新光投顧

〔投資研究報告〕

2025/6/5

雲端之外的革命：邊緣 AI 顛覆日常

目錄

一、產業鏈全景：從晶片到應用的協同生態

二、國際競爭 vs. 台灣優勢：誰能贏得下一波紅利？

三、生成式 AI 小型模型與軟體協同：讓邊緣端「會思考」

四、風險與挑戰

五、政策環境與未來場景大預測：下一個投資藍圖

六、投資人行動指引

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

從雲端到指尖，不可錯過的邊緣 AI 浪潮

2025/6/5 新光投顧

過去一談到 AI，多半想像雲端資料中心一排排伺服器與天價電費。但如今「邊緣 AI」（Edge AI）正悄悄改寫這個故事：AI 推論不再全部送到遠端，而是直接在智慧攝影機、家電、機械手臂、可穿戴裝置等終端邊緣完成，達到「低延遲」、「省頻寬」與「高隱私保護」的三重優勢。

根據 Grand View Research 報告，2024 年全球邊緣 AI 市場規模約為 **207.8 億美元**，預估在 2025 至 2030 年期間將以 **21.7%** 年複合成長率持續攀升。簡單算一下：若以 21.7% CAGR 成長，2025 年市場規模可望達到約 252.9 億美元，2030 年則有機會逼近 698 億美元。顯示各行各業對於本地化 AI 計算需求極為殷切，背後驅動因素可歸納為：

1. 即時反應需求上升：智慧攝影機須在毫秒內偵測可疑目標；無人機、機械手臂等工業設備要即時做出判斷；自駕與智慧車載需本地融合影像與感測器訊號。
2. 頻寬與營運成本壓力：若要全部原始資料往返雲端，不但頻寬佔用大，且雲端伺服器單價居高不下，邊緣 AI 可將「掏心掏肺的 Raw Data」留在本地，只上傳關鍵資訊，減少傳輸費用。
3. 隱私與資安考量：醫療影像、個人生理數據等敏感資訊，若在本地端完成初步 AI 推論，就能避免大量個資流向雲端浪潮，降低外洩風險。

市場細分與成長動力

- **硬體（AI SoC、NPU 與感測器）**：2024 年硬體端營收占整體市場比重約 **45%**，追求即時決策與低功耗。
- **軟體與服務（Edge AI 平台、模型優化、OTA 更新）**：2024 年占比約 **30%**，隨著生成式 AI 小型模型 / 嵌入式模型崛起，未來需求更大。
- **系統整合與方案（Solution & Integration）**：占比約 **25%**，涵蓋整合終端設備、雲端管理與維運服務。

細分應用領域方面：

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

1. **智慧安防與監控**：2024–2025 年出貨增速超過 **25%**，以零售門市、住宅社區、企業辦公等場景為主。
2. **智慧家電與 IoT**：搭載 NPU（神經網路處理單元，Neural Processing Unit）之智慧電視、冰箱、掃地機器人等終端數量快速攀升，2024–2025 年出貨量年增率超過 **40%**。
3. **工業自動化與機器視覺**：受惠智慧製造補貼與 5G/6G 推廣，2025 年智慧工廠邊緣 AI 方案市場規模可望達到 **15 億美元**。
4. **智慧車載與自駕系統**：隨著汽車電子化與自動駕駛趨勢，邊緣 AI 在車載領域佔比逐年提升，2025–2030 年市場複合成長率約 **22%**。

目錄

一、 產業鏈全景：從晶片到應用的協同生態
二、 國際競爭 vs. 台灣優勢：誰能贏得下一波紅利？
三、 生成式 AI 小型模型與軟體協同：讓邊緣端「會思考」
四、 風險與挑戰
五、 政策環境與未來場景大預測：下一個投資藍圖
六、 投資人行動指引

一、 產業鏈全景：從晶片到應用的協同生態

要讓邊緣 AI 真正落地，必須有一條涵蓋「晶片設計 → 矽晶圓製造 → 封裝測試 → 模型與 SDK → 系統整合 → 最終場域應用」的完整價值鏈。下文將依序解析各環節關鍵角色與代表廠商。

1.1 晶片設計（ASIC/SoC 設計）

- **核心任務**：在小體積與低功耗限制下，將神經網路推論加速器──也就是 NPU（Neural Processing Unit, 神經網路處理單元）──整合到 SoC（System-on-Chip, 系統單晶片）。設計重點在於：

    1. **效能 / 功耗比**：以 TOPS（Tera Operations Per Second，兆次運算）為單位，NPU 能力越高、功耗越低，越適合放在攝影機或物聯網裝置中跑高複雜度模型。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

2. **小型化模組化**：SoC 除了 NPU，還需內建 CPU、記憶體介面（如 LPDDR5X）、通訊模組（Wi-Fi 6/7、藍牙、5G）等，以滿足邊緣設備多元需求。

- **代表台灣廠商**：

  - **芯鼎科技（6695）**：以客製化 AI ASIC 設計見長，與台積電（TSMC）合作推出多款針對智慧安防、工業檢測、醫療影像的 NPU SoC，具備「快速驗證→矽原型→量產」三步驟優勢。
  - **瑞昱科技（2379）**：RTD 系列智慧攝影機 SoC 內建低功耗 NPU，可在休眠狀態下持續執行人形偵測等機器視覺演算法，被多家筆電、監控設備廠商採用。
  - **聯發科技（2454）**：手機 SoC 大廠，旗下 Dimensity 系列已將多核 NPU 下放到中高階手機，也推出 Genio 物聯網平台，專攻工業自動化、智慧零售、智慧醫療等邊緣場景。

- **主要國際競爭對手**：
  - **英特爾（Intel）**：以 Movidius Myriad VPU（Vision Processing Unit）著稱，主攻終端機器視覺與工業自動化，但相較台系 SoC，成本、功耗優勢略顯不足。
  - **華為海思（HiSilicon）**：Ascend 系列 NPU 在雲端訓練端表現出色，但受制於美國制裁，製程與先進封裝受限，邊緣市場佈局較艱難。
  - **英偉達（NVIDIA）**：Jetson 系列（如 Xavier NX、Orin Nano）效能領先，但功耗與單價偏高，較適合中高階機器人、自動駕駛開發者，非大規模低成本量產首選。
  - **百度（Baidu）**：Kunlun Pro AI 類似 NVIDIA Jetson，但更偏重於大模型推論，在輕量化嵌入式 NPU 領域尚未形成實質競爭力。

1.2 矽晶圓製造（Fabless → Foundry）

- **製程選擇原則**：
  - 邊緣 AI 晶片在成本與良率之間需取捨，常見製程為 **28nm、16nm、12nm** 等成熟 FinFET 製程，其中 16/12nm 製程已能提供良好的能效比與成本優勢；若要搶攻更高階智慧家電或工業設備，也可能導入 **7nm**。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- 代表晶圓代工廠：

  - **台積電（2330）**：擁有全球最頂尖製程與先進封裝（CoWoS、InFO、SiP），公布 16nm/12nm 製程已成為邊緣 AI SoC 首選，並提供客製化 eSilicon 共同設計服務。
  - **聯電（2303）**：16nm/12nm 製程成本相對低廉，良率穩定，是中小型 AI SoC 廠商的常見選擇。
  - **華邦電（2344）**：雖以記憶體代工為主，但其特色製程可協助部分 AI 晶片取得量產機會，尤其在利基型產品或低功耗傳輸晶片上具備市場空間。

1.3 封裝與測試（OSAT）

- 先進封裝技術重點：

  - **CoWoS（Chip-on-Wafer-on-Substrate）**：可將 NPU SoC、影像感測器、記憶體晶片（如 eMMC、UFS）透過高速互連堆疊，降低信號延遲並提升帶寬。
  - **SiP（System-in-Package）**：把多顆不同功能的晶片（NPU、Wi-Fi、MCU 等）封到同一基板，適合體積受限的安防攝影機或智慧門鎖。
  - **Fan-Out 封裝**：透過外扇出方式拓展 I / O 接腳，改善訊號品質與散熱，應用於高效能邊緣 AI 模組。

- 代表 OSAT 廠商（台灣）：

  - **日月光投控（3711）**：全球前三大封測龍頭，掌握 CoWoS 與 Fan-Out 技術，能協助客戶將邊緣 AI SoC 與多模組整合到一粒綜合模組中，應用於工業相機、智慧攝影機、機器人等領域。
  - **京元電子（2449）/ 力成科技（6239）**：以成熟製程封裝與自動化測試聞名，在 AI 晶片測試領域加碼投資，期望隨中低價位邊緣 AI SoC 放量而同步受惠。

1.4 模型與軟體生態（Edge AI Model & SDK）

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- 機器學習與生成式 AI 演算法日新月異，為讓終端設備「聰明起來」，需要「輕量化」且「可在邊緣硬體上跑得動」的 AI 模型，以及相互配合的開發工具與框架。

1. **小型化生成式 AI 模型（Miniaturized Generative AI Models）：**

   - **Google Gemini Nano**（僅約 200 MB）：能在手機或嵌入式 GPU/NPU 上執行自然語言理解（NLU）與生成（NLG），適用於對話式助理與即時翻譯。
   - **Meta Llama 3 Embedder**：針對嵌入式場景優化，能在 NPU 架構上完成文本嵌入（Embedding）與分類（Classification），應用在本地化推薦、即時對話。
   - **Microsoft Phi-3 Nano**：與 Azure IoT Edge 深度整合，可將語音辨識、意圖偵測等功能部署到工業現場或零售終端，並透過雲端推播模型更新。

   ✧ 註解：模型嵌入（Embedding）
   是將文字、圖像或其他高維度資料轉換為向量 (Vector) 表示，以便於在 NPU/SoC 上進行相似度計算、分類或推薦等任務。

2. **Edge AI 開發套件與框架（SDK & Framework）：**

   - **Google Edge TPU API**：讓開發者可將 TensorFlow Lite（TFLite）模型編譯為可在 Coral USB 加速器或 Edge TPU Dev Board 上執行的格式；適合打造輕量型影像辨識與自動標註系統。
   - **NVIDIA JetPack SDK**：整合 TensorRT（深度學習推論加速庫）、DeepStream（視覺智慧串流平台）、CUDA Toolkit，適用於機器視覺、智慧零售、無人機等對即時性要求極高的場景。
   - **MediaTek NeuroPilot**：可將 TensorFlow、PyTorch 模型一鍵轉換為 Genio 開發板(NPU)可執行格式，並提供韌體優化範例，降低開發門檻。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- **AWS IoT Greengrass**：可在邊緣設備上部署 AWS 機器學習模型，並透過 MQTT 或本地存取感測器資料，適合智慧製造與遠端監控。

有了這些工具，硬體與軟體才能形成協同效應。若只有「強效能 NPU」，卻缺乏易用的模型部署與更新管道，終端應用就難以大規模量產。

1.5 系統整合與場域應用（System Integration & Solutions）

- **角色定位**：系統整合商與方案服務商負責把前述 SoC、感測器、通訊模組、以及雲端平臺連結起來，打造一套「端到端」（End-to-End）的完整解決方案，並提供後續維運、模型更新與客製化服務。

- **代表廠商 / 生態系**：

  - **研華科技（2395）**：推出 ARM + NPU 架構的邊緣 AI 盒子，搭配 Advantech Edge Intelligence Suite，可用於智慧製造（即時瑕疵檢測）、智慧零售（顧客行為分析）、智慧城市（交通監控）等多種場景。
  - **樹莓派（Raspberry Pi）社群生態**：以 Pi 4、Pi 5 等開發板結合 Google Coral USB Accelerator、Intel Neural Compute Stick 等外接加速器，為 Maker 與中小企業提供入門型邊緣 AI 驗證平台。
  - **NVIDIA Jetson 生態**：Jetson Nano、Xavier NX、Orin Nano 等模組搭配 JetPack SDK，被廣泛應用在機器人、無人機、智慧門禁、智慧農業等多元市場。

比方說：

  - **智慧工廠**：研華的邊緣 AI 盒子串接機器視覺鏡頭，可在生產線上即時偵測微小裂痕或異色不良品，並透過 5G/4G 傳回雲端進行履歷管理。
  - **智慧零售**：NVIDIA Jetson Xavier NX 結合電腦視覺模型，可在無人店鋪中完成顧客進出偵測、貨架商品辨識與自動結帳，減少人力成本。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- **智慧醫療**：樹莓派結合 Google Coral Accelerator，可在偏鄉醫療站做即時超音波影像推論，初步篩檢心臟異常，並將結果加密後上傳醫院雲端。

二、 國際競爭 vs. 台灣優勢：誰能贏得下一波紅利？

在邊緣 AI 市場中，美國、中國與台灣各擁先天條件。投資人可從技術、產能、資源整合與生態合作等維度比較，幫助把握具備競爭優勢的標的。

2.1 技術與製程競爭格局

- **美國陣營（Intel、NVIDIA、Google）**：
  - **Intel Movidius Myriad VPU**：專注機器視覺終端，效能領先同級選手；但所採製程多為 16/12nm，近期 Intel 自家 7nm/5nm 製程落後台積電，良率與成本優勢不如台系大廠。
  - **NVIDIA Jetson 系列**：以 Xavier NX、Orin Nano 為代表，統合高階 GPU + NPU，可應付中高階無人機、機器人等場域；功耗與單價偏高，可能不適合大規模低成本產品。
  - **Google Coral Edge TPU**：優勢在於針對 TFLite 模型做深度優化，支援本地推論，適合 IoT 端小型影像辨識，但整體生態仍偏向雲端與美國市場。

- **中國陣營（華為海思、百度、阿里、騰訊）**：
  - **華為海思 Ascend 系列**：在雲端端與資料中心端具備領先優勢，但受制於美國對華為的制裁，先進製程與封裝技術受限，在全球供應鏈整合上面臨挑戰。
  - **百度 Kunlun Pro**：可支援邊緣行車記錄器、安防攝影機等場景，但純 AI 推論芯片生態相對封閉；需仰賴國內晶圓與封測廠。
  - **阿里 Han Gu 系列**：主打雲端 + 邊緣協同，但邊緣 AI 落地速度次於台灣及美國廠商。

- **台灣陣營（台積電、聯電、日月光、矽品等）**：
  - **製程與封裝優勢**：

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

        - **台積電（2330）** 16/12nm 製程良率高、成本相對可控；更在 CoWoS、InFO（Integrated Fan-Out）等先進封裝技術上居全球領先地位。
        - **聯電（2303）** 提供與台積電互補的成熟製程選擇，吸引許多中小型 AI SoC 客戶；**力晶（Powerchip, 6239）** 則可在特色製程上分食部分需求。
        - **日月光投控（3711）** 封包服務全球前三，2.5D/3D 封裝能把 NPU SoC 與 HBM（High-Bandwidth Memory，高頻寬記憶體）長寬整合，改善頻寬瓶頸。
    - **研發與製造一條龍生態**：
        - 台灣具備從 IC 設計、製造、封裝、測試，到 ODM / 系統整合的完整生態，對外資客戶與新創廠商具備強大吸引力。
        - 在低功耗設計與系統整合方面，台系廠商能快速協同優化散熱與電源管理，縮短從設計到量產的時程。

2.2 軟體與生態系整合比較

- **美國 / 國際陣營**：
    - **Google、NVIDIA、Microsoft** 擁有最豐富的 AI 模型與雲端服務生態（TensorFlow、CUDA、Azure），能提供從模型訓練、推論到部署的完整工具鏈。
- **中國陣營**：
    - **百度（PaddlePaddle）、阿里（Ali-DL）、騰訊（Tencent AI Lab）** 提供自有框架，但國際開源度較低，生態黏著度集中在中國本土。
- **台灣陣營**：
    - 目前以硬體和封裝見長，軟體生態尚待加速跟進；**研華** Advantech Edge Intelligence Suite 可對接國際雲巨頭，**MediaTek NeuroPilot** 則與 TensorFlow、PyTorch 等框架兼容，但在「AI 開源軟體社群」的影響力仍需擴大。

2.3 投資人觀點：關鍵競爭優勢

- **台灣在「製程 + 封裝」**環節具備全球頂尖地位，硬體成本優勢與高良率是長期競爭壁壘。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- **硬體與軟體協同仍是決勝關鍵**，投資人應留意台廠是否能與國際模型開發生態緊密結合，提供「開發工具 + 服務 + 雲端連動」的一站式方案。
- **異質整合與低功耗設計能力**，在可穿戴、智慧家電與車載等場景，正成為台灣廠商的獨特優勢。

三、 生成式 AI 小型模型與軟體協同：讓邊緣端「會思考」

邊緣 AI 要發揮真正效益，硬體效能只是一半，對「小型化模型」和「便捷開發框架」的需求日益殷切。以下三個面向，可看出台灣廠商未來可著墨的突破口。

3.1 小型生成式模型（Miniaturized Generative AI Models）

- **Google Gemini Nano**（約 200 MB）：
  - 以 Transformer 架構裁剪，能在 智慧手機或一體成型的 NPU SoC 上執行對話式翻譯、文字摘要、語音交互等功能，並保持低延遲。
- **Meta Llama 3 Embedder**：
  - 以文本嵌入技術 (Embedding) 為主，可在 NPU 或輕量級 GPU 上完成語意檢索、意圖分類，適用於本地化客服機器人或智慧推薦。
- **Microsoft Phi-3 Nano**：
  - 強調與 Azure IoT Edge 協同，可把語音助理、文字辨識、影像標註等模型部署到現場終端，且透過 OTA（Over-the-Air，無線更新）機制自動更新模型。

這些小型模型的重要意義在於：

1. 在 **NPU（神經網路處理單元）** 只有數 TOPS 推論能力的前提下，仍可執行短文本生成、多模態推論等複雜任務。
2. 減少對雲端資源依賴，使智慧家居、可穿戴、工業 IoT 等終端場景能更快普及。
3. 模型體積小、更新快速，利於廠商透過 OTA 機制持續優化用戶體驗。

3.2 邊緣 AI 開發套件與雲端協同平臺

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- **MediaTek NeuroPilot**：
  - 提供一鍵將 TensorFlow、PyTorch 模型轉成 Genio SoC 可執行格式，並附上影像分類、物件偵測等範例程式。
- **NVIDIA JetPack SDK**：
  - 整合 CUDA、TensorRT、DeepStream，支援 C++、Python 等開發語言，能快速部署機器視覺、智慧路口監控與自駕模組。
- **Google Edge TPU API**：
  - 幫助開發者將 TFLite 模型燒錄到 Coral USB Accelerator 或 Dev Board，並提供 Python / C 程式範例，適合 Maker 與小型業者做 proof-of-concept。
- **AWS IoT Greengrass**：
  - 允許在邊緣裝置上執行 AWS Lambda 函數、Docker 容器，並搭配 SageMaker 模型做本地推論，便於企業整合雲端 AWS 生態。

這些 SDK 與平臺的價值在於：

1. **縮短開發週期**：廠商可省去多數平台相容性測試的時間，快速完成邊緣 AI 部署。
2. **降低系統整合成本**：若硬體廠商能同時提供 SoC 與優化過的 SDK，就能吸引開發者快速試用並加速量產。
3. **加速生態系整合**：硬體 + 軟體 + 雲端管理三方面相互協同，才能真正支撐跨場域應用（如智慧製造、智慧城市、智慧零售等）的落地。

四、 風險與挑戰

雖然邊緣 AI 市場前景可期，但在商業化與大規模量產過程中，仍有多項風險與挑戰需要投資人關注：

1. **標準碎片化與互通性障礙**
   - **問題描述**：不同 SoC 廠商（MediaTek、NVIDIA、Realtek）、不同 AI 框架（TensorFlow Lite、TorchScript、ONNX）之間缺乏統一標準，若要讓設備 A 上的模型能在設備 B 直接「跨拷貝」運行，常需重新優化與驗證。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- **影響程度**：推動大型企業試點容易，一旦要拓展到中小客戶（中小型製造業、智慧零售店家），就面臨「一套方案只適合少數設備」的痛點，影響整體市場規模擴張。

2. **系統整合與資安防護難度**
   - **技術門檻**：邊緣裝置需同時處理感測器資料（影像、聲音、溫度等）、NPU 推論、通訊模組管理與電源調度，開發與測試週期相對長。
   - **資安風險**：若終端設備未做到韌體層加密、啟用安全啟動、Secured Boot 等機制，就容易遭受中間人攻擊（Man-in-the-Middle）或惡意韌體植入，導致資料外洩或設備被綁架。
3. **人才供給短缺與研發成本高昂**
   - **複合型人才需求**：邊緣 AI 涉及 SoC 架構、AI 模型優化、系統整合與資安防護等多領域知識，目前全球市場都存在極大缺口；台灣尤其在「AI 軟體 + 硬體整合」的跨領域人才稀缺。
   - **資金投入**：AI SoC 設計往往需要千萬至數千萬美金研發預算，若還要搭配先進封裝（CoWoS、SiP）與資安認證，整體費用高昂。若市場需求不如預期，廠商將承擔龐大庫存與技術折舊風險。
4. **法規與政策不確定性**
   - **跨國合規壓力**：入歐須符合 GDPR，入美須符合 CCPA 等個資法規，若產品含有攝影鏡頭與人臉辨識功能，更可能受美國、歐盟對人臉辨識技術的嚴格限制。
   - **地緣政治風險**：中國封測廠與台灣廠商在美中貿易摩擦中須權衡供應鏈安全；若美國再度擴大出口管制，台灣設計廠能否取得最新製程材料與設備，將影響下游晶片交期與成本。

五、 政策環境與未來場景大預測：下一個投資藍圖

5.1 各國政策趨勢與補貼方向

- **美國**
  1. **CHIPS Act（晶片與科學法案）**：提供涵蓋 1500 億美元的半導體補助與研發經費，鼓勵廠商回流美國設廠，並針對 AI 與量子計算等前沿領域額外補助。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

2. **NIST（美國國家標準暨技術研究院）**：正在制定邊緣 AI 安全與互通框架，未來常承包政府與國防單位合約的廠商需符合「可信 AI」、「安全硬體」規範。

- **中國**
  1. **十四五規劃與「2035 遠景目標」**：聚焦打造全國邊緣計算節點，目標在 2025 年前在一線城市建成千個邊緣雲伺服器。
  2. **國家級基金**：針對 AI 芯片研發與製造，持續加碼資金投入，並推動「兩化融合」示範工廠補貼。
- **歐盟**
  1. **「數位十年」計畫**：強調 AI 可信賴性與可解釋性，鼓勵成員國在 2025 年前打造至少 10 條跨國邊緣 AI 示範線路。
  2. **AI Act（歐盟人工智慧法案）**：要求高風險場景（如醫療、交通）之邊緣 AI 設備必須通過嚴格合規審查，包括資料隱私、透明度與可追溯性。
- **台灣**
  1. **「5+2 產業創新計畫」**：半導體與 AI 產業為重點，台灣政府提供研發補助、稅率減免與人才培訓經費。
  2. **科技部與經濟部合作**：推動「邊緣智能暨物聯網」示範場域，2025 年前預計在台中、南科園區設立邊緣 AI 測試示範中心。

5.2 未來生活場景大預測（2025–2030）

1. **可穿戴智慧健康貼片與眼鏡**
   - **場景描述**：
     - **智慧健康貼片**：可貼在皮膚上，用邊緣 AI 即時分析心電圖 (ECG)、血氧 (SpO_2)、體溫等生理數據，並以小型 NPU SoC 做本地推論。異常訊號可即時發警示並推送到手機 App，甚至連動醫療院所。
     - **智慧隱形眼鏡**：若在鏡片邊緣放入可居中的微型 NPU+感測器，就能即時翻譯外文招牌、識別路況，或協助色盲人士進行顏色判讀。
2. **智慧零售與無人商店**
   - **場景描述**：
     - 架上攝影機 + 邊緣 AI 推論，可即時辨別顧客拿取哪件商品、停留時間，以「預測購買意圖」；

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

- 櫃檯可結合 RFID 感測與影像辨識，在顧客結帳時自動計算金額並完成付款，免去排隊。
- 系統可透過本地化推論蒐集客流熱區圖 (Heatmap)，店家可即時調整商品陳列與促銷策略，並將行銷數據上傳雲端進行 AI 分析。

3. **智慧醫療與居家照護**
   - **場景描述**：
     - **AI 超音波掃描筆**：醫護人員只需在病患身上輕滑，即可在筆上運行邊緣 AI 模型進行初步診斷（如肝臟結節篩檢），並將結果加密後上傳醫院伺服器做二次判讀。
     - **居家健康監測站**：設置在居家環境中的多合一機台（結合心電圖、呼吸頻率、聲音偵測），由內建 NPU 在家中持續監測，若發現異常即時推播到遠端醫療平台。
4. **智慧農業與環境監控**
   - **場景描述**：
     - 植保無人機搭載多光譜攝影機 + 邊緣 AI 推論，可即時偵測病蟲害地圖，並自動規劃飛行路徑與噴藥計畫，減少農藥用量；
     - **智慧魚塘**：攝影機 + 聲音感測器連動邊緣 AI，可辨識魚群密度、活動姿態、異常行為，並同步監測水質指標，提前預警疾病。
5. **智慧交通與城市管理**
   - **場景描述**：
     - 都市道路佈建邊緣 AI 攝影機，可即時偵測違規車輛與人行走行為（如闖紅燈、逆向），並連動紅綠燈系統動態調節車流；
     - 智慧路燈內建 NPU，可透過聲音與影像辨識即時偵測火災或意外事故，並直接通知相關單位，提升公共安全。

六、 投資人行動指引

6.1 佈局「晶片設計 + AI IP」題材

- **重點標的：**
  1. 芯鼎科技（6695）：掌握客製化 AI ASIC 設計能力，若能持續推出領先架構，將受惠於多家晶圓代工廠與封測廠商訂單。
  2. 瑞昱科技（2379）：以智慧攝影機 SoC 為起點，若能將 NPU 運算能力下放到更多物聯網產品，體量有爆發潛力。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

3. 聯發科技（2454）：Dimensity 與 Genio 平台雙軌並進，本地化 NPU 運算與 5G/IoT 生態整合能力值得留意。

- **投資人觀察要點：**

1. AI IP 授權與國際生態合作：是否與 Google、NVIDIA、ARM 等開源模型或 IP 持續合作？
2. NPU 性能 / 功耗比與成本：TOPS（兆次運算）/ 瓦特耗能的數據是否優於同規格競爭者？量產成本是否具備競爭力？
3. 訂單能見度與客戶結構：是否有長期訂單佈局，且客戶涵蓋智慧家電、工業 IoT、智慧攝影機等多個領域？

6.2 佈局「晶圓代工＋OSAT」龍頭

- **重點標的：**
1. 台積電（2330）：16/12nm 成熟製程與先進封裝技術領先全球，是絕對的第一投資標的。
2. 聯電（2303）：訂單來源穩定、成本與良率兼顧，可參與中小型 AI SoC 成長紅利。
3. 日月光投控（3711）：2.5D/3D 封裝需求暴增，營收與獲利可望同步向上；京元電（2449）與力成科技（6239）則可受益於測試需求成長。

- **投資人觀察要點：**
1. 產能利用率與訂單結構：台積電與聯電是否持續提高成熟製程訂單？日月光是否新增大型 AI 客戶訂單？
2. 封裝技術升級動向：CoWoS、InFO 與 Fan-Out 新方案是否已進入量產？是否有前瞻性「System-in-Package」（SiP）專案？
3. 下游整合生態：與 AI 設計廠商、感測器廠商、系統整合商的合作是否深化？是否能提供整體方案？

6.3 佈局「系統整合與生態服務」商機

- **重點標的：**

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。

1. 研華科技（2395）：以邊緣 AI 方案著稱，未來可望在智慧製造、智慧城市及智慧醫療場域具備領先優勢。
2. NVIDIA（美股：NVDA）：Jetson 生態帶動機器視覺與自動駕駛應用，但台灣投資人可透過仰攻港股或美股連結投資。
3. 樹莓派（Raspberry Pi）生態：雖非上市公司，但可留意其合作夥伴與客戶案例，捕捉長尾市場需求。

- **投資人觀察要點：**
  1. 開發者生態：系統整合商是否提供完善的開發套件 (SDK)、技術支援與教學資源？是否能吸引大量開發者使用？
  2. 方案整合能力：是否能將硬體、模型、雲端管理與 OTA 更新一線串接，達到「一條龍」服務效果？
  3. 訂閱式服務模式：若能透過軟體/韌體訂閱收費或雲端服務提供持續營收，將有助於降低傳統硬體銷售的季節性波動。

6.4 風險控管：政策與法規、人才短缺、技術整合

1. **政策與法規風險：**
   - 美國出口管制：若對中出口規定再度收緊，台灣廠商是否能維持先進製程與高階封裝供應？
   - 歐盟 AI Act / GDPR：若欲進軍歐盟市場，必須確保邊緣 AI 設備符合資料隱私與可解釋性標準；而美國 CCPA 等法規亦不可忽視。
2. **人才短缺與研發成本：**
   - 台灣要持續挖掘具備「硬體架構 + AI 演算法 + 系統整合」能力之跨領域人才，並與學界、研發單位深化合作。
   - 公司若無法長期投入研發或無法吸引頂尖人才，可能在技術上無法與國際巨頭競爭。
3. **技術整合難度與標準碎片化：**
   - 投資人應留意標的廠商是否積極參與國際開源社群（如 Linux Foundation LF Edge、OpenNESS 等），以降低標準碎片化的影響。
   - 同時需評估廠商在垂直整合場景（如智慧製造、車載、醫療）是否具有足夠客製能力，避免脫離市場實際需求。

本內容謹提供參考並不得任意轉載，投資人於決策時應審慎衡量本身風險，新光投顧恕不負任何法律責任，亦不做任何保證。