

## 生成式AI風潮下，全球運算系統發展趨勢

李建勳

產業顧問兼副主任

產業情報研究所

財團法人資訊工業策進會

2024.04.24

# 簡報大綱

- AI PC市場與產品發展
- AI伺服器市場與產品發展
- 台灣產業發展機會
- 結論

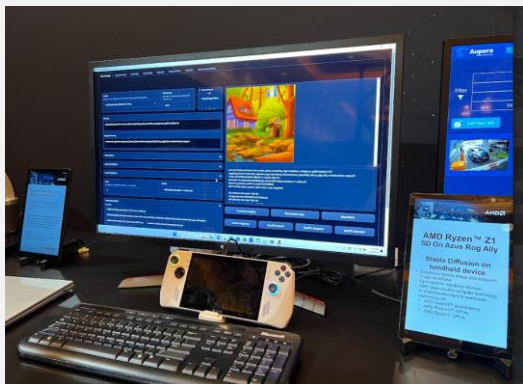
# AI PC市場發展



# 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4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

## 處理器內建AI引擎

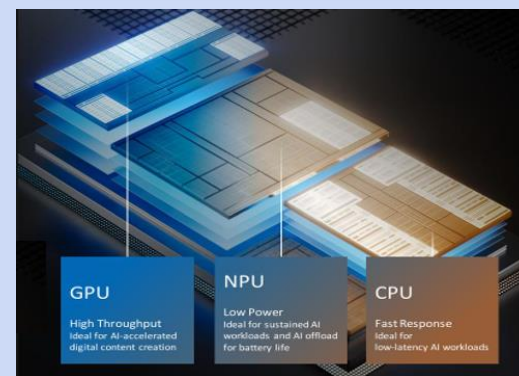
具備整合神經網絡單元(NPU)處理器，加速AI運算任務

晶片廠	內建NPU處理器	AI 算力
intel	Meteor Lake	未公告
AMD	Ryzen 8040	16 TOPS
	Ryzen 7 8700G	16 TOPS
QUALCOMM®	Snapdragon X Elite	45 TOPS
Apple	M1	11 TOPS
	M2	15.8 TOPS
	M3	18 TOPS

資料來源：各公司 · MIC整理 · 2024年4月  
 註：整理截至2024年6月各晶片廠已推出及將推出的搭載NPU之處理器

## 算力門檻

微軟正與硬體廠商合作設計AI算力超過40 TOPS PC產品，該門檻或將成為AI PC新標準



資料來源：intel · 2024年4月

備註：TOPS為Tera Operations Per Second  
 資料來源：MIC · 2024年4月



# AI應用隨使用場景不同，提供差異化的體驗服務

AI PC  
優勢

提升運算反應**速度**

節省雲端資源**成本**

提高**隱私**與安全性

## AI PC角色與功能

### 商用市場

#### 個人秘書

會議記錄與摘要  
資料辨識與除錯  
個人行程表  
Office工具使用  
...

### 遊戲市場

#### 遊戲教練

場景檢測與調教  
遊戲效能強化  
打造數位分身  
實時遊戲指導  
...

### 創作者市場

#### 工具夥伴

數位內容創作  
圖片/影片處理  
資料科學開發  
程式設計與編輯  
...

### 消費性市場

#### 朋友/閨密

設備管理  
知識管理  
記憶搜尋  
專屬聊天機器人  
...

資料來源：MIC · 2024年4月

- 未來AI PC可依循不同應用市場，提供不同AI服務體驗，成為各式輔助角色

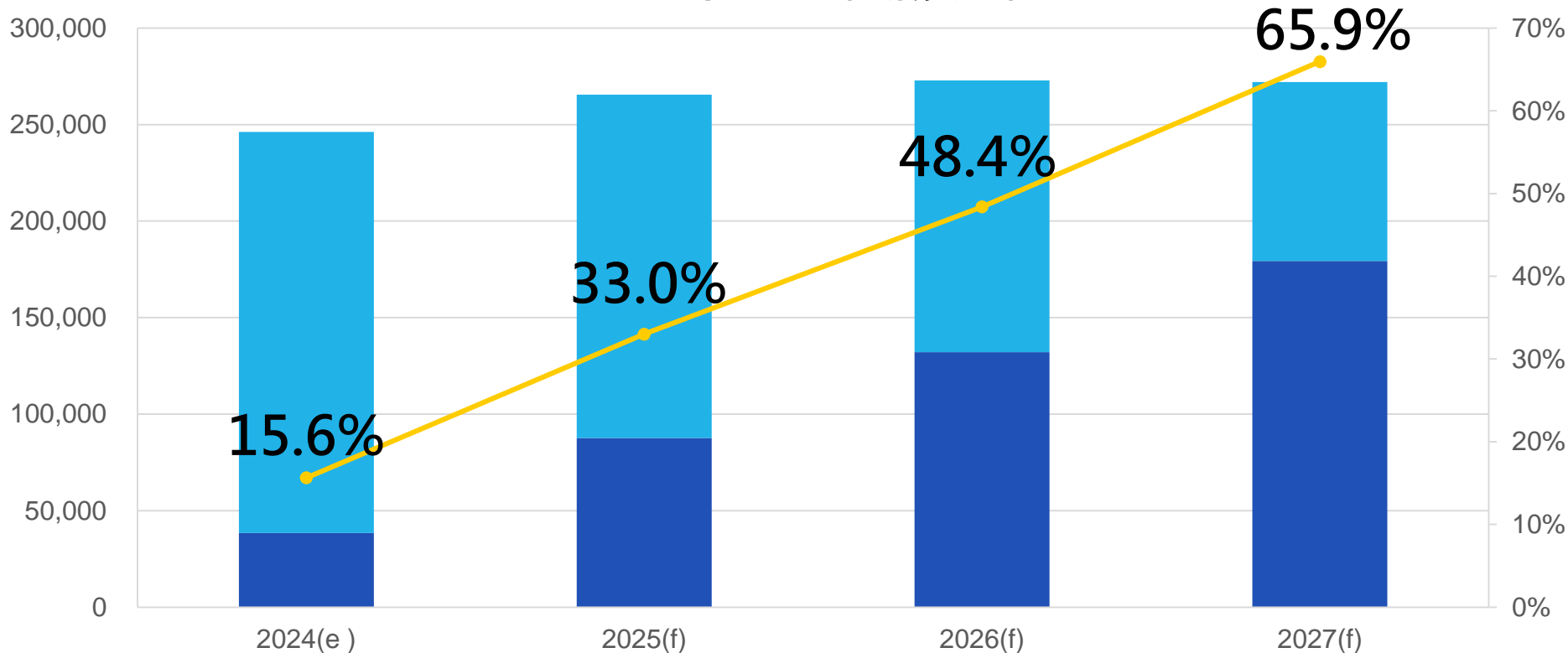


# 2027年AI PC市場滲透率近七成

單位：千台

## 全球AI PC市場滲透率

單位：%



資料來源：MIC，2024年4月

■ 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產品發展



# AI PC帶來更加混合、協作的時代

## 發展趨勢

混合異構的處理器架構

CPU



GPU



NPU

算力層

平台支援強化硬體與軟體的實踐

平台層

大模型

雲端、閉源、訓練資源大

小模型

地端、開源、客製化

模型層

雲端+地端應用開發，豐富使用體驗

應用層

硬體+軟體加成，差異化AI PC產品

終端層

## 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4月



# 處理器的異構運算為AI PC分配執行算力

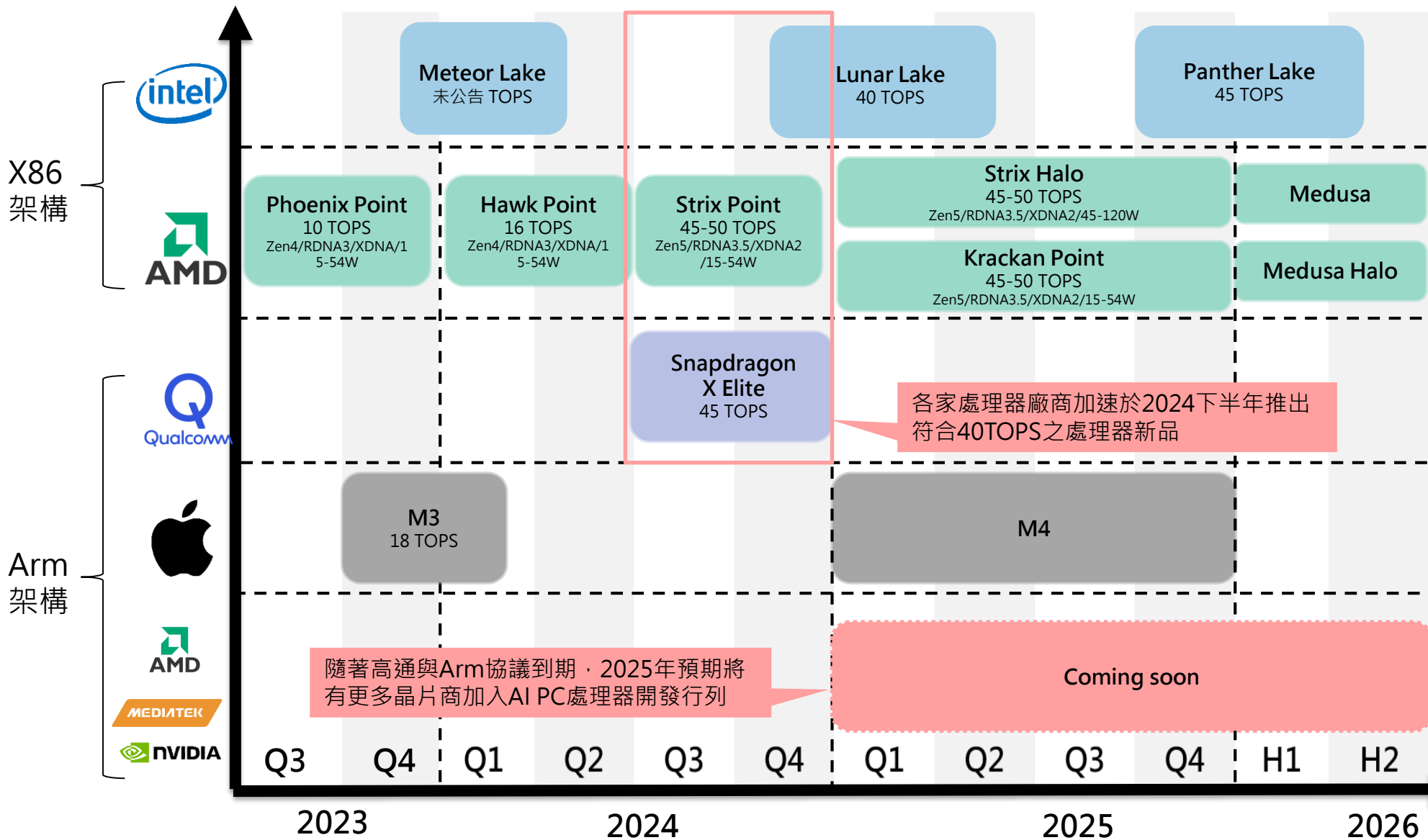
	intel	AMD			Qualcomm	Apple		
SoC	Meteor Lake	Ryzen 7040系列	Ryzen 8040系列	Ryzen 8700G、8600G系列	Snapdragon X Elite	M1	M2	M3
<b>CPU</b> 執行低延遲的AI應用，需要處理輕量且單一模型的推論	Redwood Cove + Crestmont	Zen 4 CPU			Oryon CPU	P-core+E-core		
<b>GPU</b> 需要大量平行運算和高吞吐量處理的AI任務	Xe-LPG GPU	RDNA 3			Adreno GPU	GPU		
<b>NPU</b> 具低功耗、省電優勢，負責需要持續處理的AI任務	Intel AI Boost NPU	Ryzen AI NPU			Hexagon NPU	Neural Engine		
<b>應用產品</b>	NB	NB	NB	DT	NB	NB/DT/AIO PC		

備註：各公司，MIC整理，2024年4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



# 處理器廠加速布局，2024下半年可見新品

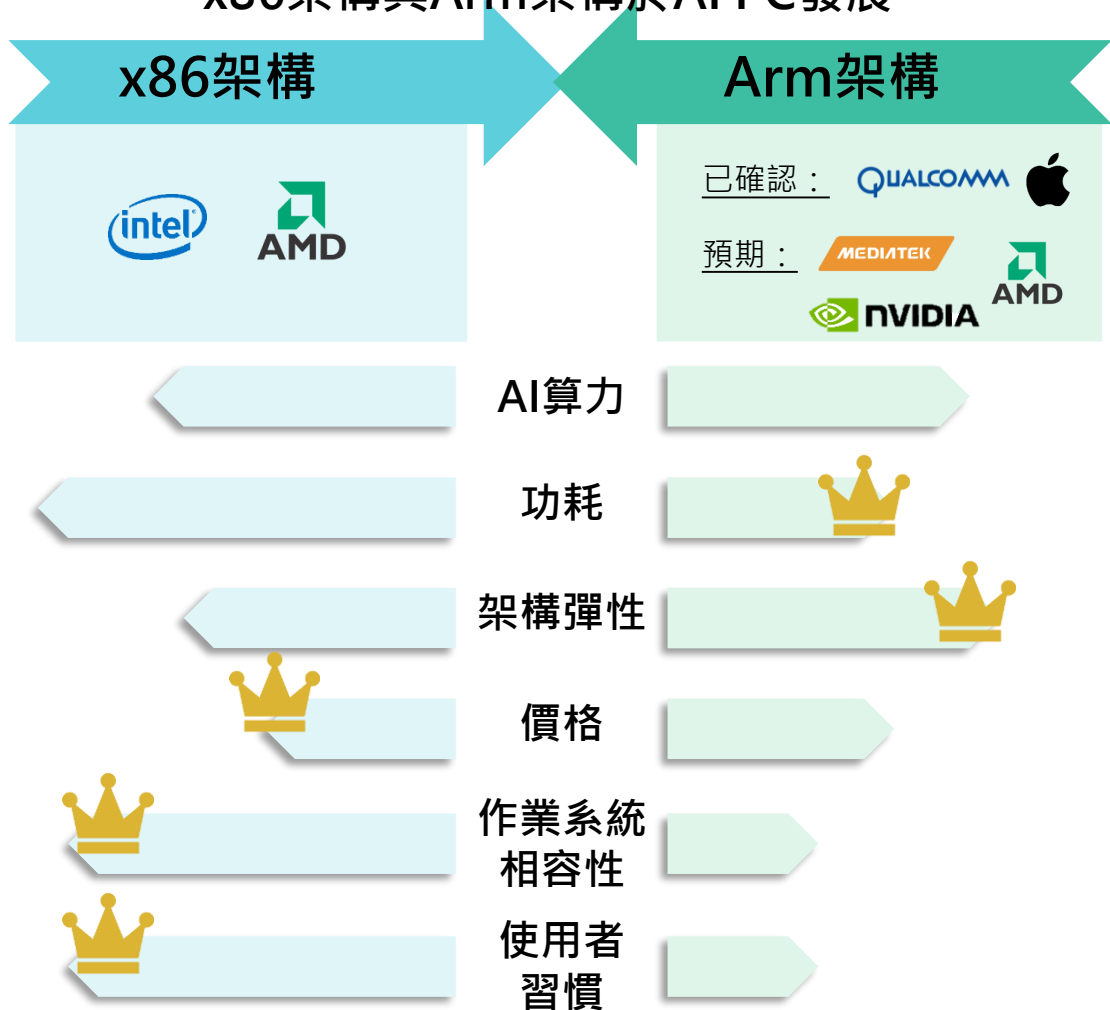


資料來源：各公司 · MIC整理 · 2024年4月

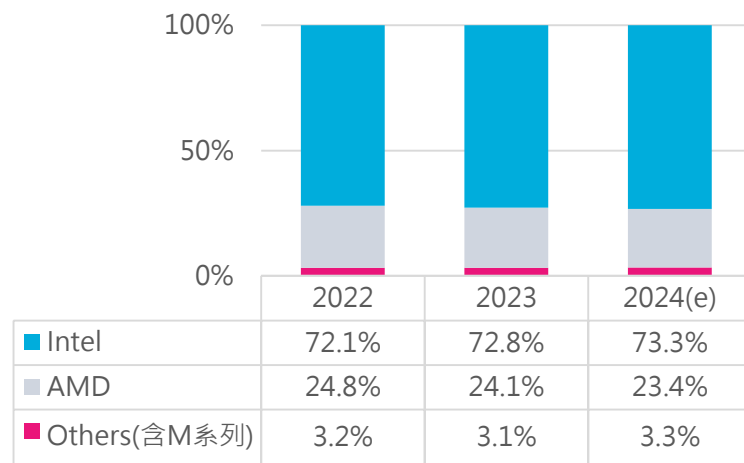
備註：僅列搭載NPU之處理器

# Arm架構雖來勢洶洶，但短期仍有考驗要解

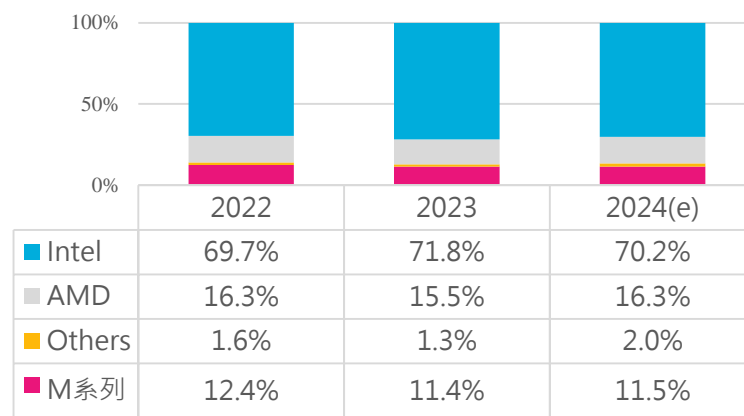
## x86架構與Arm架構於AI PC發展



全球桌機處理器搭載比重



全球筆電處理器搭載比重

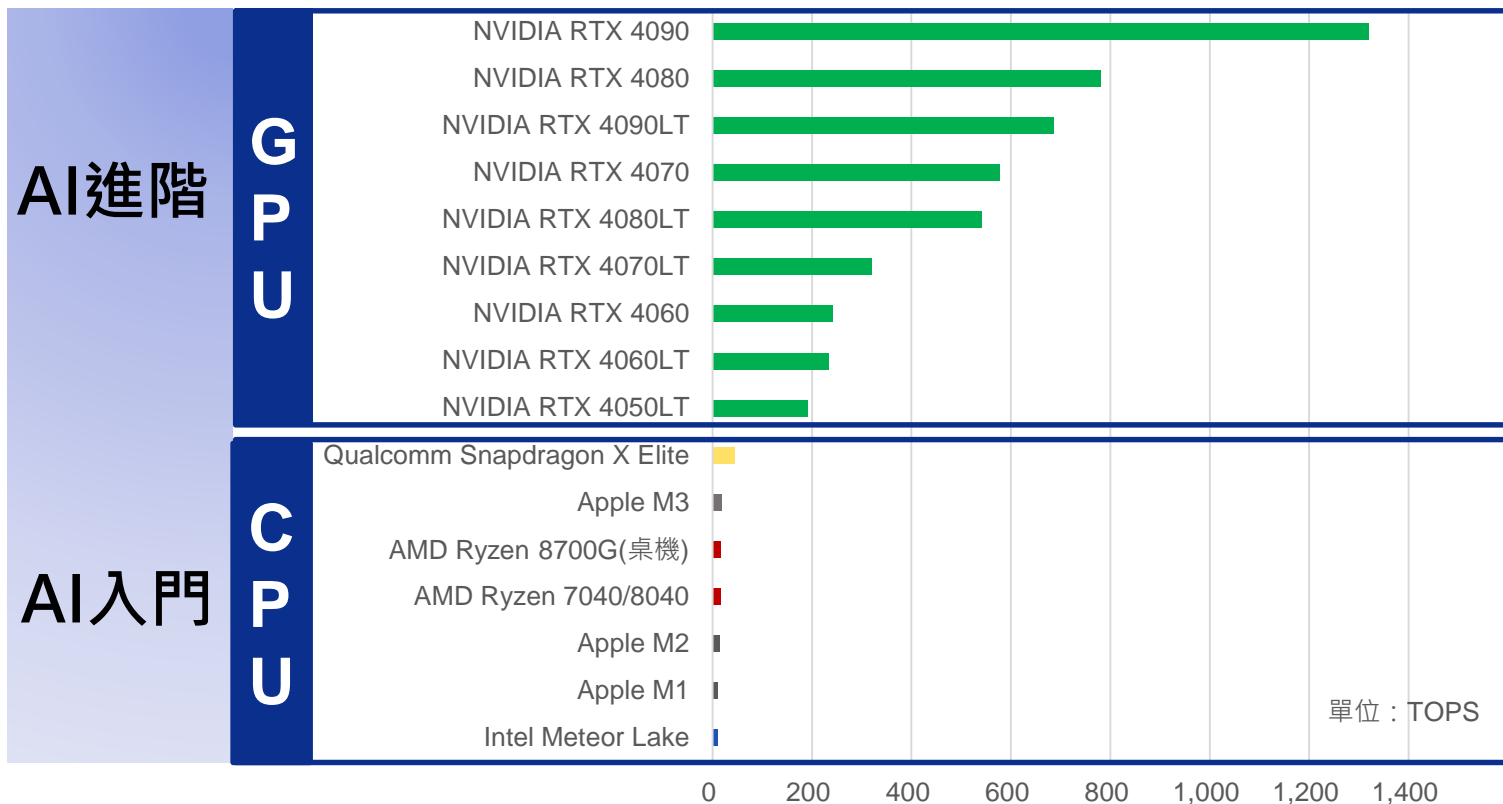


資料來源：MIC，2024年4月

- Arm架構積極展現在AI模型整合與功耗上的優勢，然與作業系統之相容性及使用者習慣，仍為Arm架構能否勝出之關鍵



# NPU支援AI PC入門算力、GPU為進階選項



399~1,600美元



375~1,000美元

資料來源：MIC · 2024年4月

- 整合NPU的CPU提供PC更強大的AI算力，現階段處理器大廠推出的處理器算力落在10-45 TOPS之間，最高約可以支持130億參數的LLM模型
- 若需要更高算力或要執行更大規模參數之生成式AI模型，可增配NVIDIA、AMD、Intel顯卡，最高可支援340億參數的LLM模型，然處理器加顯卡讓PC價格更為高昂，適合進階使用者



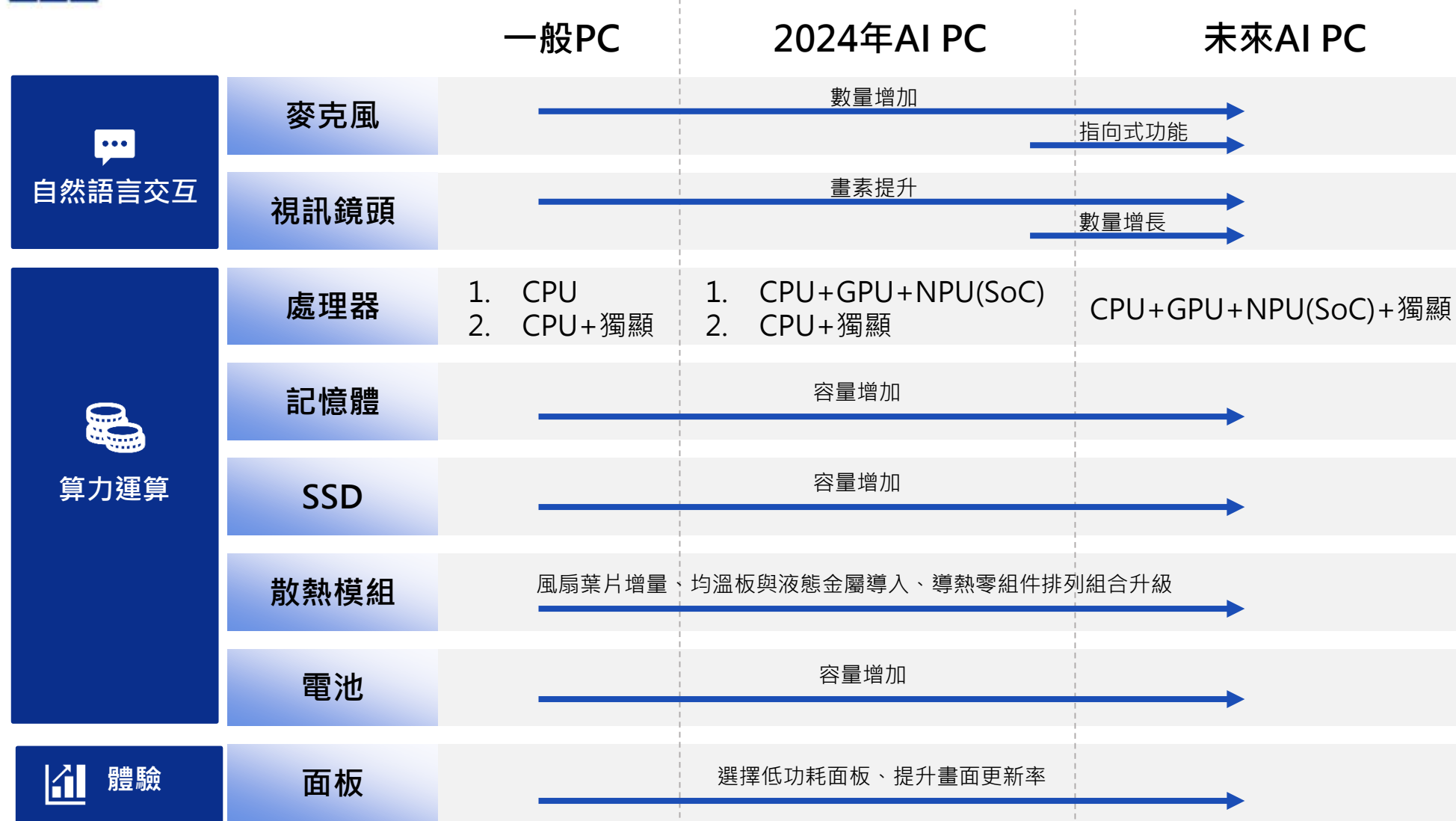
# 微軟從作業系統作為著力，加速AI PC生態系建立

	x86架構	Arm架構
<b>算力層</b>	與Intel、AMD合作，並支援NPU的AI加速 <ul style="list-style-type: none"> <li>推出DirectML，支援Intel NPU處理器</li> <li>AMD表示即將在AMD NPU處理器上啟用微軟運算驅動程式模型</li> </ul>	<ul style="list-style-type: none"> <li>Windows 11 24H2將可能率先更新至Arm架構處理器</li> <li>除高通Snapdragon X Elite外，NVIDIA和AMD將可能加入Windows on Arm陣營</li> </ul>
<b>平台層</b>	<b>Win11 24H2 即將推出</b> <ul style="list-style-type: none"> <li>配備NPU達到40 TOPS，以執行AI功能</li> <li>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</li> <li>推薦記憶體達16GB，目前AI PC運行所需最低記憶體容量為4GB</li> </ul>	
<b>模型層</b>	<b>Windows AI Studio工具，供本地端AI模型開發</b> <p>讓開發者更接近微軟所提倡的「混合循環」開發模式，可以在Azure雲端上執行模型，也可以在Windows本機邊緣運行模型，甚至是跨兩者</p>	
<b>軟體層</b>	<b>Copilot與擴充服務</b> <ul style="list-style-type: none"> <li>Windows系統支援AI助理Copilot，亦支援第三方服務外掛</li> <li>推出Copilot for Sales、Copilot for Service、Copilot Enterprise、Copilot for Finance與Copilot for Security，強化企業生產力</li> </ul>	
	<b>AI Explorer</b> <ul style="list-style-type: none"> <li>據傳隨Win11 24H2版本推出，為AI PC的關鍵功能</li> <li>用戶的歷史/時間線功能：用戶可使用自然語言搜尋其在電腦上做的所有事情，找到以前打開的對話、文件、網頁和圖像</li> </ul>	
<b>終端層</b>	<b>微軟首批AI PC</b> <p>推出Surface Pro 10和Surface Laptop 6兩款商用版AI PC，均搭載英特爾最新Core Ultra處理器，具有NPU以及今年起新增的Copilot鍵</p>	

資料來源：MIC，2024年4月

- 微軟對於AI PC著力深，除了布建Windows版本的更新外，包含算力支援、軟體開發乃至終端NB的推出均有進展

# 因應AI PC使用型態的改變，硬體規格須有所升級



資料來源：MIC · 2024年4月

- 因應人機互動模式改變、雲端與地端協作能力提升、使用體驗增長，硬體規格將有所升級，帶動供應鏈ASP增長



# 品牌廠欲透過AI應用，差異化產品體驗

使用體驗升級為最多廠商首要提供的AI服務

品牌廠希望透過開源模型訓練，打造自家AI助理

	安全性防護	音訊/視訊功能補強	運作效能管理	內容生成	個人化助理	
			Dell Optimizer/My Dell/ProSupport Suite			三大品牌廠 AI應用著力範圍廣
	HP專用AI晶片	HP Presence	HP Smart Sense		Superpower	
	Lenovo Commercial Vantage	Lenovo View/Lenovo Premium Suite/Lenovo AI智慧會議助手	Lenovo X Power/Lenovo Vantage/Lenovo Smart Power 3.0/Lenovo LA AI晶片	Lenovo AvatarMaster/Lenovo Yoga Creator Zone	Lenovo AI Now	
		Acer PurifiedVoice/Acer PurifiedView/Acer QuickPanel/Acer LiveArt	AcerSense	Acer AlterView		台灣品牌廠 重心放在AI小工具
		ASUS AiSense Camera/ASUS AI Noise-Cancelling				
		MSI AI Noise Cancellation Pro	MSI AI Engine	MSI AI Artist		
					LG AI Brain	韓系品牌廠 直搗個人助理應用發展
					Samsung Gauss	

資料來源：各公司·MIC整理·2024年4月

備註：虛線為尚未正式落地於上市產品中

# 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

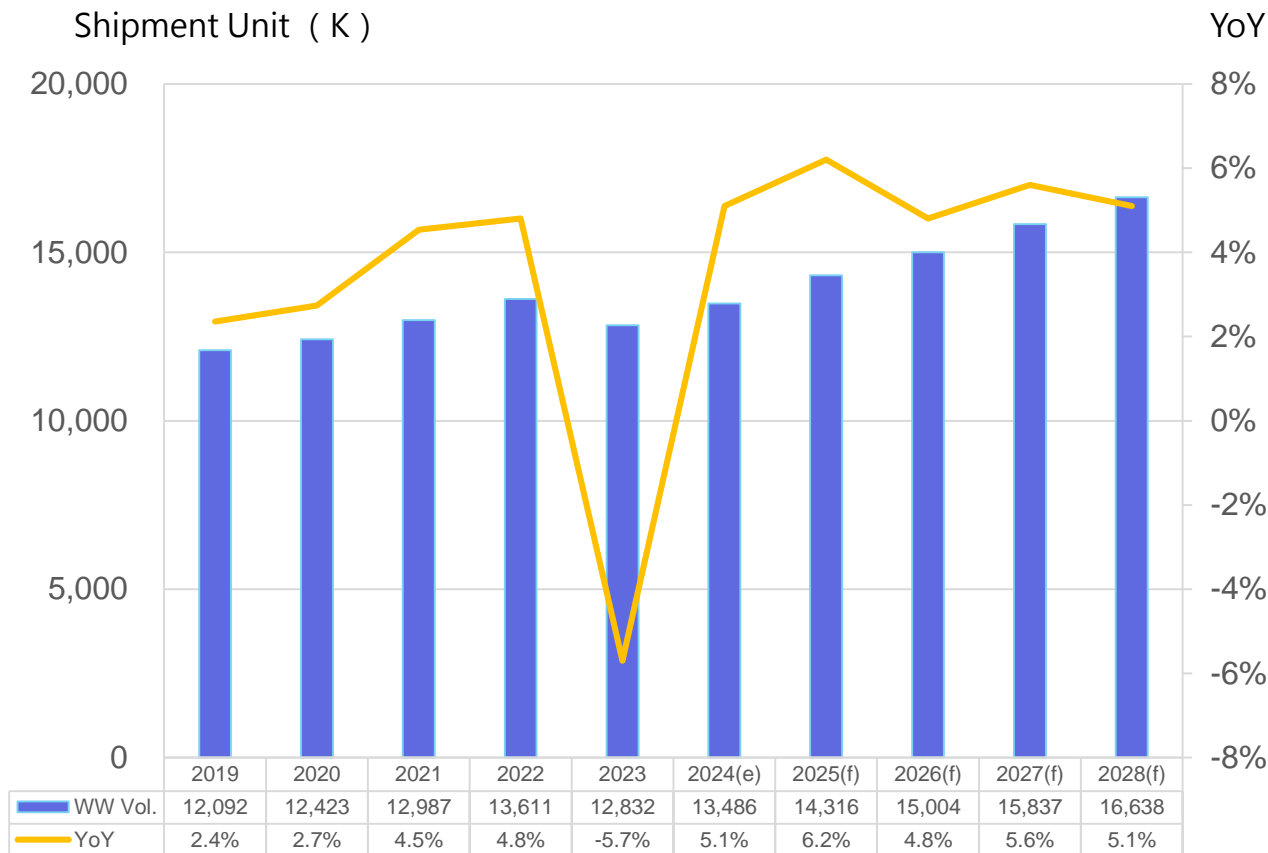
雲端服務商AI伺服器需求仍在延續

伺服器品牌商新AI伺服器產品量產

AI推論與中小模型微調需求

同樣資本支出下，一般伺服器採購率相對下滑

## 2019~2028年全球伺服器市場預測



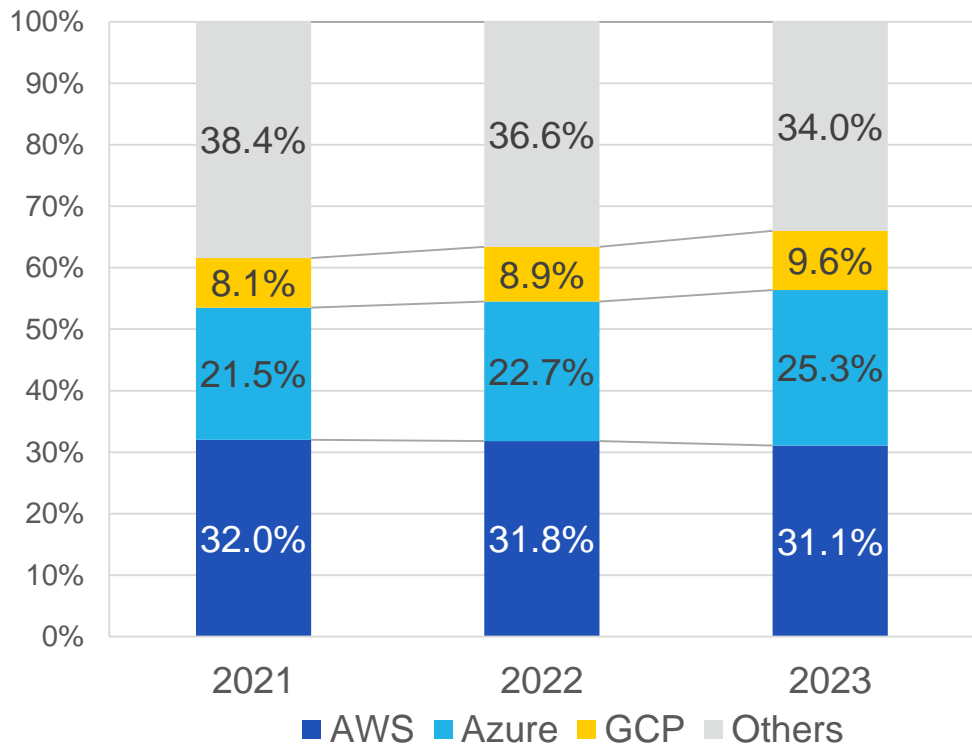
資料來源：MIC，2024年4月

- 全球伺服器市場2024年的主要驅動力為AI伺服器，不論雲端服務商、伺服器品牌商訂單均在上升
- 另外AI推論伺服器將因為企業端推論、中小模型微調 ( Fine-tune ) 需求而出貨量增加

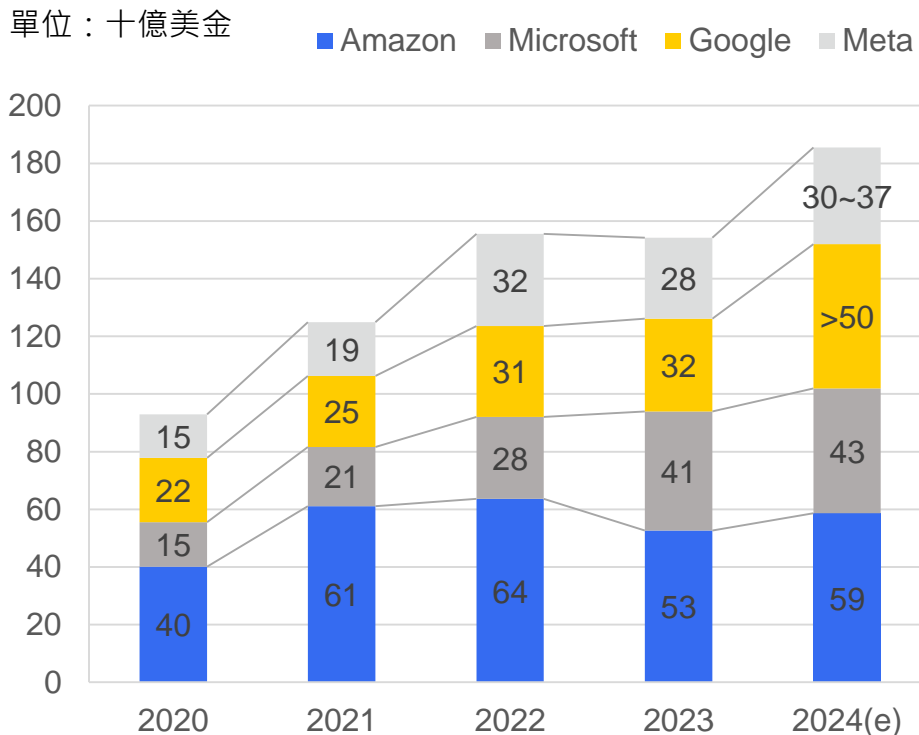


# 2024年四大雲端服務商將加大資本支出

## 2021年~2023年雲端服務商市占率



## 雲端資料中心前四大建造者資本支出



資料來源：各公司，MIC整理，2024年4月

- 2023年因生成式AI、大型語言模型需求，使Azure在市占率大幅提升，Google亦有小量成長
- 2024年預期四大CSP資本支出將進一步提升，主要支出於進行資料中心AI基礎建設



# 2024年全球AI伺服器出貨量與出貨占比遽增

## 2023-2027 全球AI伺服器占比預測



資料來源：MIC，2024年4月

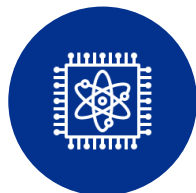
- NVIDIA H100伺服器訂單，H200、GH200預計在2024Q2量產，將使AI訓練伺服器出貨進一步提升
- 搭載 AMD MI300x、MI300A、Intel Habana Gaudi 3的伺服器同樣將在2024年放量
- 受惠於AI推論及中小模型訓練需求增加，搭載中低階GPU、FPGA的AI推論伺服器同樣將帶動整體AI伺服器出貨成長

# AI伺服器產品發展



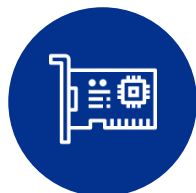
# AI全面影響資料中心不同構面的建置

## 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伺服器中記憶體暫存熱數據外，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




	Meta	Microsoft	Google	Amazon	Oracle
自研晶片	<b>MTIA (元訓練與推理加速器)</b> ：針對推理工作負載內部客製化加速器晶片	<b>Azure Cobalt</b> ：Arm架構的CPU，強化雲端運算 <b>Maia 100</b> ：為AI負載設計的AI晶片，可執行大型AI模型	<b>TPUv5e</b> ：每個Pod擁有256個晶片，AI推論較高性價比 <b>TPUv5p</b> ：每個Pod有8960個晶片，處理大型AI訓練	<b>AWS Inferentia</b> ：專注在深度學習與AI推論 <b>AWS Trainium</b> ：將重點放在AI訓練與模型建構最佳化	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
超級電腦	<b>兩個各採用24,576個NVIDIA H100 GPU的大型資料中心叢集</b>	<b>Azure Eagle</b> ：包含14,400個H100 GPU，超級電腦TOP 500排名第三	<b>A3超級電腦</b> ：最多可堆疊26,000個H100 GPU	<b>P5 Ultra Scale GPU叢集</b> ：可以堆疊20,000個H100 GPU <b>Project Ceiba</b> ：包含16,384顆GH200超級晶片	<b>OCI Super Cluster</b> ：可以擴展數萬個NVIDIA H100 GPU
GPU	預計採購35萬顆H100 GPU，於2024年底打造60萬顆GPU規模AI算力	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	2023約採購5萬顆H100 GPU，預期2024將同步擴大GPU與TPU的堆疊	2024年將採購大量GH200與H200 GPU使用於自身應用實例	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		
	<ul style="list-style-type: none"> <li>H100</li> <li>H200</li> <li>B200</li> </ul>	<ul style="list-style-type: none"> <li>L40s</li> <li>B40</li> </ul>	<ul style="list-style-type: none"> <li>H800</li> <li>H20</li> <li>L20</li> <li>L2</li> </ul>	<ul style="list-style-type: none"> <li>GH200 : Arm架構推論</li> <li>GH200 NVLink : Arm架構訓練與推論</li> <li>GB200</li> <li>GB200 NVLink 72</li> </ul>	<ul style="list-style-type: none"> <li>開發客製化晶片服務</li> </ul>
	<ul style="list-style-type: none"> <li>MI300 X</li> </ul>		<ul style="list-style-type: none"> <li>可能推出MI250降規版</li> </ul>	<ul style="list-style-type: none"> <li>MI300A APU</li> </ul>	<ul style="list-style-type: none"> <li>XILINX FPGA</li> </ul>
	<ul style="list-style-type: none"> <li>GPU Max Series ( Ponte Vecchio )</li> <li>Habana Gaudi 2、3</li> </ul>		<ul style="list-style-type: none"> <li>Gaudi 2 降規版</li> </ul>		<ul style="list-style-type: none"> <li>Intel Agilex FPGA</li> <li>Intel 客製化 ASIC</li> </ul>

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



# 資料中心數據傳輸成為AI模型訓練的關鍵

## 開發AI模型的三個階段

資料準備：蒐集和彙整輸入AI模型資料集



AI訓練：透過大量資料提供給AI模型來訓練其執行特定任務



AI推論：根據新輸入、未見過的數據進行預測或決策

需要大量數據和運算資源來支援其更新過程，AI模型從不斷收集的數據中學習以細化其參數

使用數萬台的GPU伺服器進行叢集

AI資料中心網路必須100%可靠，才能最佳化完成時間、消除延遲導致的速度降低問題

## AI資料中心網路的運作模式

### 網路拓撲架構設計 ( fabric )

使用**CLOS架構**，由多個層級的網路組成，每個層級包含一組交換機 ( switch )，具有高可靠性與擴展性。根據模型大小和GPU規模，可以使用不同層級的結構

### 流量控制和擁塞避免

具有**最佳連接數量**的適當大小的結構互連，以及**檢測和糾正流量不平衡**以避免擁塞和資料包丟失，如明確擁塞通知 ( ECN ) 與資料中心量化擁塞通知 ( DCQCN )

### 大規模建置與效能

**乙太網路**成為處理HPC和AI應用的首選開放標準解決方案，包括目前向**800 GbE** 和**資料中心橋接 ( DCB )** 的進展

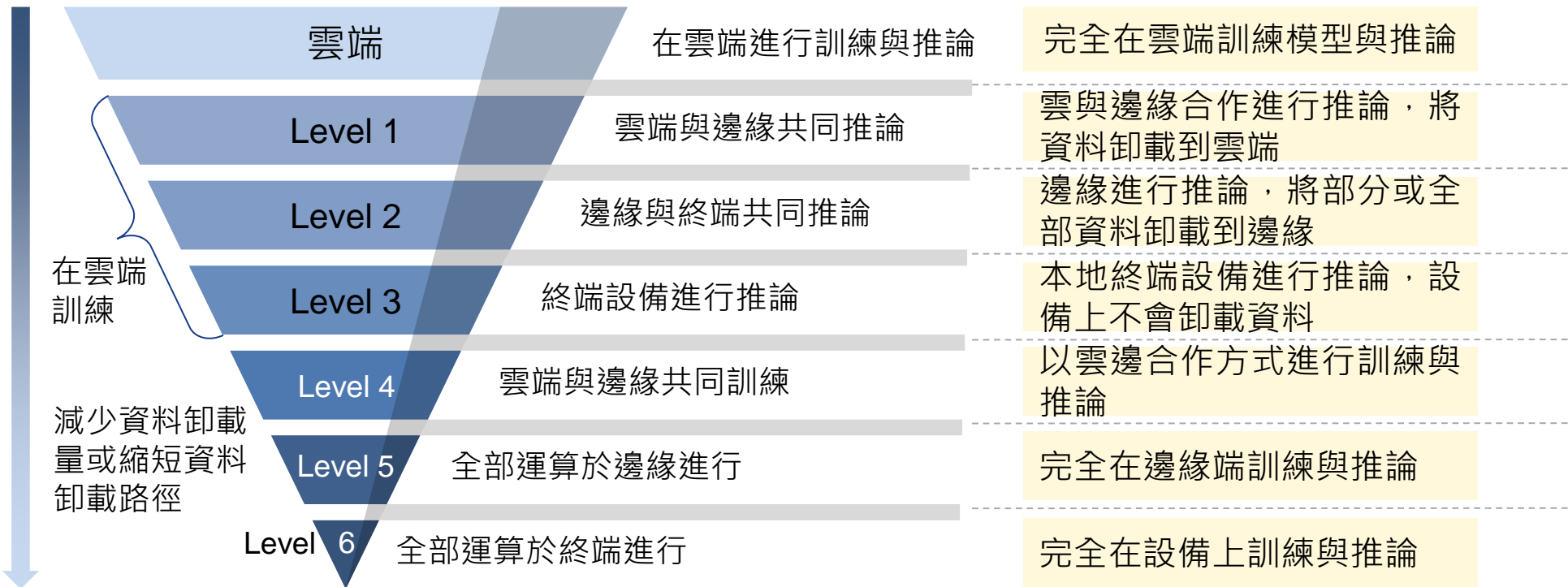
資料來源：Juniper Networks · MIC整理 · 2024年4月

- 要建置完善的資料中心的網路傳輸才能最大限度提高GPU的利用率
- 透過不同的網路架構設計，才能符合AI訓練的高速傳輸需求
- 擁塞管理、最小化延遲與最佳化作業完成時間，成為資料中心網路布建時的關注重點



# AI訓練與推論從雲逐步朝邊緣擴展

## AI運算從雲端到邊緣的六個層級



資料來源：《Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing》，MIC整理，2024年4月

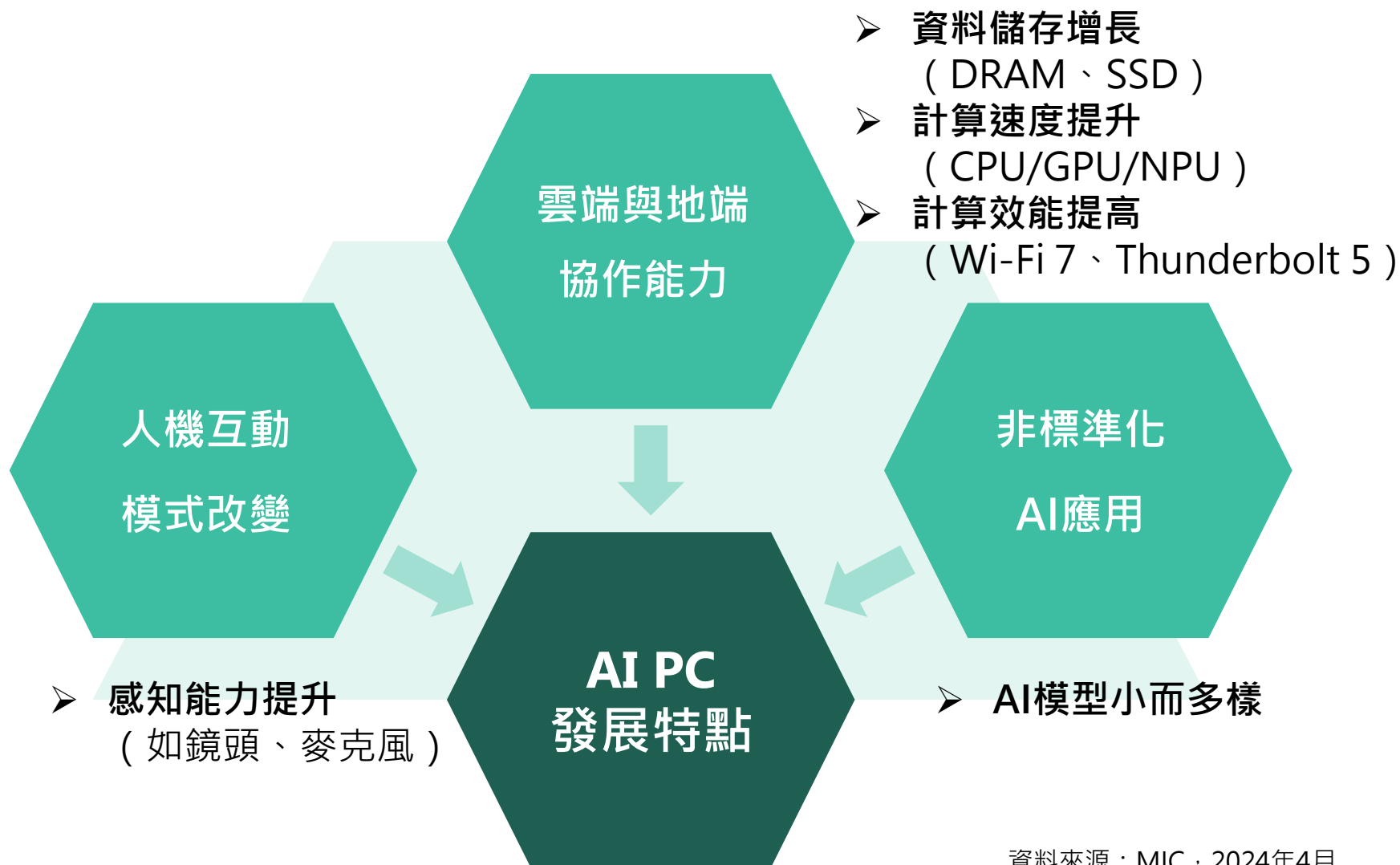
- 當前AI訓練主要集中在雲端，由雲端服務商採購AI訓練伺服器進行訓練
- 推論則有部份轉移至邊緣，由企業自身或是邊緣資料中心採購AI推論伺服器進行協助
- 預計在未來幾年內不論AI訓練與AI推論均會往邊緣端發展

# 台灣產業發展機會



# AI PC ( 1/2 ) :

## AI感知、理解與洞見，帶動軟硬體需求轉變



資料來源：MIC · 2024年4月



# AI PC ( 2/2 ) :

## 尋求AI應用利基將成為品牌廠/代工廠新機會點



### 通用性 體驗

**(一)效能提升**：透過處理器本身的效能支援及AI加速器加成，協助筆電進行畫面刷新率的動態調節與分擔

**(二)智慧化管理**：透過AI軟體導入，讓背景模糊、眼神及臉部偵測、背景噪音的消除等功能更加升級，並可自動偵測使用程式，動態調整效能配置

### 客製化 服務

**(一)商務市場建構企業私有訓練架構**：品牌廠可與作業系統商合作，針對個別公司打造獨有的AI Module，開發客製AI運算模型

**(二)消費市場進行小型模型訓練**：在本機端蒐集使用者使用資訊，透過小型訓練模型針對個人的使用習慣加以訓練並提升使用回饋

### 可推動合作方向

#### 作業系統商



提供具備AI工具、AI模型的作業系統



#### 晶片商



處理器整合AI加速器，滿足運作效能



#### 品牌商



聯手晶片商，將AI體驗內建於PC產品



#### 代工廠



與品牌廠合作，開發設計AI PC

#### 軟體開發商

- 既有PC應用程式加入AI
- 開發新的AI軟體整合方案

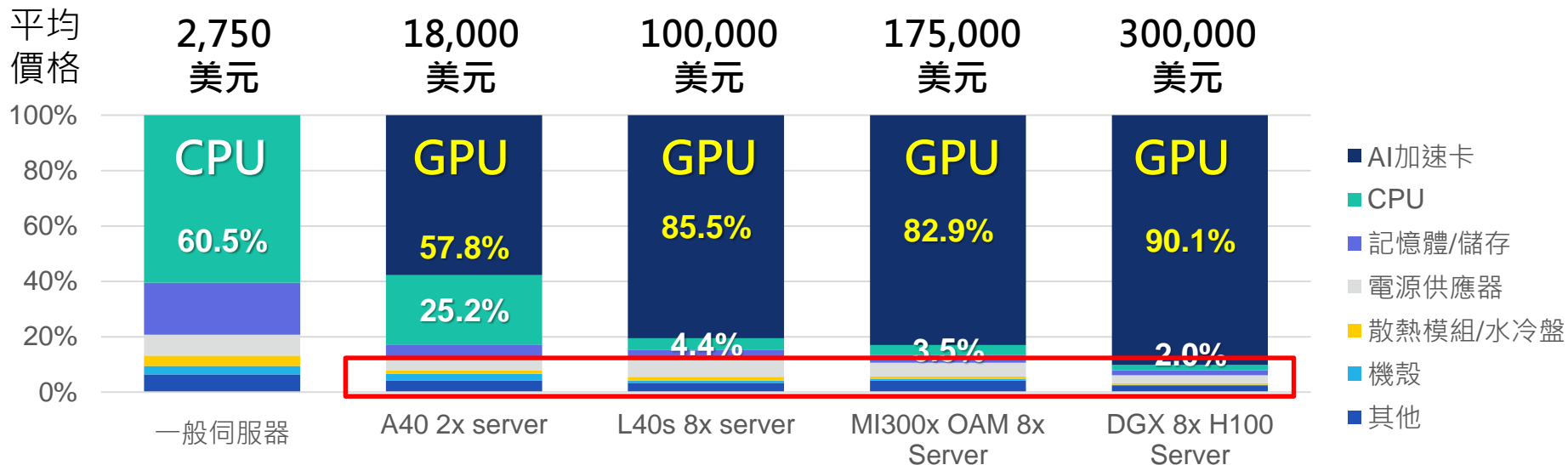
資料來源：MIC，2024年4月



# AI 伺服器 ( 1/2 ) :

## 產品規格與成本占比更加複雜

一般伺服器 VS AI伺服器BOM表價格占比



### 電源供應器

主流AI訓練伺服器搭載6顆鈦金級 3000w PSU，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1U、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表占比方面均有所不同
- 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 伺服器 ( 2/2 ) :

## AI資料中心架構促使不同類型台灣廠商升級產品規格



- 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

# 結論



# 結論

## ● AI PC產業

- ◆ **元年啟動待東風**：AI PC硬體到位，只欠軟體應用支援，2027年AI PC滲透率將有望達近七成
- ◆ **產品定位求革新**：AI PC產品定位增添混合協作功能，扮演各種貼近生活的角色
- ◆ **遊戲規則重定義**：未來AI PC的想像空間將更加豐富且多元，透過生態系擴充、產業聯盟、軟硬整合等作法，有望改變PC產業遊戲規則

## ● AI伺服器產業

- ◆ **算力需求促建置**：2024年雲端服務商建設AI資料中心以及企業採購需求，AI伺服器採購增加
- ◆ **架構調整帶商機**：AI算力需求使資料中心調整IT設備（運算、通訊、儲存等）、基礎設施（液冷散熱系統、UPS、配電裝置等）及營運（智慧化監控等）等構面之建置，帶動相關商機
- ◆ **邊緣運算更高效**：邊緣資料中心連接雲端與終端，改善AI應用的執行，提供高效可靠之應用



**MIC** 產業提昇的關鍵力量  
**Thank You**

李建勳 產業顧問兼副主任

[chienlee@iii.org.tw](mailto:chienlee@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

# 附件



# 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4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



# 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

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



- Lenovo ThinkPad X1系列筆電將攝影機、鏡頭蓋與麥克風整合成獨有多媒體列，並可選配**8MP**攝影機
- 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



機種	Prestige 16 AI Studio B1VFG-0211W	Prestige 16 AI Evo B1MG-007W	Prestige 13 AI Evo A1MG-0111W
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吋 QHD+ (2560 x 1600), 16:10, 100% DCI-P3 (typical), 400nits (typical), IPS 等級面板	16吋 QHD+ (2560 x 1600), 16:10, 100% DCI-P3 (typical), 400nits (typical), IPS 等級面板	13.3吋 2.8K (2880 x 1800), 16:10, 100% DCI-P3 (typical), OLED 面板
顯示卡	NVIDIA® GeForce RTX 4060 4GB GDDR6 筆記型電腦 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 (白光)	單色背光鍵盤 (白光)	單色背光鍵盤 (白光)
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
喇叭	2+2W Speakers / 1+ Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready	2+2W Speakers / 1+ Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready	2+2W Speakers / 1+ Audio combo jack / DTS Audio Processing / Spatial Array Microphone (3 Mic) / Hi-Res Audio Ready

- Lenovo ThinkBook 13x配備**4個音響+4麥克風陣列**
- Prestige 16 AI Studio搭載**3麥克風陣列設計+全新AI降噪技術**

資料來源：MIC，2024年4月



# 因應算力需求提升，記憶體容量成關鍵

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

## 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**
- 微軟推進Windows 11 23H2原生功能情況下，將對記憶容量的最低規格訂立為**16GB**，將為AI PC設立**記憶體最低容量標準**
- 經測試，跑llama 7b需要至少**7G**容量；llama 13b記憶體容量需增長一倍以上

## SSD容量多已搭載1TB/2TB



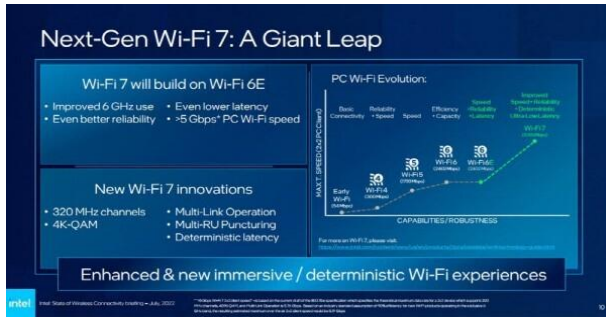
- 用以儲存大量AI學習資料的SSD在AI NB的搭載上多已搭載**1TB/2TB**容量產品
- ALIENWARE m18 R2號稱可支援最高**10TB**的儲存容量，ALIENWARE x16 R2最高也可達**8TB**

資料來源：MIC，2024年4月

# 支援雲與端的協作運算，傳輸及連線裝置效率須提升

當AI PC的價值轉化為雲端與邊緣裝置共同協作時，除了資料的計算外，雲與端間的傳輸及連線穩定性、速率都將影響體驗

## Wi-Fi 7推進速度快



- Wi-Fi 7 被稱為極高傳輸量技術，支援頻寬為Wi-Fi 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，將可幫助未來雲端與地端間的快速協作傳輸
- HP Spectre x360、MSI Summit E16 AI Studio等系列已提供支援Wi-Fi 7機種

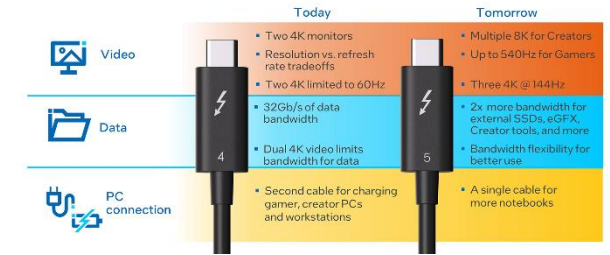
## PCIe Gen5滲透率提升

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

- Intel及AMD新處理器平台全面支援PCIe Gen5

## Thunderbolt 5啟動

New Capabilities Powered by Thunderbolt™ 5



- 2023/9 Intel正式發表Thunderbolt 5，提供每秒80 Gbps雙向頻寬，增強頻寬更高達120Gbps，提高更快且更聰明的傳輸效率

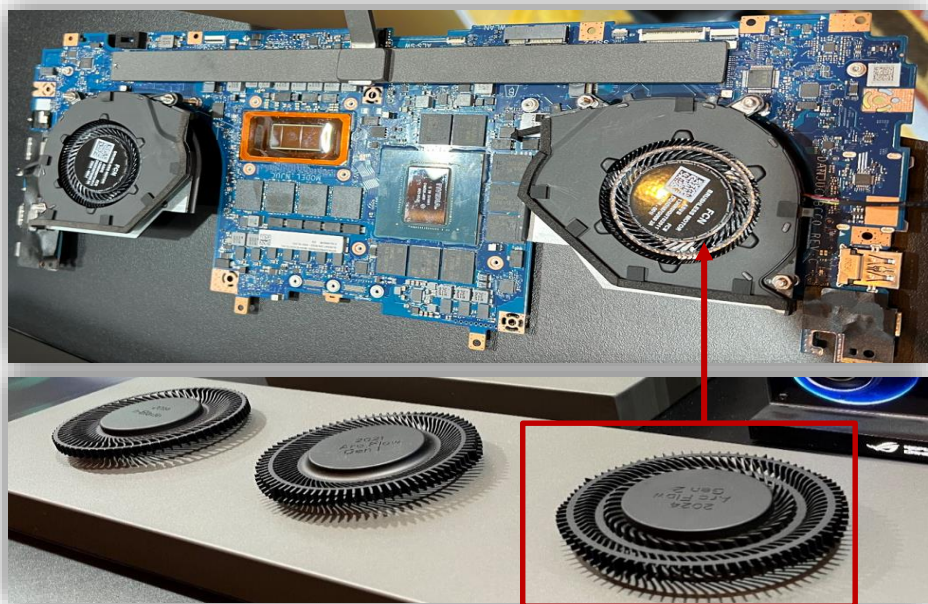
資料來源：MIC · 2024年4月



# AI運算提高PC散熱需求，散熱解決方案推陳出新

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，包含風扇、散熱片、熱導管等散熱模組件的設計重要性將隨之提升

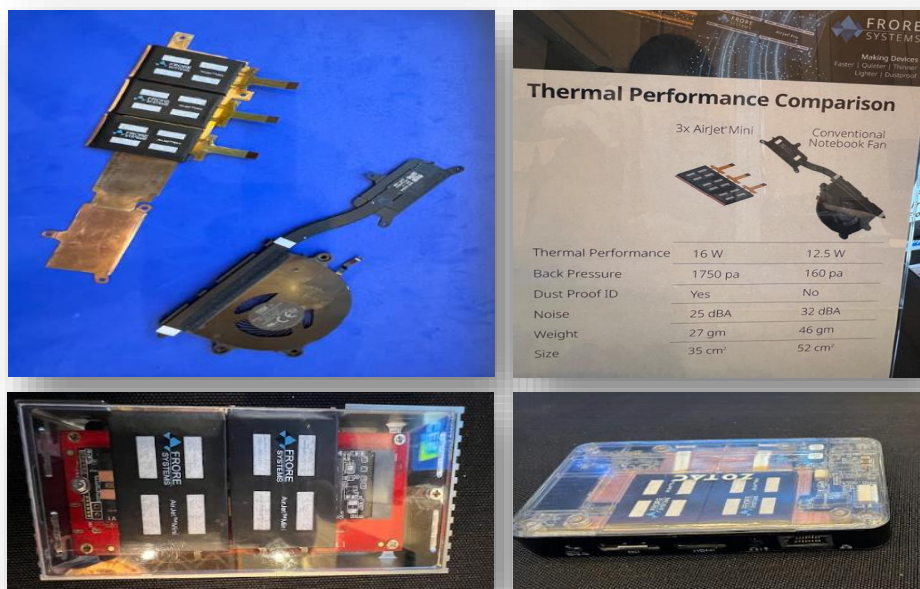
## 散熱風扇新設計



資料來源：華碩、Frore Systems，2024年4月

- 近期多家品牌廠已陸續針對筆電、桌機散熱需求提出多種解決方案，Apple MacBook Air採取金屬導板的被動式散熱，其餘多採用主動式散熱風扇
- 如華碩於2024 CES會場上展出的NB風扇技術，強化風力導流能力，更快帶走熱氣；Frore Systems推出最新一代固態主動式散熱晶片，比傳統風扇體積更輕薄，可用於NB、SSD、記憶體、手機等

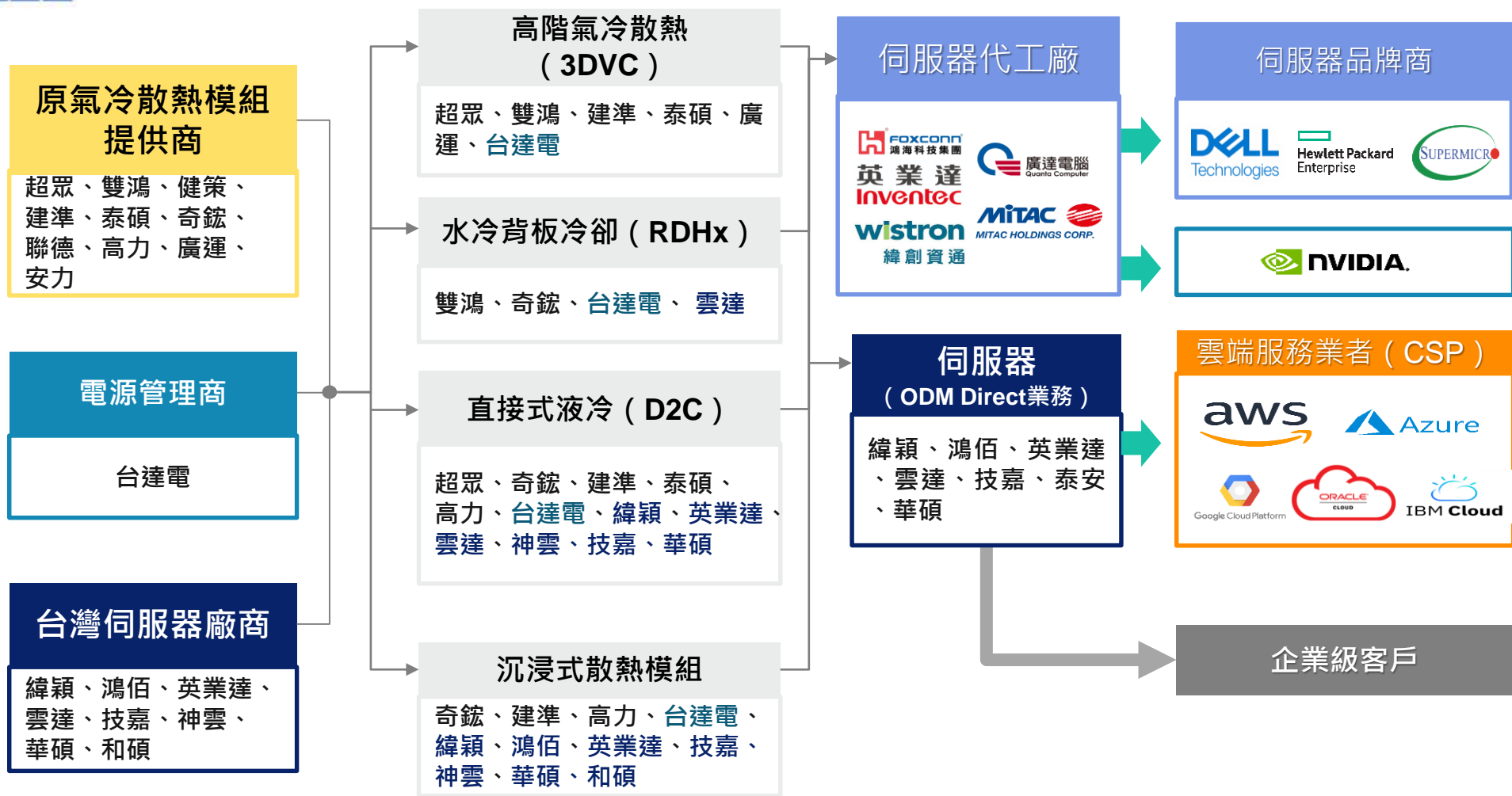
## 主動式散熱晶片新技術



Thermal Performance Comparison		
	3x AirJet Mini	Conventional Notebook Fan
Thermal Performance	16 W	12.5 W
Back Pressure	1750 pa	160 pa
Dust Proof ID	Yes	No
Noise	25 dBA	32 dBA
Weight	27 gm	46 gm
Size	35 cm <sup>2</sup>	52 cm <sup>2</sup>



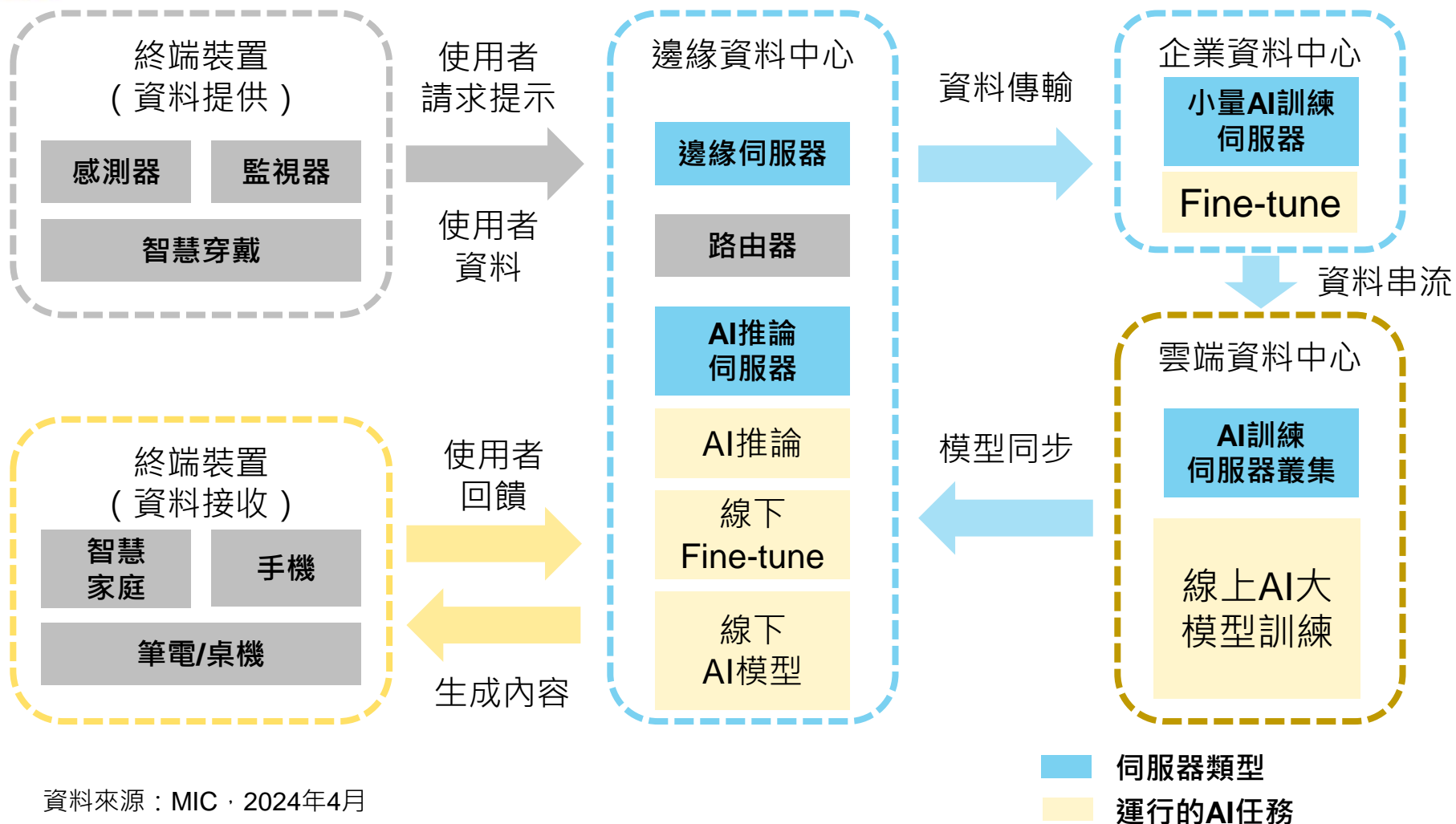
# AI伺服器帶動液冷散熱，吸引台廠進行布局



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

# 生成式AI應用將同步帶動AI訓練與推論伺服器需求



資料來源：MIC · 2024年4月

- 要使生成式AI應用運作最佳化，首先需要雲端資料中心中AI訓練伺服器叢集的AI大模型訓練
- 其次邊緣資料中心邊緣伺服器、AI推論伺服器的Fine-tune與AI推論共同協助，將同步推升產品需求