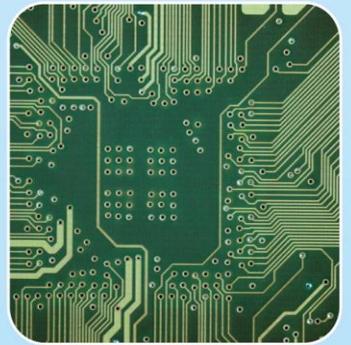


財團  
法人

中技社

# 我國主要算力盤點與產業生態 系之探討

CTCI FOUNDATION



財團法人中技社(CTCI Foundation)創立於 1959 年 10 月 12 日，以「引進科技新知，培育科技人才，協助國內外經濟建設及增進我國生產事業之生產能力」為宗旨。初期著力於石化廠之設計與監造，1979 年將工程業務分拆轉投資成立中鼎工程後，業務轉型朝向裨益產業發展之觸媒研究、污染防治與清潔生產、節能、及環保技術服務與專業諮詢。2006 年本社因應社會環境變遷的需求，在環境與能源業務方面再次轉型為智庫的型態，藉由專題研究、研討會、論壇、座談會等，以及發行相關推廣刊物與科技新知叢書，朝知識創新服務的里程碑邁進，建構資訊交流與政策研議的平台；協助公共政策之規劃研擬，間接促成產業之升級，達成環保節能與經濟繁榮兼籌並顧之目標。

本著創社初衷，為求對我們所處的環境能有更多的貢獻，本社就國內前瞻性與急迫性的能源、環境、產業、科技、社會及經濟等不同議題，邀集國內外專家進行全面的研究探討，為廣為周知，特將各議題研究成果發行專題報告，提供產官學研各界參考。

本專題報告得以出版，首先必須要感謝台東大學鄭憲宗校長擔任議題之召集人，並邀請正修科大張法憲教授、國網中心張朝亮主任及台智雲吳漢章總經理擔任此議題之顧問與諮詢委員。另邀請義守大學管理學院吳有龍院長、義守大學智慧科技學院陳啟彰副院長、國網中心蕭志梔副主任、楊嘉麗正工程師、資策會產研所魏傳虔組長、陳牧風產業分析師、陳奕伶產業分析師、台東大學資訊管理學系謝明哲副教授兼圖書館館長及台東大學綠能與資訊科技學系鍾裕峯助理教授等專家學者共同執筆撰稿，並參與多次工作會議及議題研討等工作，專報依篇章順序分別緒論（包括算力演進、研究背景與研究目的）、算力定義與影響因子、我國算力資源與盤點分析、我國算力系統生態系分析及綠算力及算力應用於永續發展，提供政府相關單位參考，極具參考價值。

**發行人：**潘文炎

**主編：**陳綠蔚、鄭憲宗

**作者：**吳有龍、陳啟彰、蕭志梔、楊嘉麗、魏傳虔、陳牧風、陳奕伶、  
謝明哲、鍾裕峯、陳世海

**執行編輯：**芮嘉瑋、陳世海

**發行者：**財團法人中技社

地址 / 106 臺北市敦化南路二段 97 號 8 樓

電話 / 886-2-2704-9805

傳真 / 886-2-2705-5044

網址 / [www.ctci.org.tw](http://www.ctci.org.tw)

本社專題報告內容已同步發行於網站中，歡迎下載參考

**發行日期：**中華民國 113 年 12 月

**ISBN：**978-626-98882-3-8

# 序

行政院於 2021-2025 年間核定的智慧國家方案，涵蓋六大核心戰略產業與 5+2 產業創新，重點在於提升數位科技能力，並以 AI 人工智能推動與算力需求作為基礎支撐。根據國際調研機構 Fortune Business Insights 預估，在生成式 AI 模型的急劇發展下，其各行各業的應用範圍也快速擴大，包括客服系統與智能推薦系統等領域，呈現爆發性成長，但對具即時性與巨量數據處理的算力需求，帶來極大壓力，促使各國競相投入資源以應對挑戰。算力的核心在於以 PC、伺服器、資料中心等為基礎的硬體整合基礎設施。在此背景下，凸顯研究算力議題之重要性，並從算力資源分佈盤點與產業生態系進行探討，以確保我國在全球 AI 競爭中的關鍵地位。

我國算力資源廣泛分布於政府部門、法人機構、企業與大專院校等單位，其各類型資料中心涵蓋高效能運算 (High-Performance Computing, HPC)、AI 伺服器及通用伺服器，並應用於科學研究、工程計算、人工智能、機器學習及大數據分析等多元需求。在當前 AI 浪潮中，算力類型從雲端延伸至邊緣與終端設備，呈現多樣化發展。本次議題亦界定算力盤點的範疇，確保盤點過程順遂進行。同時，依國際接軌的標準，盤點我國算力規模，提供政府、學術研究機構與業界參考，為算力資源的有效利用與未來發展奠定基礎。

我國半導體與資通訊產業涵蓋 IC 設計、半導體晶片製造、主機板、顯示卡、PCB、記憶體，以及全球 ODM/OEM 組裝等領域，構成了具高度競爭力的算力生態系核心環節。在美中科技對抗及紅色供應鏈限制的背景下，我國算力生態系與既有 PC 產業鏈相結合，並因 AI 浪潮的推動及換機需求的增加，預期將迎來爆發性成長。尤其，美國主要 AI 算力加速晶片製造商，如 NVIDIA 和 AMD，頻繁造訪我國並尋求合作，進一步印證我國作為全球算力製造重鎮的地位。本次討論將聚焦於整體算力生態系的產業現況，深入分析算力在製造、服務與調度等方面的發展，為我國科技產業鏈的進一步優化與國際競爭力提升提供策略參考。

總結而言，本議題旨在從兩大面向切入，分別是算力資源的盤點與算力生態系統的分析，並系統性地闡述算力定義、分類及界定可盤算力範疇，以及國內算力資源之盤點。同時，針對我國算力生態系進行闡述，並分析算力生產與服務之

現況與競爭力，以確保在 AI 科技賽道中，具備充足的計算資源，支撐 AI 技術的發展與應用。此外，綠色算力基礎設施的建置亦是 AI 算力發展之重要趨勢，本議題將一併納入相關研究與討論，為我國在 AI 時代的競爭力與永續發展提供全面支撐。

議題執行期間，為廣納產業看法，研究團隊透過訪廠匯集產業與相關專家意見，包括算力基礎建置與算力應用之廠商，計有台智雲、友崴超算、超微(AMD)、中鼎工程、台灣大哥大、意藍資訊、滿拓科技及 Ai3 人工智能，共計 8 家。中技社在召集人指導下，以平台功能結合產官學研專家學者舉辦一場綜合座談會，就算力基礎建置與產業應用發展，進行研討並提出因應之政策建言，也讓政府相關單位了解國內發展產業業者之困難與心聲，進一步意見溝通與討論，相關會議紀錄可參見附錄。另本議題團隊赴國網中心進行參訪，了解國網中心在系統整合能量與角色扮演，結合產官學研單位共同促進算力產業之發展與願景擘畫。最後，由本社負責彙編各特聘專家意見並提出最終之結論與建議。

本專題報告得以出版，首先必須要感謝台東大學鄭憲宗校長擔任議題之召集人，並邀請正修科大張法憲教授、國網中心張朝亮主任及台智雲吳漢章總經理擔任此議題之顧問與諮詢委員。另邀請義守大學管理學院吳有龍院長、義守大學智慧科技學院陳啟彰副院長、國網中心蕭志楳副主任、楊嘉麗正工程師、資策會產研所魏傳虔組長、陳牧風產業分析師、陳奕伶產業分析師、台東大學資訊管理學系謝明哲副教授兼圖書館館長及台東大學綠能與資訊科技學系鍾裕峯助理教授等專家學者共同執筆撰稿，並參與多次工作會議及議題研討等工作。

另外，感謝時任數發部數位產業署呂正華署長，於座談會中擔任座談貴賓分享「AI 政策驅動」之專題演講，並引發與會來賓踴躍討論，以及國科會科技辦公室派涂家璋副研究員出席座談會並提供國家未來算力發展方向與建言。此外，也特別感謝撥冗參與座談會及受訪之產業代表與其他專家建言與協助，包含中研院陳力俊院士、台智雲李立國策略長、超微(AMD)林建誠資深副總經理、黃偉喬資深協理、友崴超算陳武宏總經理、AI 大聯盟董偉豪祕書長、謝怡涵專案經理、中鼎工程高新事業發展暨策略管理委員會陳振欽執行長、劉國瑞經理、王範經理、Ai3 人工智能張榮貴董事長、意藍資訊楊立偉總經理、台灣大哥大王寶慶副總經理、滿拓科技吳昕益執行長。

財團法人中技社 董事長  
潘文炎  
2024 年 12 月

# 目 錄

序 .....	I
目 錄 .....	III
圖目錄 .....	V
表目錄 .....	VII
執行摘要 .....	1
一、問題及目標 .....	1
二、研究範圍及內容 .....	1
三、研究結論 .....	2
四、改善對策及建言 .....	3
第一章 緒論 .....	7
一、算力即國力之定義 .....	7
二、算力演進與超級電腦、AI 之關聯 .....	8
三、算力研究背景及其重要性 .....	12
四、研究目的 .....	20
五、小結 .....	21
第二章 算力定義與影響因子探討 .....	25
一、算力定義、類別與單位 .....	25
二、算力類型與影響因子 .....	33
三、小結 .....	36
第三章 我國算力資源與盤點分析 .....	39
一、我國算力盤點之統計資料 .....	39
二、我國 HPC 發展現況 .....	46
三、我國 HPC 算力國際競爭分析 .....	49
四、小結 .....	56
第四章 我國算力系統生態系分析 .....	59
一、算力系統生態系定義 .....	59
二、全球算力生態系發展現況 .....	60
三、台灣算力生態系發展現況 .....	69
四、我國算力基礎建設議題 .....	73
五、小結 .....	77
第五章 綠算力及算力應用於永續發展 .....	79
一、算力需求與能耗：數位時代的關鍵挑戰 .....	79
二、從綠色算力看台灣產業發展 .....	88
三、計算可持續性與應用案例探討 .....	97

四、小結 .....	111
第六章 結論與建議 .....	119
一、結論 .....	119
二、建議 .....	122
附錄 .....	125
一、綜合座談會-會議紀錄 .....	125
二、算力盤點格式範例 .....	133
三、輝達 GPU 模組 .....	134
四、專有名詞對照表 .....	138

## 圖目錄

圖 1.1、二五珠算盤.....	8
圖 1.2、帕斯卡計算機.....	9
圖 1.3、ENIAC 計算機.....	9
圖 1.4、Control Data Corporation CDC 6600.....	10
圖 1.5、台灣 AI 行動計畫 2.0.....	16
圖 1.6、數位產業跨域軟體基盤暨數位服務躍升計畫.....	17
圖 2.1、雲端、邊緣、終端設備之間的關係.....	27
圖 2.2、以硬體型態進行算力分類.....	29
圖 3.1、2011-2024 年台灣進入 Top 500 之建置單位與台數.....	49
圖 3.2、台灣歷年進入全球超級電腦 Top 500 排行榜之電腦台數.....	50
圖 3.3、2011-2024 年全球超級電腦 Top 500 排行榜之台灣 HPC 算力總量.....	51
圖 3.4、2011-2024 年全球超級電腦 Top 500 排行榜之台灣 HPC 總核心數.....	51
圖 3.5、台灣與韓國 HPC 算力比較.....	55
圖 4.1、全球算力系統生態系範疇.....	59
圖 4.2、2019~2028 年全球伺服器市場預測.....	60
圖 4.3、全球伺服器供應鏈主要業者.....	61
圖 4.4、2023 年全球前十大公有雲服務供應商.....	64
圖 4.5、生成式 AI 產業效益影響雲端軟硬體生態系.....	66
圖 4.6、2024 年全球主要國家 HPC 市場營收.....	68
圖 4.7、2024 年主要國家 HPC 產業應用分布.....	69
圖 4.8、台灣伺服器/AI 伺服器代工占全球比重.....	70
圖 4.9、2022~2027 年台灣伺服器產值預測.....	71
圖 5.1、傳統資料中心、專用 AI 資料中心和加密貨密的預估電力需求.....	80
圖 5.2、(a)AI 伺服器功率變化情形；(b)傳統伺服器功率變化情形.....	81
圖 5.3、綠能科技產業創新推動方案之願景與目標.....	87
圖 5.4、直接式與沉浸式液冷比較.....	94
圖 5.5、伺服器帶動液冷散熱台廠布局情形.....	95
圖 5.6、中華電信對不同客戶採取不同的冷卻方式.....	95
圖 5.7、中華電信 IDC 灰電與綠電使用情形.....	96
圖 5.8、台灣應積極推動儲能產業，以最大化分佈式電網之效益.....	97
圖 5.9、聯合國於 2015 年發布 17 項永續發展目標 (SDGs).....	98
圖 5.10、2024 年 SDGs 17 項目標進展評估 (百分比).....	99
圖 5.11、不同情境下預測的全球暖化程度(相對於《巴黎協定》溫度目標).....	100
圖 5.12、位於芬蘭卡亞尼市 CSC 資料中心是 LUMI 超級電腦的綠色之家.....	105
圖 5.13、Green Mountain 位於挪威埃內巴克城鎮的 OSL-Enebakk 資料中心.....	106

圖 5.14、2024 年 IMD 世界競爭力排名（前 10 名） .....	108
圖 5.15、東芝以 VPP 平台整合分布在多個家庭中的大量儲能電池.....	110

## 表目錄

表 2.1、伺服器之算力類型（以硬體分類） .....	30
表 2.2、運算單位表示法 .....	31
表 2.3、各種精度組合和其適合應用的場合 .....	32
表 3.1、台灣算力統計總覽 .....	44
表 3.2、台灣 1987-2024 年 HPC 算力發展里程碑 .....	47
表 3.3、台灣 HPC 主機建置單位.....	48
表 3.4、2024 年 6 月台灣 HPC 算力 top 500 主機清單 .....	52
表 3.5、2024 年 6 月 Top 500 排行榜各國算力排名 .....	53
表 3.6、2024 年 6 月全球超級電腦 Top 500 各國上榜台數 .....	54
表 3.7、韓國超高性能電腦法 .....	57
表 5.1、AI 伺服器與一般伺服器所需電力比較表 .....	81
表 5.2、2024 年 6 月 Green500 排名.....	83
表 5.3、台灣在 2024 年 6 月 Green500 排名.....	84
表 5.4、資料中心重要指標參考值 .....	90
表 5.5、美國 2022 晶片法案清單 .....	91
表 5.6、美國 2024 年關鍵與新興技術清單 .....	91
表 5.7、2024 年 IMD 智慧城市排名（前 20 名） .....	107



## 執行摘要

### 一、問題及目標

我國政府於 2021 至 2025 年間推動「智慧國家方案」，以六大核心戰略產業及 5+2 產業創新計畫為主軸，致力於強化數位科技能力，並以 AI 人工智能和算力需求為基礎驅動力，推動國家數位轉型與產業升級。隨著生成式 AI 技術的快速發展，文字、圖像、影音生成技術日益成熟，應用範疇涵蓋自動駕駛、醫療影像分析、自然語言處理及智能推薦系統等，促使服務業與製造業對高效算力的需求大幅提升。根據國際調研機構 Fortune Business Insights 的預測，2032 年全球 AI 市場規模將達 2.7 兆美元，年複合成長率達 20.4%。故各國正加大算力基礎設施投資，以支撐並推動數位轉型與智慧國家。另外，我國因擁有堅實的半導體及資通訊技術基礎，在全球算力產業競爭中具先天優勢，可加以探討並了解整體算力產業生態系之全貌。此外，政府現正透過推動「五大信賴產業推動方案」，積極迎接 AI 浪潮，投資 AI 數位產業，並透過新建智慧節能資料中心，爭取國際合作等擴大算力及引進低耗能方案，促進各產業智慧化轉型與國際競爭力。

本議題聚焦兩大面向：第一，盤點我國現有算力資源，包括政府部門、法人機構、企業及大專院校所擁有的伺服器算力及高效能運算（HPC）算力，並參考與揭露 TOP500 全球算力排名資訊，闡述我國 HPC 之發展現況與國際競爭分析；第二，分析我國算力產業生態系，涵蓋 IC 設計、晶片製造、主機板、記憶體等核心產業鏈。隨 AI 浪潮帶動雲端、邊緣與終端算力需求激增，以及換機潮與國際技術競爭加劇，台灣科技產業鏈的競爭力將進一步凸顯。此外，國際大廠如輝達（NVIDIA）與超微（AMD）等加強與台廠的合作，顯示台灣在全球算力製造重鎮中的地位。本議題旨在闡述算力演進與定義、界定可盤算力範疇、我國發展算力現況、算力產業生態系，包括算力生產與算力服務，供政府及產學研界參考。

### 二、研究範圍及內容

本報告第一章概述整體算力演進與算力背景，包括全球 AI 浪潮帶動算力發展、政府算力產業政策，以及台廠在算力產業之扮演重要性角色；第二章為算力定義與影響因子，闡述有關算力之基礎知識；第三章是我國算力資源與盤點分析，主要分為兩部份，一為算力盤點之統計與揭露，另一是我國 HPC 發展與國際排名分析；第四章是闡述我國算力系統生態系，其中包括算力生產與算力服務，並針對算力議題進行研析；第五章是綠算力及算力應用於永續發展，主要提及有關算力需求與能耗、結合綠色能源之綠算力與永續發展趨勢，以及闡述國外實際應用綠算力之案例；第六章結論與建議。

本社研究同仁除與召集人、研究團隊共同擬定本專題研究大綱與架構外，並負責訪談廠商、了解算力業者之運營需求及意見。

### 三、研究結論

生成式 AI 的興起與「主權商機」，加速伺服器與全球高效能運算（HPC）需求擴張，推動算力基礎建設並帶動整體算力生態系發展，為台灣產業鏈創造前所未有的機遇。在這股算力浪潮中，台廠憑藉完整的產業鏈布局，展現出關鍵優勢與競爭力。從上游的 AI 專用晶片設計與製造，到中游的關鍵零組件生產，再到下游的伺服器組裝與整合，台廠在全球扮演不可或缺的角色。上游部分，台積電以先進製程提供高效能、低功耗晶片支撐 AI 運算，聯發科等 IC 設計企業則專注於 AI 晶片研發，推出創新解決方案；中游部分，欣興、台達電等廠商在 PCB、散熱模組等領域具備技術優勢，為 AI 伺服器穩定運作提供保障；下游部分，廣達、鴻海等台廠掌握 AI 伺服器組裝核心技術，全球代工市場占比高達九成，能為企業及數據中心提供多元化客製化之運算解決方案。生成式 AI 帶動的快速發展正在推進我國半導體與資通訊產業的全面升級，讓台廠在全球算力建設與生態系中占據領先地位，進一步鞏固我國作為全球科技產業重鎮的關鍵角色。

政府在生成式 AI 浪潮中，持續推動 AI 政策，積極打造 AI 健康生態圈，以促進產業升級與發展。例如目前所推動《AI 行動計畫 2.0》，係以人才培育、基礎建設與生態系形成為核心策略。其中，人才培育是關鍵，透過校園合作、專業訓練與創新實戰計畫，打造跨領域的 AI 專業人才庫，為產業提供充足的創新動力。同時，在基礎建設方面，政府強化數位基礎設施，推動運算平台與數據整合環境的完善，以支持 AI 技術應用的快速落地。此外，通過與產業和學界的協力，深化技術開發與商品化，促進市場導入，建立完整的 AI 生態圈。另外，數發部與經濟部協同推動《AI GO》、《AI Hub》、《Fast AI》及《AI 領航》等計畫，更進一步加強數位建設、人才培訓與市場推廣，實現「產業 AI 化與 AI 產業化」的雙目標，全面推進台灣 AI 技術與產業在全球的影響力，並為數位經濟發展奠定長遠基石。此外，2024 年 7 月國科會亦透過《AI 基本法》，聚焦技術倫理、數據隱私與責任分配，建立 AI 運用之框架。

本議題透過有系統性地定義算力與類別，並界定可盤算力資源，為我國未來算力發展規劃提供基石。有關算力定義係以每秒浮點運算次數（FLOPS）為衡量標準，並考量算力數級規模較大，故單位前面會加上 Giga（ $10^9$ ）、Tera（ $10^{12}$ ）或 Peta（ $10^{15}$ ）單位予以表達，而應用範圍涵蓋 AI 模型訓練、氣象模擬及醫療分析等高效能領域。另於算力分類上，可依使用場景、應用範圍或硬體型態來進行分類，其中使用場景可分為雲端算力、邊緣算力與終端算力；而應用範圍可分為高效能運算（HPC）、AI 伺服器與通用伺服器；最後，依硬體型態分類，即以伺服器硬體進行分類，包括 HPC、大型雲端服務算力、企業型伺服器算力、邊緣算力等四類。此外，我們採硬體型態之分類方式並結合可盤資源，即政府機構、大型醫院及學術單位三類資源分布之 HPC，進行實際盤點。盤點至 2024 年 6 月，已揭露台灣 88PFLOPS 的理論算力，93%集中於政府機構，凸顯政府在科研與產業

推動中的核心地位。另外，我國更有六台設備進入 TOP500 排行，包括 Taipei-1 與其他五台國家級設施，世界排名第 17、亞洲第 4，顯示算力在國際競爭中的顯著影響力。影響算力的關鍵因素包括處理器數量、記憶體容量、存儲性能、網路架構及演算法優化，而這些元素皆為未來規劃的重要參考。

透過此次系統性的資源盤點與數據分析，不僅為台灣算力現況提供全貌與趨勢，也為後續目標設定及規模擴展奠定基礎，惟限制條件包括調查單位揭露不夠透明、企業考量隱私需求而無參加 TOP500 之算力資訊，以及算力盤點或統計時間落差狀況。我們期待在智慧國家與數位經濟願景下，未來能再透過提升算力規模與國際競爭力，我國將在全球科技創新舞台占據更關鍵的位置，並藉此機會向公眾科普算力的意涵與應用價值，深化全民對科技創新的認識與參與。

最後，隨著 AI 算力基礎建設快速發展，我國國網中心及相關業者紛紛投入資料中心之建置，但伴隨而來的是算力出口競爭與高耗能的雙重挑戰。國際市場上，全球三大雲端服務商及中國大陸低價雲服務提供者正激烈競爭，對我國算力的服務及出口構成強大壓力。而資料中心亦伴隨高能耗問題，成為各國關注焦點，其晶片技術的迭代成為解決高耗能的關鍵，例如 NVIDIA 的最新加速運算晶片 Blackwell 顯著降低 AI 大型語言模型的訓練能耗，功耗從 1000GWh 降至僅 3GWh，減少幅度達 333 倍，展現未來節能趨勢。同時，冷卻方式亦從氣冷轉向水冷與液冷，此亦成未來散熱技術發展之新產業趨勢。另外，資料中心所產生的廢熱也為能源再利用創造了新機遇，透過廢熱回收技術，將資料中心產生的熱能轉化為區域供暖、溫室加熱或水產養殖用途，不僅減少能源浪費，亦符合永續發展的需求。此外，整廠輸出的模式逐漸受到青睞，結合土木設計、廠房建設、伺服器部署及散熱系統整合，不僅可優化資料中心運營效率，亦能實現廢熱回收的經濟效益。這種整合式的基礎建設方式，為降低運營成本與環境永續發展提供了雙贏契機，同時推動產業升級，確保我國在全球 AI 算力基礎建設競爭中的優勢地位。

#### 四、改善對策及建言

我國算力產業發展仍面臨許多技術與市場之挑戰，政府已積極將算力產業列入年度施政方向，為促進國內算力產業發展，本專題報告提出下列建議方向：

##### **(一) 提出 AI 微調模型之方法與步驟，結合算力+模型+軟體解決方案，降低「燒算力」問題，協助百工百業導入生成式 AI**

生成式 AI 的蓬勃發展在全球引發熱潮，我國中小企業也意識到其潛力並躍躍欲試，然而在導入過程中卻面臨不少挑戰。這些企業往往缺乏對生成式 AI 的算力需求、晶片選擇及應用方式的全面了解，只知其強大卻不知如何有效入手。同時，全球算力晶片供應短缺導致高昂的算力成本，成為中小企業的主要障礙之一；另生成結果品質的不穩定性，更使算力消耗和成本不確定性加劇。為有效解決這些問題，導入生成式 AI 的明確方法與步驟尤為重要，政府應當提供資源與

指引，例如針對企業實際需求進行模型微調（Fine-tuning）或引入檢索增強生成（RAG）方法與平台，以提高應用的精準度與效率。然而，僅提供導入指南仍不足以突破技術門檻，算力服務業者亦可提出更全面的整合性解決方案，即結合算力、模型與應用軟體的三位一體之解決方案，降低「燒算力」本質問題（即需重複嘗試的實驗科學與結果之不確定性），以協助百工百業導入 AI。

此外，建議建置算力資源共享平台，為中小企業提供可負擔的雲端算力；同時推廣模組化算法工具，幫助企業快速部署並適應生成式 AI 模型；更應整合針對不同行業的應用軟體，將導入難度降至最低。透過這種整合性策略，能大幅降低生成式 AI 應用門檻，深化其在各產業中的實際應用價值，不僅推動技術普及與產業升級，更能打造台灣在 AI 時代的新競爭優勢。

## **(二) 研擬建置高效能運算之專法，計畫性投入 AI 高運算資料中心資源，強化 AI 算力基礎建設，維持國家競爭力**

隨著算力產業的快速發展，硬體迭代速度加快卻帶來高折舊成本的嚴峻挑戰，也就是「算力老化」問題。隨科技持續進步，許多高效能硬體在短短數年內便可能被新技術取代，這不僅加重企業財務負擔，還阻礙了 AI 高運算資料中心的建設步伐，削弱整體產業競爭力。為破解此困境，建議比照南韓於 2020 年制定的《國家超高效能電腦應用及育成法》，透過專法確保國家算力基礎建設的穩定成長。南韓依此法所律定，中央行政機關每年依據基本計畫推動高效能運算（HPC）發展，系統性支持算力產業，並以此提升國家在全球科技競賽中的地位與影響力。我國可借鏡此成功經驗，制定類似專法。另藉由該產業初期發展階段之政策輔導，包括財政補貼、稅務減免及產業聯盟等手段，降低企業升級硬體的成成本壓力，以強化民間算力基礎之建置，為 AI 相關產業提供穩固的運算支撐。這不僅能促進產業升級，更能提升國家科技實力，為經濟長期成長注入源源不絕的動能。

## **(三) 建議政府電力成長估算，納入 AI/資料中心電力需求，提出綠算力計畫，兼顧數位轉型與永續發展**

隨著人工智慧（AI）產業的快速發展，算力需求與資料中心建設量不斷攀升，電力供應成為不可忽視的核心議題。根據國際能源署（IEA）的數據，2023 年僅輝達（NVIDIA）出貨的 AI 伺服器約 10 萬台，其年耗電量已達 73 億度。考量 AI 產業的指數型成長趨勢，預期至 2026 年，其用電需求將達 2023 年的 10 倍，凸顯 AI 技術發展對能源供應帶來的壓力。另依據 2022 年國家發展委員會的資料，我國每年用電需求平均增長 2.6%，因此，建議電力規劃需依據 AI 及算力產業的實際發展狀況進行滾動式調整，確保能源供應能有效支撐產業發展。此外，電力需求應結合綠色再生能源發展趨勢，將綠能應用融入資料中心的供電方案，不僅有助於緩解能源壓力，更能推動數位轉型與永續經營的雙重目標，助力我國在 AI 浪潮中建立競爭優勢並實現永續發展願景。

**(四) 提升 TAIDE 主權 AI 計畫，建立符合國人特有文化用字及用語之多模態模型，以拓展未來百工百業應用需求**

作為我國主權 AI 發展的重要基石，TAIDE 模型依據國人特有的文化構建，展現台灣在人工智慧領域的技術實力與產業應用潛力。該模型的公開釋放，已為國內多領域帶來創新應用。另考量面對台灣以中小企業為主的經濟結構，以及服務業吸納近六成就業人口的現況，生成式 AI 於客服系統的應用成為突破口，尤其在提升服務效率與客製化溝通層面，展現出顯著價值。為進一步推動 AI 技術普及與創新，建議政府加強資源投入，研發結合用字、用語與生成圖片的多模態大語言模型，例如混合中、英、台、客、原住民用字用語，強化符合國人語言與文化特性的應用技術，支援中小企業進行數位轉型，進而提升台灣 AI 競爭力與產業創新能力，為未來創造更多可能性。

**(五) 全面提升國家公開資料之品質，以擴展算力應用與發展**

在人工智慧的廣泛應用中，除了算力基礎的支持外，高品質的大數據同樣是不可或缺的核心要素。儘管數據的規模十分重要，但唯有提升資料的品質，才能充分發揮生成式 AI 的應用潛力，進而開發出具備高度效能的 AI 應用軟體。因此，數據品質的整備成為算力價值發揮的關鍵所在，透過公開資料的優化，不僅能促進技術創新，更能提升 AI 在各行各業中的應用效能與精準度。因此，提升公開資料品質並加強產官學研的合作，促進技術整合與創新，將為算力應用的發展提供更加穩固的基礎，推動台灣 AI 產業邁向更高峰。

財團法人中技社



## 第一章 緒論

### 一、算力即國力之定義

「算力」(Computing Power)是指運算能力，特別是在電腦或者超級電腦上處理和執行複雜計算任務的能力。通常用來描述一個系統或者設備能夠在單位時間內進行多少計算工作，是衡量計算效能和速度的重要指標之一。國際間以「每秒浮點運算次數」(Floating-Point Operations Per Second, FLOPS)作為超級電腦速度單位，以1個peta FLOPS來說，等於每秒可進行 $10^{15}$ 次方、即1千兆次浮點運算。此外，算力也由處理器速度、處理器數量、所採用演算法的效率等因素決定。

簡單來說，算力能讓電腦快速的處理數據，提升人工智慧 (Artificial Intelligence, AI) 的效率和智慧程度。因此，廣達董事長林百里在出席2023台灣人工智慧年會時，就提出「AI算力即國力」的觀點，並指出市場對算力的需求，每三個月就增加一倍，速度則比兩年翻一倍的半導體摩爾定律還快。和碩董事長童子賢也提及，過去常說「算力即國力」，即AI的算力代表一個國家的實力，但也可以反過來說「國力即算力」，一旦國家經濟狀況不佳，根本無法建置動輒要價數十億美元的算力機房。雖然如此，算力的建置也需考量資訊產品變化快速，因為超級電腦一上線即進入攤提，此均為成本費用，如能透過擴大算力之應用服務，方能具體彰顯所投入算力建置之效益，並轉換成國力之一部份。故算力規模應要跟該國算力之使用人口總數及數位經濟條件，僅用各國算力規模大小作為國力之評比，顯有失真與偏誤。

算力成為當今熱門議題的背後原因可以從多個層面來看，這與當前科技發展的迅猛步伐和日益增長的計算需求密切相關。

首先，人工智慧和大數據的爆發式增長極大推動了對算力的需求。現代的AI技術，尤其是像自然語言模型、深度學習 (Deep Learning) 等，依賴於龐大的數據集和複雜的算法訓練，這需要極強的計算能力來完成。從自然語言處理、圖像識別到生成式AI等應用，無不需要大量算力來進行模型訓練和推理。此外，隨著數據科學的發展，各行各業都依賴於海量數據的處理和分析，從而為企業決策和創新提供支持，這進一步加劇了算力的需求。

其次，雲端技術的普及改變了企業獲取算力的方式。過去，企業通常需要購置自己的硬體設備來應對運算需求，而如今透過雲端服務，企業可以按需求擴展計算資源，這使得算力作為一種服務的模式變得越來越普及。這種靈活性降低了企業的技術門檻，促進了更多數位化和自動化轉型的實施，特別是在那些依賴即時數據分析和自動決策的領域中。

同時，全球技術競賽的壓力也推動算力成為關鍵議題之一。各國在超級電腦

上的競爭，已經不僅僅關乎技術領先，更涉及國家安全與經濟競爭力。美國、中國、歐盟與亞洲等經濟體都加大了對算力基礎設施的投資，俾利在全球科技競爭中保持優勢，這使得算力不僅成為商業需求，也上升到戰略資源的層面。在這樣的背景下，算力已經成為當代科技發展的基石。從雲端運算、AI 技術的普及到全球競爭，各種技術需求的增加，使得算力不再僅僅是計算機科學的話題，而是關乎全球科技創新和經濟發展的核心要素。

## 二、 算力演進與超級電腦、AI 之關聯

### (一) 電腦的演進

人類首先發明器具來幫助計算則可追溯至西元前 500 年中國的算盤，而一般所看到的算盤為「二五珠算盤」，為明代後期開始出現。採用上二珠、下五珠的形式，上一珠當五，下一珠當一。當時許多中藥店、南北行等傳統行業使用「一斤十六兩」的度量衡，因為是十六進制，又被稱為斤兩算盤，詳如圖 1.1。



資料來源：文化部典藏網

圖 1.1、二五珠算盤

然而，真正意義上的電腦發展始於 17 世紀。第 1 部機械式計算機由法國數學家帕斯卡（Blaise Pascal）在 1642 年利用齒輪推動的原理發明。其具有進位的功能，可以撥動轉盤做加 1~9 的數字，能進行加法和減法運算。稍後，德國哲學家和數學家萊布尼茨（Gottfried Wilhelm Leibniz）於 1673 年設計出能進行乘法和除法的機械計算器，如圖 1.2。

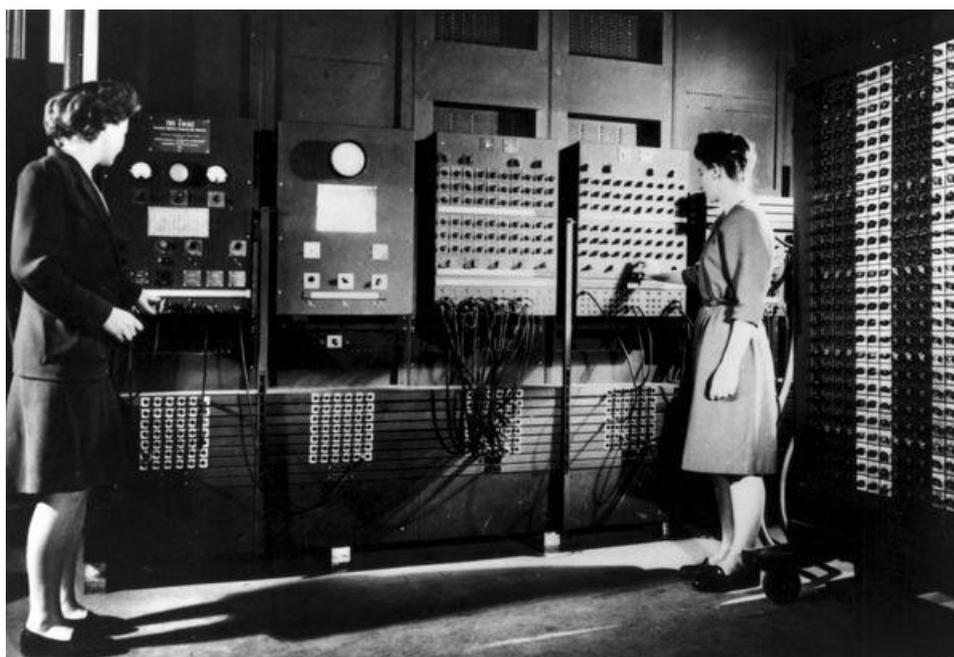


資料來源：維基百科

圖 1.2、帕斯卡計算機

20 世紀初，隨著電氣技術的發展，計算設備開始由機械轉向電子。1920 年代，美國的霍華德·艾肯 (Howard Aiken) 設計出哈佛標記 1 號 (Harvard Mark I)，這是一台基於繼電器的機械計算機。與此同時，科學家們開始探索使用真空管來構建更快、更可靠的計算設備。

1946 年，美國賓夕法尼亞大學的 ENIAC (Electronic Numerical Integrator and Computer) 問世。ENIAC 是世界上第一台通用電子數位計算機，能夠進行多種數學運算。ENIAC 由超過 17,000 個真空管構成，佔地巨大，運算速度遠超當時的任何計算設備。然而，ENIAC 的操作仍然相當複雜，因為需要手動配置電纜和開關來改變程序，如圖 1.3。



資料來源：Computer History Museum

圖 1.3、ENIAC 計算機

1950 年代，電腦的發展迎來了革命性的變革：電晶體的發明。貝爾實驗室的約翰·巴丁（John Bardeen）、威廉·肖克利（William Shockley）和沃爾特·布拉頓（Walter Brattain）於 1947 年發明了電晶體，這一發明徹底改變了電子設備。電晶體比真空管更小、更便宜，並且更加節能 and 可靠，這使得電腦變得更加實用。

到了 1960 年代，積體電路（Integrated Circuit, IC）的發明進一步推動了電腦的發展。積體電路將數百甚至數千個電晶體集成到一個小小的晶片上，大大減少了電腦的體積和成本。這一技術的突破促使超級電腦與微型電腦的誕生。尤其在超級電腦演進上，是現代科技進步的重要縮影，其能處理的龐大、複雜計算問題，體現了計算能力的飛躍和技術突破。

## （二）超級電腦的出現

1964 年美國的西蒙·克雷（Seymour Cray）設計出第一台被認為是超級電腦的 CDC 6600。這台機器具有 10 個處理單元，運算速度達每秒數百萬次運算（FLOPS），遠超過當時其他計算機。其設計理念是針對特定問題提供極高的性能，為後來的超級電腦奠定了基礎，如圖 1.4。



資料來源：Computer History Museum

**圖 1.4、Control Data Corporation CDC 6600**

1970 年代至 1980 年代，克雷公司繼續引領超級電腦的發展，推出了更快的 Cray-1 和 Cray-2。這些系統使用向量處理技術，能夠同時處理多個數據流，顯著提升了計算速度。然而，超級電腦的設計仍然受到物理限制，隨著計算需求的增加，傳統架構難以滿足日益增長的需求。因此，在 1990 年代，超級電腦的設計開始轉向大規模平行處理（Massively Parallel Processing, MPP），這使得數千甚至數萬個處理器可以同時協作完成計算任務。

進入 21 世紀，超級電腦的性能增長進入了新的階段。利用數萬個甚至數百

萬個處理器核心的分佈式計算架構，超級電腦能夠在短時間內處理海量數據，而且不僅僅追求硬體性能的提升，還強調能效比和更靈活的計算架構，以應對 AI 訓練、氣候模擬、藥物設計等複雜問題的需求。超級電腦的演進不僅反映了科技實力的提升，也推動了科學研究和工業創新的無限可能。

目前市場主流的超級電腦仍是由超大型積體電路製程製造的晶片所組成，但其容易受到散熱、製程與傳輸速度限制，運算速度還是有所極限。因此，各國嘗試研發其他媒介來做運算，如量子電腦、光學電腦、DNA 電腦等。由於不用電子做運算，具有耗熱少與大量平行處理運算能力，大幅提升了電腦的運算速度。未來，隨著技術的進步，量子電腦等可能在金融、科學模擬和人工智慧等領域發揮關鍵作用。但要普及至日常應用，尚需數十年的持續研究與投資。儘管如此，其有潛力徹底改變我們處理數據和計算的方式，推動新一輪之科技革命。

### (三) AI 與高效能的算力需求

超級電腦在當今 AI 和高效能運算 (High Performance Computing, HPC) 領域中扮演著至關重要的角色，其強大算力為大規模數據處理、科學模擬和人工智慧模型訓練提供了必要的基礎。而超級電腦正是專為此設計的，通常擁有數以萬計的處理器和大量記憶體，能夠在極短的時間內執行海量計算任務。

國家高速網路與計算中心主任張朝亮指出，超級電腦的應用範圍非常廣泛。在民生方面，它可以提升氣象預報的準確度，並進行長期的氣候測報分析，幫助受影響地區做出周全的應對。例如，在 COVID-19 疫情期間，國際間使用超級電腦來模擬咳嗽產生的氣流、飛沫擴散路徑及其影響範圍，從而制定安全的社交距離。此外，超級電腦還能加速新型藥物和疫苗的開發，進行病理影像分析和精準診斷，甚至在晶片和新材料的開發中發揮重要作用。美國國家航空暨太空總署 (NASA) 的許多航太計畫也依賴超級電腦的強大運算能力，包括地球和月球繞行軌道的設計、火箭設計、載具返回地球的落點預測，以及太陽研究等。

對於生成式 AI 來說，超級電腦提供的算力至關重要，尤其是在訓練和調整模型時。這些模型通常需要處理極其龐大的數據集，並且其內部結構可能擁有數十億甚至數百億個參數。隨著模型複雜度的增加，特別是大規模語言模型 (如 GPT 系列)，對算力的需求也呈指數級增長，訓練它們的過程需要數天甚至數週的時間。如果單靠普通的計算機，這樣的運算可能需要數月甚至數年才能完成。而超級電腦可以通過並行處理的方式，將這些運算任務分配到成千上萬個處理器上，同時進行計算，極大地縮短了訓練時間。

此外，隨著 AI 應用場景的多樣化，算力需求不僅僅局限於訓練階段，還延伸到推理階段。當 AI 模型部署到生產環境中，面對即時的數據處理需求，推理過程中仍需大量的算力來保證反應速度和準確性。例如語音識別、影像識別、自然語言處理等應用都要求即時反應，而這些應用的背後都依賴強大的運算資源來

支持模型的運行。AI 模型的性能提升需要更大規模的數據和更強大的算力來支撐，而隨著摩爾定律逐漸放緩，如何有效提升算力並優化資源利用成為業界研究的重點。

因此，超級電腦不僅僅是 AI 領域的一種工具，更是支持 AI 技術發展的核心算力基礎。隨著 AI 模型規模的持續擴大和應用場景的多元化，超級電腦在未來 AI 技術發展中的作用將會越來越重要。

### 三、 算力研究背景及其重要性

#### (一) 全球 AI 浪潮帶動台灣算力發展

隨著全球數位轉型的加速，人工智慧、大數據、區塊鏈等新興技術需求大幅增長，對高效算力的需求日益迫切。台灣具有豐富的半導體及資通訊 (ICT) 科技產業基礎，應當探討其在全球算力市場中的角色及發展潛力。尤其在 AI 淘金熱中，台灣能提供能做出高品質的半導體、計算主機，就像淘金必備的鏟子一般，非常重要，也就是具備了天然的優勢來擴展算力領域。

首先，台灣擁有世界領先的晶圓代工產業，尤其是台積電等企業在生產高性能運算晶片方面具備無可比擬的技術優勢。這些晶片不僅是 AI 運算的核心元件，還是全球最先進 AI 應用所依賴的關鍵硬體。隨著 AI 技術的發展，對於更高效能、更低功耗的算力需求也越來越高，這促使台灣的半導體企業不斷創新，開發出能應對 AI 計算需求的新型晶片，如專為 AI 訓練與推論設計的加速器晶片 (AI Accelerator Chips)。

此外，全球對雲端運算和數據中心需求的增長，也推動了台灣算力基礎設施的進一步發展。台灣的企業開始大力投資建設超大型數據中心，這些數據中心不僅為本地的 AI 應用提供算力支持，還透過國際佈局服務全球市場。國際科技巨頭如 Google、Microsoft 等公司也將台灣視為重要的算力基地，在此設立研發中心或雲端運算設施，進一步提升台灣在全球算力市場的地位。

研調機構 Fortune Business Insights 預估，全球圖形處理器 (Graphics Processing Unit, GPU) 算力服務的產值，2024 年起至 2032 年，年複合成長率高達 35.8%，也就是說，這是一個每 3 年產值就成長 1 倍的新兆元產業。Google 則是在 2024 年 7 月公布 1 份關於台灣 AI 產業發展與治理的白皮書，指出到 2030 年，台灣企業運用 AI 帶來的經濟效益將高達 3.2 兆，將超過台灣目前 GDP 的 13%。

#### (二) 台灣在全球半導體產業的領先地位

台灣擁有全球領先的半導體產業，而半導體是現代計算能力的基礎。因此，探討台灣半導體業者如何參與高效算力發展，以及其在全球算力市場中扮演的角色，對於進一步提升台灣產業競爭力至關重要。

尤其是當全球投資者將目光集中於輝達 (NVIDIA)，關注 AI 超級晶片的量

產和出貨時程時，台灣的供應鏈成為背後的關鍵推手。從 IC 設計、晶圓代工、電源、散熱、印刷電路板（Printed Circuit Board, PCB）到伺服器代工，台灣都擁有關鍵優勢，讓台灣成為名副其實的「AI 軍火庫」。蘇黎世日報記者科林（Martin Koelling）在 2024 年 6 月以「半導體生產多元化無解：沒有台灣，就沒有人工智慧」為題，專文報導了台灣科技產業與地緣政治如何影響全球。報導指出，台灣在 AI 領域的優勢短期內難以被取代，因此保護台灣的穩定和自主對全球科技產業的未來至關重要。

另一方面，隨著 GPU 超級算力的興起，台灣在資通訊領域的優勢使整個產業鏈具備強大的競爭力。輝達執行長黃仁勳讚揚台灣為全球 AI 供應鏈的核心，不僅因為台積電在全球晶片製造領域獨占鰲頭，其先進的製程技術和高效的生產能力使台灣在 AI 算力方面具有顯著優勢。此外，台灣擁有全球最完整且精密的半導體產業鏈，這不僅證明了台灣在半導體產業中的領先地位，也突顯了其在 AI 領域的強大實力。

黃仁勳指出，輝達每年營收增長超過千億美元，其中絕大部分的生產來自台灣。他表示，AI 應用使台灣受益，台灣企業可以在這波 AI 浪潮中蓬勃發展。AI 產業革命不僅能夠重新定義軟體的創造方式，還能夠塑造 AI 所能創造的軟體。在 AI 生態系中，所有台灣企業應該抓住這一契機。輝達的所有合作夥伴都在積極推出新產品，輝達每年營運成長幅度超過 1,000 億美元，其中大部分在台灣生產。AI 應用的成長機會和驅動力將使台灣受惠，台灣企業將在 AI 革命浪潮中蓬勃發展。

前科技部部長陳良基綜觀整個 AI 算力產業鏈指出，台灣最具機會的是建立一條完整的 AI 代工產業鏈。這條產業鏈包括從最上游的晶片生產、伺服器組裝、AI 運算中心建置，到下游的算力出租、AI 軟體建構、AI 模型訓練代工等服務。台灣能夠面向全球，提供一條龍服務的 AI 代工產業。而且，AI 應用於不同產業，不僅需要專屬的 AI 模型數據，還需要應對特殊任務的專用晶片。這些需求都能成為台灣既有能力的延伸。結合全球對台灣供應鏈的信賴，台灣有機會再次打造一條垂直分工的代工供應鏈，甚至可能重塑出「AI 時代的台積電」。

### **(三) 對產業生態系的影響**

算力產業並非僅限於硬體製造，更包括軟體開發、應用場景探索等多個層面。了解台灣算力產業生態系的形成、發展軌跡，以及其中的主要參與者與互動關係，對於規劃未來產業政策及技術布局有重要啟示。

例如，全球最大 AI 伺服器製造商鴻海將與輝達合作，在高雄軟體園區建設一個以超級晶片 GB200 伺服器為核心的先進算力中心，總計將配置 64 櫃（「櫃」指的是機櫃，用來放置伺服器設備）、4,608 顆 GPU，預計在 2026 年完工。借助輝達強大的 AI 技術，這個算力中心將推動鴻海的三大平台服務（即智慧製造、

智慧電動車及智慧城市)。雙方還將持續深化在 AI、電動車、智慧工廠、機器人、智慧城市等領域的合作，透過鴻海龐大的製造規模，展現 AI 所帶來的強大競爭力，為產業作出更大貢獻。鴻海表示，在高雄建立的先進運算中心將運用輝達的 AI 技術、數位孿生(Digital Twin)和機器人平台來推動技術創新。雙方將以 NVIDIA Omniverse 平台，創建數位孿生應用，並逐步導入智慧製造、智慧電動車和智慧城市平台，NVIDIA Omniverse 是 NVIDIA 發布的一個計算機圖形與仿真模擬平台。

正崙集團旗下的森崙能源將與日本優必達 (Ubitus) 攜手成立「友崙超級運算公司(友崙超算)」，專注於算力服務與 AI 應用的發展。並向華碩採購 GPU，計劃在土城民權廠建造一個 AI 運算中心，預計在 2024 第四季完工。據悉，這個新建的中心將配置 128 台 H100 伺服器，共 1,024 張 GPU 顯卡，相比於目前 NVIDIA 在台灣建立的 Taipei-1 運算中心(64 台 H100 伺服器、512 張顯卡)，算力將翻倍。未來，這個中心將成為全台最大 AI 運算資料中心，同時也擠身全球前 20 大 AI 運算中心。正崙董事長郭台強表示，AI 目前是最受關注的議題，台灣電子業也積極參與，AI 運算能力已被視為國力的象徵，因此新政府上任後也十分重視 AI 相關發展。

相較於其他資訊服務業者，建置算力中心是一項高門檻的挑戰。台灣約有 16,000 家資訊服務廠商，但根據 2024 年 2 月的台灣上市櫃資料顯示，資訊服務業上市櫃公司僅 46 家，僅占上市櫃公司總數的 2.5%。此外，資訊服務業的市值在臺灣上市櫃公司中的比重僅約 0.42%。這些數據顯示出臺灣資訊服務產業的規模較小，對其他產業的正向影響力也有限。對於中小型資訊服務業者或新創公司來說，開發高效能運算專案或生成式 AI 應用，則須面臨著算力高昂的挑戰。如同數位發展部數位產業署副署長胡貝蒂所指出，AI 算力成本高昂，企業即使有意使用，可能面臨算力不足的問題，且在專案結束後，未充分利用的剩餘算力也可能造成浪費。

為解決這些問題，數產署投入數千萬元購買輝達與超微 (AMD) 的頂級算力晶片，建立「AI 算力池」，透過算力共享的概念，讓有需求的團隊申請使用這些算力，應用範疇包括基因排序、氣候變遷和長照行為偵測等模擬運算，期望讓創新項目得以更快速推進。此外，輝達也獲得經濟部「大 A+計畫」的補助，並在臺灣設立研發中心，於 2023 年底建置完成 AI 超級電腦 TAIPEI-1。該超級電腦的 25%算力將免費提供國內產官學研與新創團隊進行研發使用。

研調機構 Fortune Business Insights 預估，從 2024 年起至 2032 年，全球 GPU 算力服務的產值年複合成長率將達到 35.8%，也就是說，這是一個每三年產值就會翻倍的新兆元產業。在過去，雲端服務市場主要由亞馬遜、微軟和 Google 三大巨頭所主導，三者共佔全球近七成的市占率。然而，隨著 AI 時代的來臨，這個市場格局有望重新改寫。更關鍵的是，這也為台灣產業的轉型帶來了重要契機。

#### (四) 政府 AI 政策的驅動與發展

自 2023 年生成式 AI 掀起全球熱潮以來，科技巨頭紛紛加大對 AI 領域的投資，各國政府也開始重視 AI 技術的發展。作為全球 AI 晶片的領導者，輝達 (NVIDIA) 提出了「主權 AI」的概念，強調各國需要建立自己的 AI 基礎設施、數據資源和人才，確保在 AI 時代的自主發展與應用，以維護國家安全、文化、語言和知識的掌控能力。基於這一理念，各國政府與科技巨頭，包括美國、歐洲、亞洲及台灣，都積極參與這場 AI 投資競賽，其中，算力已成為推動 AI 進步的核心動力。

早在 1961 年「人工智慧之父」約翰·麥卡錫 John McCarthy 就提出了「有一天運算會被歸類為一項公共事業，就像電話系統一樣」。在數位化浪潮中，算力成為像水、電一樣的公共基礎資源，資料中心、通訊網路也成為重要的公共基礎設施。對整個人類社會來說，算力不再是技術維度。它已上升到經濟學和哲學的維度，成為數位經濟時代的核心產物，也是全社會數位轉型的基石。我們每個人的生活，以及工廠企業的運作、政府部門的運作，都離不開算力。國家安全、國防建設、基礎學科研究等關鍵領域也需要大量算力。算力決定了數位經濟發展的速度和社會智慧化發展的高度，這些都需要政府政策大力的引導與促成。

行政院智慧國家推動小組在 2023 年提出「台灣 AI 行動計畫 2.0」，將以實現「以 AI 帶動產業轉型升級、以 AI 協助增進社會福祉、讓臺灣成為全球 AI 新銳」為願景，如圖 1.5 所示。從產業端出發，透過深耕 AI 技術與發展 AI 產業及產業應用 AI，帶動我國整體產業轉型升級，如圖 1.5。



資料來源：數位產業署(2024)

圖 1.5、台灣 AI 行動計畫 2.0

該計畫涵蓋五大任務，旨在提升台灣在全球人工智慧（AI）領域的競爭力。這五大任務分別是：人才培育、技術深耕、運作環境的建設、提升國際影響力，以及回應人文社會。這些任務的實施不僅將推動產業的轉型升級，還能增進社會福祉，並使台灣成為全球 AI 新銳的領導者。在五大任務中，特別強調「回應人文社會」部份，該部份體現了政府對當前國家社會面臨的挑戰的重視。例如，勞動力短缺、超高齡社會和淨零碳排等重要課題，都需要 AI 技術的支援來提出解決方案，促進社會的可持續發展。

此外，政府推出的「AI 四大計畫」也在持續推動 AI 產業的發展。其中包括「AI GO」人才培訓計畫，旨在提升 AI 相關人才的素質；「AI Hub」應用服務發展環境推動計畫，致力於建立良好的 AI 應用環境；「AI 領航計畫」則專注於引導 AI 技術的實際應用；而「Fast AI 人工智慧速捷技術深耕計畫」則著眼於加速技術的研究與開發，確保台灣在 AI 技術方面的持續創新與領先地位。另一方面，為了進一步規範和促進 AI 技術的發展，國科會已完成 AI 基本法的擬定，並已進入預告階段，未來將依循程序送至行政院進行審查。這部法律的制定，將為台灣的 AI 發展提供更加清晰的法律框架，確保技術發展的倫理性與可持續性。

數位產業署則提出「數位公共建設-數位產業跨域軟體基盤暨數位服務躍升計畫」，如圖 1.6 所示。旨在透過數位公共建設引導公私協力創新，建立 AI 算力、數據、程式碼三大軟體基盤，加速數位服務發展並降低重複開發投資，同時建立優化數位採購機制，拓展數位服務商機，並營造良好的產業發展環境，如圖 1.6。



資料來源：數位產業署(2024)

圖 1.6、數位產業跨域軟體基盤暨數位服務躍升計畫

建構多元產業應用的推動作法至關重要，即所謂的 Transparency（資料共享機制透明化）、Robustness（開發產業專用模型工具之穩健性）、Usability（產業 AI 解決方案之可用性）、Scalability（產業應用擴散之可擴展性）以及 Technology Integration（國際輸出通路及商機之技術整合）之 Trust 方案，可以做微調模型（Fine-tune）、組隊進行工具之開發、建構場域示範案例等等，以推動多元 AI 產業。通過這些 Trust 方案的實施，企業可以進行模型微調（Fine-tune），組隊進行工具開發，並建立場域示範案例，以促進多元 AI 產業的蓬勃發展。

此外，數位發展部已投入資源，啟動算力平台計畫，旨在為人工智慧（AI）算力提供示範建置模式。該計畫專為資服業者設計，並且免費開放申請，以鼓勵更多企業參與。計畫的使用週數為 6 週，期間內，參加的業者將能夠充分體驗並測試算力平台的各項功能及應用，進而了解其對業務運作的潛在影響。在計畫實施後，數位發展部將根據申請情況進行滾動檢討，持續優化建置模式和運作方式，以確保能夠滿足業者的需求，並提升整體服務品質。此舉不僅能有效提升台灣在 AI 領域的算力資源，也希望透過這個平台促進業者之間的交流與合作，進一步推動產業的數位轉型和升級。藉由算力平台計畫的推進，數位發展部期望能夠打造一個活躍的 AI 生態系，吸引更多企業進一步探索 AI 技術的應用，並創造出更多的商業機會。透過這樣的努力，將使台灣在全球 AI 競爭中具備更強的實力和影響力。

人才培育是當前推動 AI 產業發展的重點之一。具體方法包括整合企業的實戰培育、專業技能的養成以及教師的專業培訓。透過這些措施，政府與企業共同合作，期望能在一年內培養出 6000 名 AI 人才。這些人才的培育方式主要是採用帶案入場的方式，專注於五個關鍵領域：算力資源、應用平台、應用模組、教學案例以及創新解決方案。這樣的策略不僅能提升人才的實戰經驗，也能讓他們更好地理解產業需求，進而為未來的 AI 應用做好準備。

在推動 AI 應用方面，一些知名範例如智寶國際、聯發科的 DaVinci 以及鑫蘊林科等企業，均展現出相當成功的成果。推動 AI 應用的普及率是衡量成效的重要指標。根據統計，去年 AI 應用的普及率已達 32.5%，而短期目標則設定為提升至 50%。長期來看，期望能夠達到 70% 的普及率，這將對整體數位經濟的發展起到積極的推動作用。

人才、技術和產業的三位一體政策，對於數位經濟產業的發展至關重要。具體而言，建立一個良好的 AI 生態系是成功的關鍵。這包括建立平臺、開發工具和軟體，並且培養整合型人才、應用型人才及開發型人才等三類專業人才，才能確保 AI 技術的廣泛運用於各行各業。只有通過這些多元化的人才培養措施與政策，台灣才能在全球數位經濟的競爭中，建立起堅實的基礎，推動各行各業的數位轉型，並實現產業升級。未來，隨著 AI 技術的進一步普及與應用，將會為台灣的經濟發展注入新的動能，促進社會的整體福祉。

### (五) 全球算力市場競爭的挑戰與機遇

全球算力市場競爭激烈，美國、中國、歐洲等國家紛紛投入大量資源發展人工智慧、雲端計算等技術，台灣應如何應對來自全球市場的挑戰並找到自身的發展優勢，則是需要面對的重要議題。

在人才培育方面，算力市場對人才的需求呈現多樣化和高度專業化的趨勢。首先，隨著人工智慧、機器學習、大數據等技術的快速發展，企業對數據科學家、演算法工程師以及能夠設計和優化高效能計算架構的專業人才需求極為旺盛。這些人才不僅需要具備強大的數學和程式能力，還必須精通數據分析和模型構建，能夠理解並解決複雜的計算問題。此外，隨著雲端計算和分散式系統的興起，具備雲端平台開發與維護技能的雲端計算工程師也成為搶手的資源。

除了技術人才，算力市場還需要具備跨領域合作能力的人才，特別是在技術與商業之間架起橋樑。這些人才需要理解技術的商業應用，能夠促進技術創新與市場需求的對接。此外，隨著全球數據安全標準的提升，具備網路安全和數據隱私保護技能的專家也成為企業不可或缺的資產。算力市場需要的不是單一技能的人才，而是具備多領域知識、創新思維及強大協作能力的綜合型專業人才。

除了本國企業對算力市場人才的需求外，國際大廠也紛紛進駐台灣，加入搶人大戰。輝達與超微(AMD)計劃在台灣設立研發中心，因此已向經濟部申請「領航企業研發深耕計畫」(簡稱大A+計畫)。至於亞馬遜旗下的 Amazon Web Services (AWS) 則計劃設置資料中心，但未申請經濟部的補助。針對 AI 人才培育問題，立法院曾就台灣在半導體先進製程和 AI 伺服器製造上具備優勢，但在 IC 設計與生成式 AI 應用服務相關的人才卻相對不足，提出如何具體強化人才培育的課題，以及計畫中的 20 萬名 AI 人才是否包括攬才部分。經濟部在書面報告中指出，未來四年內將培育 20 萬名 AI 專業人才，涵蓋大語言模型訓練與微調等技能，以加速 AI 在各產業的應用。同時，經濟部也將推動 AI 人才認證計畫，確保台灣人才發展能跟上 AI 的全球趨勢。

此外，經濟部已核准超微(AMD)申請的 A+計畫，總經費達新台幣 86.4 億元，經濟部將補助其中的三成，約 25.9 億元，主要用於開發 AI 軟硬體的前瞻技術。由於超微的繪圖處理器和 AI 軟體平台採用開放式架構，經濟部產業技術司也爭取讓超微釋出更多合作機會，並邀請 33 家國內廠商共同參與研發。整體計畫預計可帶動 150 億元的新投資，每年為台灣培育超過 1,000 名 AI 專業人才。然而，為避免本地人才排擠效應，經濟部也要求超微引進 50% 的國外人才，以保持人才供應的平衡。

另一個關鍵挑戰則是電力需求。AI 到底消耗多少電力？根據美國新聞網站 Vox 的分析，國際能源總署(IEA)在 2024 年 1 月發布了未來兩年的全球能源使用預測，首次將資料中心、加密貨幣和 AI 相關領域的電力消耗納入評估。IEA 估

計，截至 2022 年，這三個領域的總用電量已佔全球能源需求的近 2%；然而，到 2026 年，這一需求將翻倍，相當於整個日本一年的用電量。限制 AI 發展的不僅僅是其本身的能耗，還包括相關基礎設施的能耗以及電網的承載能力。AI 的算力需要穩定且強大的電網支援，然而隨著極端天氣事件越來越頻繁，許多地區的電網變得更加脆弱。氣候變遷帶來的暖化趨勢，將導致更頻繁的極端天氣，不僅會造成用電需求急劇上升，加重電網負荷，還可能直接衝擊電網設施的運作。

根據資策會 MIC 的觀察，2024 年全球資料中心數量已達到約 8,000 多座，考慮到每個資料中心的興建週期為四年，單就全球資料中心的電力需求預期到 2028 年將翻倍，且亞洲地區的成長幅度最為顯著。資策會 MIC 的產業分析師陳奕伶指出，電力需求的急劇增長，使得資料中心在綠色發展上面臨嚴峻挑戰。如何在 AI 技術應用與永續發展之間取得平衡，成為近年來資料中心轉型的核心議題，且其影響範圍廣泛。電力問題不僅是挑戰，也為相關產業帶來了新的商機。陳奕伶進一步分析了綠色資料中心的趨勢，指出隨著資料中心逐漸朝向傳統運算與 AI 混合發展，各國針對 PUE（電力使用效率）的監管規範，以及企業推動的 ESG 減碳策略，正共同促使資料中心業者採取更環保的運作模式。透過引入綠色科技和尋找更多綠電來源這兩大途徑，創新技術商機應運而生，吸引了眾多新舊業者的參與，而台灣企業在這波趨勢中，將迎來兩大發展機會。

隨著資料中心的不斷擴張，電網也必須同步擴充，且能源來源需多樣化，以充分利用天然氣、太陽能 and 核能等資源。同時，基礎設施也必須進一步發展，以支援日益增長的電力需求。近期，微軟簽署了一項價值 100 億美元的綠色能源協議，這是目前企業在再生能源領域最大的一筆購買案，預計將掀起全球再生能源建設的熱潮。然而，隨著算力的快速發展，電力供應面臨嚴峻考驗。無論各國政府的淨零排放目標定於何年，或各國政府和企業投入多少資金參與 AI 資料中心的建設競賽，「電力開發」都將成為決定最終成敗的關鍵因素。

#### 四、研究目的

在當今科技主導的時代，算力已被視為國家實力的核心指標之一，尤其在超級電腦和 AI 的發展上表現得尤為明顯。超級電腦擁有極高的運算能力，能夠處理大量數據並進行複雜的計算和模擬，對於推動科技創新和提升經濟競爭力具有關鍵作用。與此同時，AI 的進步高度依賴於強大的算力支持。AI 模型的訓練和推理過程需要大量的計算資源，算力的提升使得 AI 系統能夠處理更複雜的任務，進行更精確的預測和決策，從而在醫療、金融、交通等多個領域中帶來顯著的效益。

另一方面，在國際競爭中，超級電腦的算力排名成為衡量國家科技實力的重要標準。美國和中國在全球超級電腦排行榜中占據領先地位，顯示出其在全球科技領域的影響力。由此可見，算力的增強不僅能推動國內科技發展，也能提升國際間的競爭力。因此，算力盤查與提升算力成為各國科技政策和產業發展的重點，

以確保在未來的科技競賽中占據有利位置。

最後，隨著數據中心與超級電腦成為現代經濟和科技發展的核心，但在高效能運算與深度學習模型訓練過程中，對算力的需求會成倍增加，進而推動電力消耗的上升。隨著數據量與算力需求的增加，如何有效地管理電力供應與強化電網，尤其在尋找可再生能源或綠電來減輕對環境的影響，也就成為各界所關切的議題。

本專題報告基於算力的重要性與影響力，乃形成本研究之背景與動機，進而擬定我國主要算力盤點與產業生態系之探討，研究目的綜整如下：

1. 算力定義與影響因子：詳細探討算力的定義、類別、單位及其在人工智慧個人電腦（AIPC）中的應用和影響因子，展示算力在現代計算系統中的核心地位。
2. 算力資源與盤點分析：提供我國算力資源盤點之具體作法，包括資源範疇、方法學及盤點結果，並期能與國際接軌。之後再進行台灣算力產業盤點，包括硬體、軟體、人才及相關基礎設施等方面的狀況，以深入了解目前產業的規模、技術水平和市場地位。
3. 算力系統生態系分析：包括界定算力生態系範疇、我國算力生態系發展現況、全球算力生態系發展現況，結果包括產值統計與類別佔比等經濟性之分析資料。再探討台灣算力產業生態系的形成與演變，分析其中各參與者（如企業、研探究機構、政府等）之間的關係與互動模式，以及其對整體產業發展的影響。
4. 綠算力與應用於永續發展：透過分析綠色算力的概念、技術實踐和應用案例，提出可行的策略和方法，以促進環境友好算力系統的發展，並推動永續發展目標的實現。
5. 目標與政策建言：根據對台灣算力產業的全面了解和分析，提出相應的政策建議，包括提升算力規模、技術創新支持、人才培養與引進、產業生態系優化等方面，以促進產業的健康發展和可持續增長。

此外，由於算力議題發展快速迭代，本稿所引用的資料和文獻主要來自今年年中左右，尤其是 TOP500 各國算力及排名部份，於今年 11 月將更新並公布新的資訊，故將產生時間上之落差。然而，這些內容仍可作為趨勢參考，並為後續規劃和推動工作提供依據。

## 五、小結

在當今科技主導的時代，算力成為國家實力的核心指標，尤其在超級電腦和人工智慧（AI）的發展上尤為突出。超級電腦擁有強大的運算能力，能處理大量數據並進行複雜的計算，對於推動科技創新和經濟競爭力至關重要。我國在所述之背景因素影響下：1. 全球 AI 浪潮帶動台灣算力發展，2. 台灣在全球半導體產業的領先地位，3. 對產業生態系的影響，4. 政府 AI 政策的驅動與發展，5. 全球算力

市場競爭的挑戰與機遇。要如何強化算力，已成為我國當前科技政策和產業發展的重點，以期能在全球競爭中繼續保持優勢。然而，隨著算力需求的激增，電力消耗也隨之上升。各界關切如何有效管理電力供應，並積極尋找可再生能源或綠電，以減輕環境負擔，這也成為全球共同面臨的重要挑戰。

## 參考文獻

1. 工商時報(2024)，慎防算力大爆發後的電力大考驗，工商時報社論，  
<https://www.chinatimes.com/opinion/20240404000185-262113?chdtv>
2. 中央社(2024)，「超級電腦」是什麼？AI 運算為何不靠它不行？台灣最新布局一次看，  
<https://www.businessweekly.com.tw/business/blog/3015825>，
3. 公視新聞網(2024)，Google 台灣 AI 產業白皮書 估 2030 年經濟效益達 3.2 兆。  
<https://news.pts.org.tw/article/707526>
4. 行政院(2024)，台灣 AI 行動計畫 2.0。行政院智慧國家推動小組，  
<https://digi.nstc.gov.tw/File/7C71629D702E2D89/e8ccec35-9e42-431c-b778-45dae073d5b5?A=C>
5. 吳家豪(2024)，全球中央 9 月號/2024 第 189 期，出版商：中央通訊社
6. 吳凱中(2024)，2024 創新論壇／和碩董座：國力即算力，代價很高，經濟日報，  
<https://money.udn.com/money/story/5648/8102211>
7. 沈柏延(2024)，壯大資訊服務與數位產業，瞄準海外市場，今周刊  
<https://www.businesstoday.com.tw/article/category/183015/post/202404090003/>
8. 侯良儒、陳慶徽(2024)，台灣有機會打造 AI 代工業！前科技部長陳良基：商機 20 年沒盡頭，商周 1911 期  
<https://www.businessweekly.com.tw/Archive/Article/Index?StrId=7010480>
9. 侯冠州(2024)，正歲建全台最大 AI 運算中心，算力是輝達 Taipei-1 的兩倍，工商時報，  
<https://www.ctee.com.tw/news/20240528700055-439901>
10. 范熒恬(2024)，AI 算力需求已超越摩爾定律？林百里說 AI 力即國力、要與之共舞，張忠謀 4 年前就神預言，大老闆金句一次看，經濟日報，  
[https://money.udn.com/money/story/5612/7453089?from=edn\\_referralnews\\_story\\_ch2187](https://money.udn.com/money/story/5612/7453089?from=edn_referralnews_story_ch2187)。
11. 徐子苓(2024)，資料中心估 4 年內電力需求翻倍，資策會：台廠有 2 大機會，自由財經，  
<https://ec.ltn.com.tw/article/breakingnews/4797421>
12. 陳亦偉(2024)，吃電怪獸來了，全球中央 9 月號/2024，189 期，出版商：中央通訊社
13. 黃嫻(2024)，無法用愛發電，高盛：七年內 AI 資料中心電力需求暴增一倍，科技新知，  
<https://technews.tw/2024/05/06/ai-drive-electricity-need-surge/>
14. 葉芳(2024)，搶攻 AI 算力商機 3 台廠出列，Smart 智富月刊 312 期
15. 劉惠琴(2024)，全球超級電腦 500 強榜單公布！美國蟬聯稱霸、台灣這 6 台上榜，自由時報，  
<https://3c.ltn.com.tw/news/58232>
16. 數位產業署(2024)，數位公共建設-數位產業跨域軟體基盤暨數位服務躍升計畫，  
<https://www-api.moda.gov.tw/File/Get/adi/zh-tw/bmLKVEp7TAcvdAb>
17. 鍾榮峰(2024)，黃仁勳：輝達每年成長逾千億美元，大部分在台灣生產。中央通訊社。  
<https://www.cna.com.tw/news/afe/202406160046.aspx>
18. 鴻海(2024)，雙方持續深化 AI、電動車、機器人、智慧城市等多領域合作，為產業做出更多貢獻，鴻海科技集團，  
<https://www.foxconn.com/zh-tw/press-center/events/foxconn-events/1349>
19. 蘇思云(2024)，採用輝達頂級晶片 數發部打造 AI 算力計畫 7 月開跑，中央

通訊社，<https://www.cna.com.tw/news/afe/202406020029.aspx>

20. TOP500(2024), TOP500 List - June 2024.

<https://www.top500.org/lists/top500/list/2024/06/>

21. Zhi, L, (2022), This article is worth reading about the “Computing Power”.  
DiskMFR. <https://www.diskmfr.com/this-article-is-worth-reading-about-the-computing-power/>

## 第二章 算力定義與影響因子探討

算力指計算系統在單位時間內處理數據和執行計算任務的能力，是現代數位時代的重要組成部分。其應用範圍廣泛，從個人電腦到超級計算機，並在人工智慧、大數據分析等領域發揮關鍵作用。算力可分為多種類型，包括雲端算力、單機超級電腦算力、邊緣及終端算力、以及人工智慧算力，這些不同類型的算力依據應用場景而設計，滿足不同需求。雲端算力具有高度靈活性，適合大數據分析、機器學習等應用，且可按需租用，降低成本；單機超級電腦則專為高密度計算設計，適合科學研究和政府機構的內部應用；邊緣和終端算力則適合需要即時處理數據的場景，如自動駕駛與智慧家居。人工智慧應用中，深度學習與推論過程通常依賴專用硬體如圖形處理單元(GPU)、特定應用積體電路(Field Programmable Gate Array, ASIC)以及現場可程式化邏輯閘陣列(Field Programmable Gate Array, FPGA)提升算力並與高效能運算(High Performance Computing, HPC)結合，形成靈活擴展的計算平台。

在電腦科學中，浮點數運算(Floating-Point Arithmetic)是一種用浮點(Floating Point, FP)方式表示實數的運算方式。浮點是一種對於實數的近似值數值表現法，由一個有效數字(即尾數)加上冪數組成，通常表示為乘以某個基數的整數次方。常用於科學計算和人工智慧領域，其中每秒浮點運算次數(FLOPS、FLOPs 或 Flop/s)是用以衡量系統性能的關鍵指標。而FLOPS依據數據精度可分為單精度(FP32)和雙精度(FP64)等不同等級。此外，算力的提升受多個關鍵因素影響，包括處理器數量與類型、RAM容量、存儲速度、網路連接以及軟體與演算法的優化。

### 一、算力定義、類別與單位

本章節先將探討算力內涵，包括定義、類別與單位說明。

#### (一) 算力定義

「算力」在中文中常指利用電腦系統進行數學運算或資料處理的能力。在英文中，可進一步分為 Computing Power 與 Computational Power，Computing Power 主要涉及計算設備或系統本身的計算能力，通常強調硬體效能，如處理速度、記憶體容量及儲存空間。相對之下，Computational Power 則著重於執行特定計算任務的能力，不僅涉及硬體，還包括軟體和演算法的效率。由於本報告是估算國內的算力統計，Computational Power 因深入影響人類資訊生活各個層面，範疇難以明確劃分，因此本研究以“Computing Power”作為評估算力的操作性定義，用以盤點與評估計算能力，定義是描述計算系統或設備在單位時間內處理資料和執行計算任務的能力。

算力是現代數位時代中不可或缺的一部分，而隨著資訊技術的飛速發展，算

力的概念已不僅僅局限於傳統的計算機，還涵蓋了廣泛的設備和系統，包括雲端伺服器、超級計算機、個人電腦、移動設備以及物聯網設備等。算力的重要性在於它決定了計算系統的效率和處理能力。強大的算力可以支持更複雜的演算法、處理更大量的數據，並在更短的時間內完成計算任務。在人工智慧、大數據分析、科學研究、金融交易等領域，算力的需求尤為突出。隨著這些應用的增多，如何提升算力成為了計算機科學與工程領域的一個核心挑戰。現代計算系統中的算力通常分為不同的層級和類型，以滿足不同的應用需求。從雲端集群到邊緣設備，每一種算力都有其獨特的應用場景和技術特性。在當前的技術環境下，算力不僅是單純的硬體性能表現，更是硬體、軟體、演算法和架構協同運作的結果。理解和掌握算力的各種類別和單位，是進行計算系統設計和優化的基礎。

## (二) 算力類別

在闡述「算力類別」時，考量兼顧分類科普性及可盤算力之框架，提出不同面向說明算力分類，一為以科普角度並採「使用情境」方式分類；另外，也有「應用範圍」分類，如高效能運算（HPC）、AI 伺服器與通用伺服器算力進行分類；最後，我們係以可盤算力角度並採「硬體型態」方式來分類，以下進行闡述說明。

### 1. 使用場景分類

#### (1) 雲端算力

雲端算力是當今計算技術中最強大且靈活的算力之一。雲端計算資源由多台超級計算機或伺服器組成，通常以叢集（Cluster）的方式呈現。這樣的結構使得雲端算力具有高度的擴展性和靈活性，能夠根據需求動態分配計算資源。雲端算力的典型應用包括大數據分析、機器學習模型訓練、複雜科學模擬和企業級應用等。

雲端算力的另一個顯著優勢在於其可近性（Accessibility）。企業和研究機構無需購置昂貴的硬體設施，只需通過網際網路即可按需求租用計算資源，這大大降低了計算的門檻。同時，雲端計算服務通常還提供各種軟體工具和框架，進一步簡化了用戶的使用過程。

隨著 5G 和物聯網技術的發展，雲端算力正逐漸與邊緣計算結合，形成一個無縫的計算環境。這種模式可以更快地響應終端設備的需求，並有效減少延遲，提高計算效率。

#### (2) 單機超級電腦算力及單機伺服器算力

單機超級電腦算力和單機伺服器算力指的是單台設備所能提供的計算能力。單機超級電腦通常配備有大量的處理器核心、大容量的 RAM 和高速存儲設備，能夠處理極其複雜的計算任務，如天氣預測、基因組分析、高能物理模擬，人工智慧模型訓練等。

與雲端算力相比，單機超級電腦的優勢在於其高度專用性和低延遲性。由於所有計算資源都集中在單台設備上，它能夠在短時間內完成高密度的計算任務，而不受網路延遲的影響。此外，單機超級電腦通常部署在有專門需求的機構內部，如科研機構、大型企業或政府部門，並針對特定的應用進行優化。

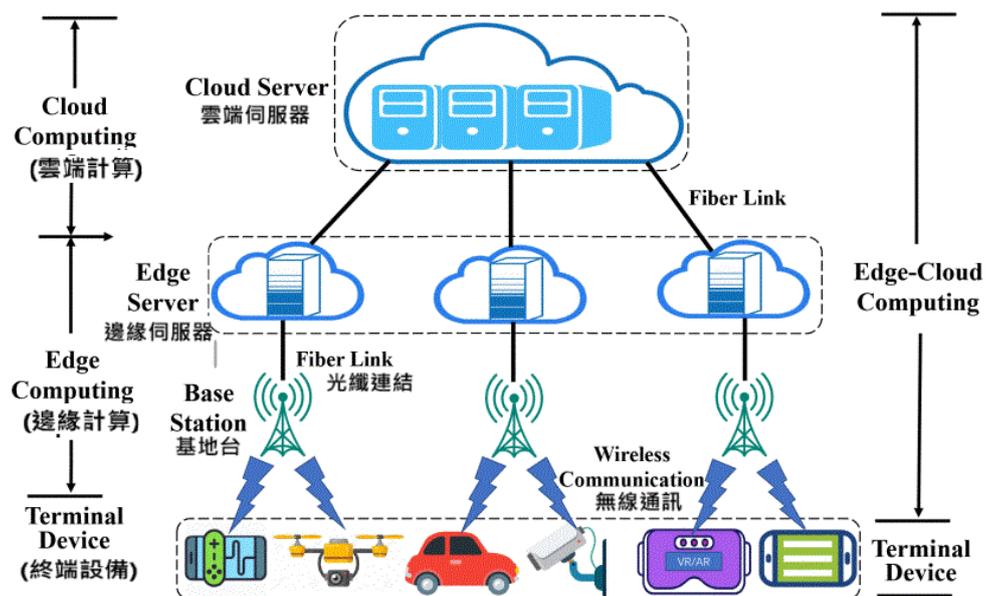
單機伺服器則廣泛應用於企業的日常業務處理中，如數據庫管理、應用服務、網路服務等。雖然單機伺服器的算力不及超級電腦，但它在穩定性、可靠性和易管理性方面具有優勢。隨著硬體技術的不斷進步，現代伺服器已經能夠在較小的物理空間內提供相當可觀的計算能力。

### (3) 邊緣及終端算力(Edge and Terminal Computing Power)

邊緣算力和終端算力是在距離數據源較近的位置進行計算，這樣的設計主要是為了降低數據傳輸的延遲，提高應用的即時性。邊緣算力通常部署在靠近用戶的網路節點上，如通信基地台、路由器或本地數據中心，以支持各種需要即時響應的應用，如自動駕駛、智慧家居、工業自動化等。

終端算力則指的是在個人設備或物聯網設備上進行的計算。隨著移動處理器的性能不斷提升，許多原本需要在伺服器上處理的任務現在可以直接在手機、平板電腦或智慧手錶等設備上完成。例如，當前的智慧手機已經具備了強大的圖像處理和人工智慧計算能力，能夠即時進行臉部識別、語音助手和增強現實應用。

儘管邊緣和終端算力的能力相對有限，但它們在降低延遲、減少網路壓力、提升用戶體驗等方面具有顯著的優勢。特別是在物聯網應用中，邊緣和終端算力的合理分佈和優化能夠顯著提升系統的整體效率。圖 2.1 說明雲端、邊緣、終端設備之間的關係。



資料來源：K. Cao et al.(2021), A Survey on Edge and Edge-Cloud Computing Assisted Cyber-Physical Systems.

圖 2.1、雲端、邊緣、終端設備之間的關係

## 2. 應用範圍分類

一般而言，算力類別也可依其應用範圍的屬性分成超級電腦、AI 伺服器與通用伺服器來進行分類。惟有關人工智慧 AI 之算力配置歸類與應用，目前已廣泛混合並運用在各種伺服器上，尤其在 AI 世代與浪潮下，其所謂之 AI 算力將更廣泛地於建置於具高效能運算（HPC）之資料中心等雲端算力、單機超級或單機伺服器算力或邊緣及終端算力，故算力分類，以現在算力發展趨勢而言，考量 HPC 未來將增添以 GPU 為核心之算力，可用更簡單地進行分類，可分為「AI 伺服器」與「通用伺服器」兩類。

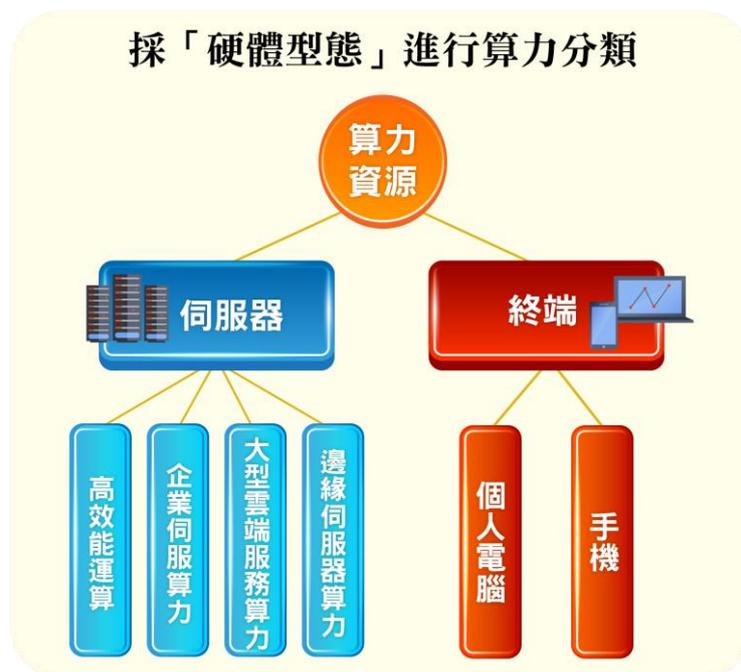
AI 伺服器係專門設計於執行人工智慧任務，而人工智慧算力（簡稱 AI 算力）是推動人工智慧應用發展的關鍵力量，尤其在深度學習與推論過程中，AI 算力的效能決定了模型的訓練與推論速度。隨著人工智慧技術的快速發展，越來越多的專用硬體被應用於 AI 算力領域，如圖形處理單元（GPU）、特定應用積體電路（ASIC）以及現場可程式化邏輯閘陣列（FPGA）。這些硬體能夠支援大規模神經網路模型運行，幫助 AI 系統高效處理大量數據。AI 算力的發展不僅依賴硬體技術的進步，也與高效能運算（HPC）和雲端算力的結合密切相關。承上，所謂 HPC 是指利用超級電腦和分散式計算資源來快速處理和解決需要大量計算能力的複雜科學、工程和數據分析任務的技術。透過這些技術的協同運作，AI 應用能夠在多樣化的環境中實現快速且準確的訓練與推論。其中 HPC 則為處理大規模數據的深度學習任務提供了強大的支援。AI 算力的應用範疇廣泛，涵蓋了自動駕駛、醫療影像分析、自然語言處理、智慧城市等前沿領域，推動了各行各業的數位化與智能化轉型。隨著技術的進一步發展，人工智慧算力將在未來成為更多創新應用的基石。

另外，有關超級電腦 HPC 在人工智慧之運用上也出現另外一種新名詞，HPC Plus Cloud，此係自 2020 年以來，隨著 AI 和 Exascale 技術的興起，在超級電腦大會（Supercomputing Conference, SC）中提出的一種異質計算架構。其結合了 HPC、雲端計算與 AI 技術，成為異構計算的主流架構，透過整合多種硬體資源（如 GPU、TPU、FPGA），大幅提升了深度學習訓練、平行處理及大規模數據處理的效率。HPC Plus 同時利用雲端計算的彈性，靈活適應不同工作負載的需求。隨著異構計算技術的快速發展，HPC Plus 已成為同時處理科學計算與 AI 工作負載的高效架構，未來將在 Exascale 計算及量子計算領域中發揮更大的作用，此發展也將逐漸模糊 HPC 與 AI 計算之間的界線。

此外，有關「通用伺服器」，主要是強調「非 AI 伺服器」，設計用於一般計算任務，如文件處理、數據存儲和網頁伺服等。而其硬體配置通常以中央處理單元（CPU）為主，支持多種應用程式的運行，並具備良好的可擴展性。雖然性能良好，但對於大量數據的平行處理能力較弱，無法有效應對 AI 相關的高運算需求，其主要應用場景包括企業數據中心、雲服務、虛擬化及傳統應用程式等。

### 3. 硬體型態分類

考慮分類「完整性」與「獨立性」，本議題於盤點時，提出一種淺顯易懂之算力（Computing Power），即以計算設備或系統之運算能力為主，著重硬體性能，其算力可依「硬體型態」分為伺服器（Server）及終端（Client）兩類，其中伺服器可分別企業型伺服器算力（Enterprise Server Computing）、大型雲端服務算力（Hypersacle Service Computing）、高效能運算（High-performance Computing）及邊緣伺服器算力（Edge Server Computing）等四類。而終端亦可分為個人電腦及手機，如圖 2.2。



資料來源：本研究整理

圖 2.2、以硬體型態進行算力分類

上述硬體分類上，其伺服器之涵蓋不同的技術需求和應用範疇，說明如下：

- (1) 高效能運算(High-Performance Computing, HPC)：適用於需要極高算力的計算密集型任務，如科學模擬、氣候預測、分子動力學、天文模擬及機器學習中的訓練階段。此類型算力依賴多核處理器、GPU、FPGA 等加速器進行運算，特點是計算核心數量龐大，處理速度極快，且要求高效的內部資料傳輸技術，常應用於解決需平行運算問題。
- (2) 企業級伺服器算力（Enterprise Computing Power）：主要負責處理企業內部計算需求，如資料庫管理、大數據分析和商業應用。此型態算力要求穩定性和高效能，支援大規模工作負載的同時維持高效資料存取。典型應用場景包括金融交易系統、電子商務平台和企業數據中心等，具備支援計算快速連接架構（Compute Express Link, CXL）之技術，為通訊高速互連業界之標準介面，用以提升記憶體與儲存裝置之間的效能。

- (3) 大型雲端服務算力 (Hyperscale Cloud Computing Power)：針對大規模雲端平台設計，提供可擴展、高彈性的算力服務。此型態算力支援全球規模的多租戶環境，能夠即時調整資源分配，以應對不同的計算需求。雲端算力常應用於大數據分析、機器學習、容器化應用和全球服務部署，主要由國際大型科技企業如 AWS、Google、Microsoft 提供，能支持數百萬個虛擬機器同時運作。
- (4) 邊緣計算算力 (Edge Computing Power)：將計算資源部署於資料來源附近，以減少以降低網路延遲和資料傳輸成本，適合即時運算需求高的應用場景，如智慧城市、物聯網 (IoT)、工業自動化以及遠程醫療等。

上述所述之算力類型，依其硬體特性與主要應用範圍及特點，此外，隨著 AI 興起，其 AI 算力亦有廣泛納入四種伺服器硬體架構之趨勢，彙整如表 2.1：

表 2.1、伺服器之算力類型 (以硬體分類)

算力類型	硬體特性	主要應用範圍及特點
高速效能運算	涵蓋計算密集型應用如模型建構和模擬，及人工智慧與機器學習。涉及技術如液體冷卻、多核 GPU 配置、和 FPGA 加速器。	常用於需要數十萬到數百萬計算核心的科學研究、科學與工程模擬、氣象預報，以及高頻交易分析等領域。
企業級伺服器算力	以 x86 架構的 CPU 為主，包括少數 ARM 或 Power 架構 CPU，配備 DDR5 記憶體，支持 CXL 技術。高效液態冷卻系統應對高負荷計算需求。	企業級資料中心使用，管理大型資料庫、處理資料密集型業務流程，如金融服務和大型零售商的庫存管理，需要極高的可靠性和連續運行能力。
大型雲端服務算力	具有擴展性和高可用性特性，常設計專屬硬體架構和優化算法以應對大規模需求。由領先企業如 AWS、Google 和 Microsoft 提供，提供數以百萬計的 VM	提供各種規模的企業在雲端進行資源密集型應用，如大資料分析、機器學習和全球服務部署。
邊緣計算算力	將計算資源部署於用戶或資料來源近處，以低至數毫秒的延遲處理資料。硬體通常配置有專門的處理器和低功耗設計。	智慧城市監控系統、工業物聯網、車聯網和遠程醫療設施，需要即時資料處理和決策，以及高度的資料隱私和安全性需求。

資料來源：Gartner 2023, Market Guide for Servers，本研究整理

此外，在終端電腦亦然。近年來隨著人工智慧技術的發展而出現的一種新型態個人電腦，稱為人工智慧個人電腦 (簡稱 AIPC)。這類計算機專門設計來處理涉及人工智慧應用的任務，如語音識別、圖像識別、自然語言處理、數據分析等。AI PC 的設計目標是在個人電腦的體積和功耗範圍內，提供強大的人工智慧計算能力，以滿足現代應用對算力日益增長的需求。

AI PC 與傳統個人電腦的主要區別在於其硬體結構和軟體配置。傳統個人電

腦主要針對通用計算任務設計，如辦公軟體、網頁瀏覽、多媒體娛樂等，而 AI PC 則專注於處理涉及大量計算和數據處理的複雜人工智慧任務。為此，AI PC 通常配備高效能的處理器和專用的 AI 硬體，例如 GPU、張量處理單元 (Tensor Processing Unit, TPU) 或其他人工智慧加速器。此外，AI PC 還預裝了各種 AI 軟體平台和工具，這些工具能夠支援自然語言處理、計算機視覺、擴增實境 (Augmented Reality, AR) 和虛擬實境 (Virtual Reality, VR) 等應用。AI 手機的新概念亦然，惟主要是在強化手機之內部硬體架構，包括算力晶片與記憶體，用以 AI 推論為主要訴求。

### (三) 算力單位

為了量化不同系統的算力性能，業界通常使用一系列標準的計量單位來表示計算系統的能力。這些單位反映了系統在處理不同類型計算任務時的性能表現。

#### 1. FLOPS (Floating Point Operations Per Second)

FLOPS，即每秒浮點運算次數，是最常用的算力單位之一，尤其在科學計算和人工智慧領域。浮點運算涉及處理帶有小數點的數字，是科學計算中必不可少的一部分。FLOPS 的數值越高，表示系統能在同一時間內處理更多的浮點運算，從而提高計算效率。根據數據的精度，FLOPS 可以分為不同的位元長度，例如單精度浮點數 (32 位元) 和雙精度浮點數 (64 位元)。FP64 (64 位元浮點數) 是許多高性能計算領域的標準數據格式，因為它能確保科學和工程計算中的精度一致性。這也是為何在評估一個超級計算機的算力時，會以 FP64 為基礎進行計算，尤其是在浮點運算性能中，FP64 作為一個重要的參考標準。

在高性能計算領域，FLOPS 通常以 TeraFLOPS (每秒兆次浮點運算, TFLOPS) 或 PetaFLOPS (每秒千兆次浮點運算, PFLOPS) 來衡量超級計算機的性能。在此次總算力統計中，採用的是 TFLOPS 64 作為基準，也就是每秒兆次 64 位元雙精度浮點數進行的運算次數。此外，TOPS (Trillions of Operations Per Second)，中文稱為「每秒兆次運算」，係以整數運算及 8 位元 (INT8) 運算為主，強調針對 AI 計算任務的優化，並簡化了傳統以浮點運算 (FLOPS, Floating Point Operations Per Second) 為主的運算能力衡量方式。表 2.2 是各種浮點運算單位所對應的數值 (即運算單位表示法)。表 2.3 是各種常用的精度組合和它們適合應用的場合。

表 2.2、運算單位表示法

單位	縮寫	次方	十位數值
FLOPS	FLOPS	$10^0$	1
MegaFLOPS	MFLOPS	$10^6$	1,000,000
GigaFLOPS	GFLOPS	$10^9$	1,000,000,000
TeraFLOPS	TFLOPS	$10^{12}$	1,000,000,000,000
PetaFLOPS	PFLOPS	$10^{15}$	1,000,000,000,000,000

表 2.3、各種精度組合和其適合應用的場合

單位	精度 (位元)	性能規模	適用場合
TFLOPS 64	64	TeraFLOPS ( $10^{12}$ )	高精度科學計算、模擬 (例如：氣候建模、物理模擬)
TFLOPS 32	32	TeraFLOPS ( $10^{12}$ )	深度學習訓練、遊戲開發、圖形計算、AI 模型訓練
GFLOPS 64	64	GigaFLOPS ( $10^9$ )	中小規模的精確計算任務，例如：學術研究中的數值計算或小型醫學圖像分析。
GFLOPS 32	32	GigaFLOPS ( $10^9$ )	一般用途運算、較輕量的機器學習任務及圖形渲染 (rendering)
FLOPS 16	16	16 位元浮點運算	神經網路加速、AI 推理 (例如：卷積神經網路)，低精度模型訓練
INT 8	8	8 位元整數運算	低功耗深度學習推理、嵌入式系統、邊緣運算中的 AI 應用

資料來源：C-COM Satellite Systems Inc. (2023)

在此特別要提到該算力數值會受位元長度影響，故在衡量算力時，應特別注意其位元長度。例如，當以 64 元位長度在衡量算力時，其算力為 100 TFLOPS，但改以 AI 運算時，位元長度為 32 位元時，其算力可能會倍增為 200 TFLOPS，如再改為 16 位元時，可能再倍增為 400 TFLOPS，故衡量算力時，其長度是相當關鍵之因子。

綜上，算力衡量可謂多樣性，依據實際應用場景需求，採用對應算力精度與運算效能。簡單地說算力可分為科學算力、AI 訓練算力和 AI 推理算力三類。科學算力如氣候預報、地震模擬、基因組研究、天文觀測數據及粒子物理學研析等，均需透過高精度數值進行分析，故一般採 64 位元；AI 訓練算力，例如當前最夯的 ChapGPT 大語言模型訓練所需算力，惟基於 AI 本質以及計算效率、資源需求現況，可降為 32 或 16 位元，以加快運算效能；最後，AI 推理算力，例如詢問 ChapGPT 時，背後推理與回應所需算力，考量回應實際資源需求與可接受容忍度，實際上採用 16 或 8 位元即可，而精度門檻最低。

## 2. IPS (Instructions Per Second)

IPS，即每秒指令數，是一種傳統的計算性能指標，用於衡量處理器在執行基本指令操作上的效率。IPS 主要應用於計算系統的基本運行性能測試中，反映了處理器在執行各種基礎任務時的能力。隨著處理器架構的不斷發展，IPS 逐漸被更為複雜的指標所取代，但在某些應用中，仍具有一定的參考價值。

### 3. Hash/s (Hash Per Second)、Sol/s (Solution Per Second)、TPS (Transactions Per Second)

Hash/s 表示每秒能進行的雜湊 (Hash) 運算次數，通常用於加密貨幣挖礦中的算力測量。Sol/s，意即每秒解算次數，主要用於加密貨幣和區塊鏈技術中的算力測量。TPS，即每秒交易數，是衡量系統處理交易能力的重要指標，尤其在金融系統和區塊鏈應用中。

#### 二、算力類型與影響因子

本節將說明有關算力類型與影響因子。

##### (一) 算力類型

在探討算力的類型時，關鍵在於理解不同計算元件的角色及其對整體系統性能的影響。隨著技術的發展，計算系統不再僅僅依賴於單一的處理單元，而是通過多種類型的處理元件來實現更高效的運算。這些元件各自專精於不同的計算任務，並在協同工作時最大化系統的整體算力。

##### 1. CPU (中央處理單元)

中央處理單元 (Central Processing Unit, CPU) 作為傳統計算系統的核心元件，負責執行大多數的基本計算任務和系統操作。CPU 的主要特點是其多用途性和靈活性，能夠高效地處理各種計算任務，從基本的算術運算到複雜的邏輯操作。然而，隨著計算需求的增加，單一的 CPU 已無法滿足所有高效能計算的需求，特別是在涉及大規模平行處理和大量數據運算的應用中。

現代的 CPU 通常具有多核心設計，每個核心能夠同時處理不同的計算任務，這種架構大大提升了系統的平行處理能力。儘管如此，CPU 在處理特定任務時，如圖像渲染 (Image Rendering)，即建立 3D 模型或場景數據轉換成 2D 圖像的過程或機器學習 (Machine Learning, ML) 模型訓練，其效率仍無法與專用處理單元相媲美。因此，CPU 在現代計算系統中更多地被用作一種通用控制單元，負責協調和管理其他專用處理器的運行。

##### 2. GPU (圖形處理單元)

GPU 最初設計的主要目的是加速圖形渲染，它在處理大規模平行計算任務時顯示出了強大的能力。與 CPU 相比，GPU 的結構更加適合於平行處理，這使得它在涉及大量相同操作的計算任務中表現出色。例如，在深度學習模型的訓練中，GPU 能夠同時處理數百甚至數千個神經網路層的運算，這極大地縮短了模型的訓練時間。

由於 GPU 在處理浮點運算方面的優勢，它成為了人工智慧和機器學習領域的首選硬體加速器。如今，許多高效能計算集群都配備了大量的 GPU 來加速數據處理任務。此外，GPU 還被廣泛應用於加密貨幣挖礦、科學模擬、圖像處理等

領域。隨著技術的不斷進步，GPU 的性能和應用範圍也在不斷擴展，成為現代計算系統中不可或缺的算力支柱。

### **3. FPGA（現場可程式化邏輯閘陣列）**

FPGA 是一種靈活的硬體加速器，其最大的特點是可以根據具體的應用需求進行硬體配置。這種靈活性使得 FPGA 在處理特定任務時具有極高的效率，特別是在需要專用硬體解決方案的場景中。FPGA 廣泛應用於電信、工業控制、軍事系統以及高頻交易等領域。

FPGA 的運行方式是通過硬體邏輯電路來實現計算，因此在速度和功耗上都有顯著優勢。相比於 CPU 和 GPU，FPGA 可以實現更低延遲的運算，這對於一些即時應用至關重要。雖然 FPGA 的開發相對複雜，需要專業的硬體設計知識，但它的可編程性和高效能使其在特定領域中有著無法替代的地位。

此外，隨著 FPGA 的開發工具和軟體支持的不斷增強，其應用門檻也在逐漸降低，越來越多的開發者開始將 FPGA 集成到其系統中，以提高計算效率和性能。未來，FPGA 有望在更多的計算領域中發揮重要作用，特別是在那些需要客製化計算解決方案的應用中。

### **4. ASIC（特定應用積體電路）**

ASIC 是一種針對特定應用設計的硬體，具有無與倫比的運算效率。由於 ASIC 是針對特定任務進行設計的，因此其性能往往遠超其他通用處理單元，如 CPU 和 GPU。然而，這種高效能的代價是缺乏靈活性，ASIC 只能在其設計任務中發揮作用，一旦應用需求發生變化，ASIC 的硬體架構無法進行調整。

ASIC 廣泛應用於需要極高效能和低功耗的場景，例如加密貨幣挖礦、網路路由器、數字信號處理器等。在加密貨幣挖礦領域，ASIC 的應用尤為突出，其專用設計使其在 hash 運算中達到最高的效率，成為了大多數挖礦設備的核心。

由於 ASIC 的開發成本高且開發週期長，通常僅在大規模生產中具有經濟效益。因此，ASIC 的應用通常集中在那些對性能和功耗有極高要求且應用範圍相對穩定的領域。

### **5. TPU（張量處理單元）：**

張量處理單元（Tensor Processing Unit, TPU）是 Google 專為人工智慧應用設計的處理器，旨在加速深度學習模型的訓練和推理過程。TPU 在設計上專注於處理張量運算，這是深度學習演算法中的核心運算形式。與 GPU 相比，TPU 在某些特定深度學習任務中的性能更為優越，特別是在大規模模型的訓練和推理過程中。

TPU 的出現標誌著人工智慧硬體的專業化發展趨勢。隨著深度學習應用的普及，對專用硬體的 demand 日益增長。TPU 以其高效能和優化的設計，成為了許多人

工智慧應用的理想選擇。目前，TPU 主要應用於 Google 的雲端服務中，但隨著人工智慧技術的發展，TPU 的應用範圍有望進一步擴展。

## 6. NPU（神經網路處理單元）

神經網路處理單元（Neural network Processing Unit, NPU）是一種專門設計用於加速神經網路推理的專用處理器。NPU 通常集成在移動設備、物聯網設備或嵌入式系統中，以實現低功耗、高效能的 AI 計算。其硬體架構針對神經網路優化，包括專用加速器單元，能快速處理矩陣運算和神經網路推理。NPU 的主要優勢在於能在設備本地進行即時 AI 推理，減少對雲端運算的依賴，適合語音識別、圖像處理等應用。然而，由於設計針對性強，NPU 的通用性較低，通常無法勝任模型訓練等更複雜的 AI 任務。

### （二）影響算力的關鍵因素

理解影響算力的各種因素對於設計和優化計算系統至關重要。算力並不僅僅取決於處理器的性能，還受到其他多個因素的影響，包括硬體配置、存儲速度、網路連接以及軟體和演算法的優化。

#### 1. 處理器數量與類型

計算系統中的處理器數量和類型是決定算力的基本因素。多核處理器和專用硬體（如 GPU、TPU 等）能夠顯著提高系統的平行計算能力，從而提升整體算力。例如，在多處理器系統中，每個處理器可以同時執行不同的計算任務，這使得系統能夠更快地完成大量的運算。

除了處理器的數量，處理器的類型和架構設計也對算力有著重要影響。不同的處理器在處理特定任務時表現各異，例如 GPU 在平行處理方面優於 CPU，而 FPGA 在專用運算方面更具效率。因此，根據具體應用需求選擇合適的處理器類型和數量，是提升算力的關鍵。

#### 2. RAM（隨機存取記憶體）

隨機存取記憶體（Random Access Memory, RAM）的容量和速度對於算力的提升有著直接影響。在處理大規模數據集或執行複雜的演算法時，充足的 RAM 能夠減少數據在處理過程中的調用時間，從而提升整體運算效率。更快的 RAM 速度則意味著處理器可以更快速地存取和處理數據，從而縮短計算時間。

在許多高效能計算應用中，RAM 的容量是瓶頸所在。特別是在處理需要大量內存的應用中，如圖像處理、視頻編碼、機器學習訓練等，充足且高速的 RAM 對於保持高效運算至關重要。

#### 3. 存儲速度與類型

除了 RAM，存儲設備的速度和類型也對算力產生重要影響。傳統的硬碟

(Hard Disk Drive, HDD)雖然容量大,但速度相對較慢,而現代的固態硬碟(Solid-state Drive, SSD)則能提供更快的讀寫速度,從而顯著提升數據存取效率。在需要頻繁讀寫大量數據的應用中,高速存儲設備能夠減少 I/O 瓶頸,從而提升整體算力。

存儲類型的選擇應根據具體應用的需求來決定。例如,在大數據分析中,使用高速的 SSD 能夠顯著提升數據處理速度;而在需要長期數據存儲的應用中,則可以考慮大容量的 HDD 以降低成本。

#### 4. 網路連接

在分散式計算系統中,網路連接的速度和穩定性對算力有著至關重要的影響。快速且穩定的網路連接能夠有效減少數據在不同計算節點之間傳輸的延遲,從而提升整體計算效率。在雲計算和邊緣計算環境中,網路連接的性能尤為重要,因為這些系統需要頻繁地進行數據交換和協同運算。

此外,在一些即時應用中,如金融交易、視頻流處理等,網路延遲可能會對系統的性能產生顯著影響。因此,優化網路架構和選擇合適的網路技術(如 5G 或高速光纖)對於提升算力至關重要。

#### 5. 軟體與演算法優化

最後,軟體和演算法的優化也是提升算力的重要因素。即使擁有強大的硬體,如果軟體和演算法未能充分利用這些資源,算力也無法達到理想的水準。高效的軟體設計和演算法優化能夠最大化硬體性能,從而提升整體系統的算力。

例如,針對 GPU 設計的平行演算法能夠充分發揮其平行處理的優勢,而針對 TPU 設計的深度學習演算法則能夠顯著提高模型訓練的效率。隨著人工智慧和高效能計算的發展,軟體和演算法優化已成為提升算力的重要手段,並且在未來將繼續發揮重要作用。

### 三、小結

本章探討了算力的定義、類別、單位及其影響因子,並指出算力作為現代計算系統的核心要素,在人工智慧、大數據及科學研究等領域的應用中具有舉足輕重的地位。隨著雲端計算和邊緣計算的發展,算力呈現出多樣化的形態,從集中式的雲端算力到分散式的邊緣與終端算力,這些不同形式的算力通過協同運作,滿足了現代應用對高效能與低延遲的需求。

算力的提升不僅依賴於硬體技術的進步,如多核心處理器、GPU、TPU 等專用硬體的使用,還需要軟體與演算法的有效配合。例如,針對特定硬體優化的演算法能夠顯著提升整體運算效率。隨著 5G、物聯網技術的普及,算力與雲端、邊緣和終端設備之間的關係將更加緊密,形成無縫的計算環境,支持各種即時應用場景,如自動駕駛和工業自動化。

未來，隨著 AI 技術的不斷發展，專用 AI 硬體如 TPU、NPU 的廣泛應用，將進一步推動算力的增長。同時，軟硬體的整合與協同優化將成為提升算力的重要手段，使得現代計算系統在性能、靈活性與可擴展性上達到新的高度。無論是科學研究還是企業應用，算力都將是驅動創新與提升效率的關鍵力量。

### 參考文獻

1. 陳世海, "AI 算力從雲端到終端台灣下一個產業發展", p7, 中技社通訊 150 期, 2024 夏季刊
2. 吳卓然, "算力網絡", 華為 Info-Finder IP 知識百科, 2024-05-13  
<https://info.support.huawei.com/info-finder/encyclopedia/zh/算力網絡.html>
3. Janus, "AI 筆電、AI PC 是什麼？有什麼功能以及規格？目前有哪些產品，什麼時候 AI PC 才是換機的成熟期？", T 客邦, 2024-01-20
4. 維基百科網頁: <https://zh.wikipedia.org/zh-tw/Wiki>
5. K. Cao, S. Hu, Y. Shi, A. W. Colombo, S. Karnouskos and X. Li, "A Survey on Edge and Edge-Cloud Computing Assisted Cyber-Physical Systems," in IEEE Transactions on Industrial Informatics, vol. 17, no. 11, pp. 7806-7819, Nov. 2021.

### 第三章 我國算力資源與盤點分析

本章概分兩部份，第一部份係針對我國算力盤點之統計資料，另外一部份係以 TOP500 的資料為基礎，闡述我國算力評估與排名之分析與探討。

#### 一、我國算力盤點之統計資料

隨著全球科技的迅猛發展，算力（Computing Power）已成為衡量一個國家科技實力的重要指標之一。從人工智慧（AI）到高效能運算（HPC），再到大數據分析，各行各業對於算力的需求日益增長。作為亞太地區科技發展的前沿，台灣在推動科技進步和經濟發展的過程中，對算力的重視與日俱增。台灣的算力不僅在學術界和醫療領域有著廣泛應用，政府機構更是在國防、氣象預測和災害預警等關鍵領域投入了大量資源。然而，要全面了解和評估台灣的算力現狀，需要進行系統的統計和分析。

此次統計旨在深入了解台灣各機構的算力配置，特別集中於公私立（技職）大學、醫學中心和政府機構三大領域。其中有關「政府機構」範疇包括「政府機關」及「由政府持股比例高之法人機構」。這些領域的算力配置對台灣的科研能力、醫療水平及國家安全有直接影響。另外，盤點時主要針對具機櫃型之伺服器，具備高效能運算（HPC），惟不包括單機電腦的顯卡算力（即插在電腦上的顯卡，通常用於遊戲或一般運算的消費級 GPU，算力較小）。

本次統計的重點在於精確地記錄和分析各機構的算力，並透過詳細的數據統計，揭露台灣在全球科技競爭中的位置。值得注意的是，由於民間企業擁有的超級算力數據涉及企業機密，因此未納入本次統計範圍。儘管如此，本次報告仍將對台灣算力的整體水平提供一個初步清晰的概況，尤其是在學術、醫療和政府領域中的表現。

此外，隨著超級計算機在全球的廣泛應用，Top 500 榜單已成為評估全球最強大計算系統的重要指標。台灣在這一榜單中的表現，特別要提有關由 NVIDIA 在台設置的 AI 超級電腦「TAIPEI-1」，進一步證明了台灣在全球計算機技術領域中的重要地位。依經濟部說明，它是捐獻 25% 算力予台灣，主要係提供給我國在學術界、非營利組織和中小企業，以提升我國整體在 AI 和高效能運算（HPC）領域的競爭力和創新能力。且於 Top 500 榜單中，Taipei 之國名隸屬於台灣，故分類上，將 Taipei-1 納入政府機構一併盤點。

透過本次報告，我們希望為台灣的科技政策制定者、研究機構和業界提供有價值的數據支持，並為未來的科技發展和算力規模提升指明方向。這些數據不僅能幫助理解當前的算力分布，也為未來的技術投資和策略調整提供科學依據。

本章節將重點介紹統計方法和結果分析，以期為台灣的算力現狀提供全面的數據支持。

## (一) 統計項目和方法說明

本次統計針對的主要數據項目包括以下幾個方面，每一項目都經過精心挑選，以確保統計數據的全面性和精確性。

### 1. 統計項目的選擇

- (1) 機構名稱：這是我們統計的基本單位，涵蓋了台灣的公私立（技職）大學、醫學中心以及政府機構。這些機構代表了台灣在學術、醫療和政府領域中的核心算力配置。我們確保所有參與統計的機構都具備一定規模的計算設備，以便數據具有代表性。範例：台灣大學、臺北榮民總醫院、國家高速網路與計算中心。
- (2) 機器名稱型號：我們詳細記錄了每一個機構所擁有的計算設備，包括其名稱和型號。這部分數據對於理解該設備的算力和應用場景至關重要。由於不同的計算設備在硬體配置和技術參數上可能存在很大差異，因此這一項目的詳細記錄有助於進一步的數據分析和性能比較。範例：NVIDIA DGX-1、PRIMEHPC FX1000、台灣杉二號（Taiwania 2）。
- (3) 機體性能：這是本次統計的核心部分，涵蓋了多個技術指標，包括計算核心數量、峰值 TFLOPS（FP64 和 FP32）、網路連結、作業系統以及消耗功率等。這些技術指標不僅展示了每台設備的硬體性能，也反映了其在不同應用場景中的潛在能力。
  - A. 計算核心數量：記錄每台機器所擁有的 GPU 或 CPU 數量。核心數量是衡量一台計算機性能的重要指標，尤其是在多核處理器普及的背景下，核心數量直接影響到計算速度和平行處理能力。範例：8x NVIDIA Tesla V100 Tensor Core GPUs（16GB）、48x ARM A64FX CPUs。
  - B. 峰值（Rpeak）TFLOPS（FP64 和 FP32）：Rpeak 這是衡量計算機運算能力的主要指標，為系統之理論性能。FP64 和 FP32 分別表示雙精度和單精度浮點運算的峰值性能。由於雙精度浮點數計算通常用於科學計算和工程模擬，而單精度浮點數計算則常用於 AI 訓練和大數據分析，因此這兩個指標對於了解機器的應用領域具有重要意義。範例：FP64：62TFLOPS，FP32：125TFLOPS。
  - C. 網路連結：記錄每台機器的網路連結技術，如 Mellanox HDR Infiniband。網路連結的性能對於高性能計算中的數據傳輸速度和平行處理效率至關重要，尤其是在大型分散式計算環境中。範例：Mellanox HDR Infiniband、Cray Aries Interconnect。
  - D. 作業系統：記錄每台機器使用的作業系統，例如 Ubuntu Server Linux OS。作業系統不僅影響機器的穩定性和安全性，也決定了其支持的軟體生態系統和

應用程式。範例：Ubuntu Server Linux OS、Red Hat Enterprise Linux。

- (4) 資料來源：為了保證數據的真實性和可靠性，我們標註了每個數據項目的來源。這些來源包括調查機構的官方網站、數位媒體報導以及相關專業人士提供的資料。我們力求在數據收集過程中保持透明度，以確保統計結果的可信度。範例：台灣大學官網，調查日期：2024年5月；Top500官方網頁 <https://top500.org/system/180215>。
- (5) 查詢時間：為了確保數據的時效性，我們記錄了每個數據項目的查詢日期。由於計算設備的性能和配置可能隨著時間變化，因此查詢時間的記錄有助於確保數據的現實意義。範例：調查日期：2024年5月。
- (6) 備註：此部分包括任何與該機器或其用途相關的附加說明。這些備註有助於補充和解釋數據中的特殊情況，如設備的特定應用場景或技術規格上的特別設置。範例：用於AI及高性能計算研究；用於氣象模擬與災害預警。上述相關欄位，可參考附錄二之算力盤點格式。

## 2. 統計方法

為了收集上述數據，我們採用了多層次的統計方法，確保數據的全面性和精確性。

- (1) 初步篩選：在統計開始前，我們對全台灣的公私立（技職）大學、醫學中心和政府機構進行了初步篩選，確定了具備大型計算設備的機構名單。這些機構被選中是基於其計算設備的規模和技術先進性，確保所選機構的數據具有代表性。
- (2) 數據收集：針對選中的機構，我們通過各種渠道進行數據收集。這些渠道包括機構的官方網站、相關技術報告、數位媒體報導以及直接與相關人士聯繫獲取的資料。在此過程中，我們特別注意數據的一致性和可比性，確保每個項目的數據來自權威來源。
- (3) 數據驗證：收集到的數據經過嚴格的驗證流程。這一過程包括與機構提供的官方數據進行比對，並通過多方來源交叉驗證。對於存在疑問或不確定的數據，我們進行了進一步的調查，確保最終統計結果的準確性。
- (4) 結果整理與報告撰寫：最後，我們將分析結果整理成報告形式，並根據不同的受眾需求進行了多層次的報告撰寫。此次報告的目的是為決策者、研究機構以及業界提供一個全面而深入的算力分布概況，並為未來的科技發展和算力投資提供科學依據。

## 3. 統計中的挑戰與應對

在進行本次統計的過程中，我們面臨了一些挑戰，包括數據收集的困難、數據來源的多樣性以及算力技術的快速變化。為了克服這些挑戰，我們採取了以下

應對措施：

- (1) 多管道數據收集：由於部分機構的數據並不公開，我們通過多管道進行數據收集，包括與相關技術人員的直接溝通，以及從公開資料中提取關鍵數據，確保統計結果的全面性。
- (2) 動態更新與持續監控：鑒於算力技術的快速變化，我們設置了動態更新機制，確保在統計期間獲取的數據反映最新的技術狀況。我們還計劃進行定期的數據更新，以持續監控台灣算力的發展趨勢。
- (3) 數據標準化與可比性確保：在數據收集過程中，我們採用了標準化的數據記錄方式，以確保來自不同來源的數據具有可比性。這包括使用統一的計算單位和技術指標，詳如附錄三之輝達各類 GPU 模組之算力資訊，並通過數據清洗和篩選來消除潛在的誤差。

透過以上的方法與策略，我們希望能夠提供一份精確、全面且具有高可信度的算力統計報告，為台灣的科技發展提供重要的參考數據。

## (二) 統計結果與分析

本次統計涵蓋了台灣的公私立(技職)大學、醫學中心和政府機構三大領域，總共盤點了 163 家公私立(技職)大學、26 家醫學中心和 15 家政府機構及 Taipei-1。然而，實際登錄數據時，我們僅登錄了算力大於 1 TFLOPS 的學校或機構。因此，最終登錄的機構數量顯示出台灣在這三大領域中高算力配置的情況。

### 1. 公私立(技職)大學

在本次統計中，我們盤點了 163 家公私立(技職)大學，但僅有 24 所學校的算力達到或超過 1 TFLOPS，符合登錄標準。這些學校集中於台灣的頂尖研究型大學及技術性大學，擁有領先的計算設備，支持各類科研項目，特別是 AI 研究、數據科學和高性能計算等前沿領域。

- (1)總算力：這 24 所學校的總算力達到了 4,358.36 TFLOPS (FP64) 和 9,777.13 TFLOPS (FP32)。這顯示出台灣的高等教育機構在算力投入上的實力，特別是在單精度計算能力上，這些學校在 AI 相關應用中的潛力巨大。
- (2)設備配置：台灣的大學計算設備主要以 NVIDIA DGX 系列為主，特別是 NVIDIA DGX-1 和 NVIDIA DGX A100 機型，這些設備專為 AI 和深度學習工作負載設計，具備強大的並行計算能力，能夠處理大量數據。
- (3)應用場景：這些學校的算力主要應用於 AI 模型訓練、基因組學研究、物理模擬、氣象預測等領域。頂尖大學如台灣大學和清華大學的計算設備不僅用於學術研究，還提供給校外企業和政府機構使用，促進了產學研合作。
- (4)發展趨勢：隨著 AI 和數據科學的快速發展，我們預計這些大學將繼續擴大

其算力規模，並引進更多尖端設備。同時，學校間的算力合作和資源共享將成為未來的趨勢，以提升整體科研效率。

## 2. 醫學中心

在醫學中心領域，我們盤點了 26 家醫學中心，但僅有 13 家醫學中心的算力達到或超過 1 TFLOPS，符合登錄標準。這些醫學中心在台灣的醫療系統中扮演著重要角色，並且已經開始大規模地將高性能計算應用於醫學影像分析、基因組學研究及臨床數據處理等領域。

(1)總算力：13 家醫學中心的總算力達到了 1,732.37 TFLOPS (FP64) 和 3,549.38 TFLOPS (FP32)。這些算力主要集中於台灣的幾家頂級醫學中心，如台北榮總、臺大醫院和成功大學附屬醫院，這些機構在全國醫療研究中具有領先地位。

(2)設備配置：這些醫學中心主要配置了 NVIDIA Tesla GPU，以及 NVIDIA DGX A100 和 H100 機型。這些設備專為醫學影像處理和基因分析等醫學應用進行了優化，能夠高效地處理大規模醫學數據。

(3)應用場景：這些醫學中心的算力應用範圍包括影像識別、疾病預測、個性化治療方案設計以及臨床決策支持系統的開發。這些應用顯著提升了醫療診斷的準確性和效率，並推動了精準醫療的發展。

(4)發展趨勢：隨著醫療數據量的持續增長以及 AI 技術在醫學領域的廣泛應用，這些醫學中心的算力需求將不斷增加。預計這些中心將繼續升級其計算設備，並加強與大學和企業的合作，以推動更多創新應用的落地。

## 3. 政府機構及 Top 500 超級電腦

在政府機構領域，我們盤點了 15 家機構及 Taipei-1，但僅有 8 家機構的算力達到或超過 1 TFLOPS，符合登錄標準。政府機構在本次統計中展現了其在算力領域的絕對領先地位，特別是台灣在 2024 年 6 月 Top 500 榜單中的表現，進一步強調了政府在高性能計算上的重大投入和技術實力。

(1)總算力：這 8 家政府機構，包括登錄於 Top 500 榜單中的 Taipei-1，共計算力達到了 81,917.49 TFLOPS (FP64) 和 153,575.36 TFLOPS (FP32)。這些數據遠遠超過大學和醫學中心，顯示出我國政府對於國土安全和重大科研項目的高度重視。

(2)Top 500 計畫與台灣的表現：Top 500 計畫自 1993 年以來，每年兩次發布全球最強大電腦系統的排名。台灣在 2024 年 6 月的榜單中共有 6 台超級電腦入榜，這些超級電腦不僅提升了台灣在全球計算機領域的地位，也為國內各種高需求計算應用提供了堅實基礎。

A.TAIPEI-1：由 NVIDIA 設置的 AI 超級電腦，排名第 38 名，為台灣最高排

名的超級電腦。TAIPEI-1 於 2023 年底建置完成，其高級運算資源部分免費提供國內各產業加強 AI 研發使用，這大大促進了台灣在 AI 技術上的突破。

B.其他上榜超級電腦：台灣的其他五台超級電腦包括中央氣象局使用的 PRIMEHPC FX1000(排名第 86 名)、PRIMEHPC FX1000(排名第 160 名)、國家高速網路與計算中心的台灣杉二號 (Taiwania 2, 排名第 106 名)、Forerunner 1 (排名第 251 名)、以及台灣杉三號 (Taiwania 3, 排名第 456 名)。這些設備主要用於氣象模擬、災害預警、國土安全 (地理資訊、環境災害等)、生醫、科學運算以及大數據分析等。

(3)設備配置：政府機構的設備以 NVIDIA DGX A100 和 DGX H100 等超級計算機為主，這些設備具備極高的計算能力和先進的網路連結技術，如 Mellanox InfiniBand。這些設備的運行效率極高，特別適合處理氣象模擬、生醫、環境、科學與大數據分析等對計算精度和速度要求極高的應用。

(4)應用場景：政府機構的算力主要應用於國土安全、氣象模擬、災害預警和重大科研項目。這些應用不僅需要大量的算力支持，還要求高度的計算精度和穩定性。因此，政府機構在這些領域的算力投入不僅確保了國家安全，也推動了科學研究的前沿發展。

(5)發展趨勢：隨著國家競爭力需求的增加和全球科技競爭的加劇，我國政府機構的算力配置將繼續擴大。我們預計未來將有更多超級計算機被引入，以支持新一代 AI 應用和高性能計算的發展。同時，政府機構也將加強與學術界和產業界的合作，以共同推動科技進步。

綜上，彙整上述學校、醫院及政府機構 (含 Taipei-1) 之算力統計表，詳如 3.1

表 3.1、台灣算力統計總覽

領域	盤點總數	FP64 TFLOPS (比例)	FP32 TFLOPS (比例)
學校	163	4,358.36 (4.95%)	9777.13 (5.86%)
醫院	26	1,732.37 (1.97%)	3,549.38 (2.13%)
政府機構 (含 Taipei-1)	15	81917.49 (93.08%)	153,575.36 (92.01%)
<b>小計</b>	<b>204</b>	<b>88,088.22 (100%)</b>	<b>166,901.87 (100%)</b>

資料來源：本研究整理

這張表格展示了我國在學校、醫院及政府機構 (含 Taipei-1) 三大領域的 FP64 和 FP32 算力分布情況，並給出了各自的比例。通過這些數據，我們可以更清楚

地看到不同領域的算力配置，並理解其在台灣整體科技實力中的位置。分述如下：

學校部份共 163 所，其算力總和為 4,358.36 TFLOPS (FP64) 和 9,777.13 TFLOPS (FP32)，分別占台灣整體算力的 4.95%和 5.86%。這顯示出台灣教育機構在 AI 和高性能計算方面的持續投入。

另外，醫院部份共 26 家，主要為醫學中心，其算力總和為 1,732.37 TFLOPS (FP64) 和 3,549.38 TFLOPS (FP32)，分別占台灣整體算力的 1.97%和 2.13%。這些算力主要應用於醫學研究和臨床診斷，對於精準醫療的發展具有重要作用。

最後，政府機構(含 Taipei-1)部份，共 15 家的算力總和為 81,917.49 TFLOPS (FP64) 和 153,575.36 TFLOPS (FP32)，這些數據遠遠超過其他兩個領域，分別占我國整體算力的 93.08%和 92.01%。這代表政府在國土安全及重大科研項目上具有強大的技術實力。

總體來看，這些數據揭示了我國在全球科技競爭中的戰略優勢，尤其是在政府機構的強力支持下，我國在 AI 和高性能計算領域的實力不容小覷。隨著未來民間企業資料中心的建成，我國的整體算力將進一步提升，為全球科技創新和發展做出更大貢獻。

#### 4. 民間企業資料中心

因應 AI 應用需求的迅速增長，我國的民間企業也開始積極投資於大型資料中心的建設。儘管此次統計主要集中於已經投入運營的機構，但我們也對即將成立的四大資料中心進行了統計，這些資料中心的算力規模顯示了未來我國在全球 AI 和高性能計算領域的潛力。

(1)四大資料中心：友歲超級運算、GMI Cloud、翔耀 (AICC 算力中心) 和鴻海這四大資料中心計劃中，依其公開媒體所揭露之算力資訊，預計於今、明完成建置，其總算力統計後高達 526,676 TFLOPS (FP64)。這一數字顯示出台灣民間企業在算力上的投資規模，特別是在 AI 和雲計算等領域的應用潛力。

(2)發展狀況：這些資料中心目前仍在建設中，尚未完全投入運營，因此並未列入本次正式統計中。然而，這些資料中心一旦投入使用，將大幅提升台灣在全球算力排名中的位置，並為各種尖端技術應用提供強大的基礎設施支持。

(3)未來展望：隨著這些資料中心的建成，台灣將在 AI、大數據分析、雲端計算等領域取得更多突破。企業間的合作也將進一步推動算力資源的共享與整合，使台灣成為亞太地區的重要技術樞紐。民間企業的積極投入顯示出市場對於算力需求的旺盛，這將促使台灣在全球科技競爭中占據更加有利的地位。

#### 5. 整體分析與未來展望

從統計結果可以看出，台灣的算力分布存在明顯的領域差異。政府機構無論在算力總量還是技術先進性上都遙遙領先，這與其在國防、氣象和重大科研項目

上的投入直接相關。相對而言，公私立（技職）大學和醫學中心的算力雖然規模較小，但它們在科研應用中的作用同樣重要，特別是在 AI 和醫學領域。

我國在 Top 500 榜單中的表現，特別是 Taipei-1 的成功，進一步鞏固了我國在全球高性能計算領域的地位。未來，隨著民間企業資料中心的建成，台灣的整體算力將得到顯著提升，並在全球科技競爭中占據更加有利的地位。透過持續的技術投入與跨領域合作，我國有望在 AI、大數據和雲計算等前瞻領域保持競爭優勢，並推動國內外技術創新與發展。

## 二、我國 HPC 發展現況

HPC 是算力中最前瞻的技術領域，是算力發展之核心，主要透過強大的計算能力來解決科學與工程中的複雜問題。HPC 資訊技術的範疇中較為專業，其底層架構和技術，往往是主流資通訊技術發展的前瞻與核心，因此是國家發展資通訊技術時，需優先關注的領域。此外，HPC 應用橫跨多個領域，除傳統用於資通訊、科學研究和工程應用外，近年隨雲端運算的快速興起，HPC 技術也逐漸被廣泛應用於各產業發展上。

由於 HPC 強大算力，常被視為國家提升科技競爭力的關鍵，因此，當國家「算力即國力」之概念時，常以 HPC 算力為核心。HPC 算力量測基準上，通常以每秒浮點運算次數（FLOPS）作為衡量標準，國際上則以全球 Top 500 超級電腦排行榜，作為衡量何謂超級電腦算力的門檻，但此門檻隨技術不斷演進，呈現快速變動狀況。如自 2019 年起，全球 Top 500 超級電腦排行榜已將每秒 PetaFLOPS（千萬億次浮點運算）作為基本門檻，PetaFLOPS 對處理器、記憶體、架構、I/O 以及網路連接等方面，相對 GigaFLOPS 有截然不同的要求，HPC 系統發展，持續進行連續性創新與突破。

隨著算力需求不斷增長，HPC 於 2022 年 6 月正式進入百萬兆級計算 Exa-scale 的時代。透過前述所提，結合異構計算技術，如 GPU 和其他專用加速器，HPC 因為大幅提升運算效率，使科學模擬、AI 應用和大規模數據處理有了突破性發展，也因此造就今日 AI 的發展。

### （一）台灣 HPC 發展歷程

台灣 HPC 發展與台灣經濟轉型息息相關。1980 年代，台灣由於勞動成本上升及石油依賴，製造業競爭力減弱，台灣開始轉向技術密集型產業，經濟轉型後，需要快速提升國內科技與人力水平，因此大量招募海外學者回台。許多海外學者回國後發現，HPC 對國家科學發展至關重要，因此 1987 年開始，學者和政府部門開始提出發展高效能計算的需求。1990 年，台灣評估顯示超級電腦的潛在用戶已達 1,000 人，達到設立國家級高速電腦中心的標準，因此於 1991 年成立「國家高速電腦中心」，做為國內主要 HPC 推動專責單位，台灣 1987-2024 年 HPC 發展里程如表 3.2。

表 3.2、台灣 1987-2024 年 HPC 算力發展里程碑

年份	里程碑
1987	國科會提出《高效能計算（HPC）可行性分析》，為台灣首度提出發展 HPC 計劃。
1988	教育部諮詢辦公室提出《超級計算機需求及設置評估》。
1991	國科會成立「國家高速電腦中心(國家高速網路與計算中心前身)」，成為台灣首個 HPC 算力專責單位。
1993	中央氣象局 HPC 主機進入全球超級電腦 Top500 第 171，使台灣成為當時亞洲領先的 HPC 國家之一。
1999	國家高速電腦中心建置 IBM SP2，使台灣首次進入全球前 100 大超級計算機榜單（第 76 名）
2002	國家高速電腦中心和中央研究院推動台灣發展格網計算(Grid Computing)，由 HPC 的應用範圍擴展到格網計算。
2003	國家高速電腦中心轉制為財團法人，更名為國家高速網路與計算中心(簡稱國網中心)，自製建置叢集電腦“福爾摩沙 1 號”，位居 Top 500 排名第 135 名。
2004	台灣線上遊戲廠商 HPC 算力進入 Top500 榜單，居 243 名。
2007	台灣共有 11 台計算主機進入 Top500 榜單，創下台灣最多台數入榜之紀錄。其中 9 台由產業界建置，包含半導體公司、系統整合商、計算機輔助工程（CAE）公司和遊戲公司。台灣在超級電腦數量和總計算能力方面達到 HPC 發展的巔峰，以建置台數排名全球第 6 和第 10。
2009	台灣無任何一台計算主機進入 Top500 榜單，故引起國內科技界針對台灣的 HPC 發展進行討論與反省
2011	國網中心建置御風者 ALPS，進入 Top500 中第 42 名，同時進入超級電腦綠色排行榜(Green 500)中的第 25 名。同時，國網中心自主開發福爾摩沙 4 號，進行 Top500 第 234 名（2011 年 11 月）
2017	國網中心建置台灣首台 PetaFLOP 級別的超級計算機台灣杉一號，Top500 中排名第 96
2018	國網中心建置國內首座 AI-HPC 主機台灣杉二號(TAIWANIA 2)，居 Top 500 中第 20 名
2023	中央氣象署二台主機入榜，於 TOP500 中排名第 8 名與 139 名
2024	台灣首次業界超級電腦 Taipei 1 於 Top 500 排名第 38 名，首次有台灣為名義之業界 HPC 進入 Top 500 前 50 名

資料來源：本研究整理

## (二) 台灣 HPC 主要建置單位

台灣 HPC 算力主要建置者，可依主機效能分為一般大學、特定研究機構、企業和國家高速網路與計算中心，如表 3.4。大學部分主要少數幾個大學研究中心自行建置，供大學內部特定研究團隊使用；特定研究機構則如中央氣象署（前身為中央氣象局）、中央研究院、國家衛生研究院等，提供對其專屬研究計畫內部使用；企業部分則自行單位使用不對外公開；國家高速網路與計算中心（簡稱國網中心）是唯一國家級 HPC 設施提供者，且為國家共用、開放之 HPC 計算資源，因此除其提供學術研究者使用外，另也開放產業應用，是國內發展高效能計算技術與服務的主要機構。

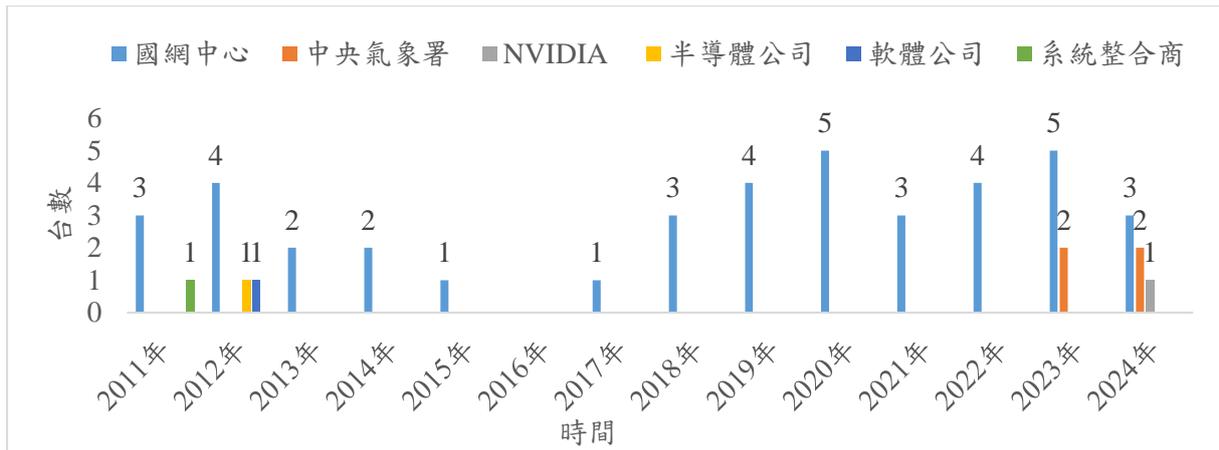
表 3.3、台灣 HPC 主機建置單位

類型	單位	說明	主要使用者
第一級 Tier 1	國家高速網路與計算中心	提供大尺度運算，為國家共用計算設施，支援學術研究及產業使用。推動台灣高效能計算技術與服務的主要單位。	一般學術研究為主，亦可提供政府或產業使用
第二級 Tier 2	大學	受台灣教育部頂尖大學計畫影響，部分國立大學在 2006-2010 年間建置大型計算設施，主要供校內小量計算使用。	特定大學內部使用，主要是小量計算及特定研究題目
專屬設施	研究機構	特定研究機構如中央氣象署、中央研究院、國衛院等，建置專屬高效能計算設施以應對特定研究主題需求。	特定研究機構為因應特定研究需求而使用
產業界	公司專屬設施	專屬於特定公司內部的計算設施，通常不對外公開。	公司內部使用，不公開

資料來源：本研究整理

由於 HPC 建置與維護成本高昂，除需各種不同領域專業人力相互合作外，從機房、網路、安全與對外服務等等門檻高，儘管許多單位有意願建置，但非一般單位有能力長期維運，因此台灣以國家實驗研究院之國家高速網路與計算中心（簡稱國網中心）為主要建置與服務者，確保國家具有長期維運 HPC 的能力，因此方能在 2011-2024 全球 Top 500 排行榜中，每年二度更新中，台灣累積共 48 次主機上過 Top 500 排行榜（註 1），其中國網中心佔 40 次，台數佔 83%，如圖 3.1。

此外，2024 年為因應 AI 應用的迅速增長，產業界也積極投資於大型資料中心的建設。包含 NVIDIA 與經濟部合作的超級電腦 Taipei-1 在高雄亞洲新灣區建置完成，算力達 22.3 PFLOPS，並在 2024 年 6 月的全球 TOP500 超級電腦榜單中排名第 38。Taipei-1 除為台灣的產官學研單位提供計算力外，還將推動生成式 AI、大型語言模型及數位孿生等前瞻技術的研發。另外友歲超級運算、GMI Cloud、翔耀（AICC 算力中心）及鴻海，也積極投入算力建置，將進一步推動台灣的 HPC 與 AI 技術發展。



資料來源：本研究整理；統計時間至 2024 年 6 月；

註 1:Top 500 排行榜每年更新二次，分別於 6 月於歐洲舉行的 ISC 會議中，以及 11 月美國超級電腦大會 (SC) 同步公布。因此本計算維累積年度總入台數，即當年度重複入榜者計算二次。

圖 3.1、2011-2024 年台灣進入 Top 500 之建置單位與台數

### 三、我國 HPC 算力國際競爭分析

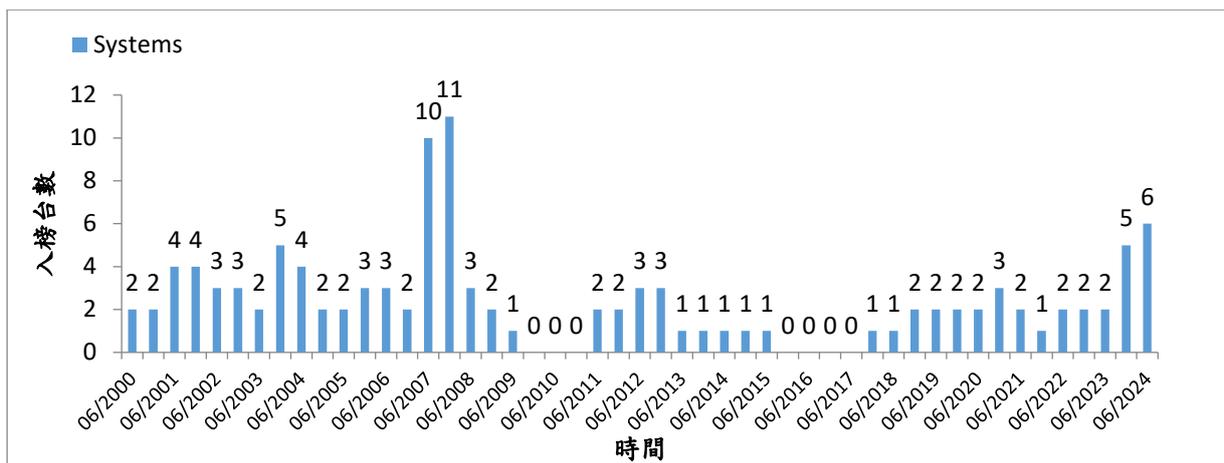
#### (一) Top 500 全球超級電腦排行榜概況

全球超級電腦 Top 500 排行榜為國際上公認之超級電腦 (Supercomputer) 排行榜，自 1993 年開始統計與發表年度全球超級電腦計算效能排名，並分別於每年 6 月及 11 月發佈排行。其根據每台超級電腦在 LINPACK Benchmark 中的效能表現進行排名，該測試主要衡量系統在解決線性方程組時的計算速度，即每秒浮點運算次數 (FLOPS)。LINPACK 基準測試，由 Jack Dongarra 等人引入，是衡量 HPC 系統計算能力的標準方法。由於 LINPACK 的廣泛使用，且能適用於多數 HPC 架構與系統效能測試，因此成為全球 TOP500 超級電腦排行榜的基準測試。LINPACK 基準測試的核心是進行 LU 分解 (Lower Upper Decomposition) 並採用部分選擇主元法 (Partial Pivoting)，並要求在計算過程中使用雙精度浮點運算 (FP64)。該測試的效能指標主要是包括效能最大之實際算力值 Rmax，其反映系統在解密集線性方程組的效率，但不完全代表整個系統的效能。

Top 500 排行榜不只反映目前國際上 HPC 算力技術方向，同時也成為各國科技實力的重要象徵，許多國家視其為提升國家競爭力的戰略資產，儘管台灣發展 HPC 算力重視實際應用效益，國際競爭僅為參考指標，然於分析 Top 500 台灣主機上榜狀況，仍對於台灣算力分析具有實際效果。

#### (二) 台灣與全球 HPC 競爭分析

截至 2024 年 6 月資料為止，我們統計台灣歷年進入全球超級電腦 Top 排行榜之電腦數，我國有 6 台進入 Top500 排行榜，詳如圖 3.2



資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

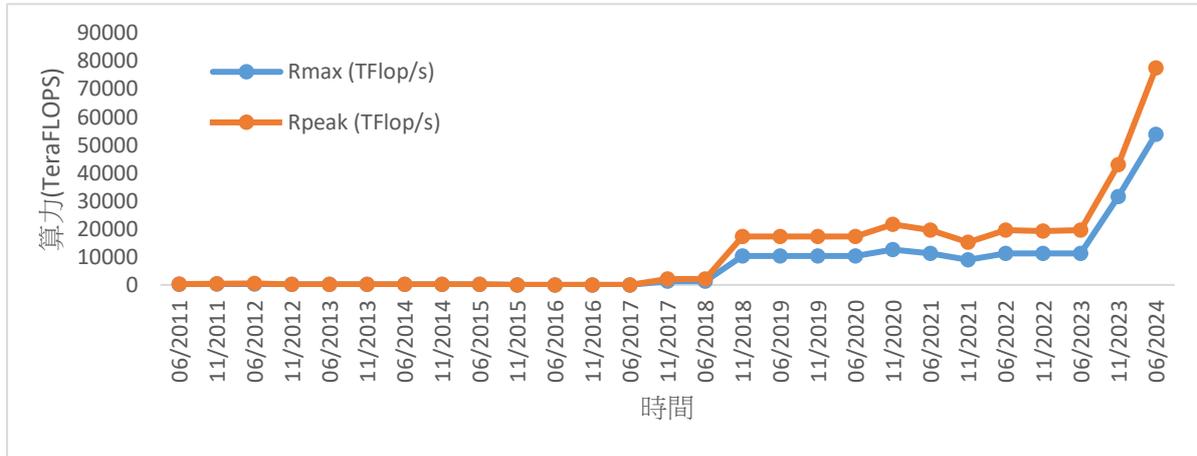
圖 3.2、台灣歷年進入全球超級電腦 Top 500 排行榜之電腦台數

從全球 Top 500 超級電腦排行榜歷年發展來看，台灣在 HPC 算力的投入呈現不穩定的狀況，主要以國家高速網路與計算中心和中央氣象署代表台灣進入 Top 500，尤其以國網中心為主要 HPC 算力提供者。國網中心於 1994 年首次入圍 Top 500，該主機於 Top500 中排名第 347 位；1996 年首次進入前 100 名，排名第 76；2002 年，台灣超級電腦排名第 60。2004 至 2006 年，隨著台灣線上遊戲產業的急劇擴張，對 HPC 的需求大幅增加，HPC 主機建置量激增。再加上叢集技術（PC Cluster）的成熟，台灣逐步以自建 PC Cluster 取代傳統的小型 HPC，許多大學的科學與工程學系也開始廣泛採用中小型 PC Cluster。2007 年是台灣 HPC 發展的全盛時期，共有 11 台超級電腦進入 Top 500，排名最高的是國網中心的 IRIS 超級電腦，位居全球第 35 名。

從圖 3.2 亦可發現，隨著全球超級電腦競爭的日益激烈，2008 年全球超級電腦正式宣告由 TeraFLOPS 進入 PetaFLOPS 時代，惟台灣的 HPC 算力在相對於國際快速發展速度，仍呈現倒退。到了 2009 年 11 月，台灣的超級電腦完全退出 Top 500 排行榜。為了維持國家的算力基礎，國網中心開始向各學術單位調查未來研究需求，並估算台灣的運算需求。隨後，國網中心在 2009 年的全國科技會議中提出建議，並成功爭取到預算，建置了 177TFLOPS 的御風者超級電腦，於 2011 年 6 月重返 Top 500，排名第 42 名。同時，該系統針對國內學術需求，設計了總容量達 PB（Petabyte）級別的儲存系統，上線僅半年便滿載，顯示出國內對算力需求的迫切性。

2018 年，國網中心建置台灣杉 1 號(Taiwania 1)，這是台灣第一台 PetaFLOPS 級別的 HPC 主機，主要以 CPU 為主。同時，國科會觀察到 AI 技術的突破性進展，加強了對 HPC 的投資，並於 2019 年由國網中心建置了台灣杉二號(Taiwania 2)，專為 AI 運算設計，主要以 GPU 為核心，也成為後來國內推動 AI 發展主要算力來源。2020 年後，台灣杉三號與四號進一步增強 CPU 為主之 HPC 算力，滿足國內科學與工程應用需求。

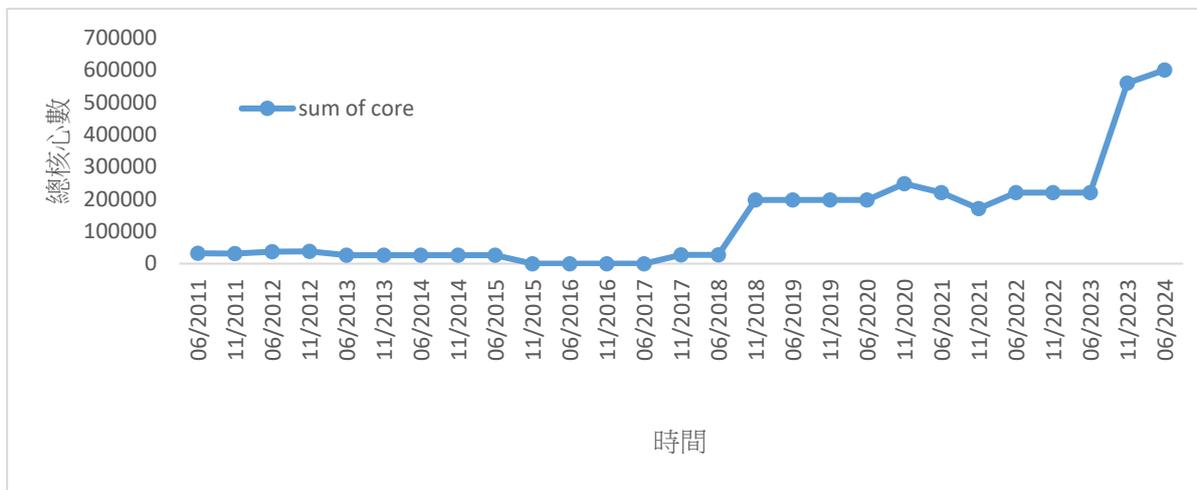
統計 2011 至 2024 年全球超級電腦 Top500 排行榜之台灣 HPC 算力總量，詳如圖 3.3。其中，Rmax 表示實際值算力，而 Rpeak 表示理論值算力。從圖 3.3 可發現台灣於 2018 年所投入之台灣杉二號後算力之突增，以及 2023 年左右，國網中心與中央氣象署等政府機關努力，並與國際大廠如輝達 (Nvidia) 合作，投入龐大資源建置台灣算力之基礎建設，大幅增加算力總量。



資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

圖 3.3、2011-2024 年全球超級電腦 Top 500 排行榜之台灣 HPC 算力總量

另由於總核心數係表示平行計算能力、可加速運算速度、資源分配具靈活性以及可處理大規模數據集等重要資訊，故亦統計台灣 HPC 總核心數，詳如圖 3.4。從圖 3.4 可發現，台灣 HPC 之總核心數亦大幅增加。



資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

圖 3.4、2011-2024 年全球超級電腦 Top 500 排行榜之台灣 HPC 總核心數

### (三) 2024 年台灣 HPC 算力現況

2024 年起，隨著大型語言模型 (LLM) 等新興技術需求的快速增長，台灣正式進入 AI 算力時代，2024 年 6 月台灣 HPC 於 Top 500 共計 6 台主機上榜，詳如表 3.4，除國網中心外，輝達公司 Nvidia 建置 Taipei-1、氣象署第六代超級電腦等總計實際值算力 Rmax 53.89 PetaFLOPS，理論值算力 Rpeak 77.56 PFLOPS，6 台

主機加總共有 600,688 個計算核心。

表 3.4、2024 年 6 月台灣 HPC 算力 top 500 主機清單

名次	建置單位	系統規格	核心數	效能 實際值 Rmax (PFLOPS)	效能 理論值 Rpeak (PFLOPS)	耗能 (kW)
38	輝達公司	TAIPEI-1 - NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR, Nvidia	40,960	22.3	34.53	-
86	中央氣象署	PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu	184,320	11.16	12.98	674
106	財團法人國家實驗 研究院國家高速網 路與計算中心	台灣杉二號 Taiwania 2 - QCT QuantaGrid D52G-4U/LC, Xeon Gold 6154 18C 3GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100 SXM2, Quanta Computer / Taiwan Fixed Network / ASUS Cloud	170,352	9	15.21	798
160	中央氣象署	PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D, Fujitsu	92,160	5.6	6.49	350
251	財團法人國家實驗 研究院國家高速網 路與計算中心	創進一號 Forerunner 1 - ASUS RS723Q-E11-RS24, Xeon Platinum 8480+ 56C 2GHz, Infiniband NDR200, ASUSTeK / ASUS Cloud / Taiwan Web Service Corporation	62,496	3.53	4	615
456	財團法人國家實驗 研究院國家高速網 路與計算中心	台灣杉三號 - QCT QuantaPlex T42D-2U/4N, Xeon Platinum 8280 28C 2.7GHz, InfiniBand HDR 100, Quanta Computer / Taiwan Fixed Network / ASUS Cloud	50,400	2.3	4.35	564

資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

此算力相對於國際上其他國家，根據 Top 500 排名之實際值算力 Rmax 總效能來看，台灣居全球第 17 位，詳如表 3.5，在亞洲名次第 4，居於中國、日本和韓國之後，韓國總效能（Rmax）是台灣的 3.5 倍。以台數來看，台灣共有 6 台主機進入 Top500，以 500 大排名上的各國進入排行榜之主機數量來說，台灣在全球排名 16，亞洲地區排名第 4（詳如表 3.6）。相對於中國以 80 台入榜主機位居全球第 2 和亞洲第 1，日本則以 29 台入榜主機排名全球、亞洲第 2，而韓國擁有 13 台入榜主機，位居全球第 7、亞洲第 3，台灣入榜主機數量與前三名差距明顯，尤其以人口與 GDP 規模計算，台灣與韓國算力仍有很明顯差距。

表 3.5、2024 年 6 月 Top 500 排行榜各國算力排名

算力效能 排名	國別	數量	前 500 大 列榜台數 (%)	效能實際值 Rmax (PFLOPS)	效能理論值 Rpeak (PFLOPS)	核心數
1	美國	171	34.2	4,408,382,050	6,968,768,506	47,911,528
2	日本	29	16	672,085,510	888,124,068	11,077,432
3	芬蘭	3	8	391,388,310	546,192,585	3,116,992
4	中國	80	5.8	356,860,484	627,663,427	21,128,112
5	義大利	11	4.8	332,618,330	436,291,218	4,079,600
6	瑞士	5	3.2	312,137,040	412,264,920	1,972,192
7	德國	40	2.6	274,903,370	421,218,896	4,399,212
8	法國	24	2.2	232,747,500	363,366,104	4,346,072
9	西班牙	3	2	221,872,600	306,102,991	1,542,016
10	南韓	13	1.8	186,445,000	246,067,510	2,293,652
11	沙烏地阿拉伯	8	1.6	99,398,340	152,763,114	2,737,652
12	英國	16	1.6	89,372,064	151,586,734	2,047,152
13	荷蘭	9	1.6	84,723,520	147,850,060	847,392
14	俄羅斯	7	1.4	73,715,000	101,737,460	741,328
15	波蘭	8	1.4	65,751,460	104,597,510	438,944
16	澳洲	5	1.2	55,638,310	74,947,805	498,848
17	台灣	6	1	<b>53,895,080</b>	<b>77,560,100</b>	<b>600,688</b>
18	巴西	8	1	53,621,650	101,912,142	806,656
19	瑞典	7	1	50,090,910	72,173,166	448,704
20	加拿大	10	0.8	41,208,360	71,911,529	845,984

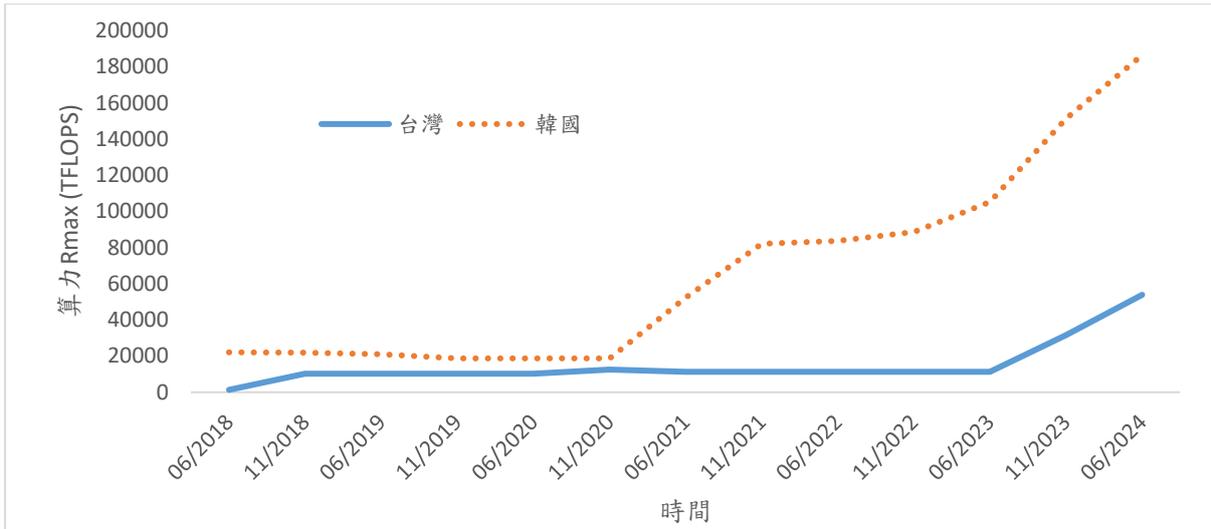
資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

表 3.6、2024 年 6 月全球超級電腦 Top 500 各國上榜台數

台數排名	國別	數量	前 500 大 列榜台數 (%)	效能實際值 Rmax (PFLOPS)	效能理論值 Rpeak (PFLOPS)	核心數
1	美國	171	34.2	4,408,382,050	6,968,768,506	47,911,528
2	中國	80	16	356,860,484	627,663,427	21,128,112
3	德國	40	8	274,903,370	421,218,896	4,399,212
4	日本	29	5.8	672,085,510	888,124,068	11,077,432
5	法國	24	4.8	232,747,500	363,366,104	4,346,072
6	英國	16	3.2	89,372,064	151,586,734	2,047,152
7	韓國	13	2.6	186,445,000	246,067,510	2,293,652
8	義大利	11	2.2	332,618,330	436,291,218	4,079,600
9	加拿大	10	2	41,208,360	71,911,529	845,984
10	荷蘭	9	1.8	84,723,520	147,850,060	847,392
11	波蘭	8	1.6	65,751,460	104,597,510	438,944
12	巴西	8	1.6	53,621,650	101,912,142	806,656
13	沙烏地阿 拉伯	8	1.6	99,398,340	152,763,114	2,737,652
14	瑞典	7	1.4	50,090,910	72,173,166	448,704
15	俄羅斯	7	1.4	73,715,000	101,737,460	741,328
<b>16</b>	<b>台灣</b>	<b>6</b>	<b>1.2</b>	<b>53,895,080</b>	<b>77,560,100</b>	<b>600,688</b>
17	澳洲	5	1	55,638,310	74,947,805	498,848
18	瑞士	5	1	312,137,040	412,264,920	1,972,192
19	挪威	5	1	15,955,310	24,963,990	425,920
20	愛爾蘭	4	0.8	11,394,210	22,450,170	290,560

資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

由於 HPC 算力通常與該國 GDP、人口數、科技人口數與其科技政策為對應指標，相對之下，韓國 HPC 算力位居亞洲第三，其人口數與 GDP 大約為台灣 2 倍，科技發展路線相近，可作為台灣比較之對象。從 2018 年迄今，韓國算力平均為台灣 4 倍，詳如圖 3.5，其中 2018 年差異達到 16 倍，2023 年 6 月達 9.3 倍。究其原因，除與其研究人口數較高有關，更重要與其國家長期發展政策有直接關係。



資料來源：全球 500 超級電腦排行榜，www.top500.org；本研究整理。統計時間至 2024 年 6 月

圖 3.5、台灣與韓國 HPC 算力比較

目前，國內普遍以進入 Top 500 排行的大型主機力，作為衡量國家之國際算力實力指標，此方法存在兩個限制。首先，進入 Top 500 排行主機，必須提交大型計算效能的實測數據，因此在進行實測時，需建置高速網路連接系統來串接所有計算節點，如 Infiniband、Slingshot、Ultra Ethernet Consortium (UEC) 等。此類設施的建置成本昂貴，因此對於非專注於大型計算服務的產業界，出於成本效益與商業機密的考量，未必會選擇提交主機到 Top 500 排行，因此 Top 500 無法完全代表一國的整體算力。其次，算力建置需要根據實際需求來考量，以確保資源投資的效益。各國在科研人口數、軟體發展等各有不同，算力的建置應與其需求匹配，而不宜單純以硬體排名進行國際間的直接比較。

#### 四、小結

##### (一) 算力盤點

本次統計報告深入分析了台灣在公私立（技職）大學、醫學中心和政府機構三大領域的算力配置，並探討了民間企業即將建設的資料中心對未來算力發展的潛在影響。透過詳細的數據收集和分析，我們可以清晰地看到台灣在全球科技競爭中的優勢與挑戰。

首先，台灣政府機構在算力領域佔據了絕對的領先地位。政府機構不僅擁有全國最大的算力總量，而且在 Top 500 榜單中的表現也尤為突出，特別是 Nvidia TAIPEI-1 的成功展示了台灣在 AI 和高性能計算領域的技術實力。這些計算資源不僅在國防和安全領域發揮關鍵作用，也為重大科研項目提供了堅實的基礎。

其次，公私立（技職）大學和醫學中心雖然在算力總量上不如政府機構，但它們的設備配置和應用場景顯示出強大的科研潛力。這些機構的算力主要應用於 AI 研究、基因組學、物理模擬和醫學影像處理等領域，對推動台灣科技進步具有重要意義。隨著 AI 技術的不斷發展，我們預計這些大學和醫學中心將在未來繼續擴大其算力資源，並加強與政府和企業的合作，促進更多創新應用的落地。

最後，民間企業對於算力的投資顯示出市場對於高性能計算和 AI 技術的強烈需求。友歲超級運算、GMI Cloud、翔耀（AICC 算力中心）和鴻海這四大資料中心的建設，將在未來大幅提升台灣的整體算力水平，並推動台灣在 AI、大數據分析和雲端計算等領域的技術進步。這些資料中心的建成，不僅將加強台灣在全球科技競爭中的地位，也將為國內外企業提供更多的技術支持，促進數字經濟的發展。

總體來說，台灣在算力領域已經具備了強大的基礎，未來的發展潛力巨大。通過持續的技術投入和跨領域的合作，台灣有望在全球高性能計算和 AI 技術的前沿保持領先地位，並推動更多創新和技術突破。未來，我們期待看到台灣在全球科技舞台上發揮更大的影響力，並為國際社會的技術進步作出更大貢獻。

##### (二) 我國 HPC 發展與排名

由於 HPC 技術更新速度極快，尤其 GPU 世代發展更快過 CPU，許多大國如美國和日本擁有多個 HPC 中心，能夠由多個單位分別建置算力基礎設施，這其國家算力發展和更新上，具備較高的靈活性和持續性。相較之下，中小型國家通常僅有 1-2 個 HPC 中心，一旦缺乏長期的建設規劃，該單位 HPC 算力面臨老化或因汰換問題，導致國家算力中斷，對國家科技穩定發展造成不利影響。因此，許多國家開始正視算力問題，並思考如何以政策與長期預算，作為推動國家 HPC 算力後盾。

例如，韓國 2020 年 12 月通過制定《國家超高效能電腦應用及育成法》，詳

如表 3.7。韓國以法律形式，強制規定中央行政機關每年需依照基本計畫推動 HPC 發展，確保國家算力的長期穩定。而台灣隨著 AI 技術的興起，國科會於 2024 年推出了《人工智慧基本法》草案，強調政府在 AI 研發和基礎建設方面的推動，並通過補助、獎勵和租稅優惠等政策促進 AI 產業的發展。但可惜針對 HPC 算力本身，台灣目前尚缺乏明確的長期政策支持。在全球 HPC 競爭日趨激烈的背景下，若要將「算力即國力」的觀念落實，政府必須進一步從法規或政策層面進行長期規劃，規劃穩定且長期的預算規劃，對此至關重要，確保 HPC 基礎設施得以持續穩定發展，以確保台灣在全球 HPC 競爭中的持續發展力。

表 3.7、韓國超高性能電腦法

名稱	《國家超高性能電腦應用及育成法》簡稱《超高性能電腦法》 《국가초고성능컴퓨터 활용 및 육성에 관한 법률》
內容	<p>第 1 條：有效建立、有系統的管理國家超高性能電腦，持續帶動科技發展</p> <p>第 3 條：國家應制訂「國家超高速運算發展政策」</p> <p>第 5 條：科技資通訊部應制訂「國家超高速運算育成基本計畫」並推動</p> <p>第 6 條：中央行政機關每年依照基本計畫訂出實行計畫並推動</p> <p>第 7 條：設置隸屬於科技資通訊部部長的「國家超高速運算委員會」</p> <p>第 9 條：為了推動國家超高速運算發展，依照總統命令設立「國家超高速運算中心」</p> <p>第 9-2 條：科技資通訊部依照總統命令，可在各領域運作「超高速運算中心」</p> <p>第 10~15 條提到擴大研發投資、取得資源、人才培育、建設尖端研究網路、資訊蒐集與普及、研發活動調查</p> <p>第 16~21 條是推動產學研與國際合作</p>

資料來源：<https://www.law.go.kr/LSW/lsInfoP.do?lsiSeq=218779#0000>，本研究整理

參考文獻

1. 科技政策研究與資訊中心: ‘全球高速計算科技發展趨勢分析’ (2009. 2009)
2. ATIP: ‘HPC in Taiwan’, in Editor (Ed.)^(Eds.): ‘Book HPC in Taiwan’ (2012, edn.), pp.
3. GPU Specs Database, <https://www.techpowerup.com/gpu-specs/>
4. Linpack, <https://top500.org/project/linpack/>
5. NVIDIA H100 Tensor Core GPU datasheet, <https://resources.nvidia.com/en-us-tensor-core/nvidia-tensor-core-gpu-datasheet>
6. NVIDIA A100 TENSOR CORE GPU datasheet, <https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet.pdf>
7. NVIDIA TESLA V100 GPU ACCELERATOR <https://images.nvidia.com/content/technologies/volta/pdf/tesla-volta-v100-datasheet-letter-fnl-web.pdf>
8. NVIDIA GB200 NVL72 <https://www.nvidia.com/en-us/data-center/gb200-nvl72/>
9. Top500 <https://top500.org/>

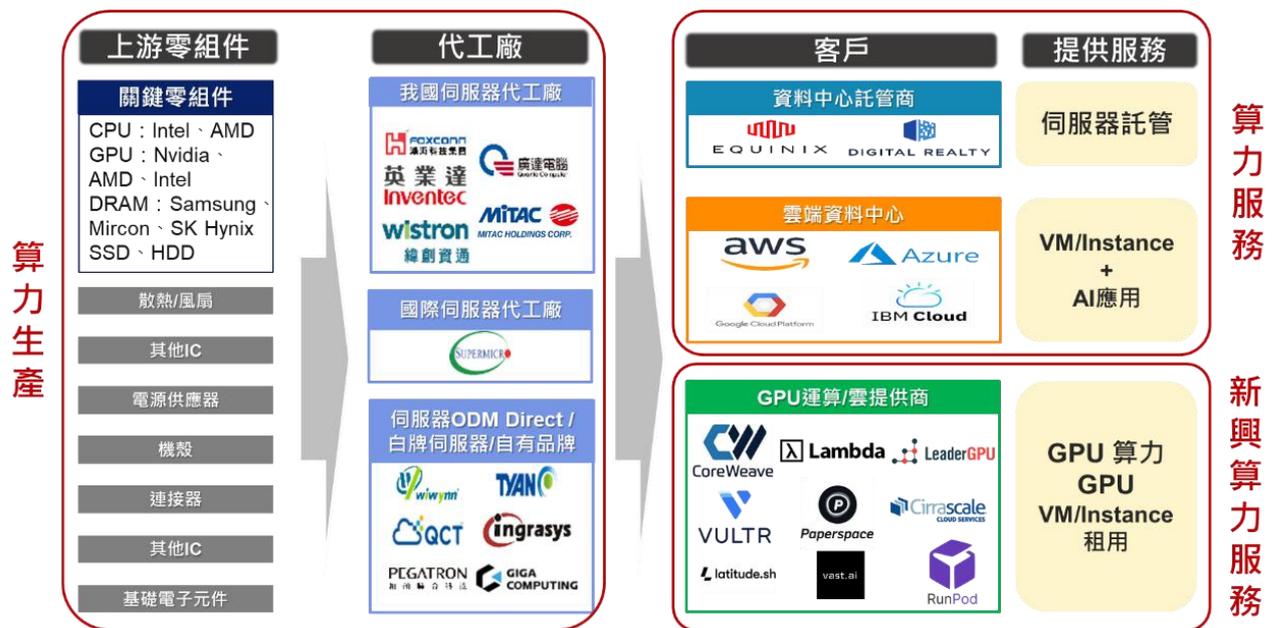
## 第四章 我國算力系統生態系分析

前面章節陸續討論算力是什麼，算力設備如何組成，以及盤點台灣算力設備。本章將從產業生態系的角度切入，延續算力的介紹與討論，讓模不著邊際的算力，透過更具體的方式，說明算力如何在產業裡發生影響。本章節將探討算力系統之生態系及全球及台灣算力生態系與資源分布之比較分析。

### 一、 算力系統生態系定義

自 2023 年初 ChatGPT 帶動生成式 AI 應用的發展，雲端服務商積極佈局大型語言模型 (Large Language Model, LLM)。至 2023 年底，雲端服務商大量採購高階 AI 伺服器來增強自身於資料中心的算力部署，持續推出新的語言模型與生成式 AI 應用。國防、科研等政府單位亦開始擴充自身的算力。自此，「算力」正式成為各主要國家在數位時代中，用以衡量資訊國力的重要指標之一。

在全球算力系統生態系當中 (如圖 4.1)，要使算力能夠實際投入到應用端，主要包含幾個關鍵角色，分別為「算力生產」、「算力服務」與「新興算力服務」。算力生產的廠商類型主要有上游零組件及伺服器代工廠，提供算力服務的廠商，是由雲端資料中心提供虛擬機器 (Virtual Machine, VM) 或應用實例 (Instance) 供企業按需租用雲端算力，而當前在 AI 算力受到重視之下，新興算力服務商提供的 GPU 算力服務開始如雨後春筍般浮現。本章將針對此三大角色在全球與台灣的重要廠商進行描繪，並且進行比較分析，藉此了解台灣在全球算力系統生態系的地位與重要性。



資料來源：MIC 整理

圖 4.1、全球算力系統生態系範疇

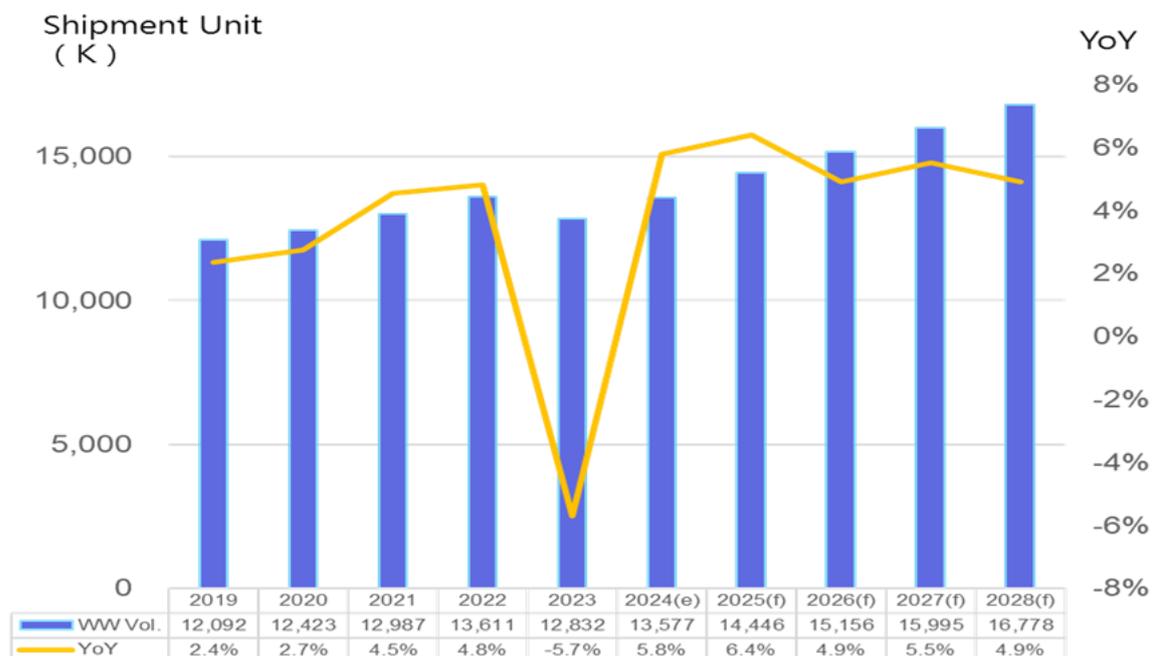
## 二、全球算力生態系發展現況

### (一) 算力生產

全球算力生產主要的參與業者包含處理器廠商、伺服器代工廠及伺服器品牌商等，以中央處理器（Central Processing Unit, CPU）為主的算力，不進行叢集架構，並由一般伺服器提供。若是以圖像處理器（Graphic Processing Unit, GPU）為主的算力，根據客戶需求進行叢集架構，用 AI 伺服器提供。高效能運算算力，則透過大量處理器的堆疊，並將伺服器運用叢集形成超級電腦架構。

受惠於生成式 AI 的應用逐漸落地，全球對於 AI 伺服器的需求大幅增加，亦促成全球整體伺服器出貨量的提升。經歷 2023 年全球伺服器市場受到俄烏戰爭衝擊企業降低資本支出，及資料中心轉型的影響，使全球整體出貨量出現罕見下滑。

2024 年受到雲端服務商持續建設資料中心並採購 AI 伺服器，GPU 運算提供商新建 AI 資料中心，及企業對於 AI 推論與中小模型微調（Fine-tune）的需求等影響，2024 年全球伺服器的出貨量（見圖 4.2）有望達到 1,357 萬台，並且至 2028 年有望提升至 1,678 萬台。

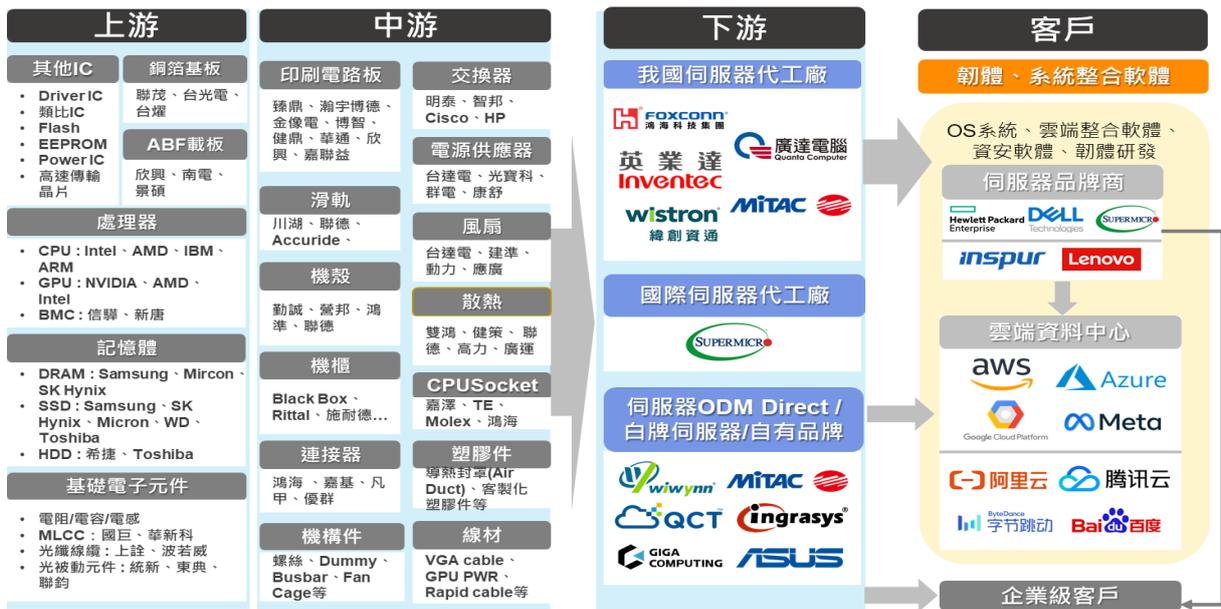


資料來源：MIC 整理

圖 4.2、2019~2028 年全球伺服器市場預測

若是細部探究算力生產的生態系，可以透過全球伺服器供應鏈來進行觀察，全球伺服器供應鏈（見圖 4.3）可以分為上游處理器、IC、記憶體與電子元件供應商；中游包含印刷電路板、機殼、連接器、電源供應器、風扇、散熱及其他零組件廠商；下游則為伺服器代工廠，依照代工性質可以分為伺服器品牌代工廠以及伺服器直接代工（ODM Direct）/白牌伺服器；客戶端包含伺服器品牌商及雲端服

務商及企業級客戶。



資料來源：MIC，2024年9月，備註：僅列舉代表性廠商

圖 4.3、全球伺服器供應鏈主要業者

### 1. 上游：處理器、IC、記憶體與電子元件

伺服器上游產品包含CPU、GPU及基板遠端控制晶片 (Baseboard Management Controller, BMC) 等處理器；驅動 IC (Driver IC)、類比 IC (Analog IC)、快閃記憶體 (Flash)、唯讀記憶體 (EEPROM)、電源管理 IC (Power IC)、高速傳輸晶片等積體電路 (IC)；銅箔基板、ABF 載板等印刷電路板及材料；動態隨機存取記憶體 (Dynamic Random Access Memory, DRAM)、固態硬碟 (Solid-state Drive, SSD)、硬碟 (Hard Disk Drive, HDD) 等記憶體；電阻/電容/電感、多層陶瓷電容器 (Multi Layer Ceramic Capacitor, MLCC)、光纖線纜、光被動元件等基礎電子元件。

在處理器方面，當前包含CPU、GPU等處理器主要提供者以美系企業為主。CPU的代表業者包含x86架構的Intel、AMD以及近年逐漸崛起的Arm架構處理器，包含AWS、NVIDIA均透過Arm架構打造自身的處理器。GPU方面當前的領導業者為NVIDIA，透過其推出的A100 (NVIDIA 2020年推出的資料中心GPU)、H100 (NVIDIA 2022年推出的資料中心GPU)等高階GPU，擁有市場八成以上的佔有率；AMD則是在近年積極研發MI250 (AMD 2022年推出的資料中心GPU)、MI300 (AMD 2024年推出的資料中心GPU)等GPU產品，希望能夠追上NVIDIA的腳步。而在BMC方面，當前是由台灣的信驊、新唐所製造，其中信驊佔有絕大多數的量。BMC作為伺服器基板的重要控制元件，是全球伺服器供應鏈的重要產品。

儘管台灣在CPU、GPU方面沒有代表性廠商，但在協助晶片與印刷電路板進

行連接的 ABF 載板，台灣有欣興、南亞電路板及景碩三家業者，其中欣興與南電的全球市佔率各自皆超過 2 成。當中欣興 ABF 載板占其營收達 80%-85%，顯示市場重要性；南電則是載板的比重超過 85%，當中包含高密度連接板與 ABF 載板。

在各類 IC 方面，主要皆放置於伺服器主機板上，透過表面貼焊技術 (Surface Mount Technology, SMT) 將其焊在主機板上。驅動 IC 中，台灣有陞達科技、茂達等企業進行布局；類比 IC 中台灣則有茂達、致新、矽力等企業提供；Flash、EEPROM 多為外國製造，由國際或台灣代理商進行購買；電源管理 IC 中，則有偉詮電、類比科、通嘉、茂達、矽力等多家廠商布局。

記憶體方面提供廠商同樣以國際企業為主，在 DRAM 方面包含三星 (Samsung)、美光 (Micron) 及 SK 海力士 (SK Hynix)；SSD 除前述三者外，亦包含威騰電子 (Western Digital)、東芝 (Toshiba) 等企業。HDD 方面則是由希捷 (Seagate)、威騰、東芝等企業進行提供。

## 2. 中游：各類零組件

伺服器產業中游的產品包含印刷電路板 (Printed Circuit Board, PCB)、滑軌、機殼、電源供應器 (Power Supply Unit, PSU)、風扇、散熱模組、CPU 插座 (CPU Socket)、連接器 (Connector)、線材 (Cable)、機構件及塑料件等產品。而若是要進行整櫃出貨的話，滑軌、機櫃及交換器等，亦需要與伺服器進行整合。

在印刷電路板方面台灣擁有多業者，包含臻鼎、瀚宇博德、金像電、博智電子、健鼎、華通、欣興及嘉聯益等企業，協助提供伺服器的主機板及其他小板的製造；機殼則是有勤誠、營邦、鴻準、聯德等企業進行布局，協助進行伺服器機殼結構的設計與製造。

電源供應器方面，包含台達電、光寶科、群電、全英及康舒等企業均有所布局，當中以台達電與光寶科為主要提供者。電源供應器隨著伺服器處理器功耗不斷提升，所需支援的瓦數亦持續增加進而帶動其發展。

風扇作為氣冷散熱的重要零組件，對於伺服器十分重要。台灣包含台達電、建準、動力科技、應廣科技均有所布局；散熱模組、均熱片等則是由超眾、雙宏、健策、建準、泰碩、奇鈺、聯德、高力及廣運負責提供，在散熱需求提升下，部分廠商亦開始研發直接式水冷及沉浸式水冷的解決方案。

其他如 CPU 插座有嘉澤、TE、Molex、鴻海等企業提供，連接器台灣則有鴻海、嘉基、凡甲及優群等。另外機構件包含螺絲、Dummy、匯流排 (Busbar)、風扇殼等；塑膠件包含導熱封罩 (Air Duct)、客制化塑膠件；線材則包含 VGA cable、GPU PWR、Rapid cable 等各式各樣的產品。

## 3. 下游：伺服器代工廠

當前全球伺服器由超過 8 成均是由台灣伺服器代工廠進行生產，並且透過兩種方式進行：過往伺服器代工廠協助伺服器品牌商進行代工，台灣主要業者包含英業達、鴻海、緯創、廣達、神雲等企業，協助伺服器品牌商客戶代工其所設計並提供規格的伺服器產品。

然而在雲端服務商崛起之下，未來能夠降低伺服器價格並且提升客制化程度，伺服器直接代工（ODM Direct）及白牌伺服器應運而生。台灣伺服器代工廠為了避免品牌商客戶的不滿，紛紛透過增設子公司或設立自有品牌的方式，來提供白牌伺服器產品。因此包含緯創分拆出來的緯穎（Wiwynn）、廣達旗下的雲達（QCT）、鴻海旗下的鴻佰科技（Ingrasys）、神雲均是透過此種方式，提供雲端服務商產品。

而在台灣白牌伺服器逐漸成熟下，廠商也開始研發自身的作業系統（Operating System, OS）、韌體系統，或是採用開源運算計畫（Open Compute Project, OCP）樣板，設計並製造伺服器直接提供給企業及用戶，變為自有品牌代表廠商如技嘉旗下的技鋼（Giga Computing）、華碩（ASUS）。

#### 4. 客戶：伺服器品牌商、雲端服務商及企業級用戶

全球伺服器供應鏈的客戶主要可以分為幾個族群，分別是美系伺服器品牌商、中系伺服器品牌商、美系雲端服務商、中系雲端服務商以及企業用戶等。美系伺服器品牌商包含 Dell、HPE、Supermicro 等全球前五的伺服器品牌商，以及 IBM、Cisco 等廠商；中系伺服器品牌商以浪潮跟聯想為主要代工廠商。

美系雲端資料中心以 Amazon、Microsoft、Google、Meta 為最重要的業者，當中 Amazon、Microsoft 及 Google 均有透過雲端平台提供雲端服務給全球的企業，Meta 則是將採購的伺服器放在自建的資料中心當中，藉此來訓練 AI 自然語言模型、改善其社交平台（包含 FB、Instagram、Threads 等）的演算法。其他廠商如 Oracle、Twitter、Uber、Apple 等亦為台灣伺服器代工場重要的合作夥伴。中系雲端服務商包含阿里巴巴（阿里雲）、騰訊、字節跳動及百度等，除台灣廠商直接代工以外，中系雲端服務商由於市場特性，許多會採買中系品牌商的伺服器。

### （二）算力服務

#### 1. 以 CPU 為主體的雲端算力服務

##### （1）全球雲端服務市場概述

隨著全球經濟數位化的持續推進，全球網路數據流量自 2019 年在 Covid-19 疫情期間成倍數成長，導因於居家上班、線上學習、影音串流、遊戲等使用量增加，推動網路數據成長，數位經濟來臨帶動雲端運算發展需求。目前全球網路數據以電信通訊、雲端服務為主要大宗，占比超過半數，其餘來自於銀行保險、證券交易、製造業、數位內容等產業。各產業領域的網路流量又以金融交易、社交

網路、數位內容（如 Netflix、YouTube、Tiktok）服務網路流量幅度最高。

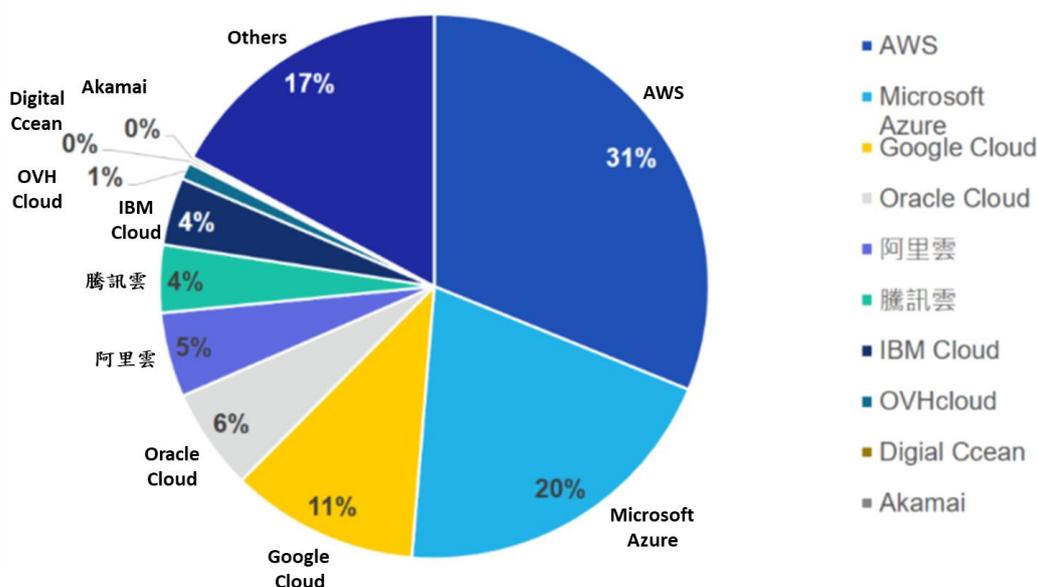
提供基礎設施即服務(Infrastructure as a Service, IaaS)、平台即服務（Platform as a Service, PaaS）為基礎的公有雲端服務模式，已是全球資訊產業領域中持續且快速發展之項目。隨著科技的不斷進步和企業對資訊科技的需求不斷增長，越來越多產業開始意識到數位化管理企業數據資料有助於提高管理效率，強化企業韌性，亦能借助雲端平台提升客戶服務品質，開拓新的商業服務機會。

在 2020 年到 2022 年全球飽受 COVID-19 影響期間，雲端服務市場成長來自於取代實體上班的數位替代需求，驅動全球企業提高對雲端運算服務使用需求。在 2022 年底 Open AI 推出 ChatGPT，生成式 AI 為雲端產業帶來「質」變，產業看到部署生成式 AI 應用，將能帶來更具效率的新生產工具商機。因此，自 2023 年起，全球雲端服務市場在生成式 AI 議題帶動下，進入新一波相比 COVID-19 期間更為劇烈、更具革命性的產業變革階段。首當其衝的是 Amazon Web Service（以下簡稱 AWS）、Google、Microsoft 等，大量採購 GPU 及自研 AI 晶片，開啟 AI 軍備競賽，雲端運算產業進入新一輪的競爭趨勢。

## (2) 雲端服務由 3 大美系公司主導市場

亞馬遜公司（Amazon）自 2006 年起最先推出雲端服務，經過多年發展，對雲端的需求已經無所不在，成為數位轉型的基礎設施。隨著市場規模成長，Microsoft、Google，中國大陸的阿里雲、騰訊雲等陸續加入雲端服務市場。2023 年全球雲端服務市場來看（見圖 4.4），主要由 AWS、Microsoft Azure、Google Cloud 美系三大巨頭包辦超過六成市場規模。

2023年公有雲大廠市場佔比



資料來源：Dgtl Infra，MIC 整理，2024 年 9 月

圖 4.4、2023 年全球前十大公有雲服務供應商

除了三大美系公司在全球雲端市場的攻城略地，中系的阿里雲、騰訊雲憑藉中國大陸及亞洲區域市場，擠進雲端服務產業排名的前十大。其他雲端服務公司如 Oracle cloud、IBM cloud 則是提供消費者處理、儲存、網路以及各種基礎運算資源的基礎架構服務，然投資金額已遠不如前三大美系公司，前十大當中唯一歐系雲服務商 OVHcloud 服務區域則以歐洲客戶為主。

由於在雲端服務產業，佈署 IaaS 算力需要投入大量資本，如在全球各地投資雲端資料中心、採購內部所需伺服器硬體設備，及開發軟體應用平台，導致其進入門檻極高，大者恆大。雲端基礎設施及軟體服務規模化後，市場逐漸形成寡占型態，較晚進入市場的 Google Cloud 仍苦苦追趕 AWS 及 Microsoft Azure，更不用說其他企業規模遠小於雲端服務三巨頭的雲端服務公司。尤其在人工智慧浪潮興起趨勢下，雲端巨頭大舉擴張全球資料中心及大肆採購 AI 伺服器運算設備，進一步墊高產業進入門檻。

### (3) 雲端服務商搶進佈署 AI 算力服務

在 2024 年，隨著生成式 AI 議題持續延燒，國際大廠 AWS、Microsoft、Google 相繼推進生成式 AI 雲端平台算力部署，投入更多資本支出用於採購大量 GPU 及自研 AI 晶片，作為 AI 訓練及 AI 推論的算力基礎，Oracle Cloud、Meta 亦跟進加碼採購 GPU 及 AI 伺服器。在國際大廠紛紛添材加火下，憑藉優勢算力推出具競爭力之 AI 大模型，並搭配 AI 平台服務，推出全新 AI 即服務 (AI as a service, AaaS) 支援企業開發及佈署生成式 AI 的資料處理。因此，雲端大廠的投入不僅是生成式 AI 能夠如此快速發展的關鍵，而各種雲端平台上的 AI 開發工具，更是讓 AI 應用能更快落地的關鍵資源。

有別於以 CPU 為主的雲端企業管理服務、數據資料庫、資料儲存與備份，AI 算力工作負載仰賴 GPU，進行 AI 訓練與 AI 應用開發。雲端服務大廠著重於強化雲端運算服務的 AI 佈局，持續拓展雲端高效能算力、增加多種生成式 AI 開發工具，與提供不同 AI 大模型選用機制，提供企業客戶一站式部署生成式 AI 所需的相關資源。Microsoft Azure 憑藉 GPT 模型之優勢，在 2023 年全球景氣低迷之際雲端部門營收逆勢成長，Google 亦推出 Gemini AI 大模型搶攻 AI 市場需求 (見圖 4.5)。Microsoft 及 Google 利用自家 AI 大模型，將生成式 AI 應用部署到的企業雲端協作工具，如 Microsoft 365 及 Google Workspace，拓展市場滲透力。而缺乏自訓大型 AI 模型的 AWS，則選擇與開源社群 Hugging Face 合作，提供企業客戶多種開源 AI 模型可供選用。



資料來源：各公司，MIC 整理，2024 年 9 月

圖 4.5、生成式 AI 產業效益影響雲端軟硬體生態系

## 2. 新興算力服務-基於 AI 運算的 GPU 算力出租

過往伺服器託管與算力租賃主要由兩大類型的廠商進行協助，分別是資料中心託管商（如 Equinix、Digital Realty 等）與雲端資料中心廠商（AWS、Azure、Google）。資料中心託管商主要提供資料中心的空間與機櫃，協助資料中心電力設備（UPS）、散熱、資安等服務的提供，伺服器則是由租賃櫃位的廠商自行採購，伺服器硬體設備所有權亦掌握於企業手中。雲端服務商則是在自建的資料中心當中放置伺服器，此些伺服器除了可以支撐雲端服務外，並透過虛擬化的方式提供虛擬機器（Virtual Machine, VM）、應用實例（Instance）給客戶進行租借。

2023 年生成式 AI 應用竄升促使各企業對 AI 算力的需求大幅提升，然而提供 AI 算力的 AI 伺服器具備單價過高、採購不易及對於其他設備規格要求更高等特點，企業要自行進行採購相對困難，且全球皆有供不應求的狀況發生。在此情形下，一批新型態的廠商逐漸浮出水面，即為租借 AI 算力、高階 GPU 算力的新興算力服務（GPU 運算）提供商。此些廠商在自身的資料中心部署大量的 AI 伺服器來集成 AI 算力，並將 AI 算力透過 VM 或 Instance 的方式提供需要使用的企業。與此同時，AI 算力增加對於資料中心託管商及雲端服務商形成不同程度的問題，讓 GPU 運算提供商找到可以切入的商機。

### (1) 國際新興 AI 算力服務提供商由四大類型業者轉型

當前新興算力服務提供商主要由四種業者轉型而成，包含裸機服務提供商、AI 演算法提供商、虛擬貨幣挖礦業者及超級電腦提供商/科研機構。裸機服務提供商過去提供 CPU、低階 GPU 的裸機伺服器（Bare Metal Server）服務，將實體伺服器擺放於廠商所建的資料中心，並租賃給客戶使用；而當前則是提供高階

GPU 算力的租借，當中代表廠商包含 VULTR、RunPod、LeaderGPU、Latitude.sh、vast.ai 等。

AI 演算法提供商則是原本提供 AI 演算法、機器學習 API 等產品，因自身擁有 AI 訓練的伺服器，逐漸開始拓展為提供高階 GPU 算力的解決方案與租借，代表廠商如 Lambda、Papersapce。虛擬貨幣挖礦者則是過去大量採購礦機來挖取虛擬貨幣，在虛擬貨幣價格大幅波動及生成式 AI 竄升的需求驅動下，轉型成為 GPU 運算的提供商，代表廠商為 Coreweave。最後則是超級電腦提供商與科研機構，過去提供自身打造或提供其他研究機構超級電腦或資料中心基礎設施，因具備設計叢集架構的能力，逐步進展到提供 GPU 算力的租借。

## (2) 多種 AI 算力採用模式

AI 算力成為要布局生成式 AI 的企業必須部署的重要基礎元件，在高階 GPU 短缺以及需求量增加下，AI 算力需求者將會有更多的採用方式。第一種模式即為向資料中心託管商租借機櫃空間，自行採買伺服器品牌商、ODM 廠商提供的 AI 伺服器產品或 NVIDIA DGX 系統。此種模式伺服器位於資料中心託管商，而伺服器為自身所擁有，建置成本是最高的，但擁有安全性最高、可以自行配置硬體設備等優勢。而若要使用雲端服務時，則可透過混合雲或託管商提供的串接服務來進行使用。

第二種模式為使用 AWS、Microsoft、Google 等雲端服務商的應用實例，伺服器位在雲端資料中心，儘管可以透過區域 (Region) 的選擇來確認伺服器位置，然伺服器主要為雲端服務商所擁有。此種模式無前期建置成本、可以直接使用，然而應用實例的租借費用較高，設備彈性較低。同時在雲端服務商的綁約策略下，要與其建立長期合約才會比較划算。而其主要優勢在於和雲端服務商其他雲端應用服務整合性高，可以一起進行採購，對於軟體設計能力較弱的公司具有吸引力。

第三種模式則為租借新興算力服務商的 AI 算力，伺服器位於 GPU 運算提供商，且伺服器為廠商所擁有。其優點包含無前期建置成本、設備彈性相對較高，可以透過廠商給予的規格選擇需要的伺服器、能夠即時使用，因此短期租用成本較低。然而當前許多廠商仍鎖定在區域進行供貨，全球布點尚未完善，因此部分國家的供給數量相對較低。另外，伺服器是共用的，在一個客戶租完之後即會開放給其他客戶進行使用，資安相對不佳，而若要使用雲端服務則需額外進行串聯。

## (三) 從全球算力市場看應用領域分布

根據 Fortune Business Insights，將 HPC 定義為具備強大處理能力的先進運算系統，具備多核心處理器 (CPU)、圖形處理單元 (GPU) 及加速器，以執行大規模模擬、數學運算、數據分析及 AI 任務。HPC 市場包括超級電腦、伺服器、儲存系統及網路設施。

隨著 2024 年 AI 發展影響全球資料中心和雲端服務，加速高效能運算系統滲

透既有運算市場。尤其數據在各領域應用的複雜性和資訊量規模成長，對平行運算 (Parallel Processing) 和高速網路傳輸等 HPC 能力的需求也在提高，且 HPC 算力部署主要位於資料中心，這也帶動市場對於資料中心的需求，以滿足如醫療業、政府等各行各業的運算需求。

美國為全球最大的 HPC 應用市場 (見圖 4.6)，中國排名第二，但與排名第三的加拿大差距不大。歐洲由德國、法國領銜，亞洲地區前五名 HPC 市場為中國、日本、印度、南韓及台灣。



資料來源：Fortune Business Insights，MIC 整理，2024 年 9 月

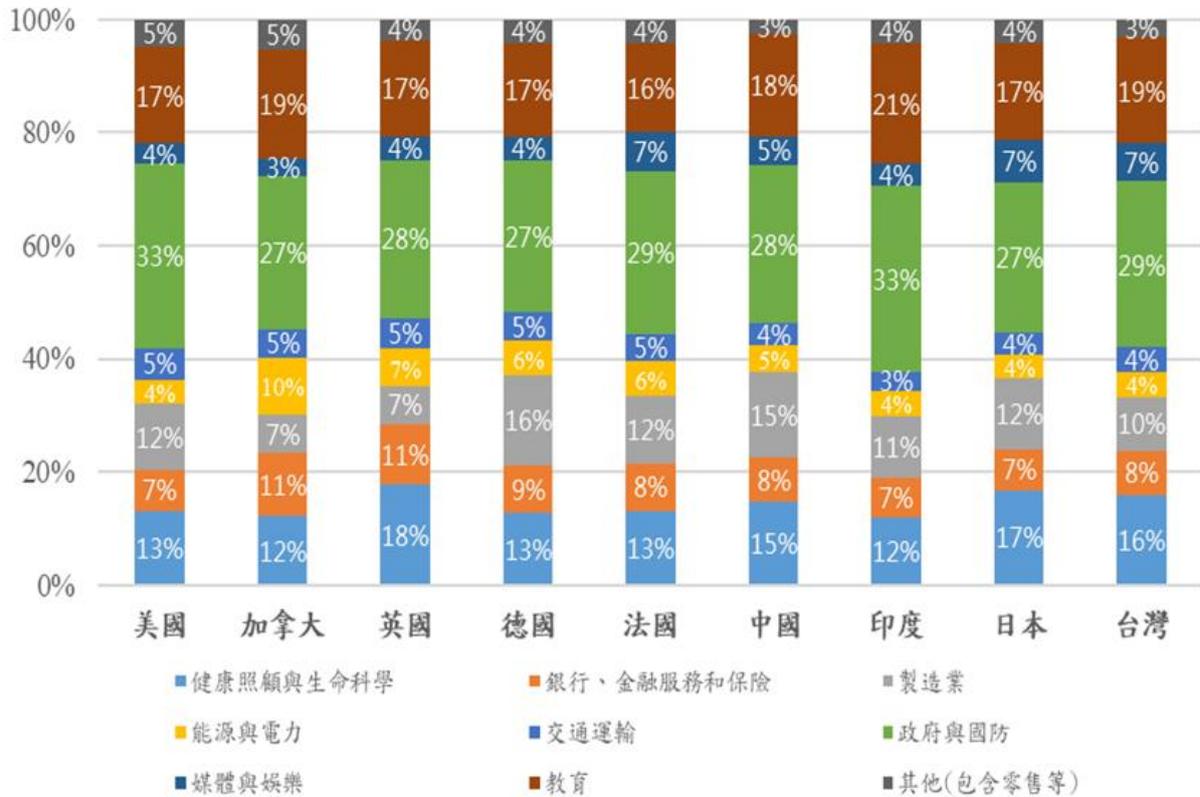
圖 4.6、2024 年全球主要國家 HPC 市場營收

### 1. HPC 產業應用分布

根據 Fortune Business Insights 的調查，2024 年高效能運算市場中以政府與國防之應用占比最高 (見圖 4.7)，排名第二為教育，排名第三為健康醫療領域，第四及第五為製造業及金融業，各國 HPC 產業市場分布排名大致相同，現階段高效能運算市場主要集中在前五大領域，且佔比超過 50%。

HPC 主要應用在政府、教育、醫療保健、金融業、製造業等領域，對於模擬和建構運算模型的需求，是驅動 HPC 市場成長的重要因素。其中政府及教育領域，有很大部分來自於各國對於超級電腦的投資，以 2024 年下半年 TOP 500 榜單來看，超級電腦入榜名單絕大部分為政府及研究機構，使用範疇以國防、航空航太、天氣預報、氣候建模、藥物發現、材料科學和工程模擬等研究發展用途為主。

2024年主要國家HPC產業應用分布



資料來源：Fortune Business Insights，MIC 整理，2024 年 9 月

圖 4.7、2024 年主要國家 HPC 產業應用分布

## 2. 生成式 AI 加速 HPC 市場成長

AI 發展影響高效能運算發展，高效能運算能處理大量數據，結合 AI 分析能力可以為多個領域的運算效能帶來進步。讓高效能運算能以更快、更準確的模擬能力，為氣象科學、生命科學和工程領域等帶來更快的新發現，在不同領域解鎖新能力。

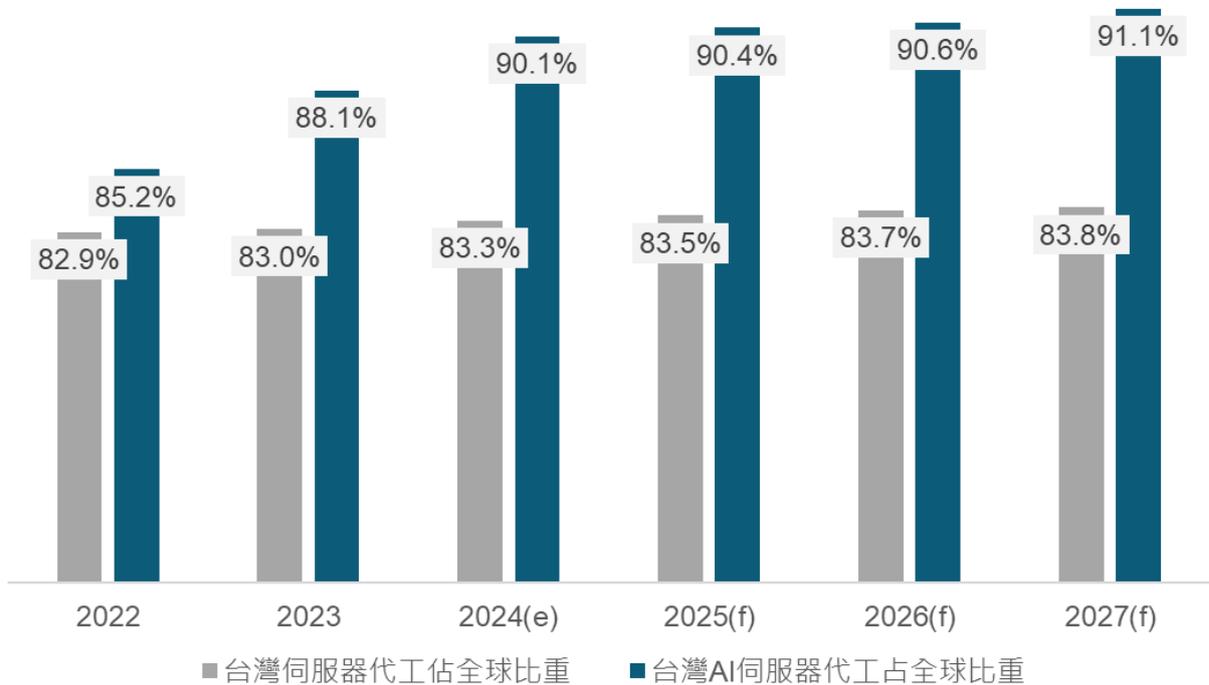
隨著 AI 算法變得越來越複雜並需要大量的運算能力，為支持模型訓練和推理，超大規模資料中心採購更多 AIGPU 及部署 AI 伺服器建構高效能運算系統，推動 2024 年高效能運算系統市場成長。高效能運算在生成式 AI 發展推波助瀾下，除了科研、醫療、材料等關鍵產業領域應用，將有助於加速各領域的垂直應用，將算力服務擴散至終端應用，如自駕車、機器人、AI 智慧工廠等。

## 三、台灣算力生態系發展現況

### (一) 算力生產

由前一節所提的全球伺服器供應鏈可以得知，在算力生產方面從伺服器主機版的 SMT 線至後段的組裝，台灣業者均占據重要的地位。2023 年全球伺服器產業代工出貨占 83%（見圖 4.8），至 2027 年受惠於 AI 推論伺服器、AI 訓練伺服器基本上均由台灣業者進行代工，將有望提升至 83.8%。AI 伺服器具有高技術進

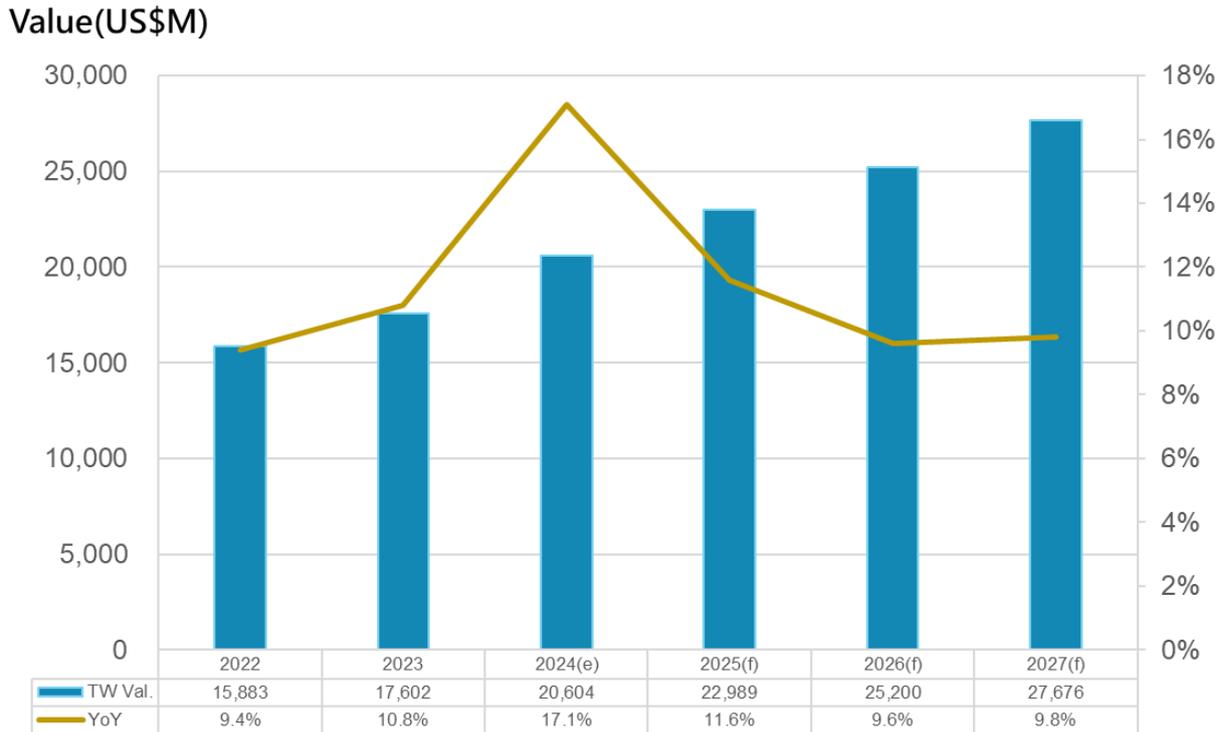
入門檻，同時美國禁令讓中國大陸業者能夠接獲的 AI 伺服器訂單大幅減少，因此對台灣廠商代工而言有望持續進行提升。預期可望由 2023 年的 88.1%，提升至 2027 年的 91.1%。



資料來源：MIC，2024 年 9 月

圖 4.8、台灣伺服器/AI 伺服器代工佔全球比重

AI 伺服器的代工佔比逐漸攀升，因為各項產品規格升級的需求促使價格更加高昂，進而讓台灣伺服器產業的平均售價（Average Selling Price, ASP）提升。當中以雲端服務商為最重要的買家，雲端服務商為了增強自身的 AI 算力，積極部署 AI 超級電腦與 AI 資料中心，將同步帶動台灣廠商的產值。除此之外，AI 伺服器促使台灣的自有品牌商有望發展，相較於純粹進行代工的模式，透過自有品牌銷售 AI 伺服器給客戶將獲得更高的利潤。因此預期台灣伺服器產值將由 2024 年的 206 億美元，上升至 2027 年的 276 億美元（見圖 4.9）。



資料來源：MIC，2024 年 9 月

圖 4.9、2022~2027 年台灣伺服器產值預測

因為受惠於 AI 伺服器的熱潮，台灣在算力生產方面主要有四大類型的客戶，首先是 Dell、HPE 等伺服器品牌商，在 2024 年持續推出新型的 AI 伺服器，以及透過超級電腦品牌如 HPE Cray 協助科研機構建超級電腦，其主要的代工廠商均為台廠。透過既往的代工模式進行。

其次是台廠和雲端服務商透過直接代工（ODM Direct）的方式進行合作，打造客製化的 AI 伺服器，包含 Amazon、Microsoft、Google、Meta 及 Oracle 等。除此之外，部分科技巨頭如 Apple、Tesla 底下的 xAI 等客戶的關係，開拓新的訂單成為布局重點。

## (二) 算力服務

### 1. 台灣雲端服務市場

台灣雲端服務市場，與國際其他地區相同，同樣是由美系 3 大雲端服務公司 AWS、Microsoft Azure、Google Cloud 所主導，市佔超過 7 成以上，本土雲端服務商台智雲後來居上，成為市占排名第四之服務供應商。早期投入國內雲端服務的本土營運商，中華電信、遠傳電信、台灣大哥大等受到國際大廠進入台灣市場後，雲端服務在台灣市場占比極為有限。以下簡要說明各家在台布局現況：

- (1) AWS 透過不斷擴大在台商業合作夥伴如伊雲谷，以及參與台灣新創園區發展，在台設立 AWS Local Zone，並於 2024 年宣布在台設立雲端區域（Region），持續保持市場領導地位。

- (2) Microsoft Azure 利用在台軟體市場的強大基礎，如生產力和協作應用程式以及數據庫管理軟體，與 Azure 進行交叉銷售，佔據台灣雲端市場的第二名。隨著微軟台灣資料中心於 2024 年啟用，有助於搶攻台灣受監管產業的市場機會。
- (3) Google 擁有完整的碳足跡追蹤解決方案與本地資料中心，且其在台資料中心高比例使用再生能源，受特定需求客戶青睞，名列第三。
- (4) 台智雲為台灣第一家純公有雲業者，由華碩投資，運算設備由國家網路中心的台灣杉二號，提供強大的超級電腦運算效能，已迅速取得部分台灣雲端市場占比。
- (5) 其他如中華電信、遠傳電信與台灣大哥大的自有品牌雲服務，相較國外公有雲業者更早進入本地企業市場，然雲端運算平台功能項目，與操作介面相較國際公雲業者較不易使用，後續成長動力有限。

由於公有雲市場有明顯的前期投入成本與規模優勢，大者恆大。再加上國際領先的雲端服務龍頭業者，如 AWS、Microsoft Azure、Google Cloud 能提供完整雲端服務，在全世界已打造數萬種解決方案，更搶先部署 AI 生態系於各自平台上，本土業者不敵國際領先業者，多改為加入三大公有雲業者的本地經銷夥伴行列。例如：中華電信銷售 AWS 雲服務產品給本地企業用戶；遠傳電信收購 Nextlink（博弘雲端科技）後，亦成為 AWS 的台灣主要經銷夥伴；另外 Microsoft、Google 也和台灣三大電信業者達成一定程度的合作關係。

目前在台灣的國際公有雲雲端服務商，提供的本土算力是以 CPU 為主的運算資源，若要租用高階 GPU 算力，需要連接到雲服務商鄰近的亞太資料中心取得 AI 算力。台灣既有的雲端服務生態，除了國內外公有雲雲端服務公司外，在台灣也衍生提供雲端服務代理銷售、諮詢服務及系統整合的服務託管供應商（Managed Service Provider, MSP）。MSP 作為國際公有雲與本土客戶的橋樑，提供維運託管或針對原廠沒有提供、而客戶有需求之雲端服務，衍生出獨家服務類型，代表之資服業者如伊雲谷、雲動力、宏基資訊、精誠資訊、邁達特等，及協助企業規劃上雲架構的系統整合業者。

## **2. 新興 AI 算力服務快速崛起**

全球新興算力服務在生成式 AI 的應用逐漸落地，大量 AI 推論、中小語言模型調校（Fine-tune）的需求下，全球新興算力服務的業者如雨後春筍般出現。新興算力服務業主透過建造 AI 資料中心，並將算力轉租給政府科研機構、雲端服務商及企業做為其主要的營運方式。

相較於歐美成熟市場，台灣在新興算力服務方面正處在萌芽的階段，因為台灣廠商對於伺服器、超級電腦乃至資料中心，本就具備設計與製造的量能，因此當前逐漸出現包含台智雲、友歲、翔耀、鴻海雲高科技等業者，緯創合作的

Zettabyte 新興算力服務公司，同樣將透過採購高階 GPU，藉此來製造高階 AI 伺服器並打造 AI 資料中心，提供本土及海外地區算力租借的服務。

此外，台灣多家企業也自行布局 AI 算力，積極採購 AI 運算系統，滿足內部研發應用所需，如鴻海高雄算力中心、金融機構、研究型醫院等，由於算力未對外提供商業服務，故未歸納在台灣算力服務生態系範疇內。

### 3. 公部門提供之算力服務

由於高階 AI 算力建置成本高昂，尤其是高效能運算系統，一台提供 AI 訓練的伺服器 NVIDIA HGX(內含 8 顆 H100 GPU 設備)，買斷售價約為 1,000 萬台幣(30 萬美元)，如果租用雲端算力，高階 GPU 單一小時租金在 23~98 美元不等。因此，資源豐沛的大型企業較能夠負擔這筆資源投入，相對來說學研機構、中小企業、新創公司則較難以負擔 AI 算力成本。

我國政府提供高階算力資源，主要由國科會、經濟部、數發部支持。提供方式，簡要說明如下：

- (1) 國科會委由轄下研究機構國網中心所營運之超級電腦，提供大專院校科學研發及產學合作項目申請使用，獲得請通過之團隊，使用超級電腦費用相較於商業營運算力設備便宜許多。
- (2) 經濟部目前算力資源來自於 NVIDIA 的捐贈，2023 年 NVIDIA 在台灣設立 Taipei-1 超級電腦，並將其中 1/4 算力回饋台灣，免費提供 2 年半算力，價值約 4 億台幣，給產官學研與新創進行研發使用。在 2024 年由經濟部協助規劃算力申請遴選，截至 2024 年 11 月底已有 9 個團隊獲選。
- (3) 數位發展部在 2024 年建立產業 AI 算力平台，今年度提供 NVIDIA H100 晶片 32 片、AMD 的 MI300X 晶片 8 片算力系統，未來將持續擴充算力資源。數發部 AI 算力平台提供國內資服業者免費算力申請，進行 AI 模型訓練，降低業者生成式 AI 模型的開發成本，以期提升我國在 AI 領域的算力資源，加快 AI 落地應用，推動產業的數位轉型和升級。

當前我國公部門及民間積極部署擴充 AI 算力，然而目前 AI 算力生態面臨的挑戰議題是，支持 AI 應用開發之算力服務需要更多技術人才。由於 AI GPU 算力以 AI 訓練、AI 應用開發為目的，相較於 CPU 運算的提供應用情境有很大不同，因此兩者間的技術轉換銜接過程中，需要更多 AI 人才與提供 AI 算力服務的資服業者，協助我國企業部署 AI 應用。

### 四、我國算力基礎建設議題

現今的科技環境中，算力被視為一種核心資源，影響著各個領域的發展。無論是人工智慧、物聯網，還是大數據分析，算力的需求持續上升，驅動著各行各業的創新和轉型。政府和企業必須重視算力的布局，以確保能夠在未來的競爭中

佔據優勢。

然而在佈建算力的過程，不論是國家層級、產業層級、甚至是企業層級，都將衍生一些議題，尚需進一步研討與面對，如國家層級有國家 AI 主權落地需求；產業層級則須面對因應算力成長而伴隨電力需求高漲與環境節能趨勢；企業層級則是需考量如何因應來自全球公有雲服務廠商的挑戰、算力不確定性與高成本的投資風險，甚至如何規劃整廠輸出與掌握廢熱商機思維等等。

### (一) 主權落地與資安驗證需求

隨著對數據保護意識的提升，企業和政府對主權落地的需求愈加強烈。例如友歲超算公司在這一背景下，致力於提升台灣本土的算力資源，以保障數據安全和自主性，並應對日益增長的算力需求。透過建設本土化的資料中心，可以減少對外部服務的依賴，確保資料的安全性和隱私性。

主權 AI 大語言模型之發展：在當今數位化快速發展的背景下，算力的應用已成為推動 AI 技術進步的重要基石。隨著公共資料的隱私性需求（如金融和健保等領域）的提升，建立具主權概念的資料中心及屬於國家自主之大語言模型，變得尤為重要。為了保障資料安全、保護本國文化、並強化自主技術建立，政府應透過公部門資源（如國網中心）建置算力，同時協助國內業者建立屬於台灣的資料中心，以擴大算力基礎建設的規模，此外，政府自 112 年即啟動「可信任生成式 AI 對話引擎」(Trustworthy AI Dialogue Engine, TAIDE) 計畫，期望帶動「數位主權」之生成式 AI 應用及促進產業發展。

在地優先與主權意識：在美中科技對抗的背景下，國內業者應該優先考量使用台灣土地和電力來建設資料中心，應優先滿足台灣當地需求的業者，實現共榮共利的生態系統，以兼顧主權 AI 產業的意識。此外，對於國內外投資者獲取的部分國家資源支持的新資料或算力中心，也應考慮優先開放給對國家科學研究有重大需求的單位或計畫。

客戶隱私與營業隱私：算力的成本和客戶個資、營業隱私的議題是擴大算力應用的重大挑戰。企業在提供算力服務時，必須重視保護客戶的隱私和商業機密，從而降低因資料外洩而引發的風險。為此，建立健全的隱私保護機制成為必要條件。企業在地端的運用必須以本地安全為首要考量，強化資安和隱私保護，避免資料外洩的風險。隨著數據安全事件的頻繁發生，企業必須採取一系列的防護措施，確保客戶資料的安全性與完整性。

數據留在台灣之必要性：台灣是否能成為數據大國，取決於資料能否有效留在本地。若資料被存放於國外，將會面臨資料外流的問題，進而影響國家安全和隱私。因此，國家需要加大對相關算力建設的投入，建立強大的本地算力基礎，以支持數據的本地化處理。

安全驗證與政府支持：算力應用的發展需要遵循安全驗證標準，例如

ISO27001，以加強算力產業的安全性。此外，政府應針對涉及民眾利益或安全的不確定事項，建立沙盒規範，並提供算力支持或補助，鼓勵企業申請並投入相關的研發，促進創新與發展。

綜之，算力與 AI 主權產業的發展不僅是技術上的挑戰，更是國家安全和經濟發展的重要議題。政府和企業需攜手合作，建立強大的資料中心，推動在地優先的發展策略，並保障客戶隱私和資料安全。只有通過共同努力，才能促進台灣在全球數位經濟中的持續發展，實現真正的共贏。

## (二) 伴隨電力需求與節能，以及效能提升趨勢

電力需求：資料中心需要電力支撐，然而，資料中心不僅涉及到技術層面，還必須考量政府、地方和業者之間的合作，以共同解決用電需求和建置電廠的問題。當然，這也涉及到南北產業的平衡發展，確保資源的合理分配和有效利用，促進整體經濟的可持續發展。

晶片節能設計：考量到算力運用所帶來的能源消耗，當前的算力晶片設計日益強調節能效益。這種節能設計不僅能減少電力消耗，還能在有限的電力資源下，有效支撐資料中心的運作。因此，選擇節能高效的晶片成為資料中心建設的重要考量之一。

液冷技術的興起：資料中心的能效指標 PUE (Power Usage Effectiveness) 反映了電力的有效利用情況。現有的冷卻技術大約有 30% 的電力消耗於氣冷散熱機制。隨著伺服器厚度的增加 (如 4U、5U，甚至未來可能達到 10U)，液冷技術將成為解決散熱問題的新趨勢。液冷不僅提高了散熱效率，也能降低整體能耗，顯示出其在未來資料中心建設中的潛力。

算力效能之提升：提升算力效能不僅僅依賴於 CPU 和 GPU 的選擇，設計上還需考量高傳輸與大容量記憶體的搭配。例如，高頻寬記憶體 (High Bandwidth Memory, HBM) 是實現高算力效能的重要組件之一。透過合理的記憶體配置，可以有效提升計算的效率，進而提升整體系統的算力效能。另外，高傳輸架構，如 NVLink72 是 NVIDIA 推出的高效互聯架構，具備極高的數據傳輸能力，專為多 GPU 系統和高性能計算設計。其架構支持每秒高達 900GB 的數據傳輸速度，遠超傳統 PCIe 介面，能顯著降低資料傳輸瓶頸，提升 GPU 之間的協同效能。這種高傳輸特性特別適合於 AI 訓練、大規模數據分析及高效能計算工作負載，為現代算力需求提供更佳的解決方案

## (三) 面對全球公有雲市場挑戰

出租算力之挑戰：算力的商業模式問題也是不容忽視的，如何有效地租賣算力給服務業者，並整合國內的產官學研力量，以採取集中式算力的方式，形成可持續的商業模式，成為關鍵挑戰。

這意味著，雖然算力的應用範圍涵蓋了各行各業，但企業需要有足夠的時間來消化這些算力資源，並在此過程中考驗自身的財務能力，特別是在建設算力的同時，如何有效地推廣和出售算力，直到達到經濟規模。

#### **(四) 算力不確定性與高成本**

算力的運行往往耗時且成本高，俗稱「燒算力」，這使得企業在投資時對回報率（ROI）有不確定性。特別是在 AI 應用尚未完全成熟的情況下，許多企業對於大規模投資感到猶豫，算力成本亦被視為重大挑戰。政府可透過鼓勵投資和政策支持，減少算力應用的不確定性，並推動產業成長。

生成式 AI 已逐步進入政府、金融和電信業等服務業，應用如 AI 代理（AI Agent）等正成為焦點。然而，製造業對於生成式 AI 的採用較慢，主要因為尚未看到具體效益，無法有效節省人力。亦即，當前算力應用仍處於起步階段，許多業者在算力服務上面臨負毛利的問題，這意味著提供算力服務的成本超過了收入。然而，這種現象在新興市場中並非罕見，算力應用的普及和生態系發展才是驅動力。政府應在此過程中提供支持，通過減稅、資源補助等措施，幫助業者渡過早期階段。

#### **(五) 整廠輸出與廢熱商機思維**

**整廠輸出：**隨著資料中心需求的增加，整廠輸出的新概念應運而生。這一模式結合了資料中心的廠房建設、機櫃散熱工程、樓地板荷重、晶片製作與組裝，最終應用於生成式 AI 的算力需求。這不僅能有效整合資源，還可拓展國際市場，形成高附加值的整廠輸出，為台灣資料中心的發展帶來全新的機遇。

**廢熱商機：**在資料中心的建設中，散熱問題不僅是技術挑戰，更是一個商業機會。建議可以結合石化產業的經驗，有效運用機櫃所產生的廢熱，轉變為有價值的資源。這樣的作法不僅能減少對環境的影響，還能夠為企業帶來新的商業機會，形成一個可持續發展的模式。

## 五、小結

綜觀全球在算力系統生態系的布局，可以發現台灣透過長年在資通訊產業所積累的研發、製造能量，使台灣在算力系統生態系具有重要的地位，在此分別對算力生產、算力服務與新興算力服務進行比較分析。

### (一) 算力生產

全球算力製造生態系當中包含上游 CPU、GPU 等處理器、其它關鍵零組件、下游伺服器代工廠及客戶端的伺服器品牌商、雲端服務商。台灣當前主要在其它關鍵零組件，以及在伺服器代工方面佔有重要的地位，隨著 AI 伺服器的蓬勃發展，台灣廠商憑藉過去製造代工所厚植的研發能量，開始強化推廣自身的 AI 伺服器品牌。

因此在伺服器品牌商方面，透過過去的伺服器 ODM 模式和台灣廠商合作密切；雲端服務商則透過直接代工模式，和台灣廠商合作生產客製化伺服器；最後則是台灣廠商透過自有品牌，將 AI 伺服器賣給全球不同類型的算力需求者。綜上所述，台灣在算力生產方面於全球具有不可替代的關鍵地位，與全球算力生產運作息息相關。

### (二) 算力服務

全球算力服務主要有兩種，一是由公部門提供之算力服務，以支持科學研究或產業研發之用，致力於科學創造之突破性發現或創新。商業市場算力服務，由大型雲端服務供應商主導，三大美系雲端服務商（AWS、Google 和 Microsoft）全球市占率超過 6 成，會因應各國產業需求進行布局，佈建雲端區域及設立資料中心提供運算服務。

台灣商業服務算力同樣以美系三大雲端服務供應商為主，近年陸續在台投資本地資料中心，以滿足台灣企業對機敏性資料的本地運算處理需求。公部門算力服務的支援系統以國網中心超級電腦為核心，近 2 年經濟部、數位部為支持企業發展所需算力，持續布局算力資源，支持關鍵產業算力需求，加速國內產業數位轉型。

台灣算力挑戰主要體現在大語言模型的訓練上，這通常需要高成本的 AI 資料中心進行處理，屬於大算力需求的範疇。隨著全球 AI 算力服務正在迅速成長，台灣業者也看到其中的商機，正在積極布局。未來兩年，這些公司將會持續進行大量的算力投資。將可預見，隨著這些 AI 資料中心的持續發展，台灣的算力應該能夠滿足日益增長的需求。

### 參考文獻

1. 陳牧風(2023)，資料中心新型態 AI 算力租借模式分析，資策會產業情報研究所(MIC)。
2. 陳牧風(2024)，AI 資料中心市場發展與關鍵議題，資策會產業情報研究所(MIC)。
3. 陳奕伶(2024)，雲端資料中心及 AI 服務趨勢分析，資策會產業情報研究所(MIC)。
4. 陳奕伶(2024)，台灣資料中心發展趨勢分析，資策會產業情報研究所(MIC)。
5. 陳牧風(2024)，臺灣代工占全球 AI 伺服器出貨達九成 帶動我國伺服器產業鏈營收成長 2024 有兩大成長動能，資策會產業情報研究所(MIC)。
6. Fortune Business Insights(2024), High Performance Computing:GlobalMarket Analysis, Insights and Forecast, 2024-2032.

## 第五章 綠算力及算力應用於永續發展

本章概分三大部分來進行闡述，分別為：算力需求與能耗、從綠色算力看台灣產業發展、以及計算可持續性與應用案例探討。

### 一、算力需求與能耗：數位時代的關鍵挑戰

在現代社會中，算力已成為不可或缺的資源。無論是從事科技研發、經濟分析，還是日常生活中的手機應用，我們都需要強大的算力來支持。然而，算力的提升並非無中生有，它需有足夠的電力。換句話說，算力和電力之間的關係是相互依存的。沒有足夠的電力，算力無從談起；沒有強大的算力，現代社會的營運和發展將受到影響。因此，算力與能源之間的平衡已成為當前最重要的議題之一。

#### (一) AI 與電力需求

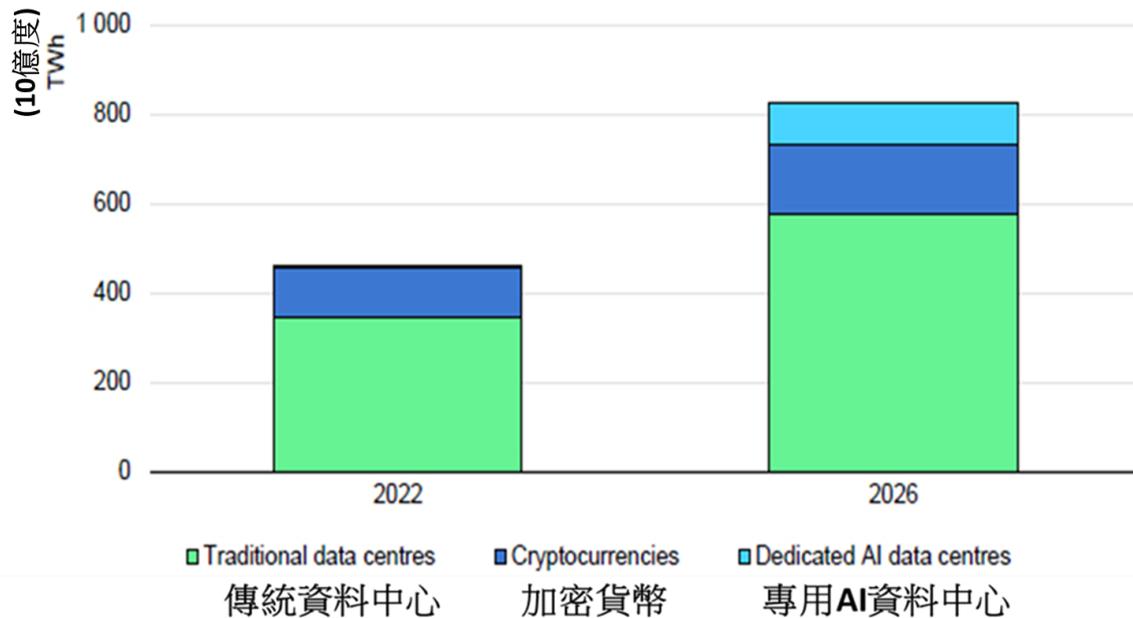
算力作為數位產業發展的主引擎，已成為國際競爭中最重要的策略資源之一。全球範圍內，超大規模的算力中心正在迅速增加，這些中心不僅推動了社會生產力的提升，也滲透到生活的各個層面。我們在享受這些技術便利的同時，也不能忽視其背後的高能耗問題。根據國際能源署（International Energy Agency, IEA）表示，2022 年資料中心、AI 與虛擬貨幣加起來消耗的電力幾乎占全球總電力需求的 2%，而且預測未來 3 年內，這三種類型的總用電量將會翻倍，相當於一個日本全國的電力消耗量。據黃召勇 2024 年在 CSR@天下的報導，美國預計到 2030 年，AI 的耗電量將佔全國供電量的 30%。同年，邱祐慶亦在遠見報導，Meta 執行長祖克柏（Mark Zuckerberg）認為「電力短缺已成為建造 AI 數據中心的最大挑戰！」，而特斯拉執行長馬斯克（Elon Musk）也認為，未來不再是「缺晶片」，而是「缺電力」。美媒《紐約客》（The New Yorker）也指出，美國家庭平均每日用電量約為 29 度，而聊天機器人 ChatGPT 每日用電量竟高達 50 萬度，約為其 1.7 萬倍。這些數據無不表明，美國即將面臨電力短缺的挑戰。然而美國的情形絕不是一個特例，是各國在推動數位轉型時不得不面對的課題。

在資料中心中，電力需求主要來自兩個方面：計算設備和冷卻降溫系統，這兩部分各占總電力消耗的 40%，其餘 20% 則來自其他相關的資訊科技（IT）設備。全球目前有超過 8000 個資料中心，主要分布於美國、歐盟和中國。國際能源署（IEA）進一步預測，這三個區域未來資料中心的數量將會爆炸性成長，用電量也將增長 30% 至 100% 不等。

DataVita 創辦人丹尼·奎恩（Danny Quinn）曾向《英國國家廣播公司》（BBC）表示：「設置傳統伺服器的標準機櫃耗電量約為 4 千瓦（kW），但設置一組 AI 伺服器機櫃的能源需求是其 20 倍，即 80 千瓦。」更引人注目的是，典型的 Google 單次搜尋平均電力需求約 0.3 瓦時（Wh），而大眾越來越依賴的 ChatGPT，單次搜尋平均電力需求高達 2.9 瓦時（Wh），是前者的 9.5 倍！因此 IEA 估算，未來

三年內，AI 產業預計將呈現指數級增長，屆時其總體耗電量至少是 2023 年需求的 10 倍。

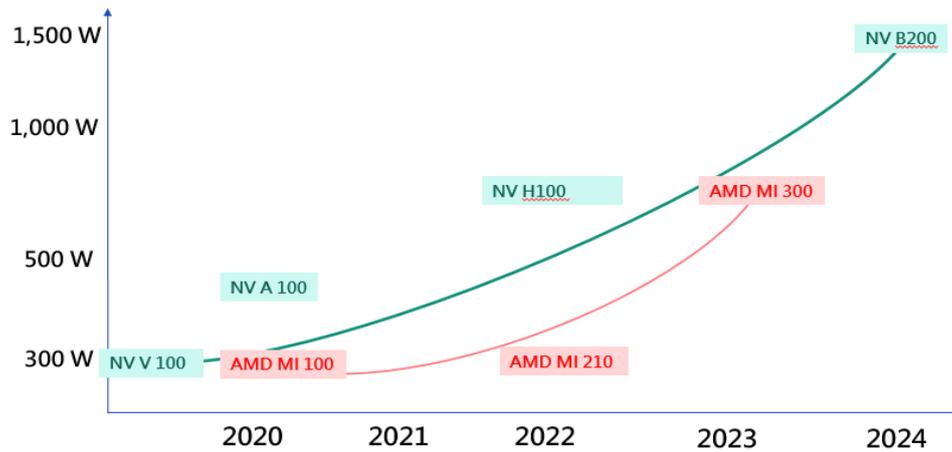
將虛擬貨幣的電力需求單獨拿出來看，根據 2022 年的數據，其佔全球年度電力需求的 0.4%。這個數字看似微小，但其實相當於荷蘭整個國家的總電力消耗，並預計到 2026 年，虛擬貨幣的用電量將增加超過 40%。圖 5.1 顯示 2022 年與 2026 年傳統資料中心、加密貨幣與專用 AI 資料中心的估計用電需求增長情形。



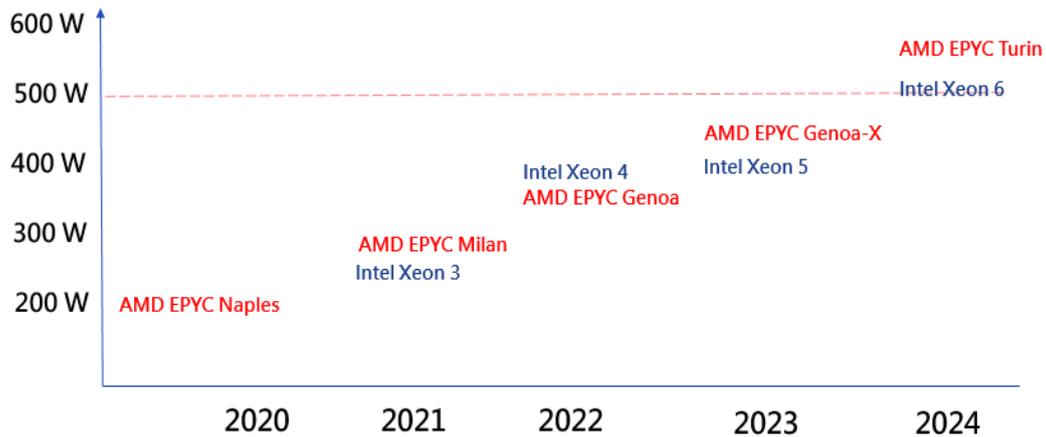
資料來源：IEA, 2024

圖 5.1、傳統資料中心、專用 AI 資料中心和加密貨幣的預估電力需求

台灣在這波 AI 發展的賽道上，並未缺席，從晶圓製造至機台組裝都可以看到台灣的身影，世界各大廠也有高度意願在台灣投入更多發展計畫。然而，AI 的發展對台灣的電力有何需求呢？經濟部能源署在 112 年全國電力資源供需報告中指出，因應 AI 科技用電需求於五年內將成長約八倍，主要原因在於 AI 伺服器需配備 AI 處理器。如圖 5.2(a)所示，NVIDIA 的 AI 處理器從 V100、A100、H100 到 B200，算力性能大幅提升下，所需功率也大幅增加。同樣地，如 AMD 所開發的 AI 處理器亦然，如 MI100、MI200 到 MI300 之功能逐年升級，進而對功率的需求也隨之增加。另外，與圖 5.2(b)相較之下，使用在傳統伺服器上的一般處理器，雖然功能逐年提升，但其所需功率增幅就相對地小。



(a)



(b)

資料來源：MIC 整理，2024 年 6 月

圖 5.2、(a)AI 伺服器功率變化情形；(b)傳統伺服器功率變化情形

此外，我們針對一般伺服器與 AI 伺服器（Dell PowerEdge XE9640）做一比較，詳如表 5.1，顯示具備 AI 功能的伺服器（由 2 個 CPU 與 2 個 GPU 組成）比傳統伺服器（由 2 個 CPU 組成）所需功率高出約 6.375 倍。

表 5.1、AI 伺服器與一般伺服器所需電力比較表

	一般伺服器	Dell PowerEdge XE9640
CPU (Intel Xeon) (W)	320	320*2
GPU (H100) (W)	無	700*2
I U伺服器總功率(W)	320	2,040

6.375倍

資料來源：MIC 整理，2024 年 6 月

如何扭轉上述三大科技高排碳的迷思？規範和技術是關鍵！2020年，歐盟開始為資料中心制定效率標準，次年推出了《氣候中和資料中心公約(Climate Neutral Data Centre Pact)》，承諾通過可衡量的目標證明能源效率、購買100%無碳能源、維修和再利用伺服器，並尋找熱回收方法，以在2030年實現碳中和。儘管這項倡議是自願性的，無法律約束力，但已吸引超過100家資料中心營運商簽署協議，包括Equinix和Digital Realty等知名企業。而在技術方面，採用高效率的冷卻系統、熱回收、通過機器學習優化伺服器運作、甚至將電力需求負載進行時間和空間上的轉移調度，都是已被運用且有效的去碳化技術。

從全球市場趨勢來看，小型資料中心逐漸轉型為超大型規模資料中心(Hyperscale Data Centers)。這類資料中心占地超過1萬平方米，電源容量至少10兆瓦(MW)，通常能夠實現最佳處理器效率，降低空閒狀態的功耗，從而提高能源效率。歐盟在其《綠色雲端運算與電子通訊研究》(Study on Greening Cloud Computing and Electronic Communications)報告中，將超大型規模資料中心視為資訊通信部門中關鍵的減碳解決方案之一。

因此，綠色算力或永續算力的概念應運而生。發展綠色算力，不僅能夠滿足不斷增長的算力需求，還能在環境保護和能源效率上獲得雙贏。其核心目的是在最大化能源使用效率的同時，最小化對環境的影響。這方面的努力主要集中在電腦晶片、系統和軟體的設計與使用上。實際上，每個人都需要綠色算力。手機使用者希望擁有高效能且長效電池的手機；政府與企業希望擁有強大且環境友善的系統；雲服務提供商更是期待能夠穩定供應全球用戶需求的電網效能。同時，各國也開始大力推動環境保護(Environment)、社會責任(Social)與公司治理(Governance)，亦即所謂的ESG倡議，淨零排放和永續發展已逐漸成為顯學。

推動綠色算力具有重要的意義。一方面，可以讓產業與世界接軌；另一方面，也有助於各行各業實現綠色轉型和升級。面對氣候變遷的挑戰，發展綠色算力顯得尤為重要。過去一個世紀，全球溫度上升了1.2°C，位於玻利維亞安地斯山的夏卡塔亞兩側山坡，曾是世界最高的滑雪坡，但由於氣候變遷和「聖嬰」現象，冰河正在迅速縮小，可能在不到五年內消失。發展綠色算力不僅可以提升算力，還能有效降低碳排放，實現資源最大化利用並減少環境影響。

總之，綠色算力不僅是技術發展的必然趨勢，更是面對環境挑戰的關鍵解決方案。透過持續推動綠色算力的發展，我們有望在提升計算能力的同時，實現可持續的能源使用，為未來的數位社會奠定堅實的基礎。

## (二) 綠色算力定義、國際趨勢及國家政策

本節將說明綠色算力、國際趨勢以及國家之政策。

### 1. 綠色算力定義

Wu Feng 是維吉尼亞理工大學的計算機科學教授，一直致力於推動綠色算力

的極限。隨著超級電腦和資料中心的發展，人們對突破其功耗極限的興趣也不斷增長。2007年11月，Wu Feng 與約30位高效能運算（HPC）領域的傑出人士合作，並收集社群的回饋意見後，推出了首創 Green500 排行榜。因此，Green500 成為關注功耗控制和效能提升人士所重視的內容。

從持續性的 Green500 排行榜可以發現，異質系統（即除了 CPU 外還配備 GPU 等加速器的系統）始終是最節能的系統搭配模式。這是因為加速器提高了大規模平行執行程式碼的能力，從而在能耗和執行速度上都有顯著改善。

就此，Green500 排行榜是節能超級計算的行業基準，開啟了超級電腦可以通過每瓦性能進行比較的新方式。具體來說，Green500 是每年兩次根據能源效率對超級電腦進行排名，名單來自超級電腦 Top500，各電腦以 Linpack（Linear System Package）基準測試進行每瓦效能的衡量。儘管任何功耗性能的指標都存在爭議，但目前 Green500，FLOPS-per-Watt 是主要的指標。其中 FLOPS（Floating-Point Operations Per Second）表示每秒執行的浮點運算次數，最近一期的 Green500 前十排名如表 5.2 所示。

表 5.2、2024 年 6 月 Green500 排名

Green500 Rank	TOP500 Rank	System	Rmax (PFLOPS)	Power (kW)	Energy Efficiency (GFLOPS/watts)
1	189	JEDI Germany	4.50	67	72.733
2	128	Isambard-AI phase United Kingdom	7.42	117	68.835
3	55	Helios GPU Poland	19.14	317	66.948
4	328	Henri United States	2.88	44	65.396
5	71	preAlps Switzerland	15.47	240	64.381
6	299	HoreKa-Teal Germany	3.12	50	62.964
7	54	Frontier TDS United States	19.20	309	62.684
8	11	Venado United States	98.51	1,662	59.287
9	20	Adastra France	46.10	921	58.021
10	28	Setonix – GPU Australia	27.16	477	56.983

註：Rmax 是表示實際算力值，以 PFlop/s（每秒千萬億次浮點運算）為單位。關於其他欄位的詳細信息，請參考 TOP500 的描述

表 5.3 為台灣在 2024 年 6 月 Green500 的排名情形，其中 Taipei-1 為台灣目前算力最強的超級電腦，但其電力消耗資料，目前無從得知。Taipei-1 定位在商業營運，但有提供部份運算資源供國內研發使用。其中創進一號為華碩打造，而台灣杉二號一半產能租給華碩旗下「台智雲」進行商業營運。其餘均為政府部門與研究單位使用。在 Green500 名單中，以美國 169 台領先各國，其次為中國大陸 80 台，除此之外，有超過 10 台以上入榜的國家有德國、日本、法國、英國、南韓、義大利、加拿大。在表 5.3 中，很清楚發現如何在追求算力的同時，亦能兼顧電力消耗的議題，美國、法國、澳洲與波蘭在此已有不錯的表現，值得台灣在發展算力時考量的方向。

表 5.3、台灣在 2024 年 6 月 Green500 排名

Green500 Rank	TOP500 Rank	System/Owner	Rmax (PFLOPS)	Power (kW)	Energy Efficiency (GFLOPS/watts)
62	86	PRIMEHPC Fx1000 氣象署	11.16	673	16.57
65	160	PRIMEHPC Fx1000 氣象署	5.6	350	15.98
78	106	Taiwania 2 台灣杉二號 國網中心	9	797	11.28
107	251	Forerunner 1 創進一號 國網中心	3.53	614	5.75
144	458	Taiwania 3 台灣杉三號 國網中心	2.29	563	1.07
308	38	Taipei-1 Nvidia	22.30		

然而隨著時代變遷，綠色算力所涉及的面向愈來愈寬，但其核心精神指的是在計算資源的生產、部署和運行過程中，致力於最大化能源使用效率並最小化環境影響的計算能力。其目標是通過創新技術和策略，提升計算效能的同時，減少能源消耗和碳排放。這涉及從硬體設計、系統架構到軟體優化等多方面改進，並包括使用可再生能源、提高設備能效、改進冷卻技術和實現資源再利用等措施，以達到永續計算(Sustainable Computing)的目的。

## 2. 國際趨勢

綠色算力，即在計算過程中追求能源使用效率最大化和對環境影響最小化的

技術和策略，已成為全球科技發展的重要方向。隨著氣候變遷問題日益嚴重，各國政府和企業逐漸認識到計算技術對能源消耗和碳排放的影響，並積極採取措施推動綠色算力的發展，以下依國際趨勢進行闡述。

- (1) 能源效率的提升：許多國家和企業致力於提升資料中心和超級計算機的能源效率。這些努力主要體現在三個方面：
  - A. 高效硬體設計：設計和使用高效能、低功耗的計算晶片和伺服器。例如，許多資料中心開始採用 ARM 架構的處理器，這些處理器以其高效能和低能耗著稱。
  - B. 改進冷卻技術：冷卻系統是資料中心能耗的主要來源之一。通過採用液冷技術和自然冷卻技術，可以大幅降低資料中心的能耗。例如，Google 位於美國喬治亞州 Douglas 郡的資料中心，使用自然冷卻技術，每年節省了大量電力。
  - C. 使用再生能源：許多科技公司承諾使用 100% 再生能源來運營其資料中心。例如，Apple、Google 和 Microsoft 等公司已經實現或正在實現其資料中心完全依靠再生能源營運。
- (2) 再生能源的廣泛應用：全球許多資料中心開始轉向使用可再生能源供電。這不僅能降低營運成本，還能減少碳排放。例如：
  - A. 美國：根據美國能源資訊管理局（Energy Information Administration, EIA）統計數據，美國於 2022 年的再生能源發電量首次超越燃煤發電，顯示再生能源已在美國獲得重視且快速發展。
  - B. 歐洲：歐洲議會 2023 年通過再生能源指令修正案，預計 2030 年再生能源占比可達 42.5%，目標是 45%，並要求政府劃設再生能源專區，專區內新設案場申請程序需在一年內完成，以加速再生能源進展。
  - C. 中國和日本等國家也在積極推動再生能源的應用，中國更是全球最大的再生能源投資國。
- (3) 新型計算架構：異質計算（Heterogeneous Computing），即結合不同類型處理器如 CPU 和 GPU 來提高計算效率，已成為節能算力的一大趨勢。這些系統能更高效地處理大規模數據和複雜計算任務。例如：
  - A. NVIDIA：NVIDIA 的 GPU 被廣泛應用於人工智慧和深度學習領域，因其在處理大規模平行計算任務方面的高效能而受到青睞。
  - B. AMD：AMD 的異質計算架構亦結合了高效的 CPU 和 GPU，提供了更高的計算效率和能耗比。
- (4) 國際合作與標準制定：各國通過合作制定全球性標準來推動綠色算力的發展。例如：

- A. 歐盟的《氣候中和資料中心公約》（Climate Neutral Data Centre Pact）承諾到 2030 年實現資料中心碳中和。這項公約吸引了眾多資料中心運營商參與，包括 Equinix 和 Digital Realty 等知名企業。
- B. 國際能源總署（IEA）發佈的報告指出，到 2025 年，再生能源將在三年內超越燃煤，成為全球最主要的電力來源，此一報告促使各國在再生能源的政策上更為積極。

### 3. 國家政策

(1) 政策推動與法規完善：我國政府已通過多項法規和政策推動再生能源的發展。例如：

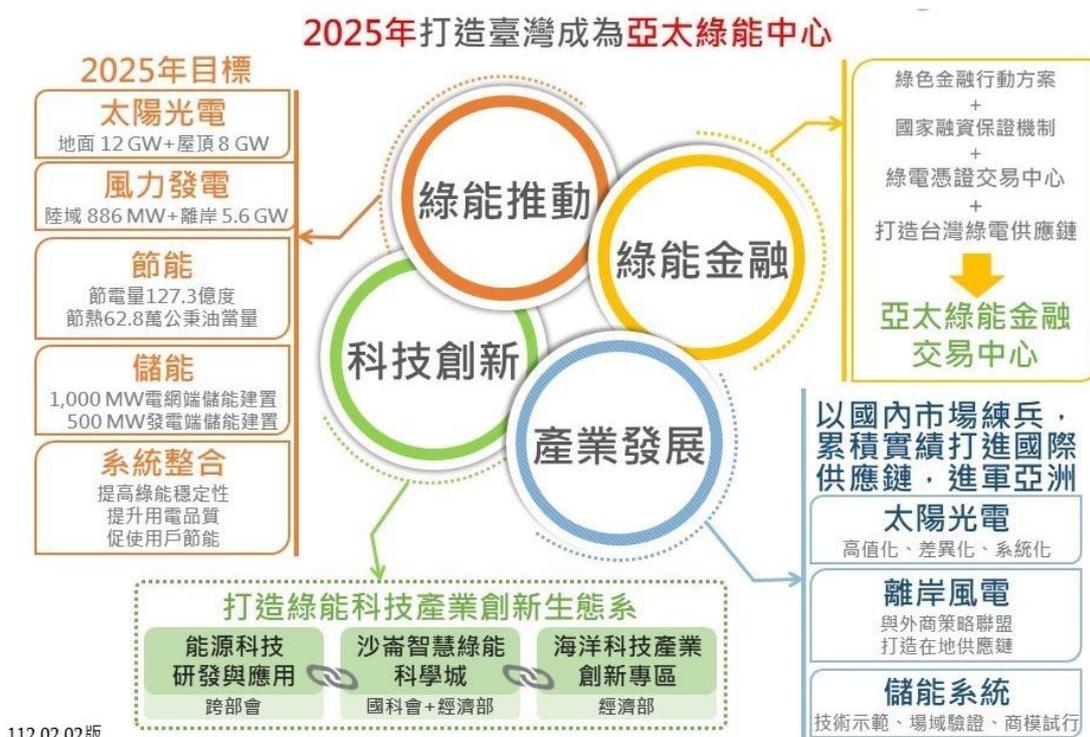
- A. 再生能源發展條例：推動再生能源發展和實現可持續能源轉型的重要法規。其主要目的是促進再生能源的發展，減少對化石燃料的依賴，降低溫室氣體排放，保障能源安全，並促進經濟和環境的永續發展。
- B. 能源轉型行動計畫：政府制定了能源轉型行動計畫，目標是在 2025 年達成 20% 的再生能源發電比例，並在 2030 年達成 30%。
- C. 臺灣 2050 淨零轉型「節能」關鍵戰略行動計畫：為了達成 2050 年淨零碳排放目標而制定的一系列具體措施和政策，如提升能源效率、發展智慧電網、促進再生能源發展、推廣低碳交通、工業減碳等議題。
- D. 電業法修正案：為了推動能源轉型、促進再生能源發展以及確保電力市場公平競爭的重要法案，其包含了開放電力市場、再生能源發展、智慧電網建設、能源效率提升、供電穩定與可靠性等議題。
- E. 綠能科技產業創新方案/前瞻基礎建設計畫—綠能建設：為加速綠能產業發展，我國政府同時於「前瞻基礎建設計畫」項下規劃綠能建設，補足未來綠能發展所需的基礎建設、技術驗證等缺口，期建構能源內需市場，帶動國內相關產業發展與創新，打造臺灣成為亞洲綠能產業發展的重要據點。

依據上面政策推動，可知政府在推動再生能源的發展非常完備，如圖 5.3 所示。綠能建設是以兼顧能源安全、環境永續及綠色經濟為目標，預計於 114 年可帶動 1.8 兆元的民間投資：

- A. 太陽光電：以 114 年設置 20 GW 為目標，其中優先推動屋頂型 8 GW，包含農業設施屋頂（如畜禽舍、農糧製銷儲設施及室內養殖）、工業屋頂（如產業園區、合法工廠、經濟部工業區）、公有屋頂（如校園屋頂及中央部會屋頂）及民間屋頂等；地面型推動 12 GW，優先推動具社會共識及無環境生態爭議之區域，包含公有閒置用地活化（如公有、國營土地、工業區土地）、不利農用土地（如汙染土地、已封閉掩埋場）及土地複合利用（如停車場、水庫、滯洪池、漁電共生、風雨球場、圳路）等。至 112 年 2 月累計裝置容

量達 10.17 GW。

- B.風力發電：打造臺灣風場具國際競爭力，建置風電發展所需水下基礎及重件碼頭，如高雄海洋科技產業創新專區及臺中港離岸風電產業專區，逐步達成至 114 年累積設置量為 6.7 GW（陸域 1.2 GW 及離岸 5.5 GW）的目標。至 112 年已設置 197 座以上離岸風力機，併網的風場將有約 2.2 GW 的風電可進入電網。
- C.沙崙智慧綠能科學城：建構創新綠能產業生態系，朝淨零碳排目標邁進，成為綠能科技產業化之示範驗證場域及產業鏈結樞紐與展示櫥窗，帶動我國綠能產業發展。
- D.綠色金融：透過綠色金融行動方案改善綠色金融體質，建立國家融資保證機制以分攤業者風險、加速產業成長，並以綠電憑證交易中心連結供給與需求，期望能打造臺灣成為亞太綠能金融交易中心。



資料來源：綠能科技產業創新推動方案願景與目標，行政院

圖 5.3、綠能科技產業創新推動方案之願景與目標

(2)技術研發與產業升級：台灣大力投資於綠色技術的研發，尤其是在半導體和電子產業領域。政府鼓勵企業研發高效能、低功耗的計算晶片和系統，提升台灣在國際市場的競爭力。例如：台積電致力於研發低功耗、高效能的半導體技術，其先進的製程技術在全球處於領先地位。鴻海積極投入電動車和智慧城市的相關技術研發，推動綠色算力的應用。台達電長期投入電源效率與儲能之研發，在國際上深具競爭力。

(3)國際合作與參與：台灣積極參與國際再生能源和綠色算力相關組織，學習先進國家的經驗，並分享自身的技術和成果。例如：

A.亞太經合組織（APEC）：台灣在 APEC 框架下，參與相關項目和會議、交流技術和經驗，藉以推動再生能源和綠色算力。

B.國際再生能源機構（IRENA）：台灣與 IRENA 合作，參與再生能源相關的研究和發展計畫，提升在國際舞台上的影響力。

(4)公共教育與宣傳：政府通過各種教育和宣傳活動，提高公眾對綠色算力和再生能源的認識，推動全民參與能源轉型和環保行動。例如：

A.再生能源教育：在各級學校推廣再生能源和環保教育，提高年輕一代對綠色科技的認識和興趣。

B.公眾活動：舉辦各種再生能源和環保主題的公眾活動，如綠色能源展覽、講座和工作坊，吸引廣大市民參與。

綠色算力的發展是全球科技進步和環境保護的重要一環。國際上，各國政府和企業正通過技術創新和政策推動，不斷提升算力的能源效率並減少碳排放。台灣在這一領域也積極採取措施，通過政策推動、技術研發、再生能源應用和國際合作等措施，致力於實現能源轉型和永續發展。這些努力不僅有助於減少對環境的衝擊，還能提升台灣在全球科技競爭中的地位。

綜上所述，綠色算力和再生能源的發展需要各界的共同努力。政府應加大對再生能源和綠色算力的政策支持和資金投入，企業應積極研發高效能、低能耗的技術，公眾應提高環保意識，積極參與永續行動。只有這樣，才能在全球氣候變遷的挑戰中取得突破，實現永續發展的目標。

## 二、從綠色算力看台灣產業發展

綠色算力應將基礎設施、算力設備、算力平台、算力應用一併考量，期許建立高效能、低污染、可持續性的綠色競爭力。根據國際能源總署（IEA）表示資料中心、AI 與虛擬貨幣的耗電量會愈來愈可觀，特別是資料中心。而資料中心能源占比又以計算設備與冷卻降溫系統為其大宗，以下從資料中心的環境永續指標與綠色算力的發展歷程來檢視台灣未來可以再努力的面向。

### （一）資料中心環境永續指標

施耐德電機（Schneider Electric）的能源管理研究中心於 2023 年提出白皮書，特別對資料中心的環境永續指標進行詳細論述。該白皮書主要提出了資料中心環境可持續性報告的指標和類別，並強調標準化這些指標對於推動該行業永續發展的重要性。其指標包括能源、溫室氣體排放、水資源、廢棄物和當地生態系統等五個類別，並提供了相關的定義和應用。與能源相關的指標包括：總能源消耗、電力使用效率（PUE）、總可再生能源消耗、可再生能源因子（REF）、能源重複

利用因子 (ERF)、以及伺服器利用率 (ITEU<sub>sv</sub>) 等。這些指標可以幫助公司評估其能源使用效率和可再生能源比例。

1. 總能源消耗 (Total Energy Consumption)：這是運營資料中心所消耗的總能源，通常是從公用事業電網中獲取的電能，也包括發電機、太陽能或風能等現場能源生產所需的能源。這個指標可以幫助追蹤能源使用情況，並在設計和運營階段減少碳排放。
2. 電力使用效率 (Power Usage Effectiveness, PUE)：這是資料中心的總能源消耗除以 IT 能源消耗的比率。PUE 是一個有效的指標，可以在設計和運營階段推動設施效率，並且可以進行不同資料中心之間的比較。
3. 總可再生能源消耗 (Total Renewable Energy Consumption)：這是資料中心機構所擁有、控制或購買用於資料中心設施的可再生能源總量。增加可再生能源比例有助於減少碳排放，這對於實現碳中和策略非常重要。
4. 可再生能源因子 (Renewable Energy Factor, REF)：這是資料中心組織所擁有和控制的再生能源與資料中心的總能源消耗之比。REF 考慮了通過可再生能源證書 (RECs) 取得的能源，可以幫助追蹤可再生能源使用情況。
5. 能源重複利用因子 (Energy Reuse Factor, ERF)：這是重複使用能源與資料中心總能源消耗的比率。ERF 的目的是推動資料中心營運商和政府尋找重新利用廢熱的方法，例如將廢熱銷售給附近的供暖系統。
6. 伺服器利用率 (IT Equipment Utilization for servers, ITEU<sub>sr</sub>)：這是資料中心所有伺服器的年平均 CPU 利用率。這個指標可以幫助資料中心營運商制定政策，以提高伺服器的能源效率，例如減少伺服器數量或增加工作負載。

除此之外，也可以參考 xUE (x 指標使用效率) 的指標，這是一個廣泛使用的術語，在資料中心環境下，常見的 xUE 指標包括：PUE (電力使用效率)、WUE (水使用效率)、CUE (碳使用效率)。其中，水使用效率 (Water Usage Effectiveness, WUE) 係用於衡量資料中心的用水效率。它是指資料中心的用水量除以 IT 能源消耗的比率，通常以 m<sup>3</sup>/MWh (每百萬瓦時的立方米) 為單位。WUE 可以幫助資料中心追蹤並減少用水量，並在設計和營運階段實現節水目標。另外，碳使用效率 (Carbon Usage Effectiveness, CUE) 則是指資料中心的碳排放量除以 IT 能源消耗的比率，通常以 kg CO<sub>2</sub>e/kWh (每千瓦時的二氧化碳當量) 為單位。這個指標有助於評估資料中心的碳足跡，並推動減少碳排放和實現碳中和的目標。

表 5.4 為重要指標的參考值，供不同資料中心間的相互比較。根據這些指標對於評估和改進資料中心的可持續性表現至關重要，有助於推動節能減排和資源有效利用。

表 5.4、資料中心重要指標參考值

指標 (metric)	定義依據 (Defined by)	最佳值 (Best-in-class value)	業界目標值 (Industry target value)
PUE	ISO/IEC 30134-2	1.1 (75%~85% load ratio)	1.2~1.3 (75%~85% load ratio)
REF	ISO/IEC 30134-3	1.0	0.75~1.0
CUE	ISO/IEC 30134-8	0.0 kg CO <sub>2</sub> e/kWh	0.1~1.2 kg CO <sub>2</sub> e/kWh
WUE	ISO/IEC 30134-9	0.0 m <sup>3</sup> /MWh	0.3~0.45 m <sup>3</sup> /MWh

資料來源：Guide to Environmental Sustainability Metrics for Data Center(Lin, Bunger, & Avelar, 2023)

## (二) 綠色算力發展歷程

回顧綠色算力的過去、現在和未來研究中，綠色算力經由開發和優化電腦晶片、系統、網路和軟體的過程，使其更高效地利用能源和減少對環境的負面影響。以下針對與資料中心有關能源部份如晶片、軟體、資料中心高效架構與技術、冷卻系統等進行說明。綠色算力中關於優化電腦晶片主要涉及到能源節約和效率提升的方面。其中包括了使用能源更高效的硬體，如處理器和記憶體，以減少能源消耗。

### 1. 資料中心高效架構與技術包括了以下內容：

- (1) 虛擬化技術：將伺服器轉換為多個虛擬伺服器，每個虛擬伺服器可以運行自己的操作系統和應用程式。研究指出，伺服器虛擬化可至少節省 50% 的能源。
- (2) 動態電壓和頻率調節：動態調整電壓和頻率以降低能源消耗。
- (3) 能源管理：利用可再生能源和動態電壓頻率調節等技術進行能源管理。

### 2. 冷卻系統包括了以下內容：

- (1) 綠色資料中心採用液冷和室外空氣冷卻等方法來降低能源使用。
- (2) 由於傳統空調系統可能是資料中心能源消耗的重要來源，因此使用自然冷源的冷卻系統可以有效地減少能源需求。
- (3) 冷卻系統的發展對於解決資料中心過熱問題並降低能源需求有著重要影響。

從以上「資料中心環境永續指標」與「綠色算力發展歷程」，在算力與能源議題中可聚焦在晶片系統效率提升、冷卻系統、以及綠色能源開發與運用。

## (三) 晶片系統效率提升

美國於 2022 年公布了「晶片法案」(CHIPS and Science Act, Creating Helpful Incentives to Produce Semiconductors and Science Act)，依預算分配顯示晶片製作與高科技研發為發展重點，如表 5.5。之後，於 2024 年公佈的「關鍵與新興技術清單」(Critical and Emerging Technologies List Update)則是以 2022 年清單為基礎

進行更新而來，如表 5.6。

表 5.5、美國 2022 晶片法案清單

序號	技術名稱
1	Advanced Computing
2	Advanced Engineering Materials
3	Advanced Gas Turbine Engine Technologies
4	Advanced Manufacturing
5	Advanced and Networked Sensing and Signature Management
6	Advanced Nuclear Energy Technologies
7	Artificial Intelligence
8	Autonomous Systems and Robotics
9	Biotechnologies
10	Communication and Networking Technologies
11	Directed Energy
12	Financial Technologies
13	Human-Machine Interfaces
14	Hypersonics
15	Networked Sensors and Sensing
16	Quantum Information Technologies
17	Renewable Energy Generation and Storage
18	Semiconductors and Microelectronics
19	Space Technologies and Systems

資料來源：CHIPS and Science Act (USA, 2022)

表 5.6、美國 2024 年關鍵與新興技術清單

序號	技術名稱
1	Advanced Computing
2	Advanced Engineering Materials
3	Advanced Gas Turbine Engine Technologies
4	Advanced and Networked Sensing and Signature Management
5	Advanced Manufacturing
6	Artificial Intelligence
7	Biotechnologies
8	Clean Energy Generation and Storage
9	Data Privacy, Data Security, and Cybersecurity Technologies
10	Directed Energy
11	Highly Automated, Autonomous, and Uncrewed Systems (UxS), and Robotics
12	Human-Machine Interfaces
13	Hypersonics
14	Integrated Communication and Networking Technologies
15	Positioning, Navigation, and Timing (PNT) Technologies
16	Quantum Information and Enabling Technologies
17	Semiconductors and Microelectronics
18	Space Technologies and Systems

資料來源：Critical and Emerging Technologies List Update( USA, 2024)

比較 2022 年與 2024 年的項目可以發現，美國 2024 年關鍵與新興技術清單「清潔能源生成與儲存」(序號 8) 係整併自 2022 年「先進核能技術」(序號 6) 與「可再生能源生成與儲存」(序號 17)。同時，2022 年有多項關鍵與新興技術，在 2024 年被升級為更具前瞻性的項目。例如，「金融技術」(序號 12) 升級為 2024 年「資料隱私、資料安全與網路安全技術」(序號 9)；「自動系統與機器人」(序號 8) 升級為 2024 年「高度自動化、自主和 UxS 無人系統及機器人技術」(序號 11)；「通訊與網路技術」(序號 10) 升級為 2024 年「整合通訊與網路技術」(序號 14)；「量子資訊技術」(序號 16) 升級至 2024 年「量子資訊及啟用技術」(序號 16)。而 2024 年序號 15 則以「定位、導航與 PNT 定時技術」取代 2022 年的「感知器」。

歐盟在 2022 年公布的「晶片法案」(European Chips Act)，提供 430 億歐元鼓勵公營與民間投資，以提升歐盟在全球晶片的市場佔有率。日本發布「下一代計算平台開發策略」、「量子技術創新策略」，開展在國家層面算力的研發部署。由此可知，各先進國家與科技龍頭企業為加速算力的提升已加快產品迭代升級。其發展重點約略可以分為晶片性能提升、運算記憶網路協同、計算平台統一調度等方面。

1. 晶片性能提升：除了晶片電路設計提升外，TSMC 半導體製程已進入 3nm 時代，與 5nm 相比，台積電電晶體邏輯密度可以提升 1.7 倍，性能有 11% 的提升。而在同等性能相比下，3nm 的功耗比 5nm 降低 25% 至 30%。Intel 的 Ponte Vecchio 加速器利用 Chiplet 技術，讓多個小晶片形成的 SiP，能夠將具備不同功能的小晶片，通過先進封裝技術整合於單一基板上，亦使得高效能解決方案成為可能。
2. 運算記憶網路協同方面：目前運算與儲存一體化的技術已由研究階段逐步進入商用化，即雲端計算和 DRAM 集成的作法，可以降低暫存器讀取的次數，進而減少延遲時間，三星 HBM2 DRAM 的技術即實現了此一可能。
3. 計算平台統一調度：利用統一的程式框架以實現不同架構運算平台的控制，在時間與空間的維度進行切分與調度。Intel 推出跨平台的 oneAPI 即為簡化跨平台架構之間的程式編程，方便 Intel 自身設備與其它廠商的晶片配合使用，以優化工作負載。

晶片系統效能的提升與系統架構的優化，將會是資料中心在節能與能源效率提升最核心的一環，也是目前全世界正努力的焦點之一。

#### (四) 冷卻系統

為達綠色算力之高效低碳目標，各階段均應通過不同的技術進行 PUE 的優化。如發電系統、輸配電系統、冷卻系統、負載系統、能源分配系統等，必須一體化監控管理，方能達到節能最佳化的訴求。算力最為關鍵的即為資料中心，在

耗能盤點下資料中心為耗能最大來源之一，為妥善處理此問題，各先進國家與企業均採取積極策略。

美國制定了資料中心整合計畫，以優化現有資料中心能源效率；歐洲亦提出「歐洲氣候中立資料中心公約」，期望到 2025 年歐洲資料中心使用再生能源電力要達到 75%，在 2030 年達到 100%。日本於 2020 年 12 月公告「綠色成長策略」明確表示在 2030 年之前資料中心的能耗要降低 30% 以上，並且資料中心的部份電力要轉換為可再生能源。日本在 2022 年頒布「能源合理化使用和非化石能源轉換法」，要求營運商提高資料中心設備的能源效率並降低 PUE 值。另外，德國在 2023 年公布「能源效率法案」，要求 2026 年 7 月必須將資料中心的 PUE 控制在 1.3 以下，同時要求一定規模的企業必須披露廢熱與 PUE 等相關訊息。此外，北美洲的 Google 資料中心，透過創新的市電供給、熱空氣隔離、水側節能等技術與大量的營運優化，成功地將 PUE 值降低到 1.12。同時，利用深度學習演算法不斷偵測 IT 設備能耗、室外氣溫以及製冷等設備的運轉情形，每 30 秒計算一次 PUE，依此建立系統運轉模型，進而改善資料中心的能源效率。

Meta 靠近瑞典北部的北極圈資料中心，通過數個巨型風扇引入室外極地的自然冷風為伺服器降溫，形成自然冷卻，其 PUE 能效值維持在 1.1 以下。Microsoft 位於新加坡的資料中心，通過淨化處理水的再生水與機械冷卻來保持伺服器在適當的溫度下運行。中國大陸以打造「零碳資料中心」為發展終極目標，善用天然資源，如「數字青海綠色大資料中心」，其地處青藏高原東北部的青海省，水量豐富、陽光充足、風力好、地又廣，同時氣候乾燥、冷涼。基於氣候優勢，青海建立的資料中心 PUE 在 1.2 以下，與大陸其他地區相比運營成本降低三成。

從以上國家與企業推動的方向來看，為了讓伺服器避免運轉產生的高溫而影響服務提供，冷卻系統在資料中心的重要性不言可喻。以下為資料中心常用的冷卻系統之技術分類：

1. 空氣冷卻系統：採用空氣作為主要冷卻介質，使用空調機組或通風系統進行熱量傳遞和散出。包括中央式空調、機櫃專用空調等。
2. 水冷系統：使用水作為冷卻介質，通過熱交換器將熱量從設備傳遞到水中。有集中式水冷和機櫃內部水冷兩種方式。
3. 液浸式冷卻系統：將設備直接浸入特殊冷卻液中進行散熱，可以大幅提高散熱效率。
4. 自然通風冷卻系統：利用自然對流和輻射效應進行散熱，無需機械設備，具有可靠性高、能耗低等優點。
5. 熱管冷卻系統：利用熱管原理將熱量從熱源傳遞到冷端，具有高效熱傳導、無需電力驅動等優點。

6. 製冷系統：採用製冷設備如冷水機組、製冷機等實現主動製冷，可精確控制出口冷空氣溫度。

傳統上以氣冷式系統為主，但在 PUE 與 ESG 之壓力下，其使用上備受挑戰。隨之而起的液冷式系統將更受歡迎，如 Nvidia 在 2024 Taipei Computex 推出的 GB200，其熱功耗（Thermal Design Power, TDP）超過 1000W，GB200 NVL72 基於 1U 標準機殼設計，將由氣冷散熱模式改為液冷散熱。預計 2025 年起，液冷及相關機構件產品在資料中心的滲透率將大幅起飛。

液冷技術為利用冷卻液降溫，常見的伺服器液冷技術可分為直接液冷（Direct Liquid Cooling, DLC）及浸沒式液冷（Immersion Cooling）。直接式液冷可運用機櫃方式提供，小型資料中心、企業級資料中心與大型資料中心均可導入；至於浸沒式液冷則適用於雲端資料中心，或是單櫃使用於邊緣資料中心，如圖 5.4 所示。



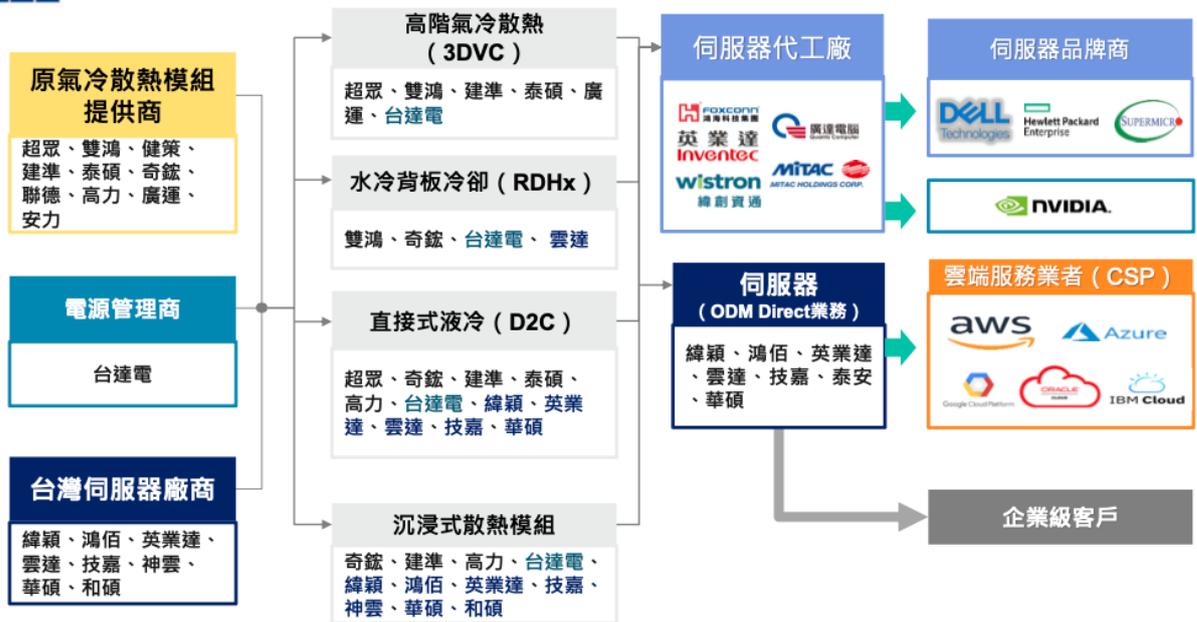
資料來源：Supermicro、GRC，MIC 整理，2024 年 4 月

**圖 5.4、直接式與浸沒式液冷比較**

台灣既有的散熱模組提供商、電源管理與伺服器廠商，紛紛開始布局液冷散熱，並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，詳如圖 5.5。



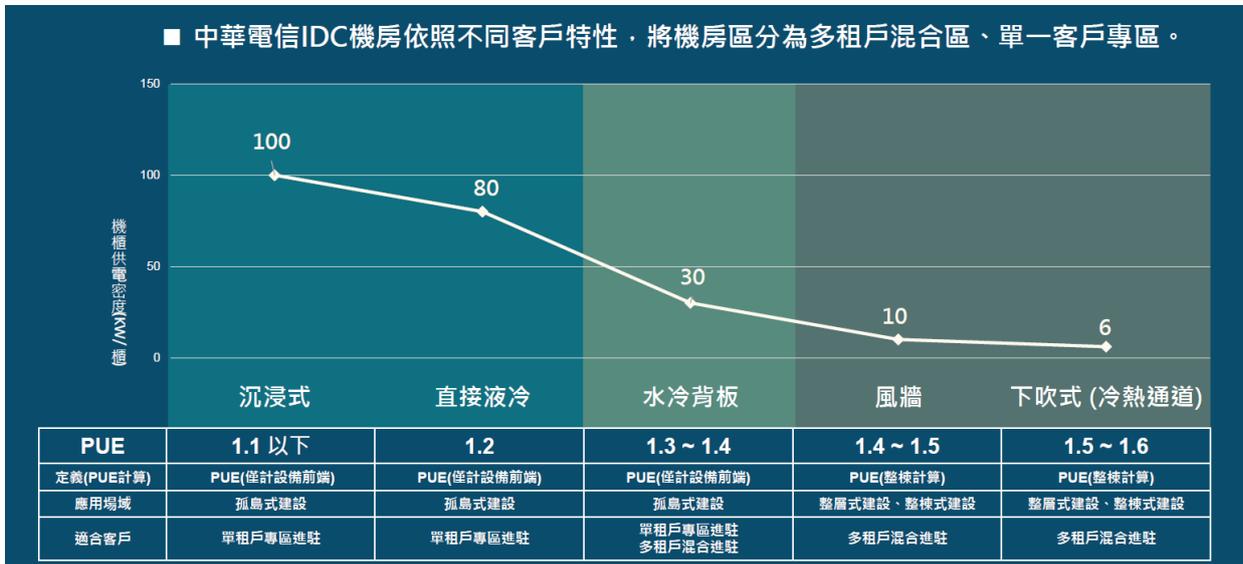
## AI伺服器帶動液冷散熱吸引台廠進行布局



資料來源：各公司，MIC 整理，2024 年 4 月

圖 5.5、伺服器帶動液冷散熱台廠布局情形

我們也看到中華電信針對不同客戶特性，將機房區分為多租戶混合區、單一客戶專區，所採取不同之冷卻方式與對應之 PUE 目標，詳如圖 5.6 所示。



資料來源：企業機房論壇/資料中心的綠色奇”機”，中華電信，Digitime，2024 年 5 月

圖 5.6、中華電信對不同客戶採取不同的冷卻方式

另外，從 Meta 與中國大陸「東數西算」給我們的啟示，善用各地區的優勢條件建立分佈式電廠，積極建設算力調度平台。以實現跨區域、跨領域、跨機房的動態調配以達最佳化之效益，亦是可供參考之方向。

### (五) 綠色能源開發與運用

根據過去經驗，台灣每年用電需求平均增長 1.8%，而隨著 AI 應用的迅速普及和全球廠商紛紛在台灣設立 AI 相關研發中心，未來用電需求將更加迫切。如何在滿足不斷增長的電力需求與實現淨零碳排放之間取得平衡，已成為一個亟需解決的重要政策議題。

在國際上，美國的資料中心在使用綠色能源方面非常積極，Amazon、Apple、Google、Meta、Microsoft、IBM 等公司為致力於碳中和，其資料中心運用了大量全美國企業可使用的可再生能源，且全球大型資料中心正朝著使用 100% 可再生能源運行的方向前進。根據國際能源總署（IEA）的預測，到 2025 年，再生能源將在三年內超越燃煤，成為全球最主要的電力來源。美國 2022 年的再生能源發電量首次超越燃煤發電，顯示出再生能源在該國的快速發展。國際再生能源機構（IRENA）認為，為應對氣候變遷，到 2030 年，再生能源的年度新增裝置容量需達到現在的三倍。

台灣也在能源轉型方面取得了顯著進展，2022 年綠色能源首次超越核電，再生能源在台灣能源結構中崛起。「再生能源發展條例」中要求新增建或改建的建築物達一定規模以上者，必須安裝一定容量的太陽光電發電設備，促進分散式再生能源的發展。同時，台灣的風電產業也快速發展，裝機容量已超過 2.1GW，為能源結構帶來了重要的變革。過去五年，台灣再生能源裝置容量的年均增長率達到 21.9%，高於全球平均水平，再生能源發電占比也達到 10%，顯示了台灣能源轉型的積極成果。如圖 5.7 亦可發現，為配合國家政策與國際趨勢中華電信逐年減少灰電的使用量，而提高綠電使用，IDC 機房目標於 2030 年 100% 使用再生能源。



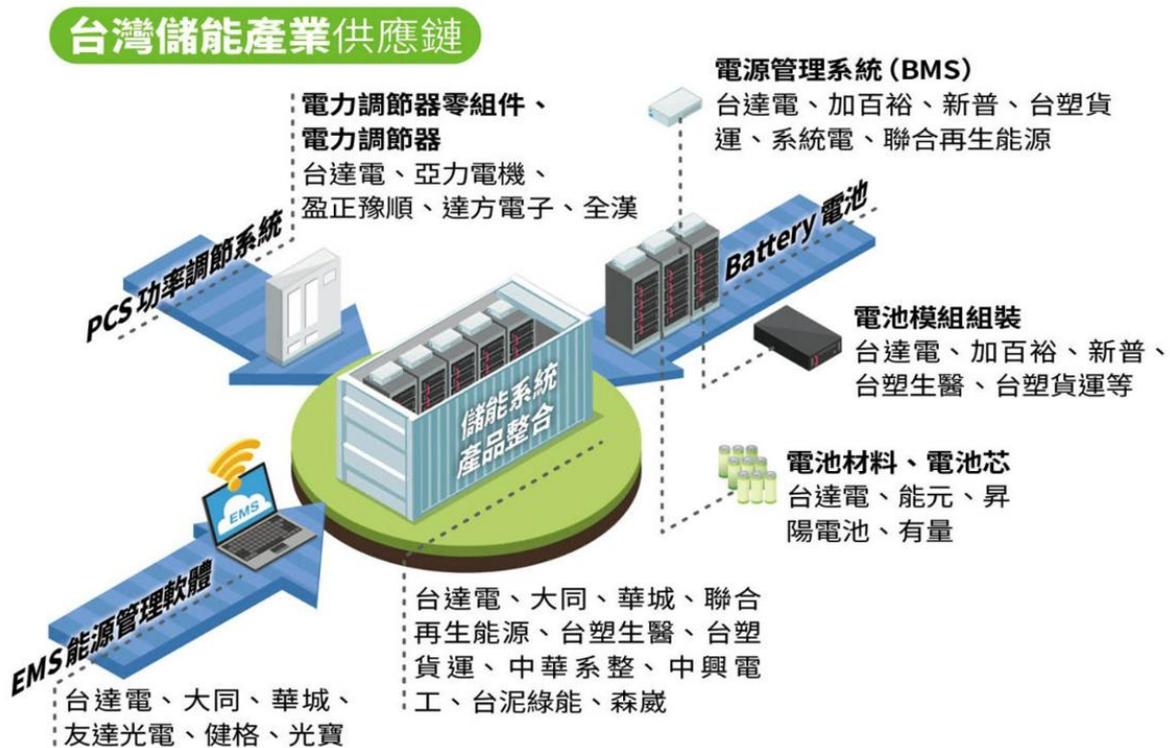
資料來源：企業機房論壇/資料中心的綠色奇「機」，中華電信，Digitime，2024 年 5 月

圖 5.7、中華電信 IDC 灰電與綠電使用情形

在未來以著經濟部於 2023 年 4 月核定之「臺灣 2050 淨零轉型「節能」關鍵

戰略行動計畫』為方針，配合綠色算力需求之趨勢有以下建議：

1. 應透過新興科技重新擬定基載電力的配比。
2. 完善布建 AMI 智慧電表，經由智慧電網架構透過能源物聯網與邊緣運算之技術，精進發展台灣在地化能源供需最佳化之智慧能管系統。
3. 深耕散熱「下水道工程」，研發掌握多元冷卻技術，降低建置成本與用電效率 PUE 向 1 邁進。
4. 創新低碳製程與高效率設備技術研發，以加速綠能產業發展。
5. 除了太陽光電、風力發電之外，尚可善用台灣既有天然資源進行綠色能源發電之研究，如地熱、小水力、洋流、生質能。透過遍地開花之策略，積小成多可以為能源調度帶來更彈性的選擇，以強化台灣電網之韌性。除此之外，氫能源的發展亦不可忽視。
6. 為了平緩再生能源間歇性特質，應積極推動儲能產業，包含電源管理系統、電池模組、電池材料與電池芯、功率調節系統等等，如圖 5.8 所示。
7. 建立更親民公開可靠的電業交易平台，以鼓勵更多分佈式能源業者的投入。



資料來源：儲能聯盟、工研院產科國際所(IEK)

圖 5.8、台灣應積極推動儲能產業，以最大化分佈式電網之效益

### 三、計算可持續性與應用案例探討

本節將針對全球永續發展趨勢、計算可持續性、資通技術及人工智能對推動

永續發展目標的影響，以及相關應用案例與未來展望，進行說明。

### (一) 全球永續發展趨勢

1987年，世界環境與發展委員會（The World Commission on Environment and Development）在聯合國發布的一份報告《我們共同的未來》（Our Common Future）中，除對地球現況和人類長久生計表達嚴重關切外，並奠立了「永續發展」（Sustainable Development）是在滿足當代需求的同時，不損及後代滿足其需求能力之發展的共識。後來，聯合國在2015年發布「2030年永續發展方針」（UN, 2015），進一步為人類和地球的和平與繁榮提供了現在和未來的共享藍圖。其所揭示的17項永續發展目標（Sustainable Development Goals, SDGs），不僅是世界各國應該努力的目標，同時也逐漸成為企業、學校、和個人實踐永續發展的共同指引，如圖5.9。



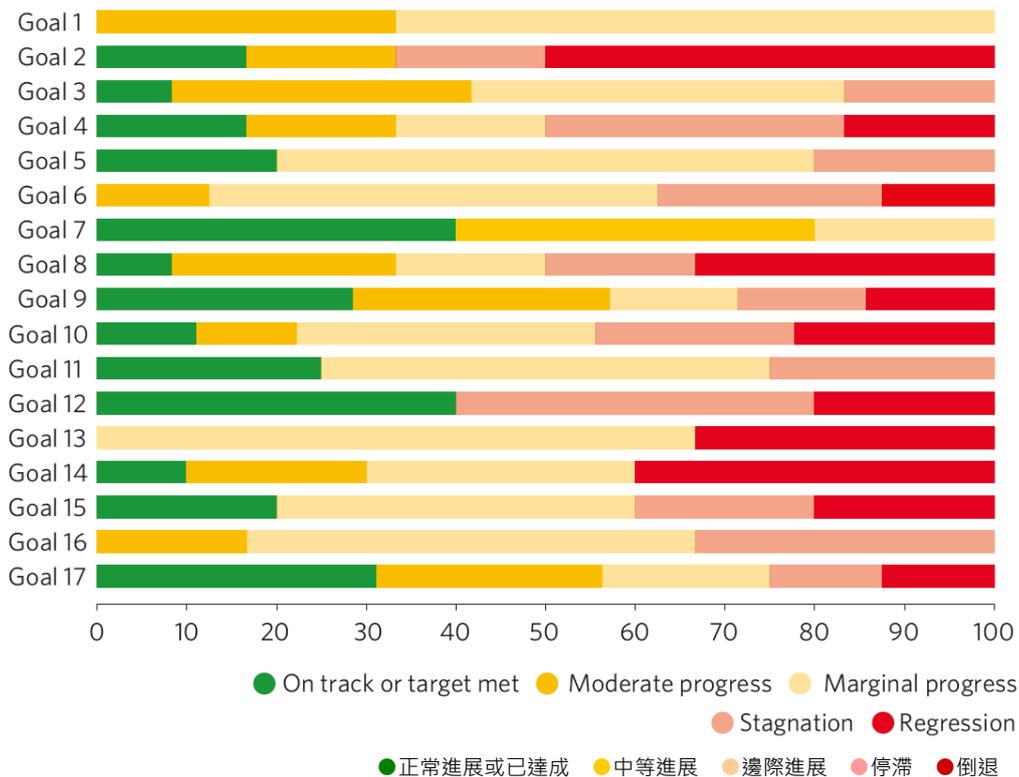
資料來源：聯合國永續發展目標網站

圖 5.9、聯合國於 2015 年發布 17 項永續發展目標（SDGs）

然而，聯合國在 2023 年和 2024 年的《永續發展目標報告》（The Sustainable Development Goals Report）中指出，永續發展目標陷入了困境，且 2024 年的進展評估顯示，全球在實現 2030 年議程方面嚴重失敗。當前只有 17% 的永續發展目標在進展上處於正軌，接近一半顯示微弱或中等進展，超過三分之一的目標則停滯不前或低於 2015 年基準水平，詳見圖 5.10。

新冠疫情、衝突、地緣政治緊張局勢和氣候危機對目標的實現造成了嚴峻挑戰。全球極端貧困率因疫情上升，預計到 2030 年仍有 5.9 億人生活在極端貧困中；而世界的飢餓情形處於 2005 年以來最糟狀況，饑餓人數在 2023 年達到 7.33 億，亟需改善食品系統的可持續性與韌性。教育總體進展也面臨挑戰，尤其是在

數學和閱讀技能上降幅顯著，到 2030 年，約 8400 萬兒童將失學。女性在政治和經濟領域的參與仍需加速，儘管法律改革有所進展，但不能跟上人口增長的步伐。全球尚有 6.85 億人未獲得電力，並且食品浪費和電子垃圾的增長未得到妥善處理。氣候行動方面，全球氣溫持續上升，二氧化碳含量已達到 200 萬年來從未見過的水平。現行各種短視的經濟和政治制度對自然的戰爭愈演愈烈；可能還需要 25 年的時間才能制止毀林行為，而全球大量物種正面臨滅絕的威脅。面臨如此嚴峻的挑戰，各國政府更應付出前所未有的努力，加強國際合作，加速政策創新與資金流動。企業、民間社會、科學界、青年、地方當局等各個方面需要建立與永續發展目標有關的全球行動聯盟，以確保所有人都能共享永續發展的成果。

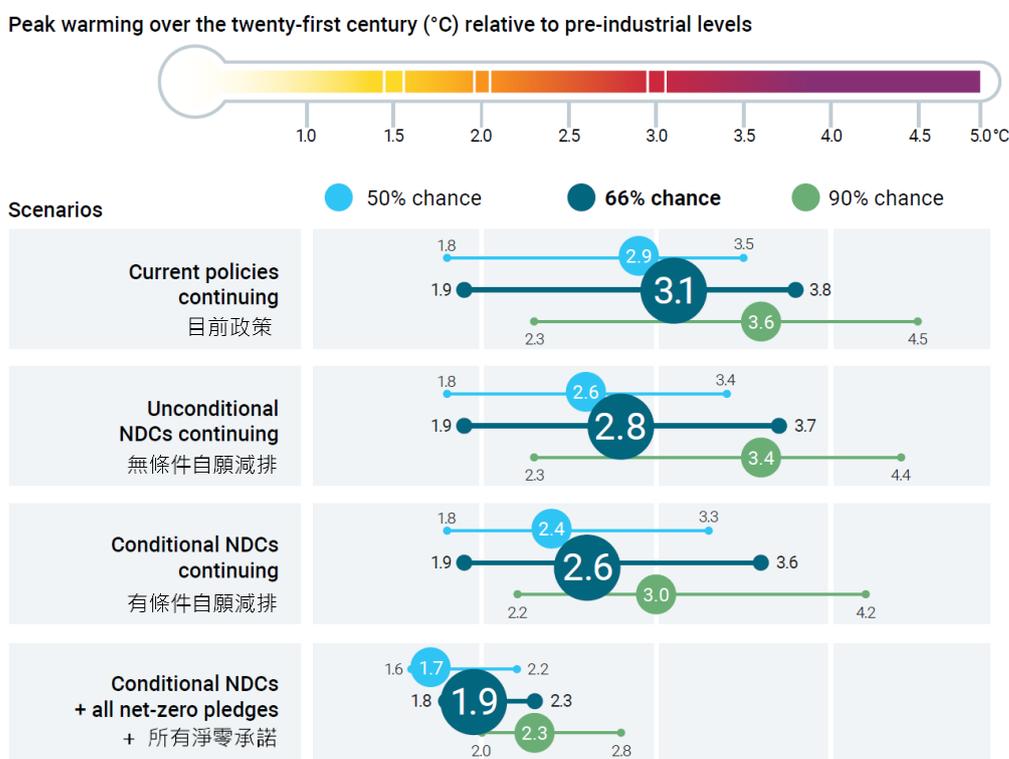


資料來源：聯合國 2024 年《永續發展目標報告》

圖 5.10、2024 年 SDGs 17 項目標進展評估（百分比）

此外，聯合國環境規劃署（UNEP）在 2023 年和 2024 年《排放差距報告》（Emissions Gap Report）中亦指出，全球溫室氣體排放在 2023 年創下 570 億公噸的新紀錄，較 2022 年增加 1.3%，其中發電部門仍是最大的排放來源。G20 國的排放在 2023 年亦有所上升，未來必須加速減排以達成目標。除非各國近十年進一步減少溫室氣體排放量，否則全球溫度將比工業革命前高出 2.5°C 至 2.9°C。這將使人類提早面臨 IPCC（政府間氣候變化專門委員會）科學家於 2014 年在溫室氣體排放缺口第五次評估報告(The Fifth Assessment Report, AR5)所預言的氣候變遷災難，包括：「全球冰河體積進一步減少」、「海平面上升速度加快」、「乾溼區與乾濕季降雨對比增加」、「海洋持續吸收二氧化碳增加海洋酸化」、以及「氣候變遷將影響碳循環，進而加速大氣中二氧化碳濃度的增加」等。

圖 5.11 顯示了在不同情境下預測的全球變暖程度，相對於《巴黎協定》的溫度目標。該圖表展示了根據目前政策、無條件 NDCs (Unconditional Nationally Determined Contributions, 是指各國在《巴黎協定》框架下所提交的自願減排目標)、有條件 NDCs (Conditional Nationally Determined Contributions, 是指各國設定的減排目標, 但這些目標的實現依賴於外部資源或支持, 例如國際援助、技術轉讓或資金支持)、以及所有淨零承諾的情境, 各個預測的溫度上升情境。數據顯示, 若按照當前政策的延續, 全球變暖的最高預測可以達到 3.1°C, 而根據無條件和有條件 NDCs 情境, 預測分別為 2.8°C 和 2.6°C。這些結果顯示, 在達成 1.5°C 的目標幾乎為零的情況下, 若不加強行動, 全球變暖仍會高於這一目標。唯一接近《巴黎協定》目標的情境是所有國家完全實施的最為嚴格的承諾, 這預測在世紀末將全球變暖限制在 1.9°C。



資料來源：聯合國環境規劃署 (UNEP) 2024 年《排放差距報告》

圖 5.11、不同情境下預測的全球暖化程度(相對於《巴黎協定》溫度目標)

因此,「SDG13 氣候行動」即鼓勵各國透過五個細項指標,採取緊急措施以因應氣候變遷及其影響,包括:「加強對氣候災害的抗災能力和調適能力」、「將氣候變遷措施納入政策和規劃」、「建立應對氣候變化的知識和能力」、「執行聯合國氣候變遷框架公約」、以及「促進建立機制,提高規劃和管理能力」。

「SDG11 永續城鄉」則鼓勵各國透過十個細項指標,促使城市與人類居住具包容、安全、韌性及可持續性,包括:「安全可負擔的住宅」、「負擔得起和永續的運輸系統」、「包容性永續城市化」、「保護世界文化和自然遺產」、「減少自然災害

的不利影響」、「減少都市對環境的影響」、「提供安全和包容的綠色和公共空間」、「強化國家和地區發展規劃」、「執行包容、資源效率和減少災害風險的政策」、以及「支持最低度開發國家進行永續和有彈性的建築」。

而透過永續發展教育 (Education for Sustainable Development, ESD)，使所有年齡層的學習者在應對相互關聯的全球挑戰時(包括氣候變遷、生物多樣性喪失、資源的不可持續利用、以及不平等)，能夠做出明智的決定，並採取個人和集體行動來改變社會和關懷地球，則是落實這 17 項永續發展目標的關鍵促動力。同時有助於在 21 世紀的今日，帶動全體人類朝向健康永續的生活型態 (Lifestyles of Health and Sustainability, LOHAS) 轉型，親身實踐健康的飲食、生活、身心探索與個人成長，並且致力於永續維護地球上的各種生命。

## (二) 計算可持續性

受到聯合國《我們共同的未來》啟迪，以及後來 SDGs 17 項永續發展目標的助力，近年來計算可持續性(Computational Sustainability)因應全球永續發展倡議，成為一個國際關注的新興領域。該領域研究者主要是藉由數學、計算機科學和資訊科學等方法，致力於在未來人類福祉的基礎上平衡社會、經濟和環境資源。例如，平衡環境與社會經濟需求、生物多樣性和保護、以及再生能源與材料等。美國康乃爾大學 Gomes 教授將計算可持續性定義為：「一個新的跨學科研究領域，總體目標是開發計算模型、方法和工具，以幫助管理永續發展的環境、經濟和社會需求之間的平衡」。因此，Chatterjee and Rao 認為計算可持續性可以包含計算系統中的可持續性問題，即所謂的「綠色資訊技術 (Green Information Technology)」或「永續計算 (Sustainable Computing)」。

Gomes 教授與多位學者並於 2019 年提出了具有前瞻性倡議，主張計算機科學家能夠並且應該對解決人類面臨的主要社會和環境挑戰作出重要貢獻，以追求永續的未來；而計算可持續性領域匯集相關努力的同時，也豐富了計算機科學。計算可持續性可以應用計算技術和運算思維，例如：人工智能、數據與機器學習、動態模式與模擬、賽局理論、優化、最佳化、多代理人系統、群眾外包與公民科學等，來促進可持續性，以及減少計算技術本身對環境不利影響的相關問題。重要的是，一大群計算機科學家與更多來自社會、環境和自然科學的領域專家合作，能夠以在較小或較少的跨學科環境中不可能實現的方式推動計算可持續性。他們希望在計算可持續性方面的工作能激勵更多的計算科學家追求具有廣泛社會影響力的舉措。

德國哥廷根大學的 Kolbe 教授和柏林工業大學的 Zarnekow 教授早於 2009 年即指出，永續發展的概念對於資通技術 (Information and Communication Technology, ICT) 產業而言仍然是較新的觀念，且缺乏整合經濟、生態與社會層面的 IT 管理框架。這兩位學者強調，制定永續資訊管理框架對於指導 ICT 產業及政策制定者至關重要。同時，永續資訊管理的研究可以聚焦在以下幾個領域：

綠色 IT 管理理念、基於永續 IT 管理的競爭優勢、綠色 IT 策略調整、記分卡和成熟度模型的開發，以及永續發展報告（包含 KPI 的定義與衡量）。

可持續性 (Sustainability) 是一個跨域整合的多維概念，通常以整體可持續性為中心，並以三個相交的圓圈，表示環境永續、經濟永續和社會永續等三大支柱。這三大支柱旨在促進和諧互動，在保護自然系統的同時實現經濟增長並維持社會公平。早在 1980 年，國際自然保護聯盟 (International Union for Conservation of Nature, IUCN) 就提出「世界自然保護策略」。這份報告影響了《我們共同的未來》，並為現在的永續發展共同原則奠定了基礎。這些發展促使一些國際組織在 1980 年代中期採取重大舉措，加強努力將研究和開發納入有效策略，以實現永續發展。此後，許多學者進一步擴展了這一概念，認為永續發展應當包括經濟增長、社會進步和環境保護的三重底線。

政策制定和治理是實現可持續性目標的關鍵，這包括國際間的合作、國家級政策調整，以及地方政府和社區的參與。Hansen 等人指出政策在促進環境創新和系統性變革的重要性。要解決生態挑戰，僅靠技術創新是不夠的，還需經濟、社會和制度的全面變革。政策可以透過激勵機制和法規來推動創新，並利用多樣的工具如經濟激勵、規範和支持研發。同時，透過多方協作，包括政府、業界、學界和公民社會的合作，對實現永續發展至關重要。可持續性的應用遍及不同行業和領域。在企業層面，企業社會責任 (Corporate Social Responsibility, CSR) 被認為是實現永續發展的有力工具之一，因其能夠促進企業行為與社會需求及環境保護的統一。在城市發展方面，智慧城市戰略正日益受到重視，這有助於資源的有效管理和城市居民生活質量的提升。

近年來，可持續性已擴展至各個領域，包括科技領域。在軟體工程中，可持續性被視為一個新興的品質目標，旨在促進長期維護和節能。Calero 和 Piattini 於 2017 年提出了一個框架，從技術、經濟、社會、環境和個人經驗等多個面向全面評估軟體的可持續性。其中，技術可持續性涉及軟體的長期運行能力，包括可維護性、可擴展性和可靠性，確保其能適應未來的技術變化。經濟可持續性強調成本效益分析，確保軟體在整個生命週期中提供經濟價值，並最佳化資源利用，維持長期財務可行性。社會可持續性則關注軟體對用戶和社會的影響，包括可及性和滿足社會需求的能力。環境可持續性考量開發、部署和運行過程中的環境影響，如能源消耗和碳足跡，因此可持續的軟體設計應致力於減少生態負面影響。最後，個人可持續性聚焦於軟體對用戶和開發者的影響，包括使用方便性和工作壓力，以提升用戶體驗和開發者的幸福感。

### **(三) 資通技術及人工智能對推動永續發展目標之影響**

隨著全球對 SDGs 的重視，資通技術 (ICT) 和人工智能 (Artificial Intelligence, AI) 已成為實現這些目標的重要助力。根據國際電信聯盟 (International Telecommunication Union, ITU) 於 2021 年的報告，ICT 在推動 SDGs 方面具有關

鍵作用，特別是加速工業化、創新及基礎建設(SDG 9)的實現。對於經濟欠發達國家來說，ICT 的發展能提升經濟競爭力、改善金融包容性、減少貧困並提高健康水平。

在農業、教育、醫療和金融等領域，ICT 的應用同樣至關重要。透過數位金融服務的推廣，ICT 幫助減少貧困和飢餓；而改善連接性則提升了遠距醫療和健康資訊的品質。此外，數位技能的提升促進了教育的普及和更多工作機會的創造。ICT 還支持性別平等、清潔能源、減少不平等以及智慧城市等發展目標。儘管在 SDGs 的 169 個細項目標中，ICT 未被廣泛提及，但聯合國仍強調其在各目標實現中的重要性。Adam 和 Alhassan 於 2021 年透過分析 143 個國家的資料說明，ICT 的發展不僅推動經濟增長，還在社會和環境方面發揮了積極作用，而人類發展進一步增強了 ICT 對 SDGs 的貢獻。

但是，ICT 對 SDGs 仍然存在潛在的負面影響。根據 Rothe 等人於 2022 年報告分析，儘管 120 篇聯合國的 ICT 相關出版物強調了 SDGs 的全局性和全球性，但在實施發展策略時，對 ICT 的附帶效應和潛在負面影響的考量卻相對不足。為了真正達成 SDGs，ICT 的應用需考慮其在環境、社會和經濟等多個發展領域的綜合影響。因此，未來的 ICT4D (ICT for Development) 研究應重點識別並減少 ICT 帶來的負面外部性，同時促進正面協同效應。

隨著 AI 的發展，基於 AI 的創新將在多個領域進一步影響永續發展，包括改善環境、促進經濟成長和增進社會福祉。根據 Nahar 於 2024 年的研究，利用系統動力學建模對 2022 至 2030 年期間的 22 個國家進行預測，結果顯示 AI 的影響因 SDGs 而異，對環境結果、生產力、包容性和公平性具雙重影響。AI 可以優化能源系統、提高效率及減少浪費，同時有效預測自然災害，並改善農業和水資源管理，這些都支持永續發展並推動循環經濟及可持續產品的開發。然而，AI 在推進永續發展的過程中也可能加劇財富不平等。因此，在推進 AI 技術時，必須尋求經濟、環境和社會之間的平衡，以實現更全面的永續發展。

此外，Gupta 和 Degbelo 於 2022 年探討了 AI 在推動 SDG 11 永續城鄉方面的角色。AI 在廢棄物管理、空氣品質監測及災害應變管理等領域的應用表現出顯著的潛力，但目前大部分 AI 項目仍是為市民提供服務，而非與市民協作，因此公民參與不足。事實上，市民參與對 SDGs 的推進至關重要，只有透過積極的公民參與，AI 系統才能更有效地支持政策制定，增強信任和包容性。

Vinuesa 等人於 2020 年時，透過共識專家調查指出，AI 可以促進 SDGs 134 個細項目標的實現，但同時也可能阻礙 59 個細項目標，突顯了 AI 在永續發展中的雙面性。AI 的應用範圍廣泛，其正面影響包括在醫療診斷、農業生產力和資源使用效率等方面的潛力。然而，負面影響則包括倫理問題、隱私侵犯及因技術進步造成的就業不平等。研究強調，為了讓 AI 技術真正促進永續發展，必須及時制定和執行監管框架，以確保 AI 的透明性、安全性和倫理合規性。如果缺乏這

些監管，AI 的快速發展可能會引發透明度不足、安全隱憂及倫理標準缺失等問題。

同樣，Theodorou 等人於 2022 年的研究分析了北歐國家的 AI 策略，發現儘管北歐國家在 SDGs 推動上表現卓越，其 AI 策略文件對 SDGs 的提及頻率卻與其他國家無顯著差異。值得注意的是，這些文件對性別平等（SDG 5）和減少不平等（SDG 10）的內容明顯不足，且對 AI 在環境影響方面的關注亦不夠。因此，研究指出，缺乏對負責任 AI 原則的重視將限制 AI 在達成 SDGs 中的潛力。

Jungwirth 和 Haluza 於 2023 年進一步指出，AI 在推動 SDGs 時有雙重性：一方面，AI 能促進社會、經濟和環境的發展；另一方面，如果不當使用，AI 可能成為阻礙因素。他們利用 GPT-3 模型探討了 AI 在教育、健康和通信等領域的潛力，指出 AI 能通過數據分析預測貧困區域和優化教育資源分配，同時也面臨數位落差、隱私侵犯和倫理問題等風險。缺乏適當監管的 AI 應用可能導致透明度、安全性和倫理標準不足。

綜上所述，ICT 和 AI 在推動 SDGs 方面展現出巨大的潛力。這些技術不僅提升了社會福祉，還促進了經濟成長和環境保護。然而，與之伴隨的數位落差、個人資料保護及演算法偏見等挑戰亟需重視。AI 和 ICT 一樣，屬於資訊時代由計算機科學領域專家所開發的人工系統。儘管 AI 能夠模擬人類智能過程，不可取代人類的心靈、智慧。

為了確保科技應用能支持永續發展的長期目標，我們必須加強監管、推動多方合作並提高公民參與。隨著人們對生命本質的認識加深，以永續發展教育（ESD）啟發生命智慧將成為實現 SDGs 的根本解決之道。許多社區開始重視在地文化及其在永續發展中的角色，呼籲保護傳統知識和技藝，並將其融入現代經濟，以促進文化認同和社會凝聚力。同時，許多藝術家和文化工作者透過作品融合永續發展議題，促進公眾對 SDGs 的實踐。聯合國教科文組織（United Nations Educational, Scientific and Cultural Organization, UNESCO）也提倡將「文化和創意產業視為永續發展的一部分」。因此，善用 AI 和 ICT 於文化的永續發展，不僅有助於文化的保存和傳播，還能結合永續行銷（Sustainability Marketing），擴大參與並持續促進永續發展。

#### （四）應用案例與未來展望

##### 1. 高效能運算與綠色資料中心的結合

高效能運算（HPC）與綠色資料中心的融合已成為資訊科技領域的一項重要趨勢。位於芬蘭東北部卡亞尼市（Kajaani）的 CSC 資料中心是世界最具生態效率的資料中心之一，它在歐洲高效能運算的發展中扮演著關鍵角色（如圖 5.12）。該資料中心的 LUMI 超級電腦於 2024 年被評選為全球第十二台綠色超級電腦，並由來自芬蘭、比利時、捷克、丹麥、愛沙尼亞、冰島、荷蘭、挪威、波蘭、瑞

典和瑞士等十一個歐洲國家的 LUMI 聯盟共同運營。LUMI 的持續計算能力達到 380 PetaFLOPS，相當於 150 萬台最新型筆記型電腦的綜合性能，這使其能夠有效支援氣候變遷、AI 等多個資料密集型的尖端研究。



資料來源：歐洲高性能計算共同企業（EuroHPC JU）網站

圖 5.12、位於芬蘭卡亞尼市 CSC 資料中心是 LUMI 超級電腦的綠色之家

作為歐洲高性能計算共同企業（EuroHPC）的重要成員，LUMI 的設計與運營充分體現了高效能計算與環境可持續性的完美結合。該超級電腦完全依賴 100% 可再生的水力發電，並將產生的廢熱用於為當地數百戶家庭提供暖氣，顯著減少二氧化碳排放。LUMI 不僅為歐洲的科研與創新提供了世界級的計算資源，還促進了數據驅動經濟的發展，進一步強化了歐洲在全球高效能運算領域的領先地位。

另一個值得參考的典範是位於挪威埃內巴克城鎮的 Green Mountain 資料中心（如圖 5.13）。隨著全球對永續發展的重視，Green Mountain 憑藉其創新的營運模式和先進的技術能力，有效負責任地對待環境和資源，在可再生能源、熱能回收和廢棄物管理等各方面表現卓越。該中心利用挪威寒冷潮濕的氣候進行免費冷卻（Free Cooling），透過自然冷卻空氣與水，取代了傳統需耗費大量電力的機械制冷系統，顯著降低能源消耗。例如，其 SVG-Rennesøy 設施的冷水來源於相鄰的峽灣，利用重力引流，穩定維持 8°C 的水溫，使 PUE（電力使用效率）降低至 1.2 或以下，遠低於業界平均水平。

在人才發展方面，Green Mountain 充分認識到人力資源是企業成功的關鍵，因此實施了一系列人才與能力發展策略，包括能力矩陣、培訓需求識別和人才發展計畫。學徒計畫為年輕人才提供專業培訓，以確保他們能持續吸收新知並貢獻於資料中心的運營。而員工在「綠色學院」系列訓練中的參與，也進一步培養了持續學習的文化，提升了整體營運效率。

此外，Green Mountain 全面採用挪威的水力發電，100%依靠可再生能源，並且幾乎零碳足跡，且具有強大的供電安全性。挪威在能源自給自足上表現優越，具備極佳的水力發電儲存條件，確保在需求高峰期仍能穩定供應。客戶在 Green Mountain 運行的資料中心中，隨時感受到來自清潔能源的穩定電力支持。



資料來源：Green Mountain 資料中心網站

圖 5.13、Green Mountain 位於挪威埃內巴克城鎮的 OSL-Enebakk 資料中心

整體而言，Green Mountain 透過整合先進的冷卻技術、員工發展策略和對可再生能源的承諾，不僅展示了如何有效運作高效能運算資料中心，還樹立了全球綠色數據中心的標竿。隨著各項可持續性計畫的推進，Green Mountain 正穩步邁向成為未來綠色數據中心的典範，促進行業夥伴共同推動數據中心的數位轉型與永續發展。

## 2. 城鄉數據智慧治理

城鄉數據智慧治理的核心在於構建一個跨單位的數據整合、流通與共享平台，旨在提升城鄉之間的協同管理與運作效率。該平台結合 AI、IoT 及資料倉儲（Data Warehouse, DW）等先進技術，實現即時監測和基礎設施管理，促進更有效的決策和分析。透過數位轉型，這一治理模式不僅改善公共政策的決策品質，還能活化產業應用，最終達到促進城鄉永續發展的目標。

具體而言，這一模式確保政府部門、企業和民間之間的數據共享，有效打破資訊孤島，形成全面的發展規劃。IoT 技術使基礎設施的運行狀況（例如水管和交通系統）能夠進行實時監控，從而能夠早期發現並解決潛在問題。同時，AI 技術為政府提供基於數據的決策支持，幫助制定符合實際需求的政策，促進產業創新，推動智能解決方案的實施，進一步增強產業的競爭力和效率。因此，城鄉數據智慧治理實現了數位技術與城鄉發展的深度結合，確保在經濟、社會及環境等多個方面達成可持續成長。

根據瑞士洛桑管理學院（International Institute for Management Development,

IMD) 在 2024 年發布的《IMD 智慧城市指數》，在 142 個接受評比的城市中，蘇黎世、奧斯陸、坎培拉、日內瓦、新加坡、哥本哈根、洛桑、倫敦、赫爾辛基、阿布達比、斯德哥爾摩、杜拜、北京、漢堡、布拉格、台北、首爾、阿姆斯特丹、上海、和香港名列前 20 (見表 5.7)。2024 年的數據顯示，與 2023 年相比，前 20 名城市大致相同，仍包括蘇黎世、奧斯陸和新加坡等，顯示智慧城市的全球格局變化緩慢。但值得肯定的是，臺北市首次進入前 20 名，躍居第 16。排名前 40 的城市共同特點在於，其社會和經濟環境相對可預測，並積極推動各項城市生活措施。然而，這些城市在基礎設施和安全等關鍵領域仍面臨挑戰，需要制定有效的策略以因應未來的不確定性。此外，數位轉型成為城市未來發展的關鍵，城市應重視信任和治理的建立，以實現綠色、數位化及以人為本的發展目標。

表 5.7、2024 年 IMD 智慧城市排名 (前 20 名)

City	Smart City Rank 2024	Smart City Rating 2024	Structure 2024	Technology 2024	Smart City Rank 2023	Change
Zurich	1	AAA	AAA	AA	1	—
Oslo	2	AA	AA	A	2	—
Canberra	3	AA	AAA	A	3	—
Geneva	4	AAA	AAA	AA	9	+5▲
Singapore	5	A	A	A	7	+2▲
Copenhagen	6	AA	AA	A	4	-2▼
Lausanne	7	AA	AA	A	5	-2▼
London	8	A	BBB	AA	6	-2▼
Helsinki	9	AA	AA	A	8	-1▼
Abu Dhabi	10	BB	BB	BB	13	+3▲
Stockholm	11	A	A	A	10	-1▼
Dubai	12	BB	BB	BB	17	+5▲
Beijing	13	BB	BB	BB	12	-1▼
Hamburg	14	BBB	BBB	BBB	11	-3▼
Prague	15	A	A	A	14	-1▼
Taipei City	16	A	BBB	A	29	+13▲
Seoul	17	AA	BBB	AAA	16	-1▼
Amsterdam	18	A	BBB	A	15	-3▼
Shanghai	19	BB	BB	BB	25	+6▲
Hong Kong	20	A	BBB	AAA	19	-1▼

資料來源：IMD Smart City Index 2024

進一步觀察瑞士洛桑管理學院 (IMD) 2024 年發布的《IMD 世界競爭力年報》，在 67 個接受評比的國家中，新加坡、瑞士和丹麥分別位居前三名 (見圖 5.14)。對照表 5.7 的前 20 名智慧城市，可以發現這三個國家的城鄉數據智慧治理模式非常值得參考。而臺灣在 IMD 世界競爭力位居第八，在超過 2,000 萬人口的經濟體中，已連續四年榮登世界第一，發展潛力顯著。

## 2024 COMPETITIVENESS RANKING

			Score		
01	Singapore		100.00	↗	3
02	Switzerland		97.55	↗	1
03	Denmark		97.07	↙	2
04	Ireland		91.86	↙	2
05	Hong Kong SAR		91.49	↗	2
06	Sweden		90.30	↗	2
07	UAE		89.75	↗	3
08	Taiwan (Chinese Taipei)		88.50	↙	2
09	Netherlands		86.94	↙	4
10	Norway		86.22	↗	4

資料來源：2024 年 IMD 世界競爭力年報

圖 5.14、2024 年 IMD 世界競爭力排名（前 10 名）

瑞士在智慧城市發展方面取得卓越成就，特別是在技術整合和提升城市生活品質方面。根據 2022 年的調查，超過 50 個瑞士城市和市鎮積極推動智慧城市概念，並採用 LoRaWAN（Long Range Wide Area Network）低功耗廣域網路協定與技術支持物聯網應用，這有助於提升城市運行效率和可持續性。蘇黎世、洛桑和日內瓦等城市有效利用 LoRaWAN 進行空氣品質監測和停車空間優化，改善城市功能，推動智慧生活。蘇黎世在治理、活動、健康和 safety 等領域的優異表現，於 2023 年和 2024 年皆名列 IMD 智慧城市指數全球榜首；洛桑在開放空間和公共交通在線服務方面獲得好評；日內瓦則專注於教育、城市綠化和市民參與，利用 LoRaWAN 監控環境變數，推動其智慧城市計畫。

新加坡的智慧城市（Smart City）方案則以實現全面數位化為願景，屬於國家級計畫，旨在透過數位轉型來提升健康、交通、城市生活、政府服務和商業等領域，並保持在 IMD 智慧城市指數及世界數位競爭力排名中的領先地位。該計畫以「數位化為核心，服務以人為本」為理念，推動數位經濟成長，吸引外資並創造就業，特別強調數位社會的包容性和全民數位素養的提升。透過政策和立法支持，該計畫促進公私合作，促進技術創新。新加坡推出了如 Project Pensive、SELENA+ 及 OneService Chatbot 等 AI 應用工具，並建立 SGTraDex 和 SGFinDex 等公共數位基礎設施，優化數據共享和財務資訊整合。安全系統如 ScamShield 及 SMS Sender ID Registry 有效防範詐騙，而 CentEx 和 LifeSG 等平台則提升公共服務效率。Singpass 及 OpenCerts 確保數位交易和學術認證的安全性，這些措施進一步鞏固了新加坡在全球智慧城市中的領導地位，增強了居民在數位轉型過程中的便利與安全感。

在台北，台北大數據中心在智慧城市治理中發揮了關鍵作用。該中心透過先進的數據策略推動城市管理的數位化和智能化轉型，專注於跨域數據整合，匯集

來自物聯網、業務統計和空間資訊等多方資料，提供即時且全面的支持。這幫助市政府各部門快速回應市民需求和城市變化。在交通管理和公共安全領域，該中心運用大數據技術深入分析，協助政府制定精準政策。同時，市政府通過設立資料治理委員會保障數據安全與隱私，進一步確保數據應用的合法性。中心的智慧儀表板透過模組化、客製化的數據展示工具提升了部門協作與效率，尤其在疫情治理中，透過整合全球和本地醫療數據有力支持防疫指揮，因此獲得國際認可。在綠色交通方面，台北大數據中心協助優化 YouBike 運用並選擇新站點，支持達成 2030 年 70% 綠色運輸目標。作為一個強大的數據運用平台和創新引擎，該中心將持續探索大數據的潛力，發展智慧城市解決方案，全面升級城市治理，提升市民生活品質。

另一方面，台東正在透過數位科技推動智慧城市建設，將智慧治理深植於城市的發展 DNA 之中。台東縣政府運用先進技術和數據平台，提升公共服務效率強化市民生活品質，含蓋交通管理、環境監測和社區參與等各個面向，這不僅促進了當地經濟的永續發展，還提升了居民的社會參與感和滿意度。台東縣政府推出的 TTPush 平台作為推進數位治理城市的重要工具，因其卓越表現而榮獲 2023 智慧城市創新應用獎。此平台促進了數位點數經濟，激勵居民和商家積極參與數位經濟活動，提升地方消費和社區參與。

算力在城鄉數據智慧治理中扮演著重要角色，負責數據的處理與分析，支持即時監測和智能應用，並促進社會參與和創新。隨著 IoT 設備的普及，強大的算力變得越來越重要，成為快速數據處理和分析的必要條件。高效的數據分析不僅提高了政府管理效率，還能確保資源配置更精準。然而，隨著數據量的快速增加，僅依賴數據進行決策可能導致誤解，不利於城市治理。因此，英國倫敦大學學院（University College London, UCL）Michael Batty 教授，也是著名的城市規劃學者，特別強調大數據需與理論結合才能得出真正有意義的結論，並在政策制定時採取合理的參考。

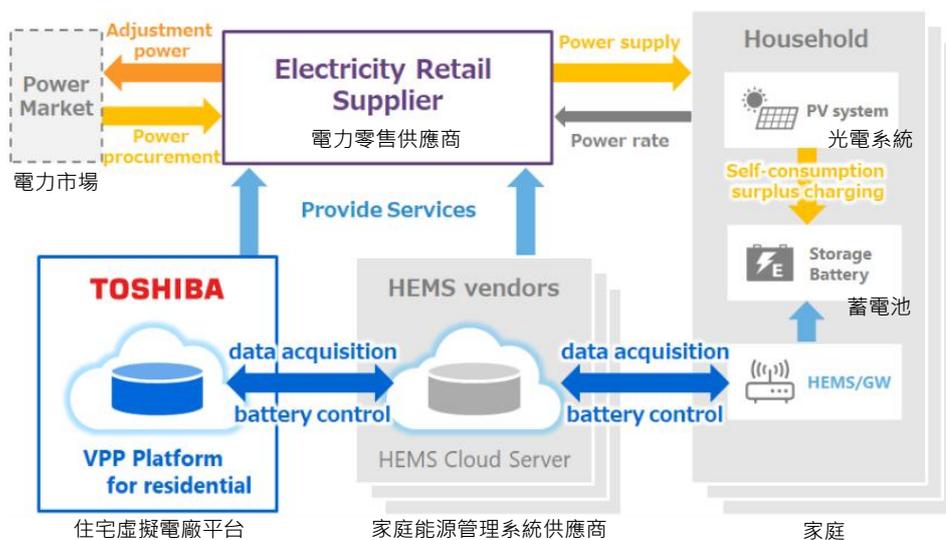
### 3. 分散式能源管理

隨著全球對可再生能源的重視，社區能源正迅速成為推動公正和彈性能源系統的重要力量。這些社區能源項目通常由當地居民設立和擁有，包括屋頂太陽能、地方風力發電和水力發電，服務於鄰里、合作社及小型企業，使居民成為能源的生產者和消費者，並促進低碳轉型。分散式能源資源管理系統（Distributed Energy Resource Management System, DERMS）結合硬體和軟體，能實現多種資源的即時通信和控制，幫助平衡能源供需及提高運營效率。DERMS 功能廣泛，包括資源監控、消費模式預測及設備行為控制，對能源轉型至關重要。在能源系統從集中生產向包含分散資產的複雜模式轉變的過程中，DERMS 可分為網關型和雲端型，前者提供低延遲及離線能力，後者則透過應用程式介面（Application Programming Interface, API）連接資源。

依 Market Research Network 於 2024 研究報告指出，在德國 DERMS 按照應用分為幾個主要領域，包括住宅、商業、工業、機構和公用事業等。隨著可再生能源倡議的興起，越來越多的家庭和企業採用 DERMS 以優化能源消耗並整合可再生能源資源。另外，依據 Verified Market Reports 於 2024 年研究報告指出，德國的 DERMS 市場估計達到 16 億美元，預計到 2030 年將增長至 45 億美元，年均成長率為 15.8%。DERMS 是一種高階平台，能夠即時管理、控制和優化分散能源資源，透過先進的算法和數據分析來處理多個能源資源與主電網之間的複雜關係。主要應用包括監測、預測能源生產與消費、促進可再生能源的整合，以及提高電網的穩定性和韌性。

目前市場趨勢顯示，DERMS 重點在提高電網效率及技術整合，特別是 AI 和機器學習技術的運用。雖然 DERMS 面臨許多挑戰，例如政策支持不足、資金獲取困難等，但仍有不少成功案例。市場中主要參與者包括西門子、NEC、施耐德電氣、伊頓、GE、ABB 和甲骨文等公司。這些公司在技術創新、業務策略和市場地位等面向持續推動 DERMS 的發展，並致力於創造更安全、高效的能源系統。

在日本，虛擬電廠（Virtual Power Plant, VPP）則自 2019 年起步以來取得可觀成果。東芝與東京電力的合作推動，使得虛擬電廠技術在橫濱市落地，並獲得電源開發公司（J-Power）的加入，有效促進可再生能源的整合。虛擬電廠利用智慧電網和蓄電系統，能夠即時調整電力的供需，當再生能源過量時儲存多餘電力，而在需求增加時釋放儲存的電力，不僅提高了系統的彈性，還降低了停電風險。如圖 5.15 所示，東芝以住宅虛擬電廠（VPP）平台實現先進的群組控制，將分布在多個家庭中的大量儲能電池進行整合，並支持多家廠商的家庭能源管理系統（Home Energy Management System, HEMS），使東芝能持續為電力零售供應商提供多樣化的選擇。



資料來源：TOSHIBA Global 網站

圖 5.15、東芝以 VPP 平台整合分布在多個家庭中的大量儲能電池

此外，日本政府透過虛擬電廠和蓄電池補助計畫來增強電力供需的靈活性。這些補助方案針對不同類型的蓄電池，包括災害應變型、NETWORK 型和頻率控制型，提升了電力管理的效率和調整能力。同時，透過需量反應機制，居民在需求過高時可減少耗電，從而獲得經濟回報。在推動可再生能源與提升電力系統韌性方面，日本的努力顯著且有具體成效。

工研院於 2024 年與日本東芝能源系統株式會社及台灣東芝電子零組件股份有限公司簽署合作協議，計劃共同開發和驗證虛擬電廠技術，為滿足不斷增長的電力需求和提高系統穩定性提供支持。工研院的「聚合多元資源」虛擬電廠驗證平台正透過資源整合與智慧預測技術進行能源調配，未來將擴大應用範圍。

綜合來看，這些措施展示了分散式能源管理的巨大潛力，促進可再生能源的利用，提高整體能源系統的靈活性和可持續性。在此過程中，算力的提升扮演著關鍵角色，支持數據處理及資源優化，並推動區塊鏈技術在分散式能源管理中的應用，最終實現能源的可持續利用，確保未來能源需求的穩定供應。

#### 四、小結

隨著數位時代的發展，算力已成為科技研發和日常生活的關鍵資源，卻也伴隨著對電力的高需求，成為環境保護的挑戰。為了平衡算力需求與能源消耗，綠色算力應運而生，旨在提高能源使用效率同時減少環境影響。根據國際能源署（IEA）的預測，資料中心、人工智能及虛擬貨幣的電力需求將在未來幾年翻倍，進一步引發電力短缺的問題。全球各國政府及企業正積極推動綠色算力的發展，採用高效硬體設計、改進冷卻技術及推廣再生能源，以提升能源效率及降低碳排放。同時，各方也在研發新型計算架構，如異質計算架構，以提高計算效率和能耗比，並促進再生能源的廣泛應用，降低資料中心的碳排放。為推動綠色算力的全球發展，各國也正積極合作，制定相關的國際標準。

在此背景下，我國政府也正透過再生能源發展條例及能源轉型行動計畫等政策，推動綠色算力的發展，促進環境永續和技術創新，並提升國際競爭力。臺灣在半導體和電腦製造方面的優勢使其能夠發展綠色算力產業，包括研發高性能低功耗的晶片和高效冷卻系統，並提供綠色資料中心解決方案。綠色算力的發展需整合考量基礎設施、算力設備、平台與應用，建立高效能、低污染和可持續的競爭力。資料中心的能耗主要來自計算設備和冷卻系統，為此施耐德電機提出的白皮書詳細探討了資料中心的環境可持續性指標，協助企業評估能源使用的效率及可再生能源的比例。

回顧綠色算力的發展歷程，不難發現，透過晶片、系統、網路和軟體的優化，可提升能源使用效率並減少環境影響。伺服器虛擬化和動態電壓調節等技術已被廣泛應用，而液冷和自然通風等冷卻系統的推廣也有效降低了能耗。許多先進國家均在推動資料中心的能源轉型和再生能源運用方面表現活躍，舉例來說，芬蘭

的 CSC 資料中心擁有全球第十二台綠色超級電腦 LUMI，完全依賴可再生水力發電並回收廢熱供應當地暖氣，降低碳排放。同樣，挪威的 Green Mountain 資料中心則利用自然冷卻技術並重視人才發展，達成幾乎零碳的運行。

台灣在追求綠色能源開發方面亦已展現出顯著進展，近年來的法規促進了再生能源的普及。尤其是太陽能與風力發電的快速增長反映了台灣政府的政策支持。展望未來，台灣應透過新興科技調整基載電力比例，完善智慧電網與能源管理系統，研究多元冷卻技術，並推動儲能產業，最終實現能源轉型和永續發展的目標。

在永續發展的背景下，計算可持續性（Computational Sustainability）成為一個新興研究領域，旨在通過數學和計算科學平衡社會、經濟和環境資源，並且探討 ICT 和 AI 如何推動工業化和基礎建設。儘管 ICT 和 AI 對經濟增長和社會福祉有積極影響，仍需注意其可能帶來的負面影響。因此，為確保科技支持永續發展目標，需加強監管與合作，提高公民參與及推動永續發展教育，以實現更廣泛的社會認同。

城鄉數據智慧治理作為計算可持續性的一個典型案例，同時也是綠色算力的重要應用，旨在建立數據整合與共享的平台，利用 AI、IoT 和資料倉儲提升管理效率，實現即時監測和基礎設施管理，從而改善公共政策與促進永續發展。根據 IMD 2024 年智慧城市指數，蘇黎世、奧斯陸和新加坡名列前茅，台北則首次進入前 20 名，顯示數位轉型對城市發展的重要性。瑞士利用 LoRaWAN 技術改善城市功能，新加坡推動全面數位化，而台北則透過大數據中心支持城市治理，台東則提升公共服務和市民參與。算力在數據治理中扮演關鍵角色，確保即時分析與高效決策，並結合數據與理論達成有意義的結論

此外，隨著全球對可再生能源的重視，社區能源迅速成為促進公正和彈性能源系統的力量，通常由當地居民設立並運營。分散式能源資源管理系統(DERMS)結合硬體和軟體來即時管理資源以平衡供需，並預測德國的市場將大幅成長。另外，在日本，虛擬電廠（VPP）運用智慧電網調整電力供需，並透過政府補助增強管理效率。這些措施展示了分散式能源管理的潛力，提升能源系統的靈活性和可持續性，算力在數據處理和資源優化中扮演關鍵角色，支持未來能源需求的穩定供應。

總結來說，促進綠色算力的發展，積極推動系統化技術創新，以及在文化的永續發展中善用 AI 和 ICT，將是各國達成永續發展目標的關鍵驅動力。而透過永續發展教育（ESD）來啟發生命智慧，則可為實現永續發展提供根本解決之道。

參考文獻

1. 工研院 (2024), 工研院攜手日本東芝 以虛擬電廠打造臺灣電網韌性。  
[https://www.itri.org.tw/ListStyle.aspx?DisplayStyle=01\\_content&SiteID=1&MmmlID=1036276263153520257&MGID=113071712583137739](https://www.itri.org.tw/ListStyle.aspx?DisplayStyle=01_content&SiteID=1&MmmlID=1036276263153520257&MGID=113071712583137739)
2. 台達基金會 (2024), 崛起中的吃電巨獸, 資料中心、ai 及虛擬貨幣用電攀升。  
<https://brandnews.deltaww.com/Home/GreenLifeDetail/12522>
3. 再生能源資訊網 (2019), 日本虛擬電廠市場漸成形 海外企業加入競逐。  
<https://www.re.org.tw/news/more.aspx?cid=199&id=2819>
4. 邱祐慶 (2024), Ai 發展最怕什麼? 祖克柏、馬斯克都直指是「缺電」。  
<https://www.gvm.com.tw/article/112164>
5. 星島 (2024), ChatGPT 日耗電量超過美國家庭 1.7 萬倍 黃仁勳: AI 盡頭是光伏和儲能。  
<https://std.stheadline.com/realtime/article/1986230>
6. 陳政伶 (2024), 生成式 AI 浪潮興起: AI 有多耗電、耗水?。  
<https://www.deltafoundation.org.tw/blogdetail/8476>
7. 黃昭勇 (2024), AI 耗電也耗水! 微軟、Google、Meta 紛設定 2030 水資源正效益目標, 創造環境投資契機。  
<https://csr.cw.com.tw/article/43752>
8. 經濟部 (2023), 前瞻基礎建設計畫—綠能建設。  
<https://www.ey.gov.tw/Page/5A8A0CB5B41DA11E/604ea908-2610-41e0-a8a0-2dfe7230708e>
9. 經濟部 (2024), 112 年度全國電力資源供需報告。
10. 臺北大數據中心 (2024), 城市的數據掌握。  
<https://tuic.gov.taipei/zh/about>
11. 臺東縣政府 (2022), 臺東 TTPush 非同凡響, 數位治理再下一城! 榮獲 2023 智慧城市創新應用獎!。  
[https://www.taitung.gov.tw/News\\_Content.aspx?n=13370&s=120825](https://www.taitung.gov.tw/News_Content.aspx?n=13370&s=120825)
12. 環境資訊中心 (2021), 虛擬電廠實驗計劃? 日本政府三種蓄電池補助解析。  
<https://e-info.org.tw/node/231369>
13. Adam, I. O. & Alhassan, D. (2021). Examining the link between information and communication technologies and the UN sustainable development goals, *International Journal of Technology Management & Sustainable Development*, 20(1), 21–41.
14. Batty, M. (2013). Big data, smart cities and city planning, *Dialogues in Human Geography*, 3(3), 274–279.
15. Calero, C. & Piattini, M. (2017). Puzzling out software sustainability, *Sustainable Computing: Informatics and Systems*, 16, 117–124.

16. Carroll, A. B. (1991). The pyramid of corporate social responsibility: Toward the moral management of organizational stakeholders, *Business Horizons*, 34(4), 39–48.
17. Chatterjee, D. & Rao, S. (2021). Computational sustainability: A socio-technical perspective, *ACM Computing Surveys*, 53(5), 1–29.
18. Clark, W. C. & Dickson, N. M. (2003). Sustainability science: The emerging research program. *Proceedings of the National Academy of Sciences*, 100(14), 8059–8061.
19. GONGRESS.GOV (2022). H.R.4346 - chips and science act, 117th Congress (2021-2022). Retrieved September 2024, from <https://www.congress.gov/bill/117th-congress/house-bill/4346>
20. Elkington, J. (1998). *Cannibals with forks: The triple bottom line of 21st century business* New Society Publishers .
21. EU (2023). Regulation (EU) 2023/1781 of the european parliament and of the council of 13 september 2023 establishing a framework of measures for strengthening europe’s semiconductor ecosystem and amending regulation (EU) 2021/694 (chips act) (text with EEA relevance).
22. Gomes, C. P. (2011). Computational sustainability. In *Proceedings of the International Symposium on Intelligent Data Analysis (IDA’11)*, 8.
23. Gomes, C. P., Dietterich, T., Barrett, C., Conrad, J., Dilkina, B., & Ermon, S. (2019). Computational sustainability: Computing for a better world and a sustainable future, 62(9), 56–65.
24. Green Mountain (2024). Green Mountain’s Data Center Sites among the Greenest Worldwide. Retrieved September 2024, from <https://www.azrieliigroup.com/properties/green-mountain-data-centers/>
25. gridX (2024). Distributed energy resource management system (DERMS). Retrieved September 2024, from <https://www.gridx.ai/knowledge/distributed-energy-resource-management-system-derms>
26. Gupta, S. & Degbelo, A. (2022). An empirical analysis of AI contributions to sustainable cities (SDG11), in Mazzi, F. and Floridi, L. (eds) *The Ethics of Artificial Intelligence for the Sustainable Development Goals*.
27. Hansen, M. T., Norberg-Bohm, V., & Bromley, D. W. (2002). Environmental innovation and systemic change: The role of policy, *Technology Analysis & Strategic Management*, 14(4), 387–404.
28. Hollands, R. G. (2008). Will the real smart city please stand up? *City*, 12(3), 303–320.
29. ICS (2009). Institute for Computational Sustainability. Retrieved June 2024, from

- <https://computational-sustainability.cis.cornell.edu/index.php>
30. IEA (2024). Electricity, 2024 Analysis and forecast to 2026, IPCC, 2014, The IPCC's fifth assessment report (AR5). Retrieved June 2024, from <https://www.ipcc.ch/assessment-report/ar5/>
  31. IMD (2024a). IMD Smart City Index 2024, IMD World Competitiveness Center.
  32. IMD (2024b). 2024 Rankings out of 67 countries, IMD World Competitiveness Center. Retrieved November 2024, from <https://www.imd.org/centers/wcc/world-competitiveness-center/>
  33. IPCC (2014). The IPCC's Fifth Assessment Report (AR5). Retrieved June 2024, from <https://www.ipcc.ch/assessment-report/ar5/>
  34. ITU (2021). Digital technologies to achieve the UN SDGs. Retrieved June 2024, from <https://www.itu.int/en/mediacentre/backgrounders/Pages/icts-to-achieve-the-united-nations-sustainable-development-goals.aspx>
  35. IUCN & WWF (1980). World Conservation Strategy: Living Resource Conservation for Sustainable Development. Gland, Switzerland: IUCN. Retrieved September 2024, from <https://portals.iucn.org/library/efiles/documents/wcs-004.pdf>
  36. Jungwirth, D. & Haluza, D. (2023). Artificial intelligence and the sustainable development goals: An exploratory study in the context of the society domain, *Journal of Software Engineering and Applications*, 16, 91–112.
  37. Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage Publications.
  38. Kolbe, L. M. & Zarnekow, R. (2009). Studie: Nachhaltigkeit und Green IT in IT-Organisationen. Retrieved November 2024, from [https://www.static.tu.berlin/fileadmin/www/10002432/Dokumente/Forschung/Green\\_IT/Studie\\_Nachhaltigkeit\\_und\\_Green\\_IT\\_TUB-UG\\_2009.pdf](https://www.static.tu.berlin/fileadmin/www/10002432/Dokumente/Forschung/Green_IT/Studie_Nachhaltigkeit_und_Green_IT_TUB-UG_2009.pdf)
  39. Lin, P., Bungler, R., & Avelar, V. (2023). Guide to environmental sustainability metrics for data centers. Schneider Electric – Energy Management Research Center White Paper 67 Version 2.
  40. LUMI (2024). The European High-Performance Computing Joint Undertaking (EuroHPC JU). Retrieved September 2024, from <https://www.lumi-supercomputer.eu/about-lumi/>
  41. Market Research Network (2024). Germany distributed energy resources management system market by application. Retrieved September 2024, from <https://www.linkedin.com/pulse/germany-distributed-energy-resources-management-id9of>
  42. Matuszak, J. (2023). Swiss cities pioneering smart tech. Retrieved September

- 2024, from <https://knowhow.distrelec.com/internet-of-things/swiss-cities-pioneering-smart-tech/>
43. Meadowcroft, J. (2007). Who is in charge here? governance for sustainable development in a complex world. *Journal of Environmental Policy & Planning*, 9(3-4), 299–314.
  44. Nahar, S. (2024). Modeling the effects of artificial intelligence (AI)-based innovation on sustainable development goals (SDGs): Applying a system dynamics perspective in a cross-country setting. *Technological Forecasting & Social Change* 201, 123203.
  45. NSTC (2024). Critical and Emerging Technologies List Update. A Report by the Fast Track Action Subcommittee on Critical and Emerging Technologies of the National Science and Technology Council.
  46. Purvis, B., Mao, Y., & Robinson, D. (2019). Three pillars of sustainability: In search of conceptual origins. *Sustainability Science*, 14(3), 681–695.
  47. Ray, P. H. & Anderson, S. R. (2000). *The cultural creatives : How 50 million people are changing the world*. New York: Three Rivers Press.
  48. Rosanis, D. (2020). *Extreme cities*, Director, Video/DVD.
  49. Rothe, F.F., Audenhovea, L.V., & Loisena, J. (2022)., ICT for development and the novel principles of the Sustainable Development Goals. *Third World Quarterly*, 43(6), 1495–1514.
  50. Sharma, S., Hsu, C., & Feng, W. (2006). Making a case for a Green500 list. *Proceedings 20th IEEE International Parallel & Distributed Processing Symposium*, 8.
  51. Paul, S. G., Saha, A., Arefin, M. S., Bhuiyan, T., Biswas, A. A., Reza, A. W. (2023). A Comprehensive Review of Green Computing: Past, Present, and Future Research. *IEEE Access*, 11, 87445-87494.
  52. Smart Nation Singapore (2023). A Smart Nation that we can trust. Retrieved November 2024, from <https://www.smartnation.gov.sg/>
  53. Snaith, M. (2024). Power to the People: An Overview of Community Energy. September 2024, from <https://earth.org/power-to-the-people-an-overview-of-community-energy/>
  54. Theodorou<sup>1</sup>, A., Nieves<sup>1</sup>, J. C., & Dignum<sup>1</sup>, V. (2022). Good AI for Good: How AI Strategies of the Nordic Countries Address the Sustainable Development Goals. *AIofAI 2022: 2nd Workshop on Adverse Impacts and Collateral Effects of Artificial Intelligence Technologies*.
  55. TOSHIBA Global (2024). Residential VPP Platform. Retrieved November 2024, from <https://www.spinex-marketplace.toshiba/en/services/residential-vpp>

56. UN (1987). Report of the World Commission on Environment and Development: Our Common Future. Retrieved June 2024, from <http://www.un-documents.net/our-common-future.pdf>
57. UN (2015). Transforming our world: the 2030 Agenda for Sustainable Development. Retrieved September 2024, from <https://sdgs.un.org/2030agenda>
58. UN (2023). The Sustainable Development Goals Report 2023: Special edition. Retrieved September 2024, from <https://unstats.un.org/sdgs/report/2023/>
59. UN (2024). The Sustainable Development Goals Report 2024. Retrieved November 2024, from <https://unstats.un.org/sdgs/report/2024/>
60. UNEP (2023). Emissions Gap Report 2023. Retrieved September 2024. from <https://www.unep.org/resources/emissions-gap-report-2023>
61. UNEP (2024). Emissions Gap Report 2024. Retrieved November 2024. from <https://www.unep.org/resources/emissions-gap-report-2024>
62. UNESCO (2024a). Culture and Sustainable Development. Retrieved September 2024, from <https://www.unesco.org/en/sustainable-development/culture>
63. UNESCO (2024b). What you need to know about education for sustainable development. Retrieved September 2024, from <https://www.unesco.org/en/sustainable-development/education/need-know?hub=72522>
64. Verified Market Reports (2024). Global distributed energy resources management system market by type (software, service), by application (commercial, residential), by geographic scope and forecast. Retrieved September 2024, from [https://www.verifiedmarketreports.com/product/distributed-energy-resources-management-system-market/?trk=article-ssr-frontend-pulse\\_little-text-block/](https://www.verifiedmarketreports.com/product/distributed-energy-resources-management-system-market/?trk=article-ssr-frontend-pulse_little-text-block/)
65. Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark M. and Nerini, F. F. (2020). The role of artificial intelligence in achieving the sustainable development goals. *Nature Communications*, 11(1), 233..



## 第六章 結論與建議

在全球科技競爭日益激烈的背景下，算力已成為國家實力的重要象徵。隨著人工智慧、物聯網、5G 等尖端技術的快速發展，對高效算力的需求持續攀升，成為推動科技創新與經濟成長的關鍵動力。而台灣憑藉其在半導體與 ICT 產業的領先地位，具備了強化算力的先天優勢。本議題針對算力的基本內涵、盤點我國主要算力資源與排名、算力生態系分析、算力與電力能耗需求與綠算力的趨勢以及座談會之各項結論與建議，以進一步提升算力應用並確保產業發展，本議題之相關結論與建議說明如后：

### 一、 結論

#### (一) 生成式 AI 興起與主權商機，帶動全球算力基礎建設，推動整體算力生態系，台廠迎來新契機

生成式 AI 的興起與各國強調「主權 AI」之商機，加速全球對高效能運算（HPC）需求與 AI 伺服器的擴張，推動整體算力生態系發展，也為台灣相關產業帶來重要機遇。在此產業鏈中，台灣廠商分布於上游、中游及下游，扮演不可或缺的角色。上游產業部份，包括 AI 專用晶片的設計與製造，台積電等領導企業以先進製程提供高效能、低功耗的晶片支撐 AI 運算需求。台灣的 IC 設計公司如聯發科及其他專注於 AI 晶片的企業，也在開發相關解決方案。而中游產業部份，涵蓋電子零組件製造，包括印刷電路板（PCB）、散熱模組、風扇、連接器等關鍵零組件，為 AI 伺服器提供穩定運作的基礎。台廠如欣興、台達電等在精密製造和散熱技術領域具有競爭優勢。此外，在下游產業部份，則專注於 AI 伺服器的組裝與整合。包括廣達、鴻海等大廠在全球伺服器市場占據領先地位，尤其在全球 AI 伺服器代工市場上，占比將達 9 成，提供全球量身客製化之 AI 運算解決方案，滿足各種企業與數據中心的需求。

總之，生成式 AI 的快速發展與「主權商機」，正推動台灣半導體及資通訊產業的再次爆發，帶動全產業鏈的升級與成長，強化台灣在全球科技產業中的關鍵地位。

#### (二) 政府持續推動 AI 政策，從產業化到 AI 基本法，打造 AI 健康生態圈，促進產業發展

隨著全球 AI 技術的蓬勃發展，各國紛紛投入資源推動 AI 產業化，搶占數位經濟的制高點。台灣政府為因應此趨勢，積極推動相關政策，例如《AI 行動計畫 2.0》、《AI 基本法》等等。其中，《AI 行動計畫 2.0》奠定了發展人工智慧的政策基石，聚焦於三大核心方向，首先為人才培育，即透過校園合作、專業訓練及創新實戰計畫，培養具備跨領域能力的 AI 專業人才，為產業發展提供充沛的人力資源。再者為技術深耕，即鼓勵學術研究機構與產業攜手合作，共同開發具市場

價值的前瞻性 AI 技術，強化台灣在全球供應鏈中的角色。最後則為國際影響力部份，期以參與國際標準制定及策略合作為目標，提升台灣在全球 AI 技術與應用生態圈中的影響力。另外，為因應 AI 技術的快速發展及應用需求，國科會於 2024 年 7 月 1 日預告《AI 基本法》，並送交行政院審查。此法案旨在建立完整的法規架構，包括技術倫理、數據隱私及責任分配等核心議題，為 AI 技術的安全落地與大規模應用提供法治保障。這不僅能降低技術應用帶來的風險，亦為未來 AI 與產業的深度結合奠定法治基礎。

此外，數發部與經濟部密切合作，並共同規劃的四大 AI 計畫包括《AI GO-AI 產業實戰應用人才淬煉計畫》、《AI Hub-AI 智慧應用服務發展環境推動計畫》、《AI 領航-AI 領航推動計畫》、《Fast AI-人工智慧速捷技術深耕計畫》，其重點為數位基礎建設的規劃與推動，以確保 AI 技術應用所需的基礎設施，如運算平台和數據整合環境，得以完善。此外，亦聚焦技術環境建設與技術深耕，並搭配人才培訓、技術商品化及市場推廣，全力打造健康的 AI 生態圈，促進產業升級與技術發展，達到「產業 AI 化與 AI 產業化」之數位經濟發展，實現數位經濟的長遠目標。

### (三) 系統性定義算力與類別，界定可盤算力資源，研析盤點結果與目標，策進未來算力規模

隨著人工智慧、大數據、雲端運算等技術的蓬勃發展，「算力」成為驅動科技創新的核心資源之一。透過科普方式，有系統性地介紹算力的定義、分類方式與關鍵影響因素，並透過現有資源的盤點結果，探討台灣未來算力規模的發展方向。

算力定義與單位部份，指的是計算設備執行數值運算的能力，通常以每秒浮點運算次數 (Floating Point Operations per Second, FLOPS) 來衡量，另考量算力數級規模較大，故表達時，通常會加上 Giga ( $10^9$ )、Tera ( $10^{12}$ ) 或 Peta ( $10^{15}$ )，即 GFLOPS、TFLOPS 或 PFLOPS。例如，每秒千兆次運算 (PFLOPS) 常見於人工智慧模型訓練、氣象模擬及精準醫療分析等高需求應用場景，其高效能為現代科技創新提供了重要基石。

在算力分類部份，為了全面理解算力的應用特性與技術分類，我們可以從使用場景、應用範圍與硬體型態三個層面進行闡述。從使用場景來看，算力可以分為雲端算力、邊緣算力與終端算力。雲端算力集中於大型數據中心，適合高效能遠程運算及 AI 大模型訓練；邊緣算力部署於設備端，適合低延遲即時運算需求及 AI 微調模型及客製化的模型訓練，如智慧交通管理；終端算力則存在於個人裝置內，用於基本數據處理及 AI 推理；應用範圍上，算力可區分為 HPC、AI 伺服器與通用伺服器。HPC 主要用於科學研究、高性能模擬與工程分析，AI 伺服器專為人工智慧模型訓練與推論設計，而通用伺服器則服務於多元應用場景的中等算力需求；最後，我們也用「硬體型態」包括 HPC、大型雲端服務算力、企業型伺服器算力、邊緣算力等四類，並結合可盤算力資源，分別政府機構 (含 Taipei-

1)、大型醫院及大專院校，來進行算力分類與說明。

統計至 2024 年 6 月份算力盤點資料，我們共盤點 204 個單位，考量全面性資訊之揭露原則，算力係以 Rpeak 即理論值為統計基礎，而本次盤點結果約 88PFLOPS，從資料顯示，其佔比最大是政府機構，約占 93%，顯示政府帶動國家科研及產業之重要性。若僅計算進入 TOP500 之排名，我國共有六台，包括 Taipei-1、氣象署 2 台及國網中心 3 台，大約 77PFLOPS，世界排名 17 名，亞洲排名第 4 名。相較全部算力來看，該六台算力約佔所盤點算力之 9 成，表示純以 TOP500 算力排行資料，具有一定可信度。若結合其它產業經濟成長數據與施政目標，亦可納入推估未來國家算力規模之參考資料。而影響算力的關鍵因素，亦系統性地闡述包括處理器數量、隨機存取記憶體、存儲速度與類型、網路連接及軟體與演算法優化等。

透過此次資源盤點與數據分析，台灣已初步建立具規模的算力基礎，並在全球算力競爭中占有一席之地。結合產業經濟成長數據與施政目標，未來可進一步擴大算力規模，支撐智慧國家與數位經濟的發展。此外，透過本次專報，讓一般民眾得以一窺算力科技之大門，同時也達到教育性之目的。

#### **(四) AI 基礎建設帶來新契機，算力出口與能源需求新挑戰，晶片迭代與散熱新趨勢，整廠輸出與廢熱新商機**

隨著 AI 基礎建設加速推進，台灣的國家高速網路中心與民間業者紛紛投入資料中心的建設，但也帶來全新挑戰。首先是算力輸出的國際市場之競爭壓力，如全球三大雲端服務商，以及來自中國大陸的低價雲服務提供者，對台灣算力出口構成激烈挑戰。同時，AI 伺服器的高耗能特性，使電力需求問題尤為嚴峻。在全球範圍內，AI 運算中心的興建蔚為潮流，但伴隨著巨大的能源消耗需求。為了應對這些挑戰，科技大廠如 NVIDIA 與 AMD 持續推進算力晶片技術迭代，以降低單位算力的功耗。例如，NVIDIA 執行長黃仁勳提到，透過其最新加速運算晶片 Blackwell，可將訓練 AI 大型語言模型的能耗從 1000GWh 大幅降低至 3GWh，減少 333 倍，充分展現未來晶片功耗改善與散熱技術進步的新趨勢。

在資料中心建設中，廢熱回收逐漸成為能源利用的新焦點。由於資料中心的運作會產生大量廢熱，其能量可被轉化用於區域供暖、溫室加熱或水產養殖，不僅有效減少能源浪費，也符合永續發展趨勢。然而，廢熱回收需要克服低溫熱能回收的技術門檻，以及與需求方距離所帶來的能量損耗問題。企業可通過高效熱交換器技術，並與用熱單位合作，形成完整的回收與再利用體系。更進一步，採取整廠輸出的模式，即結合資料中心之土木設計與廠房建置、以及廠房內之伺服器與整體之散熱系統等，不僅能實現廢熱回收的經濟效益，也為降低資料中心營運成本開創新商機，促進產業升級與環境永續的雙贏發展。

## 二、建議

### (一) 提出 AI 微調模型之方法與步驟，結合算力+模型+軟體解決方案，降低「燒算力」問題，協助百工百業導入生成式 AI

生成式 AI 的應用正在全球掀起一股浪潮，台灣的中小企業也看到了其潛力，躍躍欲試，但仍面臨諸多挑戰。這些企業對於生成式 AI 的算力需求、算力晶片的選擇與應用方式多半缺乏了解，只知其強大卻不知如何入手。同時，全球算力晶片供不應求，更使高昂的算力成本成為一大阻礙。此外，每次 AI 生成結果的品質不穩定，也加劇了實驗性操作帶來的算力不確定性，推高了實施成本，俗稱「燒算力」。因此，建議政府提供相關資源與專業指引，協助各行各業，特別是中小企業，進行生成式 AI 的導入規劃。這可以包括針對特定需求進行微調模型 (Fine-tuning) 設計，或採用檢索增強生成 (Retrieval Augmented Generation, RAG) 方法，幫助企業更精準地滿足實際應用需求。

此外，面對生成式 AI 應用門檻高的問題，僅靠方法與步驟的指引仍不足以全面降低障礙，建議應進一步提供整合性的算力、模型與應用軟體三位一體解決方案。例如，通過算力資源共享平台，為中小企業提供可負擔的雲端算力；同時引進模組化算法工具，讓企業能快速部署 AI 模型；並整合適用於各行業的應用軟體，將技術門檻降至最低。透過這種方式，不僅能深化生成式 AI 在百工百業中的應用，更能促進技術普及與產業升級，打造台灣在 AI 時代的競爭新優勢。

### (二) 研擬建置高效能運算之專法，計畫性投入 AI 高運算資料中心資源，強化 AI 算力基礎建設，維持國家競爭力

算力產業的發展速度驚人，但也帶來了硬體快速迭代更新與高折舊成本的挑戰，易言之，算力老化問題。隨著科技迅猛進步，許多高效能硬體在短短幾年內可能被淘汰，企業因此承受了巨大的財務壓力。這不僅阻礙了 AI 高運算資料中心的建設步伐，也削弱了整體產業的競爭力。

為應對這一問題，政府有必要研擬專法，系統性投入 AI 算力基礎建設，通過財政補貼、稅務減免及產業聯盟等誘因，幫助企業降低硬體升級的成本負擔，促進算力產業的可持續發展。例如，南韓於 2020 年制定的《國家超高效能電腦應用及育成法》便是一個成功範例，該法要求中央行政機關每年依照基本計畫推動高效能運算 (HPC) 的發展，確保國家算力基礎建設穩定成長，從而提升國家在科技競賽中的競爭力。台灣可以借鑑此經驗，透過專法推動算力產業升級，搭建穩健的 AI 基礎建設，支撐 AI 相關產業的蓬勃發展，並為經濟增長提供強大動力。

此外，從算力規模的角度來看，大語言模型的訓練高度依賴高成本的大規模算力支持，通常需要機櫃型資料中心。然而，台灣目前的主要算力需求集中於中小企業應用所需之模型微調 (Fine-tuning) 與推論 (Inference)。這類應用對算力

的需求相對較低，成本也較為低廉，因此現階段國內的算力資源尚能應付。然而，隨著技術的不斷進步以及應用範疇的持續擴展，特別是在多模態應用中整合文字、語音與影像進行推論與生成的場景，未來的算力需求勢必大幅增加，這將成為推動算力成長的關鍵趨勢。正如當年高速公路建設初期，因規模龐大而一度被質疑是否有必要，但正因高速公路的興建，不僅帶動經濟起飛，更在經濟成長後滿足激增的交通需求。如今，交通需求的持續增長更進一步催生增建高速公路或拓寬路幅的必要。同樣地，算力的基礎建設也需超前部署，方能支撐未來數位轉型的需求，進一步提升國家競爭力。

### **(三) 估算政府電力成長，納入 AI/資料中心電力需求，提出綠算力計畫，兼顧數位轉型與永續發展**

隨著人工智慧（AI）產業快速發展，算力需求和資料中心建置量持續提升，電力供應成為不可忽視的關鍵議題。根據國際能源署（IEA）的估算，僅 2023 年輝達（NVIDIA）出貨的 AI 伺服器約 10 萬台，年耗電量已達 73 億度。考量 AI 產業的指數型成長趨勢，預期至 2026 年，其用電需求至少將達到 2023 年的 10 倍，這顯示 AI 技術與基礎設施發展對能源供應的壓力將愈加沉重，如何因應電力需求成為迫切課題。

依據國家發展委員會 2022 年所提供數據，我國每年用電需求平均成長 2.6% ±0.1%，其電力規劃必須充分納入資料中心的建置需求，方能支撐 AI 產業的長遠發展。同時，綠色能源應融入資料中心的供電方案，成為因應能源挑戰與永續發展的核心策略。透過推動綠能應用與創新能源技術，我國可在滿足 AI 產業需求的同時，實現數位轉型與永續經營的雙重目標。

### **(四) 提升 TAIDE 主權 AI 計畫，建立符合國人特有文化用字及用語之多模態模型，以拓展未來百工百業之應用需求**

目前，TAIDE 模型作為我國主權 AI 發展的重要基石，是依據國人特有文化用字構建的大語言模型，不僅展現了台灣在人工智慧領域的技術實力，也具備極大的產業應用潛力。該模型已公開釋放，供國內各界使用，並成功延伸至多個亮點應用案例，例如台南大學開發的「台語對話生成式 AI 機器人」，以及陽明交通大學結合「台客語 TAIDE 大語言模型」與語音 AI 介面的技術開創。未來，期待持續推動更具創新性和多模態特徵的主權 AI 大語言模型（如混合中、英、台、客、原住民用字用語），滿足國人多元語言與文化需求，拓展更多應用可能。

台灣的經濟結構以中小企業為主，占全體企業 98% 以上，服務業則吸納近六成的就業人口。因此，生成式 AI 在客服系統的應用將是第一波突破口，特別是在提升服務效率和客製化溝通上展現價值。建議政府持續投入資源，推進結合用字、用語與生成圖片的多模態大語言模型研發，並進一步強化符合國人語言文化特性的技術應用。這不僅有助於中小企業的數位轉型，更能全面提升台灣的 AI 競

爭力與產業創新能力。

**(五) 全面提升國家公開資料之品質，以擴展算力應用與發展**

人工智慧的廣泛應用，不僅依賴算力基礎，更有賴於高品質的大數據支持。數據的規模固然重要，但唯有提升資料品質，才能有效發揮生成式 AI 的應用潛力，開發出優質的 AI 應用軟體。高品質數據的整備是算力價值發揮的核心要素，透過公開資料的優化，不僅能促進技術創新，亦能提升 AI 應用於各產業的效能與精準度。因此，建議全面提升公開資料品質，並持續加強強化產官學研的合作，促進技術整合與創新，為算力應用的發展提供更堅實的基礎，推動台灣 AI 產業向前邁進。

## 附錄

### 一、綜合座談會-會議紀錄

中技社「我國主要算力盤點與產業生態系之探討」議題  
 擴大算力應用面，提升算力規模與生態系發展  
 會議紀要

- 一、時間：113年8月12日（星期一）下午14:00~16:30
- 二、地點：中技社8樓大會議室（台北市敦化南路二段97號8樓）  
 視訊連結如下：<https://meet.google.com/wxj-cyoh-rhx>
- 三、主席：
- |      |            |
|------|------------|
| 中技社  | 陳綠蔚執行長     |
| 台東大學 | 鄭憲宗校長(召集人) |
| 台智雲  | 吳漢章總經理     |
- 四、出席人員：
- |          |           |
|----------|-----------|
| 數位產業署    | 呂正華署長     |
| 國科會科技辦公室 | 涂家瑋副研究員   |
| 台智雲      | 李立國策略長    |
| 友歲超算     | 陳武宏總經理    |
| 超微(AMD)  | 林建誠資深副總經理 |
| 超微(AMD)  | 黃偉喬資深協理   |
| AI大聯盟    | 董偉豪祕書長    |
| 中鼎工程     | 陳振欽委員會執行長 |
| 中鼎工程     | 王範經理      |
| 台灣大哥大    | 王寶慶副總經理   |
| 意藍資訊     | 楊立偉總經理    |
| 滿拓科技     | 吳昕益執行長    |
- 五、列席人員：
- |             |         |
|-------------|---------|
| 正修科大電子系     | 張法憲教授   |
| 義守大學管理學院    | 吳有龍院長   |
| 義守大學智慧科技學院  | 陳啟彰副院長  |
| 國網中心        | 蕭志楳副主任  |
| 國網中心高速計算應用組 | 楊嘉麗正工程師 |

資策會產研所(MIC)

魏傳虔組長

陳奕伶產業分析師

陳牧風產業分析師

台東大學綠資系

鍾裕峯助理教授

中技社經產中心

楊顯整主任

中技社科工中心

芮嘉瑋代主任

陳世海研究員

黃鎧輝研究員

郭華軒研究員

## 六、會議提報資料：

### 七、會議紀要

- (一) 隨著數位轉型的加速，算力已成為現代科技發展的基石。它不僅僅是一種資源，更是未來創新和發展的推動力。無論是人工智慧、物聯網，還是大數據分析，這些前瞻技術的發展，都依賴於強大的算力支撐。
- (二) 本次邀請數發部數位產業署呂署長專題演講，演講題目：「AI 政策驅動與發展」，以及邀請國科會科技辦公室官員代表共同參與討論。此外，討論提綱有二個，分別為「算力基礎建置與運營議題」及「算力應用擴展、評估利基與生態系議題」。其中，「算力基礎設施的建置和運營」至關重要，這不僅涉及到硬體設備的配置，更包括運營維護、能源管理以及可持續發展等多方面的挑戰。而「算力應用的擴展與生態系的發展」也是我們必須關注的重點，在技術快速變化的時代，我們需要不斷評估各種算力應用的利基，從而挖掘更多的潛在價值。更重要的是，建立一個良性的生態系，讓產官學研產業之間能夠形成有效的合作，共同推動技術應用的邊界。
- (三) 算力政策與指標(數發部/國科會/國網中心)
  1. 數發部服務範疇：隨著數位發展，擴及各行各業，範圍除工廠登記證外，服務業也包括在內。
  2. 政策形成與 AI 演進：是結合「人才」、「技術」與「產業」相關配合而成。而隨著生成式 AI 興起，其人的角色愈來愈少，而 AI 的角色將扮演更吃重的角色，即來將大量出現 AI 代理(AI Agents)。
  3. AI 計畫與法規
    - (1) AI 行動計畫 2.0：政府持續推動 AI 行動計畫 2.0，其五大任務，包括人才、技術深耕、運作環境、提升國際影響力及回應人文社會，以達成產業轉型升級、增進社會福祉與全球 AI 新銳之願景。此外，

在五大任務中，特別強調的是回應人文社會，解決國家社會面臨挑戰，如勞動缺乏、超高齡的社會、淨零碳排等重要課題。

(2)AI 四大計畫：AIGO 之人才培訓計畫、AIHub 之應用服務發展環境推動計畫、AI 領航計畫及 Fast A 人工智慧速技術深耕計畫等四個計畫，目前均持續推動 AI 之產業發展。

(3)AI 基本法：國科會已完成擬定 AI 基本法，進行預告中，未來將循程序送至行政院進行審查。

#### 4.AI 生態系與數位經濟

(1)結合政府相關部會(國科會、數發部與經濟部)共同打照 AI 園區，而數發部其輔導數位產業發展，關注之三大焦點:建構 AI 生態系、充實 AI 人才及產業 AI 化。另外，2026 年數位經濟之產值目標達到兆元以上。

(2)結合 AI 能量登錄機制：協助國內 AI 業者，以國際開源框架為基礎，開發 AI 工具且強化技術能量，並透過登錄機制，以逐步打造本土 AI 生態系。

5.AI 產業推動規劃、目標與核心策略：三大推動重點在於培育產業即戰人才、建構多元產業運用及開發技術及驗證。而其推動背後之四大策略，其具體為軟體的基盤化、應用的規模化、還有生態的商業化，以及合作的國際化。而上項所提及四大策略係基於 AI 行動計畫 2.0 加上產業 AI 化與 AI 產業化為核心概念所擬定。

6.數位產業基建計畫：數發部積極建立數位產業基建計畫，其重點在於軟體基盤共享平台，包括算力共享平台，還有程式碼共享平台，還有信數據信任的共享平台。

7.建構多元產應用之推動作法：即所謂的 Transparency (資料共享機制透明化)、Robustness (開發產業專用模型工具之穩健性)、Usability (產業 AI 解決方案之可用性)、Scalability (產業應用擴散之可擴展性)以及 Technology Integration (國際輸出通路及商機之技術整合)之 Trust 方案，可以做微調模型 (fine-tune)、組隊進行工具之開發、建構場域示範案例等等，以推動多元 AI 產業。

8.AI 應用與普及率指標：相關 AI 應用知名範例，如智寶國際、聯發科 DaVinci 及鑫蘊林科，均相當成功。而推動 AI 應用其普及率是重要指標，去年其普及率達 32.5%，短期目標為提升至 50%，長期目標為達到 70%。

9.核心技術與建立領域模型：GAI 之技術核心在於透過 no code/low code 概念建構 AI 平台與工具。數發部之目標為建立 10 個專屬領域模型，即透過 AI 賦能應用平台與工具，使線上賦能應用，而線

下韌性應用，可廣泛應用於製造業/數位經濟產業。

10. 算力平台計畫：數發部已投入資源，為 AI 算力提供示範建置模式。該計畫係免費開放資服業者申請，使用週數為 6 週。後續視申請情形滾動檢討建置。

11. 人才、技術、產業三位一體之政策：人才、技術、產業都是數位經濟產業發展之關鍵重要，而具體而言就是好的 AI 的生態系，建立平臺、工具與軟體，以及培養整合型人才、應用人才及開發型人才等三種人才，方能廣泛運用在百工百業之領域。其中，人才培育是重點，其具體方法包括整合企業之實戰培育、專業技能養成以及教師培育。此外，也透過工協會、國際大廠的連結及國內 SI 廠商，期望一年培養 6000 人次之 AI 人才。而打造之產業 AI 人才係透過帶案入場的方式，並聚焦在算力資源、應用平台、應用模組、教學案例，以及新創解方等五個面向之人才培育。

#### 12. AI 產業政策

(1) 就政府而言，我們必需克服電力的問題與投資效益。此外，政府投入特定產業之資源，其過程是需要協調的，同時，其中需要兼顧提升產業國產化及避免變成保護主義。此外，在行政院經濟發展委員會，設定三個主題，分別為創新經濟、均衡台灣與發展成長。其中，有關均衡台灣部份，主要是解決水電基礎設施產業的問題，以及建立智科技生態系，包括桃竹苗大矽谷推動方案與智慧科技大南方產業生態系計畫，以達到均衡台灣。此外，有關創新經濟部份，將協助 AI 產業化與產業 AI 化之議題，具體而言，主要係解決資料、算力及模型之問題。

(2) 主權產業：國家會持續推動 TAIDE 模型，完善臺灣 AI 基礎建設，並打造可信任 AI 對話引擎。

(3) 微調模型產業：有關科研、氣象、醫療，以及各企業所要導入之微調模型(Fine-tuning)之各類算力需求，國科會及相關政府均有考量，並提出各項施政計畫，據以執行。

13. 國家高速運算的預審法：國網中心因應「算力老化」，將提出「國家高速運算的預審法」，即透過長期規劃，制訂算力基礎設施之建置，以及相對應之人才培育。

#### 14. 國家算力之目標

(1) 國網中心規劃，並依「晶創計畫」，有計畫性建置算力中心，今年可達 16PFLOPS，明年達到 100 多 PFLOPS，進而於晶創計畫結束時，能達到 300PFLOPS，以達到亞洲第三。此外，亦將 30% 算力提供予產業界，以培植新創公司。

(2)算力議題要有多元指標：算力硬體建置不要只有一項指標，FLOPS，應該針對算力建立多元指標，創造出新的賽道。

15.政策支援：針對有能力自建算力之企業或針對 AI 生成式 AI 技術有所創新和突破的企業，政府應提供稅收減免、研發補助等激勵措施。

#### (四) 算力基礎建設與節能散熱趨勢

1.百工百業的算力應用之驅動：算力是國力，而隨著算力應用之擴大加深算力需求，已是不可回頭的趨勢。

2.晶片節能設計：考量算力耗能議題，故目前算力晶片，均採節能設計，以滿足在有限電力資源下，設置資料中心/數據中心。

3.液冷為未來散熱趨勢：目前資料中心之 PUE 是有效運用電力之指標，惟電力約有 30%是耗在算力之氣冷散熱機制，而隨著科技進步，其伺服器之厚度已達到 4U、5U，未來可能達到 10U，亦即透過液冷方式進行冷卻方式也是未來之散熱趨勢。

4.資料中心廢熱商機新思維：目前建置資料中心，除樓地板荷重問題外，散熱亦是重點。建議可結合石化產業經驗，有效運用機櫃「廢熱」，除避免成為嫌惡設施之環保議題，同時也是廢熱商機之思維。

5.綠算力之前瞻思維：資料中心所衍生出散熱議題確實是一個研究重點，尤其是資料中心能帶來之產業效益，卻也帶來環保議題，在本次議題研究上，研究團隊於年度算力議題專刊中，將針對綠算力，即結合算力與綠色能源的概念與實例，提供國內建置資料中心之前瞻思維與參考。

#### (五) 算力中心建置議題

1.目前政府機構或企業之各項資料都有保密與隱私需求，且有主權落地之強列需求，目前全球三大雲端服務業者(包括 AWS、Google 及 Microsoft)均屬於美國企業，並非主權落地之台廠，故友歲超算公司已積極建置全台最大算力提供之業者，以強化主權產業運營。

2.國內資料中心建置者之兩大挑戰：即電力需求(如台電已公開說明，資料中心建置為桃園以南)，即建置算力需要政府之電力支持，另一為賣算力(租算力給資服業者)，期整合國內產官學研力量，採集中式算力，形成商業模式，提高算力使用效益。其中算力建置與租賣算力確實是挑戰，雖然算力應用是百工百業都需要，但需要時間，故建算力的同時，是考驗口袋深度，也就是如何賣算力，撐到具經濟規模。

3.整廠輸出新概念：結合資料中心之廠房、機櫃散熱工程、樓地板荷

重、晶片製作、組裝及最後之生成式 AI 算力之應用端，其市場亦擴及國外，此具高附加價之「整廠輸出」提供資料中心廠房建置之新概念。

4. 算力老化：算力產業的硬體更新速度非常快，硬體很快就會被折舊和淘汰，造成高昂的折舊成本。而該議題確實是挑戰，主要係因科技變動太快，算力建置後產生之規格淘汰或折舊，且如果算力應用還沒到位，就產生巨大的成本壓力，故實需政府透過誘因，強化算力產業之發展。
5. 資料中心設置與產業平衡發展：資料中心所需之電力資源，是政府、地方及業者一起共同解決之「用電需求」與「建置電廠」議題，此外，也會考量我國南北產業之平衡發展。

#### (六) 我國算力主權與隱私安全

1. 算力與 AI 主權產業：算力應用之資料具主權概念，如金融、健保等公共資料，其資料具隱私性，建議政府除透過國網中心建置算力外，亦應協助國內業者建立屬於國內之資料中心，如可仿臨近日本扮演更積極角色，提供政策補助與提供土地資源，來支撐算力基礎建設，以擴大該產業之生態系規模，促進國內 AI 主權產業之發展。
2. 主權產業與在地優先：目前在美中科技對抗下，國內相關業者已考量資料中心所使用為台灣土地與電力，故兼顧主權 AI 產業意識，拒絕向大陸業者提供算力是正確選擇，更進一步地，應該優先提供給台灣當地需求業者才有道理，共榮共利共生，建立算力之生態系。此外，針對國內外投資並獲取部份國家資源支持之新資料或算力中心，可考慮優先開放給國家科學研究有重大需求之單位或計畫。
3. 客戶隱私與營業隱私：算力成本與客戶個資或營業隱私議題，是擴大算力應用很大的議題。
4. 安全地端與隱私需求：企業於地端運用時，首重本地端安全落地，以強化資安與隱私，避免資料外洩。
5. 台灣是數據大國還是數據移民地：資料要留在台灣，國家就需要建置相關算力，如將資料放在國外，就會有資料外流問題。
6. 安全建立並通過 AI 驗證規範：算力應用之發展，在地端部份，需要透過類似資訊之安全驗證，如 ISO27001，以強化算力產業安全，並輔導成長。此外，政府機關可針對重要施政、涉及民眾或安全的等不確定，由機關訂定沙盒之規範，輔以算力支持或補助，開放企業申請並投入研發。

### (七) 我國算力足夠與替代性

1. 國內算力應用：算力不足，是指大語言訓練的算力，主要為機櫃型之資料中心範疇，屬於大算力需求，成本很高；而就台灣國內應用面，主要為微調模型(Fine-tuning)及算力推論(Inference)範疇，屬於小算力，成本相較為低，故我國的算力應該足夠，建議強化產官學研共同合作，擴大算力應用。
2. 國內四大資料中心建置：依目前已揭露之新聞報導，目前包括鴻海、友歲超算、翔耀實業及 GMI Cloud，已依公告之訊息，初步盤點其算力總和，並納入研究議題報告。此外，上述公司於今、明兩年均配合未來發展趨勢，大量投資算力之建置，故算力應該是足夠的。
3. CPU 是可用算力：算力不是只有 GPU 可以執行，而 CPU 也是一種可用算力，為隨手節點(Hand as a Node)之算力，當實際模型較小，或參數較小時，亦可用合適之 CPU 代替算力，並省下昂貴 GPU 費用。
4. 算力效能：算力效能除了 CPU、GPU 等晶片外，設計上亦需搭配高傳輸與記憶體較大之記憶體，如高頻寬記憶體(High Bandwidth Memory, HBM) 記憶體容量也是重要的零組件，方能達到較高的算力效能。

### (八) 算力應用之評估與利基

1. 「算力+模型+軟體應用(包含資安)」之算力商機：現階段各行各業都想要運用，惟算力成本、模型與相關配套軟體撰寫之各種技術門檻，故可綁住層級方式，即整合算力、合適的演算法，以及軟體應用與資安防護，進行垂直整合，提供業者一種全方位之解決方案(Toal solution)，應是很好的算力商機。
2. 算力應用之利基評估面向：主要算力來源可概三個面向，分別為自建自用(如台積電、聯發科等)、三大公有雲(AWS、Google Cloud、Microsoft Azure)及國內資料中心建置(如友歲超算)。此外，其算力應用利基評估時，建議可朝技術合宜(Technology Fit)及市場合宜(Market Fit)兩個因子來進行探討。
3. 國內資料中心市場之侷限性問題：我們會依客戶的需求，提供合適的模型並搭配三大雲服務商為落地之主要考量，惟就搭配國內資料中心而言，將同時存在國內大語言模型之競爭力不足與市場之侷限性問題。此外，一般而言，我們會依客戶較喜歡之大語言模型(如 OpenAI)結合客戶端之資料，進行客製化微調，其所需要的算力是小算力。
4. AI 推論之正確性與反應時間：隨著算力應用之展開，所要關注為

AI 推論的正確性與反應時間。

6.生成式 AI 應用與效益問題：目前生成式 AI 應用部份，以政府、金融及電信業等服務業會比較有意願，可應用之程式如 AI 代理(AI Agent)，而製造業反而比較慢，其具體原因是沒有效益，即無法節省人力。

7.算力負毛利：意指於提供算力服務時，提供者的毛利率為負數，即算力服務的成本超過了收入。主要係『應用面』為算力規模與生態系發展之驅動力，惟當算力應用尚未普及時，考量整體產業及新興市場的發展，算力負毛利是業者經營的一種策略思維。另外，目前算力應用在啟步階段，而且是有未來式發展的空間，但不可能一直是負毛利，只是需要時間。

8.燒算力與算力之不確定性：算力執行時，往往相當耗時，成本很高，是一種燒算力，且由於算力之應用面尚未展開，故各行各業於算力投資時，其投資報酬率(ROI)具不確定性，會覺得算力很貴。此外，由於 AI 高度之不確定性，產業初期需有負毛利概念，故政府的角色是很重要的。例如政府在鼓勵資安投入時，就投入大量資源，有步驟地提供資源(如減少稅金)予資服業者。

(九) 建置算力之單一窗口：考量資料中心之投資，涉及多個政府單位（包括地方政府），建議政府應整合成單一窗口，由一個主管單位協助並解決業者在建置時之各項問題。此外，數發部成立，主要是因為台灣是硬體製造大國，而隨著 AI 世代興起，我們也要發展軟體，成為我國發展之雙引擎。故有關單一窗口部份，數位產業署可以協助相關業者處理建置之窒礙。

(十) 資料品質問題：資料對 AI 發展也是重點，除了要有大數據外，同時，也需要有高品質的數據，目前我國公開之大數據資料均具品質不佳的問題，此外，有時也難以獲得數據，需政府有相對應之窗口，提供一站式協助與解決。

(十一)座談會議內容原則：希望在座各位之業者，能暢所欲言，本社將各位之發言內容做好記錄，提供各位確認。此外，也會依以往做法，保護各位之發言。

(十二)本次座談收穫非常多，各位如有新的想法或建議，隨時都可以提出來，我們納入本議題報告。此外，企業除了要考量利潤外，永續經營也是很重要觀念，以達到共好共榮的目標。

八、散會：下午 16:30

二、 算力盤點格式範例

單位	機器名稱	機體性能				系統	功率	資料來源
		計算核心 (GPU/CPU)	算力理論值 (Rpeak) (TFLOPs)	網路連結	算力實際值 (Rmax) (TFLOPs)			
中央氣象署	PRIMEHPC FX1000, A64FX 48C	A64FX 48C 2.2GHz(184,320 core)	12,980	Tofu interconnect D	11,160	氣象用	673.56 kw	<a href="https://top500.org/system/180216/">https://top500.org/system/180216/</a>
...								

### 三、輝達 GPU 模組

#### (一) A100 datasheet



#### The Most Powerful Compute Platform for Every Workload

The NVIDIA A100 Tensor Core GPU delivers unprecedented acceleration—at every scale—to power the world’s highest-performing elastic data centers for AI, data analytics, and high-performance computing (HPC) applications. As the engine of the NVIDIA data center platform, A100 provides up to 20X higher performance over the prior NVIDIA Volta™ generation. A100 can efficiently scale up or be partitioned into seven isolated GPU instances with Multi-Instance GPU (MIG), providing a unified platform that enables elastic data centers to dynamically adjust to shifting workload demands.

NVIDIA A100 Tensor Core technology supports a broad range of math precisions, providing a single accelerator for every workload. The latest generation A100 80GB doubles GPU memory and debuts the world’s fastest memory bandwidth at 2 terabytes per second (TB/s), speeding time to solution for the largest models and most massive datasets.

A100 is part of the complete NVIDIA data center solution that incorporates building blocks across hardware, networking, software, libraries, and optimized AI models and applications from the NVIDIA NGC™ catalog. Representing the most powerful end-to-end AI and HPC platform for data centers, it allows researchers to deliver real-world results and deploy solutions into production at scale.

#### NVIDIA A100 TENSOR CORE GPU SPECIFICATIONS (SXM4 AND PCIe FORM FACTORS)

	A100 40GB PCIe	A100 80GB PCIe	A100 40GB SXM	A100 80GB SXM
FP64	9.7 TFLOPS			
FP64 Tensor Core	19.5 TFLOPS			
FP32	19.5 TFLOPS			
Tensor Float 32 (TF32)	156 TFLOPS   312 TFLOPS*			
BFLOAT16 Tensor Core	312 TFLOPS   624 TFLOPS*			
FP16 Tensor Core	312 TFLOPS   624 TFLOPS*			
INT8 Tensor Core	624 TOPS   1248 TOPS*			
GPU Memory	40GB HBM2	80GB HBM2e	40GB HBM2	80GB HBM2e
GPU Memory Bandwidth	1,555GB/s	1,935GB/s	1,555GB/s	2,039GB/s
Max Thermal Design Power (TDP)	250W	300W	400W	400W
Multi-Instance GPU	Up to 7 MIGs @ 5GB	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 5GB	Up to 7 MIGs @ 10GB
Form Factor	PCIe		SXM	
Interconnect	NVIDIA® NVLink® Bridge for 2 GPUs: 600GB/s** PCIe Gen4: 64GB/s		NVLink: 600GB/s PCIe Gen4: 64GB/s	
Server Options	Partner and NVIDIA-Certified Systems™ with 1-8 GPUs		NVIDIA HGX™ A100-Partner and NVIDIA-Certified Systems with 4, 8, or 16 GPUs NVIDIA DGX™ A100 with 8 GPUs	

\* With sparsity

\*\* SXM4 GPUs via HGX A100 server boards; PCIe GPUs via NVLink Bridge for up to two GPUs

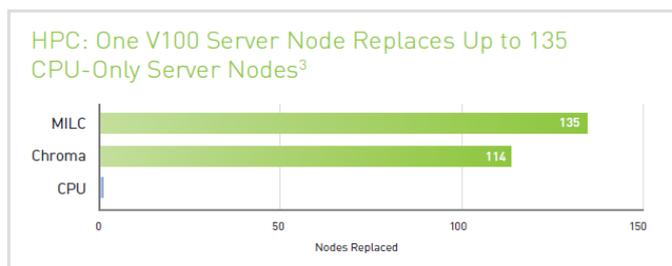
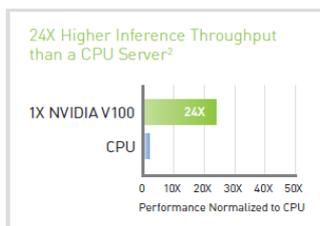
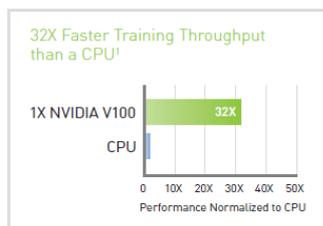
資料來源：<https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet.pdf>

(二) V100 datasheet



The World's Most Powerful GPU

The NVIDIA® V100 Tensor Core GPU is the world's most powerful accelerator for deep learning, machine learning, high-performance computing (HPC), and graphics. Powered by NVIDIA Volta™, a single V100 Tensor Core GPU offers the performance of nearly 32 CPUs—enabling researchers to tackle challenges that were once unsolvable. The V100 won MLPerf, the first industry-wide AI benchmark, validating itself as the world's most powerful, scalable, and versatile computing platform.



SPECIFICATIONS

	V100 PCIe	V100 SXM2	V100S PCIe
GPU Architecture	NVIDIA Volta		
NVIDIA Tensor Cores	640		
NVIDIA CUDA® Cores	5,120		
Double-Precision Performance	7 TFLOPS	7.8 TFLOPS	8.2 TFLOPS
Single-Precision Performance	14 TFLOPS	15.7 TFLOPS	16.4 TFLOPS
Tensor Performance	112 TFLOPS	125 TFLOPS	130 TFLOPS
GPU Memory	32 GB /16 GB HBM2		32 GB HBM2
Memory Bandwidth	900 GB/sec		1134 GB/sec
ECC	Yes		
Interconnect Bandwidth	32 GB/sec	300 GB/sec	32 GB/sec
System Interface	PCIe Gen3	NVIDIA NVLink™	PCIe Gen3
Form Factor	PCIe Full Height/Length	SXM2	PCIe Full Height/Length
Max Power Consumption	250 W	300 W	250 W
Thermal Solution	Passive		
Compute APIs	CUDA, DirectCompute, OpenCL™, OpenACC®		

資料來源：<https://images.nvidia.com/content/technologies/volta/pdf/tesla-volta-v100-datasheet-letter-fnl-web.pdf>

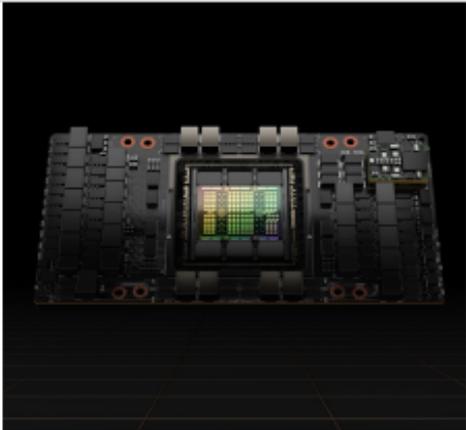
### (三) H100 datasheet

Datasheet



## NVIDIA H100 Tensor Core GPU

Exceptional performance, scalability, and security for every data center.



#### Technical Specifications

FP64	34 teraFLOPS	26 teraFLOPS	68 teraFLOPS
FP64 Tensor Core	67 teraFLOPS	51 teraFLOPS	134 teraFLOPS
FP32	67 teraFLOPS	51 teraFLOPS	134 teraFLOPS
TF32 Tensor Core	989 teraFLOPS <sup>2</sup>	756 teraFLOPS <sup>2</sup>	1,979 teraFLOPS <sup>2</sup>
BFLOAT16 Tensor Core	1,979 teraFLOPS <sup>2</sup>	1,513 teraFLOPS <sup>2</sup>	3,958 teraFLOPS <sup>2</sup>
FP16 Tensor Core	1,979 teraFLOPS <sup>2</sup>	1,513 teraFLOPS <sup>2</sup>	3,958 teraFLOPS <sup>2</sup>
FP8 Tensor Core	3,958 teraFLOPS <sup>2</sup>	3,026 teraFLOPS <sup>2</sup>	7,916 teraFLOPS <sup>2</sup>
INT8 Tensor Core	3,958 TOPS <sup>2</sup>	3,026 TOPS <sup>2</sup>	7,916 TOPS <sup>2</sup>
GPU memory	80GB	80GB	188GB
GPU memory bandwidth	3.35TB/s	2TB/s	7.8TB/s <sup>3</sup>
Decoders	7 NVDEC 7 JPEG	7 NVDEC 7 JPEG	14 NVDEC 14 JPEG
Max thermal design power (TDP)	Up to 700W (configurable)	300-350W (configurable)	2x 350-400W (configurable)
Multi-instance GPUs	Up to 7 MIGs @ 10GB each	Up to 7 MIGs @ 10GB each	Up to 14 MIGs @ 12GB each
Form factor	SXM	PCIe > dual-slot > air-cooled	2x PCIe > dual-slot > air-cooled
Interconnect	NVLink: > 900GB/s PCIe > Gen5: 128GB/s	NVLink: > 600GB/s PCIe > Gen5: 128GB/s	NVLink: > 600GB/s PCIe > Gen5: 128GB/s
Server options	NVIDIA HGX™ H100 partner and NVIDIA-Certified Systems™ with 4 or 8 GPUs  NVIDIA DGX™ H100 with 8 GPUs	Partner and NVIDIA-Certified Systems with 1-8 GPUs	Partner and NVIDIA-Certified Systems with 2-4 pairs

資料來源：<https://resources.nvidia.com/en-us-tensor-core/nvidia-tensor-core-gpu-datasheet>

## (四) GB200 datasheet

## Specifications

GB200 NVL72<sup>1</sup> Specs

	GB200 NVL72	GB200 Grace Blackwell Superchip
<b>Configuration</b>	36 Grace CPU : 72 Blackwell GPUs	1 Grace CPU : 2 Blackwell GPU
<b>FP4 Tensor Core<sup>2</sup></b>	1,440 PFLOPS	40 PFLOPS
<b>FP8/FP6 Tensor Core<sup>2</sup></b>	720 PFLOPS	20 PFLOPS
<b>INT8 Tensor Core<sup>2</sup></b>	720 POPS	20 POPS
<b>FP16/BF16 Tensor Core<sup>2</sup></b>	360 PFLOPS	10 PFLOPS
<b>TF32 Tensor Core</b>	180 PFLOPS	5 PFLOPS
<b>FP32</b>	6,480 TFLOPS	180 TFLOPS
<b>FP64</b>	3,240 TFLOPS	90 TFLOPS
<b>FP64 Tensor Core</b>	3,240 TFLOPS	90 TFLOPS
<b>GPU Memory   Bandwidth</b>	Up to 13.5 TB HBM3e   576 TB/s	Up to 384 GB HBM3e   16 TB/s
<b>NVLink Bandwidth</b>	130TB/s	3.6TB/s
<b>CPU Core Count</b>	2,592 Arm® Neoverse V2 cores	72 Arm Neoverse V2 cores
<b>CPU Memory   Bandwidth</b>	Up to 17 TB LPDDR5X   Up to 18.4 TB/s	Up to 480GB LPDDR5X   Up to 512 GB/s

資料來源：<https://www.nvidia.com/en-us/data-center/gb200-nvl72/>

四、專有名詞對照表

專有名詞	英文	中文
AWS	Amazon Web Service	亞馬遜雲端服務
API	Application Programming Interface	應用程式介面
AI	Artificial Intelligence	人工智慧
AR	Augmented Reality	擴增實境
ASP	Average Selling Price	平均售價
BMC	Baseboard Management Controller	基板遠端控制晶片
CUE	Carbon Usage Effectiveness	碳使用效率
CPU	Central Processing Unit	中央處理單元
CXL	Compute Express Link	計算快速連接架構
CSR	Corporate Social Responsibility	企業社會責任
DW	Data Warehouse	資料倉儲
DLC	Direct Liquid Cooling	直接液冷
DERMS	Distributed Energy Resource Management System	分散式能源資源管理系統
DRAM	Dynamic Random Access Memory	動態隨機存取記憶體
ESD	Education for Sustainable Development	永續發展教育
ERF	Energy Reuse Factor	能源重複利用因子
FP	Floating Point	浮點
FP32	Floating Point 32	單精度
FP64	Floating Point 64	雙精度
FLOPS	Floating-Point Operations Per Second	每秒浮點運算次數
Flop/s	Floating-Point Operations Per Second	每秒浮點運算次數
GPU	Graphics Processing Unit	圖形處理器
HDD	Hard Disk Drive	硬碟
Hash/s	Hash Per Second	每秒能進行的 hash 運算次數
HPC	High Performance Computing	高效能運算
IaaS	Infrastructure as a Service	基礎設施即服務
IC	Integrated Circuit	積體電路
ITU	International Telecommunication Union	國際電信聯盟
ITEUsr	IT Equipment Utilization for servers	伺服器利用率
LLM	Large Language Model	大型語言模型
LOHAS	Lifestyles of Health and Sustainability	健康永續的生活型態
MSP	Managed Service Provider	託管式服務供應商
MPP	Massively Parallel Processing	大規模平行處理
MLCC	Multi Layer Ceramic Capacitor	多層陶瓷電容器
NPU	Neural network Processing Unit	神經網路處理單元

專有名詞	英文	中文
PaaS	Platform as a Service	平台即服務
PSU	Power Supply Unit	電源供應器
PUE	Power Usage Effectiveness	電力使用效率
PCB	Printed Circuit Board	印刷電路板
RAM	Random Access Memory	隨機存取記憶體
REF	Renewable Energy Factor	可再生能源因子
SaaS	Soft as a Service	軟體即服務
SSD	Solid-State Drive	固態硬碟
Sol/s	Solution Per Second	每秒解算次數
SC	Supercomputing Conference	超級電腦大會
TPU	Tensor Processing Unit	張量處理單元
TDP	Thermal Design Power	熱功耗
TPS	Transactions Per Second	每秒交易數
TOPS	Trillions of Operations Per Second	每秒兆次運算
UNESCO	United Nations Educational, Scientific and Cultural Organization	聯合國教科文組織
VM	Virtual Machine	虛擬機器
VPP	Virtual Power Plant	虛擬電廠
VR	Virtual Reality	虛擬實境
WUE	Water Usage Effectiveness	水使用效率
xUE	x Usage Effectiveness	x 指標使用效率



## 國家圖書館出版品預行編目(CIP)資料

我國主要算力盤點與產業生態系之探討 /吳有龍、陳啟彰、蕭志  
 梲、楊嘉麗、魏傳虔、陳牧風、陳奕伶、謝明哲、鍾裕峯、陳世海  
 作. -- 初版. -- 臺北市：財團法人中技社，民 113.12  
 140 面；21×29.7 公分  
 ISBN 978-626-98882-3-8 (平裝)

1. 人工智慧；2. 雲端運算；3. 資料中心；4. 產業發展

312

113019430

**著作權聲明©財團法人中技社**

本出版品的著作權屬於財團法人中技社（或其授權人）所享有，您得依著作權法規定引用本出版品內容，或於教育或非營利目的之範圍內利用本出版品全部或部分內容，惟須註明出處、作者。財團法人中技社感謝您提供給我們任何以本出版品作為資料來源出版的相關出版品。

未取得財團法人中技社書面同意，禁止改作、使用或轉售本手冊於任何其他商業用途。

**免責聲明**

本出版品並不代表財團法人中技社之立場、觀點或政策，僅為智庫研究成果之發表。財團法人中技社並不擔保本出版品內容之正確性、完整性、及時性或其他任何具體效益，您同意如因本出版品內容而為任何決策，相關風險及責任由您自行承擔，並不對財團法人中技社為任何主張。



財團  
法人 **中技社**

**CTCI FOUNDATION**

106 台北市敦化南路2段97號8樓

Tel : 02-2704-9805~7 Fax : 02-2705-5044

<http://www.ctci.org.tw>



使用再生紙印製