

解析三菱綜合研究所《生成式 AI 普及對日本電力需求的影響》

葉詠綺¹

一、前言

隨著生成式 AI 的發展，各國擔憂資料中心和 ICT 產業的電力需求將大幅增加。如日本正在擬定的「第七次能源基本計畫」，亦視電力需求增長為關鍵議題。日本三菱綜合研究所(Mitsubishi Research Institute, MRI)於 2024 年 8 月發表《生成式 AI 普及對日本電力需求的影響》報告，根據生成式 AI 在 ICT 產業²和資料中心的應用，以「數據處理量」和「用電效率³」估算至 2040 年的電力需求。

MRI 預測到 2040 年，網路流量將比 2020 年爆發性成長約 348 倍，因此資料中心的計算量也會隨之增加。隨著處理的資料量增加，用電也將依比例增加。「數據處理量」增加會使電力需求增加，而「用電效率」提升能抑制電力使用量，未來 ICT 產業的用電量將取決於「數據處理量」和「用電效率」之間的平衡。

二、資料中心的數據處理量

資料中心數據處理量取決於生成式 AI 模型大小，OpenAI 的 GPT-4 和 Google 的 Gemini Ultra 等大型 AI 模型的參數數量估計超過 1 兆個。Meta 的 Llama3 的模型擁有 80 億和 700 億參數兩種版本，相較之下規模較小。MRI 將其分為以下 3 種 AI 模型應用情境，推估不同情境下的數據處理量。結果顯示 2040 年大型 AI 模型情境的數據處理量為適材適用情境的 14 倍、小型 AI 模型情境的 90 倍(圖

¹ 工業技術研究院/綠能與環境研究所/能源與淨零策略研究室 Email: itriB30189@itri.org.tw

² ICT 產業包括資料中心和網路。

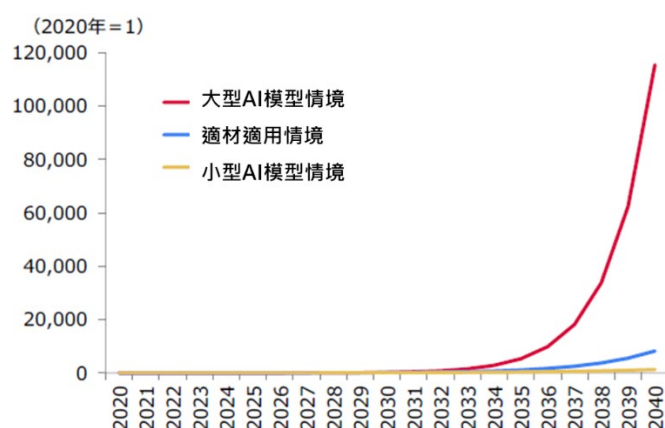
³ 單位用電量所能處理的資料量。

1)。

(1) **大型AI模型情境**：優先採用大型AI模型(假設未來會出現數兆至數十兆參數之模型)。

(2) **適材適用情境**：根據不同用途選擇合適的AI模型。

(3) **小型AI模型情境**：除了進行複雜處理外，皆優先採用小型AI模型(數百億參數之模型)。



資料來源：MRI (2024)

圖 1：各情境至 2040 年數據處理量

三、ICT 產業的用電效率

ICT 產業用電效率取決於半導體技術的進步，尤其是「先進封裝技術」、「光電融合技術⁴」和「AI 專用晶片」能夠顯著提升資料中心的用電效率，降低 AI 運行時的電力需求。MRI 將其分為以下 2 種半導體技術發展情境，推估不同情境下的電力效率變化。結果顯示在 2040 年達成先進封裝和光電融合技術的情境下，用電效率將比 2020 年提高約 600 倍(圖 2)，若再加上高效率 AI 專用晶片情境，

⁴ 「光電融合技術」指結合光學技術與電子技術，可以提高用電效率。

用電效率將進一步提升。

(1) 情境A：達成先進封裝技術。

(2) 情境B：情境A加上達成光電融合技術。

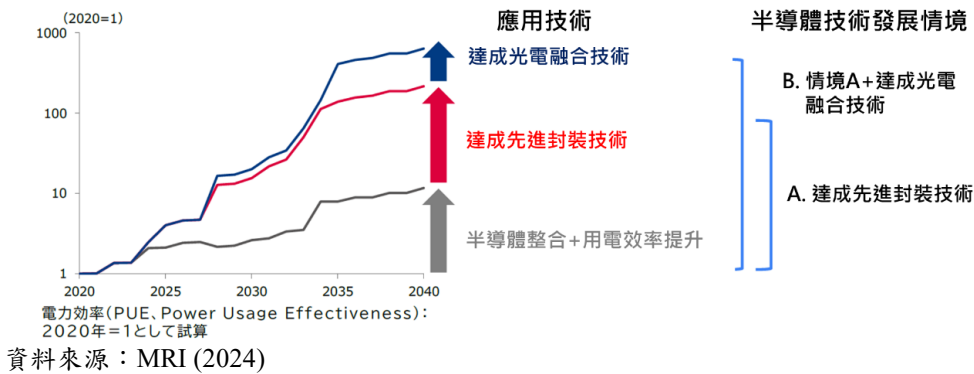


圖 2：半導體技術發展情境與資料中心用電效率預測(不包含 AI 專用晶片)

四、ICT 產業的用電量推估

根據數據處理量的 3 種情境、用電效率的 2 種情境，再搭配高、低效率 AI 專用晶片 2 種情境，一共 12 種情境估算 ICT 產業的用電量。

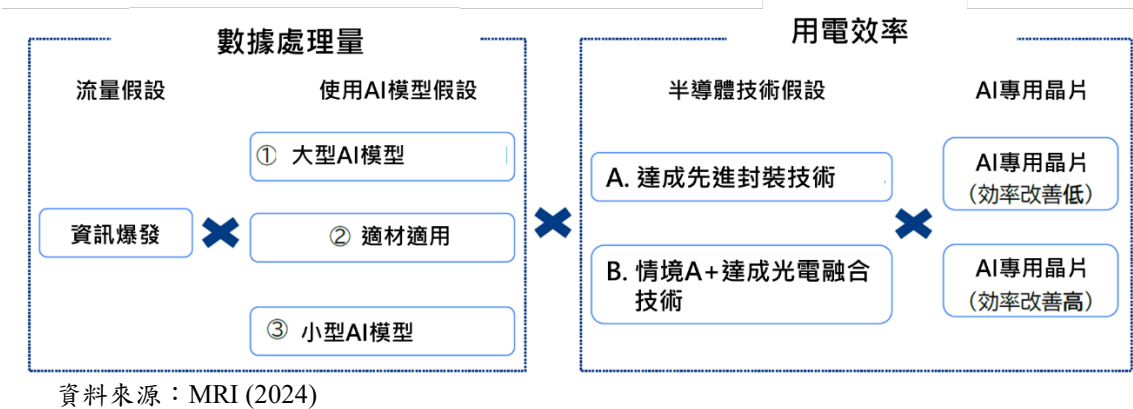
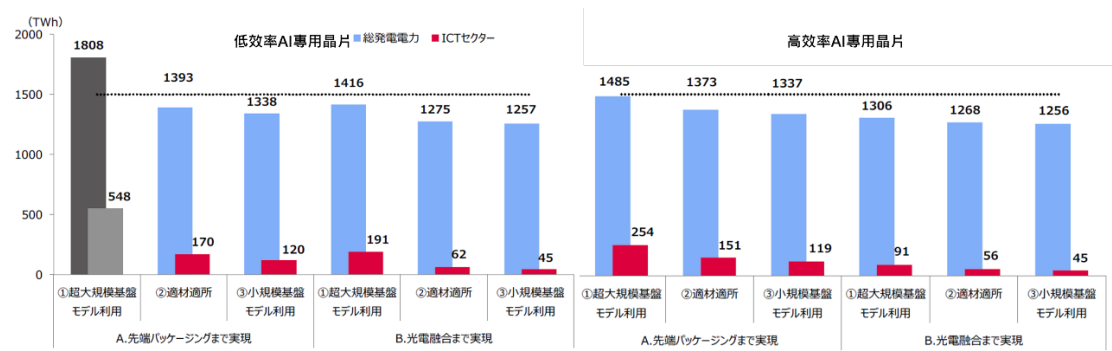


圖 3：12 種情境組合估算 ICT 產業的用電量

依據既有電力設備投資計畫，MRI 估計日本在 2040 年的最大供電能力可達 1,500 TWh(約 2020 年的 1.5 倍)。在 12 種情境中，僅[低效率 AI 專用晶片*大型 AI 模型情境*情境 A]的用電需求高達 1,808 TWh(圖 4)，最大供電能力 1,500 TWh

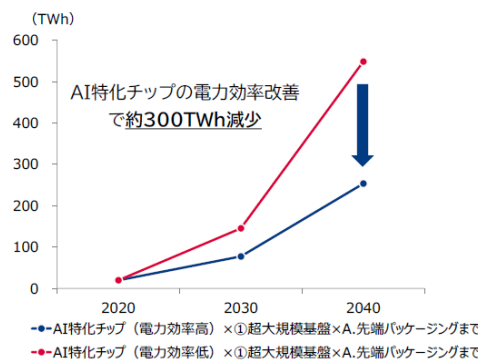
將無法滿足此情境的發展需求。



資料來源：MRI (2024)

圖 4：預計發電量與 12 種情境之 ICT 產業的用電量

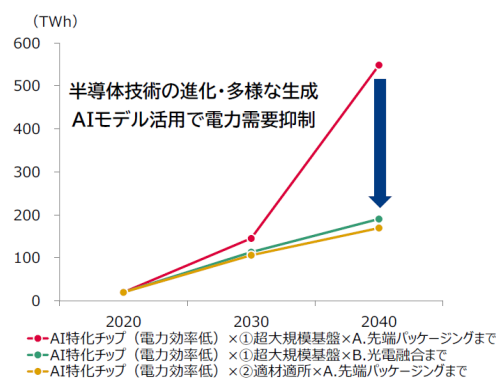
(1) 探討 AI 專用晶片效率對電力需求之變化：在 2040 年[低效率 AI 專用晶片*大型 AI 模型情境*情境 A]的情境組合下用電量最高，為 2020 年的 27 倍，與高效率 AI 專用晶片的情境組合相比，相差 300 TWh(圖 5)。



資料來源：MRI (2024)

圖 5：AI 專用晶片效率差異對電力需求之變化

(2) 探討 AI 模型和半導體技術對電力需求之變化：在使用低效率 AI 專用晶片的情況下，2040 年[低效率 AI 專用晶片*大型 AI 模型情境*情境 B]和[低效率 AI 專用晶片*適材適用情境*情境 A]的組合用電量都能控制在 2020 年的 10~27 倍左右(圖 6)。



資料來源：MRI (2024)

圖 6：AI 模型和半導體技術對電力需求之變化

根據以上情境假設組合估算 ICT 產業的用電量，結果顯示即便 AI 專用晶片的發展不如預期，透過積極發展光電融合技術，並結合「適材適用」的 AI 使用模式，仍然能夠有效控制 ICT 產業的電力需求成長。估計 2040 年 ICT 產業的電力需求為 2020 年的 2~27 倍。

五、生成式 AI 的使用情境與建議

MRI 在探討生成式 AI 的使用情境，依據對大型 AI 模型的依賴程度從高到低分為資料量爆發情境、適材適用情境、優先節電情境。並提出各情境的應對措施。

- (1) **資料量爆發情境：**為了滿足大型AI模型產生的龐大電力需求，應對措施為增加國內電力供應或將部分生成式AI處理工作移轉到海外資料中心。需特別注意增加電力供應可能會有能源短缺的風險，上述2種方式也會導致日本國際收支產生負面影響。
- (2) **適材適用情境：**隨著生成式AI應用範圍越來越多元，需因應使用目的來選擇適用的AI模型，減少過度使用大型AI模型而導致電力資源的浪費。應對措施包括大型AI模型在超大型雲端中運行，並充分利用AI專用晶片、小型AI模型

在邊緣雲端⁵或終端設備上運行。無論是大型科技公司、日本國內企業都能有各自合適的發展方向，對日本AI發展帶來多樣性。

- (3) **優先節電情境：**隨著對碳中和的需求提升，火力發電的使用將有所限制，導致電力吃緊，故優先考慮使用較省電的小型AI模型。應對措施為提升小型AI模型的效率，包括開發專用的小型AI模型，以及將大型AI模型壓縮成小型AI模型。預計2040年ICT產業的電力消耗量可以控制在45 TWh左右(約2020年的2倍)。

若要降低 ICT 產業的電力需求，除了研發先進半導體技術外，MRI 建議日本應追求「適材適用」的 AI 使用模式，在性能和電力需求間取得平衡，並為日本企業創造新的發展機會、提升國際競爭力。

參考文獻

日本三菱綜合研究所(Mitsubishi Research Institute, MRI) (2024)，生成式 AI 的普及對日本電力需求的影響。<https://reurl.cc/Ll6Q2X>

⁵ 「邊緣雲端」指放置在非核心地區(如東京、大阪)的雲端運算。