雖然結合AI與邊緣運算十分具有意義，但硬體和AI軟體面臨著多重挑戰，包括功耗、處理能力、資料儲存和安全性…

根據統計資料庫Statista估計，物聯網(IoT)裝置——包括行動電話、虛擬助理、筆記型電腦、平板電腦、建築物感測器、無人機、安全攝影機和穿戴式健身感測器等——數量可望在2025年增加到超過700億部，並帶動邊緣運算(edge computing)應用隨之成長。另一家市場研究機構Tratica的資料預計，全球人工智慧(AI)邊緣裝置的數量將從2018年的1.614億部迅速增加到2025年約26億部。

IoT裝置在許多領域的應用相當多樣化，包括零售、醫療保健、工業、航太國防、運輸、設施維護、能源、製造、供應鏈物流以及智慧城市等。每一個IoT裝置都可以連續收集資料，因此需要快速地分析，以實現即時決策，特別是針對像自動駕駛車、智慧電網、遠端手術、石油鑽井平台以及甚至是軍用無人機等應用。

**邊緣運算vs雲端運算**

傳統上，雲端運算是IoT裝置分析和預測的模型。在中央雲端運算模型中，資料從終端用戶裝置(「邊緣」)發送到雲端進行分析，然後再將決策回傳到裝置執行。雖然在中央運算模型的資料中心具有處理和儲存資料的巨大能力，但它們的維護成本高昂且耗電量巨大。

在邊緣和雲端之間的傳輸資料不僅代價昂貴、耗時並且會產生延遲。此外，資料傳輸所需的能量也將超過低功耗無線IoT裝置所能支援的能量。而且從後勤、營運或財務上來說，當所收集的資訊中只有小部份可能有用時，也沒有必要將所有的資料都傳到雲端。最後，資料傳輸還可能對資料的完整性和安全性產生不利影響。

相較之下，透過邊緣運算，可以只在IoT裝置上收集和分析資料，以進行快速推論(或決策)，然後只需將少量有用的資料傳送到雲端。邊緣運算提供了幾個優點。由於無需將資料從IoT裝置傳送到雲端，因此其延遲時間、頻寬消耗和成本都將較低，而且可以根據資料分析迅速決策。

此外，即使系統處於離線狀態，邊緣運算也可以持續運作，而且，即時資料處理還可使其更輕鬆地決定傳送哪些資料到雲端進一步分析。



**開發AI邊緣的挑戰**

雖然將AI與邊緣運算相結合別具意義，但硬體和AI軟體面臨著多重挑戰。

第一項挑戰是處理和功耗。AI包括訓練和推論軟體。透過訓練教會模型辨識相關參數，使其得以解讀資料。推論則是讓模型根據學習進行預測。

在雲端運算方面，耗電量高的訓練是在雲端進行的，然後再將訓練好的軟體部署到邊緣，用於相對低功耗的預測(或推論)任務。在邊緣運算方面，這種訓練則轉移到邊緣，因而對邊緣硬體的處理能力提出了更高的要求。對於IoT裝置而言，能耗的增加帶來了更大的問題，需要重新權衡處理能力和功耗需求的關係。

由於邊緣裝置會保留大部份數據而僅將其中的一小部份傳送到雲端，資料的儲存和安全成為第二項挑戰。此外，裝置需要為學習和推論儲存相關參數。第三項挑戰則是IoT裝置的數量龐大，而且目前還缺乏安全標準。

因此，業界科技公司需要開發處理能力更高、功耗更低的硬體，以及能夠更有效執行學習和推論的軟體。此外，IoT的應用與特定的場景和產業有關，因此需要專門為其建立強大的生態系統和開發環境。

**開發AI邊緣的最新進展**

致力於IoT邊緣硬體的大、中、小型公司包括BrainChip (Akida神經形態SoC)、CEVA (NeuPro系列)、Google (Edge TPU)、GreenWave (AI處理器GAP8)、華為(Ascend晶片) 、英特爾(Xeon)、Nvidia (Jetson TX2)、高通(視覺智慧平台)以及意法半導體(STM32微控制器)。

小型公司傾向於關注IoT邊緣軟體。有些公司專注於像Ekkono、FogHorn和Swim (一種基於雲端的POS)等學習，而其他公司則針對推論，例如瑞薩(e-AI)等。還有許多公司開發同時具有這兩種功能的軟體，例如亞馬遜(AWS Greengrass機器學習推論模型)、BrainChip (Studio軟體)、Google (Cloud IoT Edge)、華為(Atlas平台)和IBM (Watson IoT平台)。

大型科技公司則處於建構生態系統的最佳位置，協助開發人員創建特定產業和場景的解決方案。這些公司包括Google (AI平台)、華為(MindSpore)、IBM (Watson)、英特爾(AI開發人員計畫)和微軟(Azure)，以及像IoT Hub、Azure Databricks、ML Studio和Power BI等企業級IoT建構模組。

然而，還有一些較小規模的公司正在創造這一生態系統，例如BrainChip的Akida開發環境。此外，開放霧聯盟(OpenFog Consortium)等產業組織，以及包括Living Edge Lab、ETSI Multi-access Edge Computing和EdgeX Foundry等開源專案也都在為此生態系統做出貢獻。此外，包括高通、微軟和英特爾等主導企業則積極推動大規模的業界合作，與各產業領域的合作夥伴共同努力。

**結語**

透過專用的硬體、軟體和開發環境，邊緣運算可以提高營運的可靠性，實現即時預測以及提高資料的安全性。5G技術可望降低延遲、增強覆蓋範圍和反應能力，此外，量子運算(quantum computing)還能夠加速運算，進一步增進邊緣運算的效率。

然而，在整個邊緣裝置的網路上有效分配處理的需求，將會帶來一大挑戰。此外，對於避免系統故障和最佳化機器學習來說，高效率的任務排程也至關重要。未來，預計將會出現功耗要求更低的更強大處理晶片，屆時，基於AI的邊緣運算將會真正大顯身手。

(參考原文：[How AI changes the future of edge computing](https://www.electronicproducts.com/Internet_of_Things/How_AI_changes_the_future_of_edge_computing.aspx)，by John Koon)