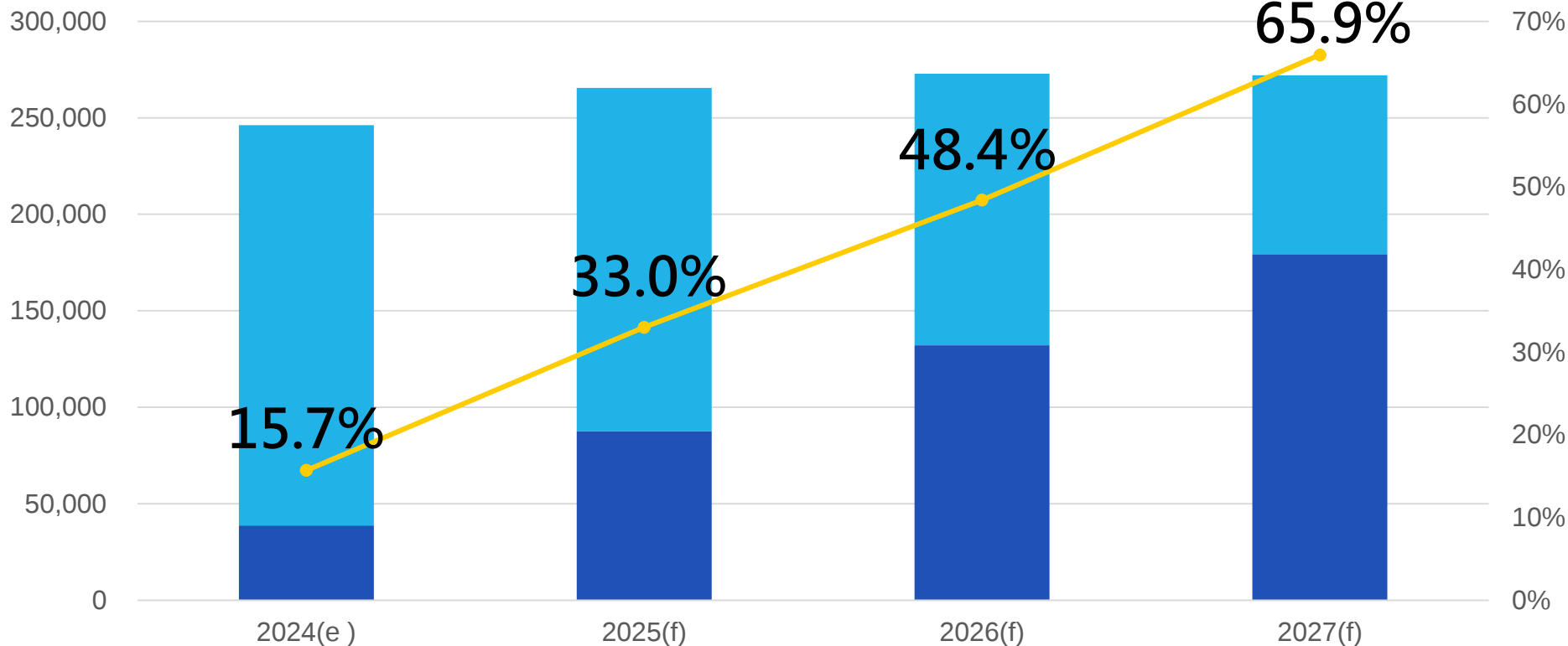


2027年AI PC市場滲透率近七成

單位：千台

全球AI PC市場滲透率

單位：%



資料來源：MIC，2024年7月

註：此處AI PC泛指搭載NPU的PC產品

■ AI PC ■ 非AI PC —●— AI PC 滲透率

- 2024年為AI PC元年，在軟體應用尚未到位情況下，AI PC滲透率仍低
- 隨著晶片商推出面向高中低階的處理器，以及AI軟體應用多元化，2027年AI PC滲透率將達近七成



AI PC帶來更加混合、協作的時代

發展趨勢



算力層

混合異構的處理器架構

CPU



GPU



NPU



平台層

平台支援強化硬體與軟體的實踐



模型層

大模型

雲端、閉源、訓練資源大

小模型

地端、開源、客製化



應用層

雲端+地端應用開發，豐富使用體驗



終端層

硬體+軟體加成，差異化AI PC產品

關鍵議題

異質運算為Arm架構帶來利基
處理器廠商競爭將更為激烈

平台對軟硬體的相容性重要度提升
作業系統商各擁陣營打造AI PC生態系

對應不同情境的AI運算
雲、地共存之AI模型因應而生

廣泛且無窮可能的AI應用
讓產業供應鏈需強化開源工具與資源的釋出

硬體規格升級與軟體應用多元性
使品牌廠的產品價值性增長

資料來源：MIC，2024年7月



處理器的異構運算為AI PC分配執行算力

SoC	intel	AMD	Qualcomm	Apple	HUAWEI
	Lunar Lake	Ryzen AI 300系列 針對AI PC產品系列	Snapdragon X Elite/ X Plus	M4	麒麟 9000C
	CPU 執行低延遲的AI應用，需要處理輕量且單一模型的推論	Lion Cove+ Skymont 效能與能耗平衡	Zen 5 CPU	Oryon CPU	P-core+E-core
	GPU 需要大量平行運算和高吞吐量處理的AI任務	Xe2 GPU	RDNA 3.5	Adreno GPU	GPU
	NPU 具低功耗、省電優勢，負責需要持續處理的AI任務	Intel AI Boost NPU	XDNA 2 NPU算力當前最強	Hexagon NPU 強調每瓦效能表現	Neural Engine
應用產品	NB	NB	NB	NB/DT/AIO PC	NB/DT PC
上市時間	2024年第三季	2024年7月	2024年6月18日 最早面向市場	預計2024年第四季	預計2024年9月

備註：各公司，MIC整理，2024年7月

- CPU+GPU+NPU將分別處理不同的工作，且對於需執行大規模資料運算與推導的AI應用而言，較低功耗的NPU成為重要運算架構
- 筆電產品因需考慮攜便性及續航力，因此行動處理器的混合異構將先行導入；未來因應ESG需求，桌機的低功耗執行仍為必須



PC作業系統商陸續更新作業系統及其AI支援



Windows 11 24H2



推出時程

- 兩階段，率先更新至Arm架構Windows裝置
- 1) 2024年6月：預計先進行基礎更新，並搭載驍龍X Elite的新PC一起推出
 - 2) 2024年9/10月：向現有Win11 PC用戶推出



AI應用

- 提升搜尋效能、快速啟動各項計畫或工作流程、精準把握上下文
- 更智慧的歷史記錄功能
- Windows搜尋功能支援自然語言
- 超級解析度 (Super Resolution)
- 增強版的即時翻譯字幕功能



硬體規範

- 配備NPU達到**40 TOPS**，以執行AI功能
- CPU必須支援POPCNT指令，並沿用舊版Windows 11要求：TPM 2.0、安全啟動等
- 推薦記憶體達16GB



HarmonyOS NEXT

- 2024年6月推出開發者Beta版
2024年8月推出消費者Beta版
2024年第四季推出正式版

內建盤古大模型，具備AI能力

- 手繪線稿、填色、塗鴉、照片擴圖等生成式AI創作
- AI聲音修復(年底上線)
- 「小藝看世界」協助讀圖功能

- 首款原生裝置為新一代旗艦手機 Mate 70 系列
- Pura 70、Mate 60 以及 Mate X5、MatePad Pro 11/13 平板等產品將首波支援更新



MacOS 15(Sequoia)

2024年6月



自動摘要網頁文字重點與資訊、提供內容總結目錄等



整合ChaGPT 4O



使用 Siri、打字或語音，生成使用者想要的表情符號



透過自然語言來搜尋特定的照片，並提供圈選清除功能編輯照片



協助修改措辭將文章改寫、自動檢查文本語法、用詞並提供建議

macOS 15支援硬體

Mac：iMac Pro 延伸到 2017 年的機種，其餘Mac僅 2019 年之機型才能升級

- 作業系統商預計於2024下半年推出新一代系統平台，以因應AI PC在軟硬體間的搭配與支援

資料來源：MIC，2024年7月



AI PC提供本地端運算豐富體驗、雲端為輔

影像生成

圖像生成

文字創作

程式碼生成

OpenAI
SORA

Gemma

Stable Diffusion
LLAMA 2

OpenAI
GPT - 4
MISTRAL
AI_

通义千问
智谱·AI



雲端運算



地端運算

具備強大可擴展之算力

資訊具即時性

優勢

免費

個人資料隱私性高

傳輸延遲低(距離短)

持續付費訂閱

個人資料隱私性低

傳輸延遲高(距離長)

劣勢

算力受限於PC裝置

受限模型更新時間，資訊有時間性

資料來源：MIC · 2024年7月

- 雲端與地端的模型訓練及推導有其優劣，可根據於不同情境的使用需求進行選擇



產業生態系以開源模式，共創多元AI應用

開源
模式



軟體
模型

- 和Meta合作，在OPENVINO上放LLAMA2
- Phi-2小語言模型

- Ferret、MGIE等AI開源模型

- 盤古大語言模型

- Gemma

- AI Hub模型庫

- Ryzen AI軟體
- LM Studio

- NVIDIA TensorRT軟體
- NVIDIA RTX技術(Project G-Assist、NVIDIA ACE等)



開發
框架與
資料庫

- Azure
- Windows AI Studio工具
- DirectML
- ONNX Runtime

- MLX(ML Explore)
- Core ML
- Create ML
- SiriKit工具

- HarmonyOS SDK
- 倉頡程式語言

- localllm

- Open VINO
- ONNX Runtime
- DirectML
- WEBNN

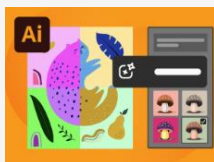
- Qualcomm AI Stack

- AMD software

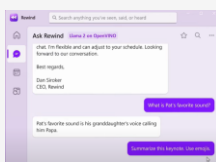
- CUDA
- NVIDIA AI Workbench
- TensorRT-LLM for Windows開源函式庫



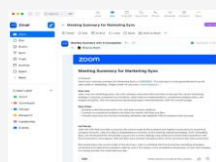
AI
應用



生成式重新上色



個人化AI助理



會議摘要



動畫生成



文生圖

...Others

資料來源：MIC，2024年7月

● 供應鏈大廠擴大釋出開源架構及工具，以協助更多軟硬體廠商加入生態系，共同開發創新AI服務

因應AI PC使用型態的改變，硬體規格須有所升級

 算力運算	處理器	1. CPU 2. CPU+獨顯	1. CPU+GPU+NPU(SoC) 2. CPU+獨顯	基礎AI PC仰賴NPU；進階AI PC則需加上GPU協助
	記憶體	容量增加		為記憶使用者習慣、同時因應AI模型的複雜化需求，PC記憶體容量及頻寬的重要性將最先被凸顯
	SSD	容量增加		
	散熱模組	風扇葉片增量、均溫板與液態金屬導入、導熱零組件排列組合升級		隨著資料量的增長及AI運算時間的提升，散熱設計的需求將隨之提高
	傳輸裝置	效率提升		AI PC仰賴雲與端協作，雙邊的傳輸及連線穩定性、速率都將影響體驗
 自然語言交互	麥克風	數量增加		人機交互、虛實邊界的模糊，使人與PC間的交互關係更加多元，扮演「感知」角色的音/視訊設備將迎來升級/用量提升等契機
	視訊鏡頭	畫素提升		
 體驗	面板	選擇低功耗面板、提升畫面更新率		在搭配高算力晶片下，耗能將成為負擔，面板廠配合推出高省電LTPS面板
	電池	容量提升		電池容量增長有助提升續航力表現

資料來源：MIC，2024年7月

- 因應AI PC運算、感知及體驗的改變，硬體規格將有所升級，帶動供應鏈ASP增長

全球AI伺服器市場發展



2024年全球伺服器市場藉由AI伺服器重新回溫

影響全球伺服器市場因素

雲端服務商持續擴增東南亞、南亞的資料中心

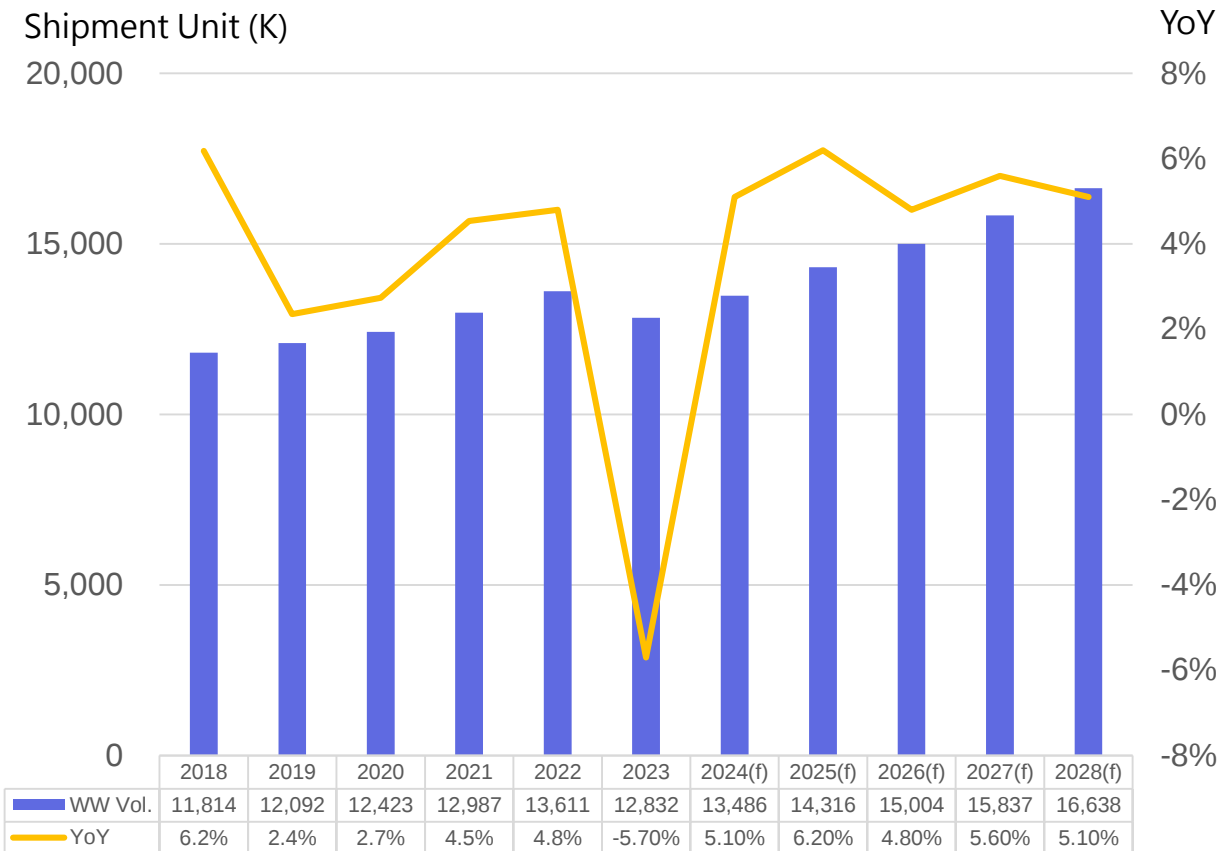
雲端服務商AI伺服器需求仍在延續

伺服器品牌商新AI伺服器產品量產

AI推論伺服器因推論與中小模型微調需求，出貨將逐年增加

同樣資本支出下，一般伺服器採購率相對下滑

2018~2028年全球伺服器市場預測



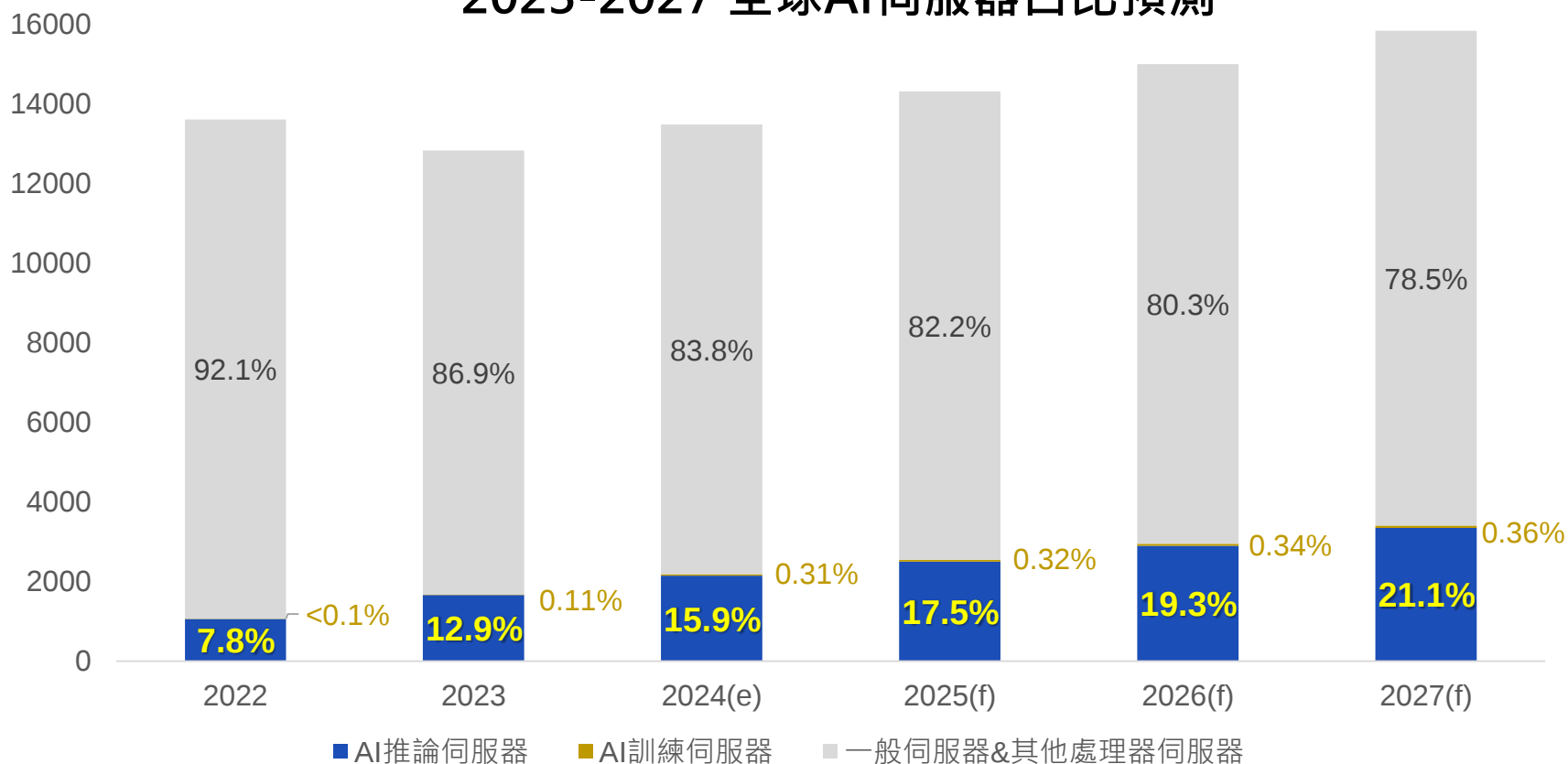
資料來源：MIC，2024年4月

- 全球伺服器市場2024年的主要驅動力為AI伺服器的需求，不論雲端服務商、伺服器品牌商的AI伺服器訂單均仍在上升，另外雲端服務商加速布局亞洲資料中心，以東南亞、印度為主要投資地點
- 另外AI推論伺服器將因為企業端推論、中小模型微調 (Fine-tune) 需求而出貨量提升，然於此同時在同樣的資本支出下，AI伺服器高昂的價格將會壓縮一般通用伺服器的出貨



2024年全球AI伺服器出貨量與出貨佔比遽增

2023-2027 全球AI伺服器占比預測



資料來源：MIC，2024年4月

- NVIDIA H100伺服器訂單，H200、GH200預計在2024Q2量產，將使AI訓練伺服器出貨進一步提升
- 搭載 AMD MI300x、MI300A、Intel Habana Gaudi 3的伺服器同樣將在2024年放量
- AI推論及中小模型訓練，NVIDIA中低階GPU伺服器、Xilinx FPGA同樣將帶動AI伺服器出貨成長



AI推論、AI訓練伺服器的作用與差異

一般、AI推論、AI訓練伺服器需求情境

特徵 / 差異	一般伺服器	AI推論伺服器	AI訓練伺服器
主要運算單元	CPU	低階GPU、FPGA、ASIC	高階GPU
運算設備體積	高度以1U、2U為主	高度以1U、2U為主	根據應用，高度可達4U、5U、7U
出貨數量	最多，一般企業均會進行採用	次多，在未來小型AI模型蓬勃發展，有望進一步提升	最少，價格高昂，有大型語言訓練、AI訓練需求的企業才會採購
出貨方式	單台、整櫃出貨	單台、整櫃出貨	除單台、整櫃出貨外，透過超級電腦方式提供(DGX GH200)
伺服器的價格	1,500-3,000 (美元)	根據加速卡，價格範圍較大 3,000-20,000 (美元)	15萬以上 (美元) DGX 8x H100價格高達 20萬美元
總耗電量/熱能	最低，平均功耗200w-1,000w不等	平均功耗2,000-4,000w	最高，DGX A100 6.5KW DGX H100 10.2KW
代表性產品	一般市售伺服器	可搭載NVIDIA L4、A2等GPU AMD Xilinx Alveo 加速卡、 Intel Gaudi2和Greco的伺服器	可搭載NVIDIA H100 & A100、 AMD MI200&MI300、Intel Ponte Vecchio GPU的伺服器
作用/應用場景	數據量小，但應用多元且複雜的場景	人臉辨識 & 車牌辨識等，需要迅速進行大量推論的場景	AI大型語言模型訓練，數量龐大且須長時間訓練的數據

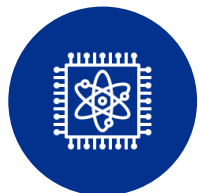
資料來源：MIC，2024年

- 一般、AI推論及AI訓練伺服器，各自符合其應用場景。在ChatGPT帶動的AIGC浪潮下，AI訓練伺服器的需求增加，未來可望帶動AI推論伺服器需求



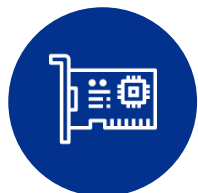
AI全面影響資料中心不同構面的建置

AI對資料中心IT設備的影響



運算

資料中心集成大量GPU、ASIC、FPGA等AI晶片。為符合AI運算的需求，AI伺服器需透過更緊密的方式進行叢集



網通

為符合AI訓練數據快速傳輸需求使用智慧網卡（NIC）、DPU來協助網路傳輸，並透過交換器、複雜的光纖結構來使用乙太網路（Ethernet）或Infiniband



儲存

AI訓練的過程中，除了AI伺服器中記憶體의暫存儲存熱數據外。大量的冷數據需要透過儲存伺服器來存放，增加HDD、SSD的需求

AI對資料中心基礎設施的影響



基礎設施

AI運算需耗用更多電力，因此UPS、配電裝置（PDU）、備援電池（BBU）均需要進行調整

冷卻方面包含氣冷風扇牆、冰水主機，通風管道，以及直接、沉浸式液體冷卻的導入



營運

AI對資料中心營運的影響

使用AI來營運資料中心，可以通過動態設置和自適應功能實現更好的工作負載。並透過機器學習（ML）來最佳化資料中心監測模型，進行預測性分析



雲端服務商積極於資料中心堆疊AI算力

自研晶片

超級電腦

GPU

Meta

MTIA (元訓練與推理加速器)：針對推理工作負載內部客製化加速器晶片

兩個各採用**24,576個Nvidia H100 GPU**的大型資料中心叢集

預計採購35萬顆H100 GPU，於2024年底打造60萬顆GPU規模的AI算力

Microsoft

Azure Cobalt：Arm架構的CPU，強化雲端運算

Maia 100：為AI負載設計的AI晶片，可執行大型AI模型

Azure Eagle：包含14,400個H100 GPU，超級電腦TOP 500排名第三

2024預計採購8萬台H100 AI伺服器（64萬顆），並將採購下一代B100 GPU

Google

TPUv5e：每個Pod擁有256個晶片，AI推論較高性價比

TPUv5p：每個Pod有8960個晶片，處理大型AI訓練

A3超級電腦：最多可堆疊26,000個H100 GPU

2023約採購5萬顆H100 GPU，預期2024將同步擴大GPU與TPU的堆疊

Amazon

AWS Inferentia：專注在深度學習與AI推論

AWS Trainium：將重點放在AI訓練與模型建構最佳化

P5 Ultra Scale GPU叢集：可以堆疊20,000個H100 GPU
Project Ceiba：包含16,384顆GH200超級晶片

2024年將採購大量GH200與H200 GPU使用於自身應用實例

Oracle

Oracle因自研晶片需要數年時間才能完成，決定不投入自研晶片，使用NVIDIA、AMD GPU

OCI Super Cluster：可以擴展數萬個NVIDIA H100 GPU

2024年除NVIDIA GPU外，AMD MI300X成重要採購目標

資料來源：各公司，MIC整理，2024年4月

- 雲端服務商積極部署自身的AI算力，透過自研AI加速晶片來符合自身雲端服務的客製化需求
- 此外，和伺服器品牌商、處理器廠商合作打造AI超級電腦，協助自身與客戶進行AI模型訓練



處理器廠商AI加速晶片進入全新的競爭態勢

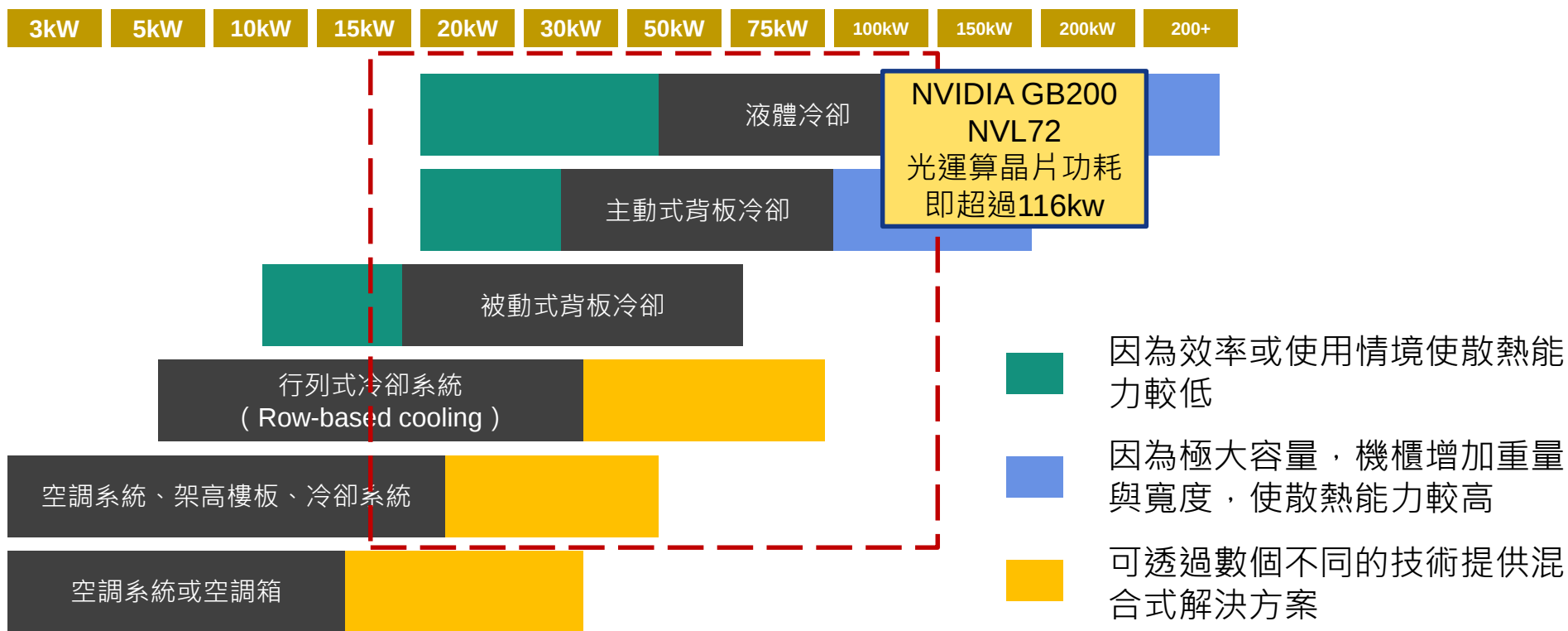
Vendor	GPU			超級晶片、APU	FPGA、ASIC
	AI訓練	AI推論	特殊規格		
	- H100 - H200 - B200	- L40s - B40	- H800 - H20 - L20 - L2	- GH200：Arm架構推論 - GH200 NVLink：Arm架構訓練與推論 - GB200 - GB200 NVLink 36/72	開發客製化晶片服務
	MI300 X		可能推出MI250降規版	MI300A APU	XILINX FPGA
	- GPU Max Series (Ponte Vecchio) - Habana Gaudi 2、3		Gaudi 2 降規版		- Intel Agilex FPGA - Intel 客製化 ASIC

資料來源：各公司，MIC整理，2024年4月

- 各處理器廠商為符合客戶各種應用場域的AI運算需求，開始擴增資料中心AI加速晶片的產品規格
- GPU包含AI訓練、AI推論專用及面對中國大陸市場限制的特殊規格
- 將CPU與GPU封裝在一起的超級晶片、APU亦成為新的模式，強調使CPU和GPU之間的互聯更加緊密



AI資料中心耗能增加使液體冷卻成為必備方案



資料來源：Vertiv，MIC整理，2024年4月

*紅色框為液體冷卻可行選項

- 當機架密度超過20kW 時，基於空氣的冷卻系統就會失去效力，此時液體冷卻就成為可行的方法
- 節能背板冷卻 (RDHx) 是一項成熟的技術，可為管理20 kW 以上的密度提供可行的解決方案
- 此技術不會將液體直接輸送到伺服器內部，但用液體高熱傳導性，採用相似於直接液體冷卻的做法



AI算力需求大幅提升改變伺服器託管型態

上游

處理器



- CPU : Grace、Grace Hopper超級晶片
- GPU : A100、H100

下游

我國伺服器代工廠



國際伺服器代工廠



伺服器ODM Direct / 白牌伺服器/自有品牌



客戶

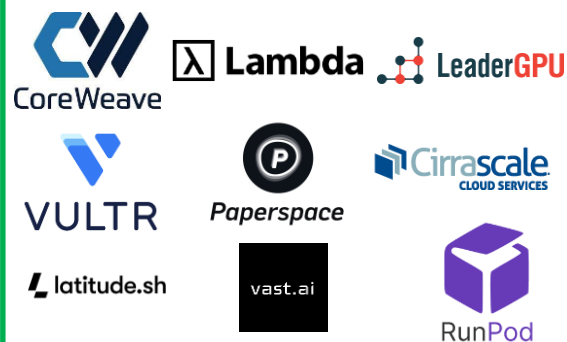
資料中心託管商



雲端資料中心



GPU運算/雲提供商



提供服務

伺服器託管

VM/Instance
+
AI應用

GPU 算力
GPU
VM/Instance
租用

台灣產業供應鏈發展機會

-AI PC產業-



中高階產品定位優先，AI NB售價較一般性NB高

搭載NPU之主流AI NB規格



處理器

搭配Intel Core™ Ultra系列的產品
(部分為搭載AMD Ryzen 7040/8040系列)

面板

OLED面板、120HZ更新率
(部分搭載Mini LED、更新率最高達165Hz/240Hz)

記憶體

32GB DDR5+1TB SSD
(最高可達96GB DDR5；10TB SSD)

電池

多數在70Whr-80Whr區間
(最高搭載容量為99Whr)

價格區間 主流規格落在4萬-6萬元新台幣 (最低約2萬5千元；最高則約9.3萬元)

資料來源：各公司，MIC整理，2024年5月

- 因應AI NB體驗提升，在面板搭載換上顯色度較佳的OLED，以及較順暢畫面體驗的高更新率螢幕
- 因應AI運算的資料大小及運行時間，記憶體與電池等零組件規格提升
- AI NB優先搶攻中高階產品定位市場，產品售價較一般性NB高出2-3成，然在AI應用尚未完善時，考驗消費者的接受程度及市場滲透率進展



因應算力需求提升，處理器與記憶體為首要關鍵



處理器

1. CPU
2. CPU+獨顯

1. CPU+GPU+NPU(SoC)
2. CPU+獨顯

基礎AI PC仰賴NPU；進階AI PC則需加上GPU協助

記憶體

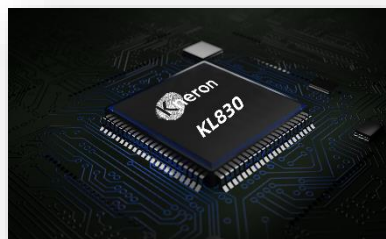
容量增加

為記憶使用者習慣、同時因應AI模型的複雜化需求，PC記憶體容量及頻寬的重要性將最先被凸顯

SSD

容量增加

處理器搭載方式多元



耐能第三代 NPU 晶片 KL830，主打應用於AI PC，並已獲四家品牌廠導入



電競與創作者筆電強調CPU加上GPU，可讓AI算力明顯提升

DRAM升級、容量提升

系列	Intel			AMD		
	Alder Lake	Raptor Lake	Meteor Lake	Ryzen 6000	Ryzen 7000	Ryzen 8040
推出時間	1H22	1H23	4Q23	1Q22	1Q23	4Q23
DRAM 支援	DDR4/DDR5	DDR4/DDR5	DDR5	DDR5	DDR5	DDR5

- Intel與AMD新處理器平台全面支援容量更高、速度更快的DDR5，有助**提升DDR5**滲透率
- 微軟、Google均設定記憶體容量的最低規格限制，迫使AI PC設立記憶體**最低容量標準**

SSD容量增加至512GB以上

用以儲存大量AI學習資料的SSD在AI NB的搭載上多已搭載1TB/2TB容量產品



AI運算需求增長，散熱與傳輸方案推陳出新



算力運算

散熱模組

風扇葉片增量、均溫板與液態金屬導入、
導熱零組件排列組合升級

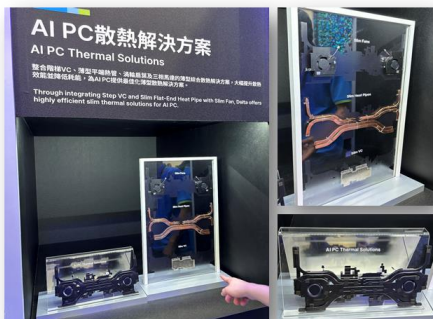
隨著資料量的增長及AI運算時間的提升，散熱設計的需求將隨之提高

傳輸裝置

效率提升

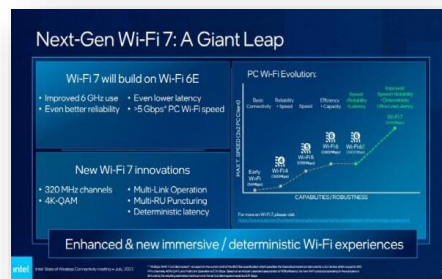
AI PC仰賴雲與端協作，雙邊的傳輸及連線穩定性、速率都將影響體驗

散熱設計組合多



台達電從散熱管、VC板及風扇設計著手，打造**薄型散熱**解決方案

WiFi 7推進速度快



Wifi 7支援頻寬為WiFi 6/6E的一倍，可達更低延遲、更快速度、更穩定連接及降低干擾，

主動式散熱晶片新技術



FrORE Systems 推出新一代固態**主動式散熱晶片**，聯手Nvidia展示搭載的AI電腦Jetson Orin系列

PCIe Gen5滲透率提升

	Intel		AMD	
系列	Raptor Lake	Meteor Lake	Ryzen 7000	Ryzen 8040
PCIe 支援	Gen 5	Gen 5	Gen 5	Gen 5

因應運算及資料傳輸規模的龐大，AI PC將可能導入PCIe Gen4甚至Gen5的傳輸裝置

資料來源：各公司，MIC整理，2024年7月



人機互動模式重視感知需求， 音訊與視訊設備升級/用量增加

自然語言交互

麥克風

數量增加

視訊鏡頭

畫素提升、偵測功能準確化

人機交互、虛實邊界的模糊，使人與PC間的交互關係更加多元，扮演「感知」角色的音/視訊設備將迎來升級/用量提升等契機。

NB鏡頭升級



HP Spectre x360系列
具備**9MP攝影鏡頭**，並
具有硬體支援的低光調
節功能

NB麥克風數量增加

	Prestige 16 AI Studio 8115G/2011W	Prestige 16 AI Evo 8115G/2011W	Prestige 16 AI Evo AI MG/2011W
處理器	Intel® Core™ Ultra 7 155H	Intel® Core™ Ultra 7 155H	Intel® Core™ Ultra 7 155H
作業系統	Windows 11 Home	Windows 11 Home	Windows 11 Home
晶片組	Integrated SoC	Integrated SoC	Integrated SoC
記憶體	LPDDR5 onboard 32GB	LPDDR5 onboard 32GB	LPDDR5 onboard 32GB
螢幕	16.1" QHD+ (2560 x 1600), 16:10, 100% DCI-P3 (typical), 400nits (typical), IPS 雙面螢幕	16.1" QHD+ (2560 x 1600), 16:10, 100% DCI-P3 (typical), 400nits (typical), IPS 雙面螢幕	13.3" 2.8K (2880 x 1800), 16:10, 100% DCI-P3 (typical), OLED 螢幕
顯示卡	NVIDIA® GeForce RTX 4060 8GB GDDR6 雙記憶體板 GPU	Intel® Arc™ Graphics	Intel® Arc™ Graphics
儲存裝置	1TB NVMe PCIe SSD	1TB NVMe PCIe SSD	1TB NVMe PCIe SSD
鍵盤	單色背光鍵盤 (由光)	單色背光鍵盤 (由光)	單色背光鍵盤 (由光)
網路	Gigabit Ethernet / Intel® Killer™ Wi-Fi 7 BE1750, 藍牙 v5.4	Gigabit Ethernet / Intel® Killer™ Wi-Fi 7 BE1750, 藍牙 v5.4	Intel® Killer™ Wi-Fi 7 BE1750, 藍牙 v5.4
音訊	2x 2W Speakers / 1x Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready	2x 2W Speakers / 1x Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready	2x 2W Speakers / 1x Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready

Prestige 16 AI Studio
搭載**3麥克風陣列**設計+
全新AI降噪技術

PC鏡頭晶片提升使用體驗



Synaptics在Computex展示智慧感測解決方案與**Astra系列晶片**的PC攝影機，可達到反應更快速、更準確的人體存在偵測(HPD)，也可節省電力並支援頭部姿勢追蹤、手視辨識功能等

資料來源：各公司，MIC整理，2024年7月



續航力表現增強，提升AI PC體驗感受



體驗

面板

選擇低功耗面板、提升畫面效果

在搭配高算力晶片下，耗能將成為負擔，面板的省電成大勢所趨

電池

容量提升

電池容量增長有助提升續航力表現

OLED面板強調顯示效果

AI NB搭載或可選配OLED面板，以提供使用者較佳的顯色度，因應圖像生成等照片、影片的呈現；同時，相較LCD，OLED亮度通常較一般LCD更低，因此較為省電

NB電池容量提升

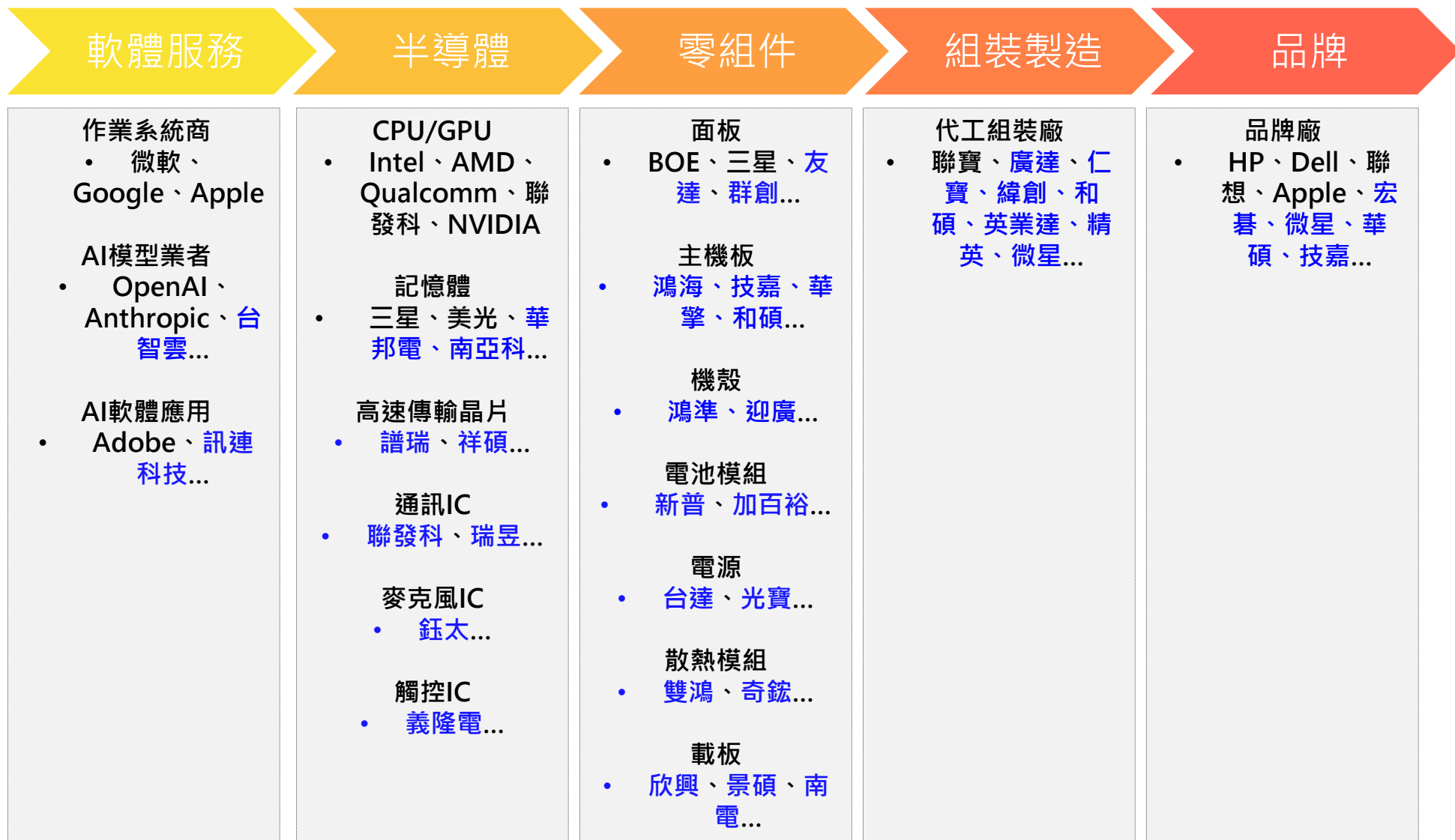
筆電機種	電池容量
Acer Swift 14 AI	75Wh
Acer TravelMate P6 14	65Wh
ASUS Zenbook S 16	78Wh
MSI Prestige 16 AI Evo	99.9Wh

電池容量的提升可使筆電在運算需求增加的同時，維持較長的電池續航力，目前主流AI NB的電池容量，大多落在**70Wh-80Wh**之間

資料來源：各公司，MIC整理，2024年7月



AI PC市場帶動供應鏈廠商產品加值



資料來源：MIC，2024年7月

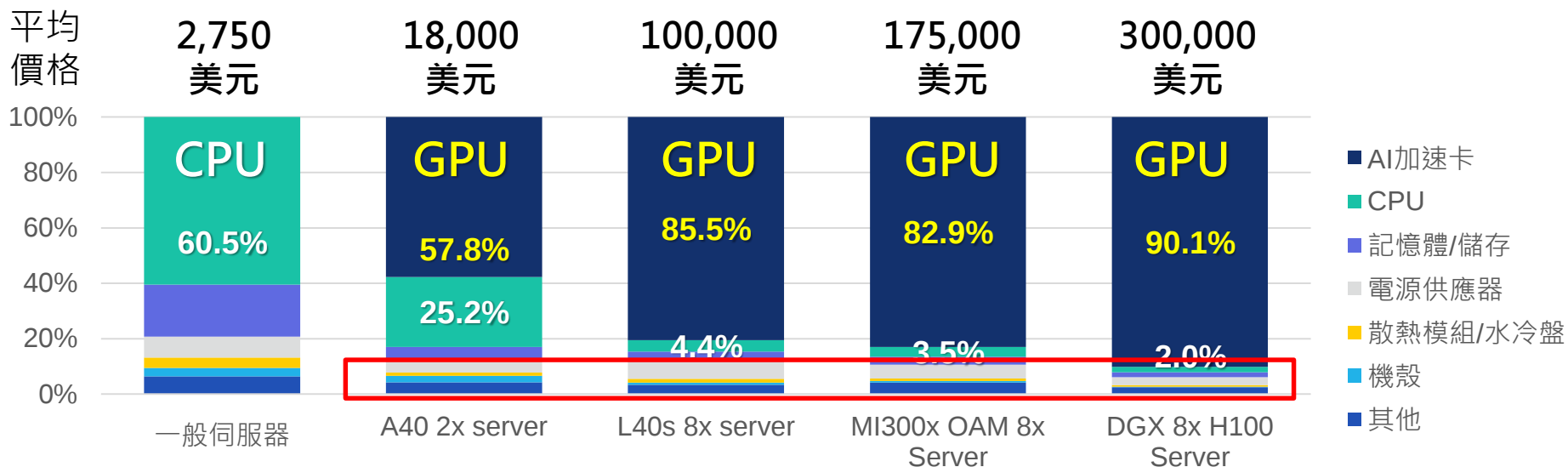
台灣產業供應鏈發展機會

-AI伺服器產業-



AI伺服器規格與成本佔比更加複雜

一般伺服器 VS AI伺服器BOM表價格佔比



電源供應器

主流AI訓練伺服器搭載6顆鈦金級 3000wPSU，當前廠商持續研發4000w~6000w的PSU

機殼

AI訓練伺服器機殼架構與高度調整，由1、2U變為7U，使機殼整體單價增加

散熱模組/水冷盤

如氣冷所需的3DVC與直接式液冷所需的水冷盤

其他

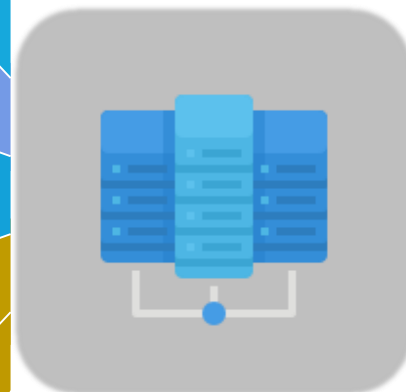
AI伺服器所需BMC數量增加、PCB單板層數增加

資料來源：MIC，2024年4月

- AI伺服器的規格逐漸多元化，在最終售價以及BOM表佔比方面均有所不同
- 儘管伺服器零組件的單價均上升，伺服器當中GPU成本比重仍在提升



AI資料中心架構促使不同類型台灣廠商升級產品規格

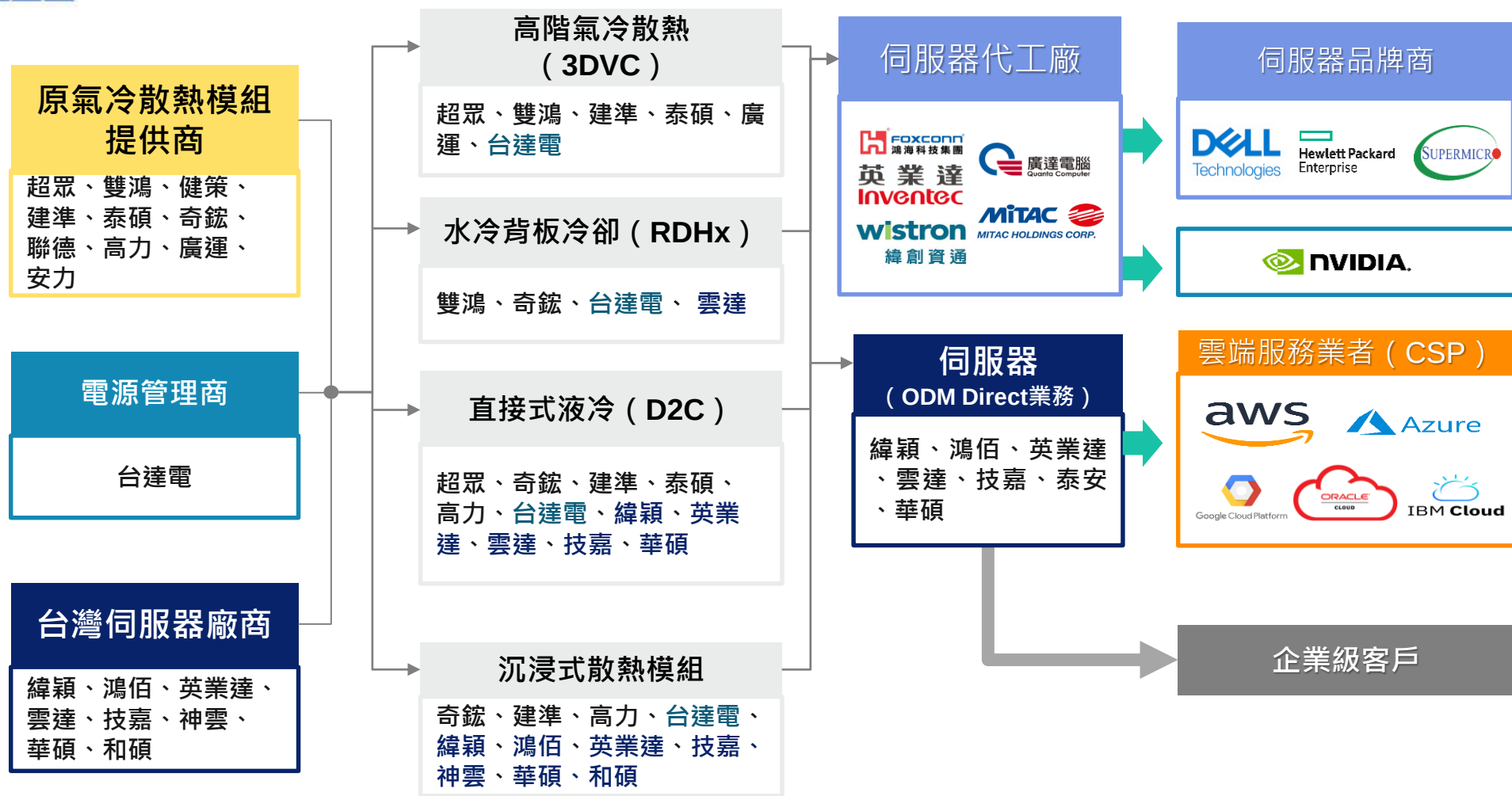


資料來源：MIC，2024年4月

- 台灣企業在AI硬體耕耘多年，AI資料中心架構主要影響AI伺服器發展
- 另外對於網通設備、儲存設備、散熱與基礎設施相關廠商均將造成影響



AI伺服器帶動液冷散熱吸引台廠進行布局



資料來源：各公司，MIC整理，2024年4月

- 台灣既有的散熱模組提供商、電源管理商與伺服器廠商，紛紛開始布局液冷散熱
- 透過不同的散熱產品組合，滿足客戶需求並提供更完整的AI硬體解決方案

結論



結論(1/2)

- **AI PC議題火熱然應用趨勢未明，2024年PC市場回溫有限**

- ◆ 儘管全球經濟情勢仍未明朗，2024年全球PC市場表現將在多項正向因素支持下，略好於2023年。正面因素如Windows 10將停止支援、商用PC更新週期來臨等，帶來正面影響效應，而AI PC的問世更為PC市場需求帶來一劑強心針，不過仍需軟體支援尋找應用利基，以利滲透率的加速

- **區域市場復甦不同調，新興市場成長看好**

- ◆ 儘管週期換機將有望帶動PC市場需求增長，然全球景氣回溫復甦進度影響PC採購力道。歐美市場通膨趨緩，消費動能在2023下半年已略見好轉，預期2024年PC市場將能有正向表現；新興市場如印度，消費力道受惠數位需求增長而維持強勁；中國大陸則仍陷經濟疲弱，有待復甦

- **處理器三強競逐AI PC主導權，處理器算力競爭白熱化**

- ◆ Intel、AMD及Qualcomm相繼發布具AI運算功能行動處理器，三強將在2024年以AI算力決一勝負；桌機部分，Intel、AMD將發布新架構處理器，在2024年下半年競逐消費級PC產品算力寶座



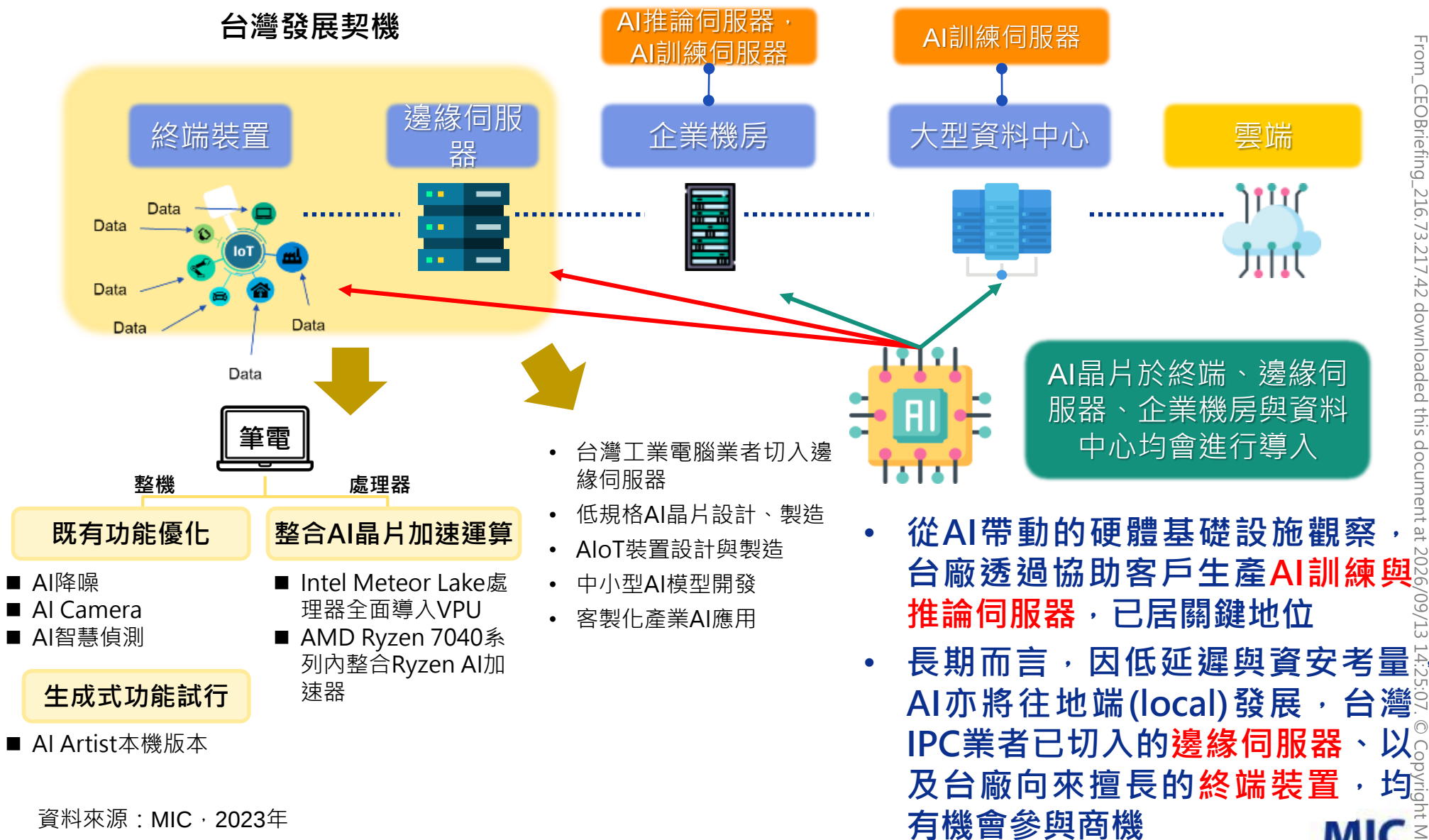
結論(2/2)

- 2024年AI訓練與AI推論伺服器同步發展，帶動全球伺服器市場
 - ◆ 繼雲端服務商之後，伺服器品牌商亦開始推出AI伺服器產品。然因NVIDIA H100 GPU的產能，供不應求，使得AI伺服器訂單延遲至2024年。展望2024年，各企業將持續搶購高階AI訓練伺服器，同時中小生成式AI模型落地，將有望促成AI推論伺服器的發展，帶動全球伺服器市場出貨
- 2024年伺服器處理器市場更加複雜，AMD市占持續上升，更多廠商進入Arm架構生態圈
 - ◆ 2024年伺服器處理器在CPU方面，受惠於產品性價比優勢，AMD市佔將會持續上升。另外Arm架構生態圈吸引許多希望自研處理器的廠商，包含導入多年的AWS、Ampere，研發Grace CPU的NVIDIA，以及Microsoft發表的新處理器Cobalt，預期未來競爭將更加激烈
- 液冷散熱技術已成AI資料中心機房發展重點
 - ◆ 運算晶片功率高漲，且伺服器運算晶片承載密度提高，傳統空氣冷卻已不敷需求，液體冷卻系統將成為高階運算系統解決方案，也讓資料中心機房更節能永續
 - ◆ 資料中心引入液冷系統連帶影響資料中心機房布建，伺服器及基礎設備商傾向共同合作，推出整合性解決方案服務，方便客戶採購並提高產品競爭優勢



未來產業發展重點及建議：供應鏈觀點

台灣發展契機



資料來源：MIC，2023年



MIC[®] 產業提昇的關鍵力量

Thank You

魏傳虔 產業顧問兼組長 chriswei@iii.org.tw

黃 馨 資深產業分析師 angelahuang@iii.org.tw

陳牧風 產業分析師 stephenchen@iii.org.tw

陳奕伶 產業分析師 ylchen@iii.org.tw

陳彥蓉 產業分析師 norachen@iii.org.tw

產業情報研究所

智慧財產權暨引用聲明

- 本活動所提供之講義內容或其他文件資料，均受著作權法之保護，非經資策會或其他相關權利人之事前書面同意，任何人不得以任何形式為重製、轉載、傳輸或其他任何商業用途之行為
- 本講義內容所引用之各公司名稱、商標與產品示意照片之所有權皆屬各公司所有
- 本講義全部或部分內容為資策會產業情報研究所整理及分析所得，由於產業變動快速，資策會並不保證本活動所使用之研究方法及研究成果於未來或其他狀況下仍具備正確性與完整性，請台端於引用時，務必注意發布日期、立論之假設及當時情境