

# 【新興領域：5月焦點3】AI晶片大戰 誰與爭鋒

人工智慧 (AI) 被視為第四次工業革命核心，預估到 2020 年，市場商機規模高達 3,000 億美元。AI 晶片更是這場革命中最重要的角色之一，除了傳統晶片大廠外，也吸引科技巨頭、AI 新創企業爭相投入研發與商業化量產的行列。

本文主要的目的在介紹 AI 晶片大戰中，值得關注的晶片與科技大廠，以及夾著深度學習演算法優勢，加入 AI 晶片大戰的新創企業，引領讀者掌握最新 AI 晶片的發展進度。

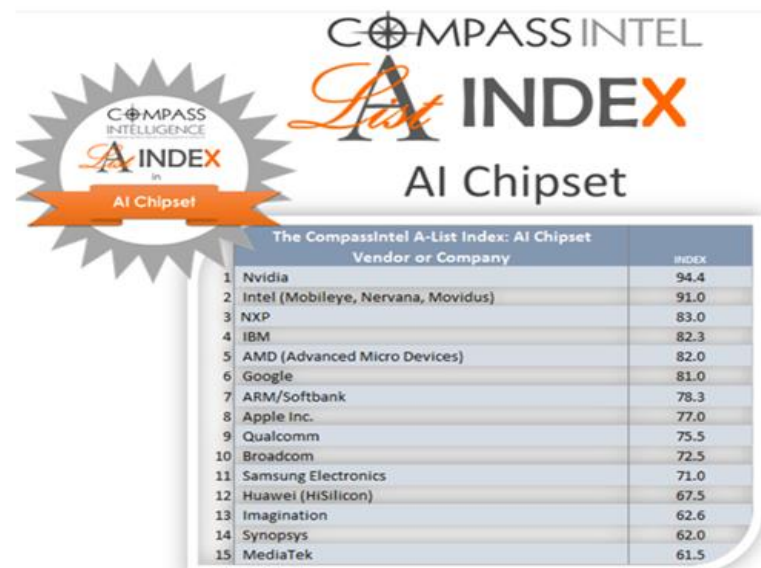
林秀英

台灣經濟研究院研六所副  
研究員



人工智慧 (AI) 被視為第四次工業革命核心，預估到 2020 年，市場商機規模高達 3,000 億美元，AI 晶片更是這場革命中最重要的角色之一，除了傳統晶片大廠外，也吸引各大科技廠、AI 新創企業爭相投入研發與商業化量產的行列。

最近，市場研究顧問公司 Compass Intelligence 對全球 100 多家 AI 晶片企業進行了排名 (AI Chipset Index)，發佈全球前 15 大人工智慧(AI)晶片企業排名表“A\_List”中，名依序為輝達 (Nvidia)、英特爾(Intel)、恩智浦 (NXP)、IBM、AMD、Google、ARM/軟銀、蘋果、高通、博通、三星電子、華為 (海思)、Imagination、新思科技 (Synopsys) 和聯發科。



Source: Compass Intelligence, LLC

## AI 晶片定義

AI Chipset Index 的 A-list 包括提供 AI 晶片組的軟體和硬體公司。所謂「AI 晶片組」(AI Chipset) 包括中央處理器 (CPU)、圖形處理器 (GPU)、神經網路處理器 (NNP)、專用積體電路 (ASIC)、現場可編程邏輯閘陣列 (FPGA)、精簡指令集計算機 (RISC) 處理器與加速器等。有些晶片組針對邊緣處理或設備，有些則是針對雲端運算中使用的服務加速器，另外一些是針對機器視覺和自動駕駛平台等用途的晶片。

## AI Chipset Index 的 A-list 評分標準

該榜單的評分標準包括：

1. 供應商指標 (Vendor Metrics) (30 分)：公司收入和 AI 收入 (或範圍)、併購可能性、員工或資源、成長潛力、商業化所需的時間。
2. 產品和客戶指標 (Product and Client Metrics) (30 分)：產品組合實力、特點強度、創新水準、客戶成功案例、市場覆蓋率、品牌強度。
3. 經濟指標 (Economic Metrics) (25 分)：市場廣度、潛在市場、可獲得市場；彈性水準、市場占有率。
4. 其他 (共 15 分)：領導能力、近期新聞和活動、Google Mentions 和 AI

以下介紹 AI Chipset Index 的 A-list 前 10 大企業：

### NVIDIA

NVIDIA (輝達) 自 1999 年發明 GPU 起，成功驅動 PC 電競市場的成長，並重新定義現代電腦顯示卡產品，掀起電腦平行運算革命。GPU 的浮點運算、同步並行運算是非常適用於近年崛起的人工智慧深度學習神經網路，因此成為 AI 晶片領頭軍。NVIDIA 目標正朝向遊

戲 (市場規模 1,000 億美元)、人工智慧 (市場規模 3 兆美元) 和運輸行業 (市場規模 10 兆美元) 三路發展。

NVIDIA 在 2018 年 1 月 CES 大展中，針對自動駕駛汽車領域推出全新 AI 處理器 Xavier，擁有 90 億個以上電晶體的 Xavier，是迄今最複雜的系統單晶片，代表著兩千餘名 NVIDIA 工程師投入四年光陰與二十億美元研發經費的心血。

Xavier 的架構為一個訂制的 8 核心 CPU、一個新的 512 核心 Volta GPU、一個全新深度學習加速器、全新電腦視覺加速器，還有全新 8K HDR 影片處理器。低耗能的 DRIVE Xavier 有著更高的處理能力，每秒可處理 30 兆次的運算作業，耗電量只有 30 瓦，比前一代架構的節能效率還要高出 15 倍。

Xavier 是 NVIDIA DRIVE Pegasus 人工智慧運算平台的重要關鍵之一，是全球首款為 Level 5 全自動駕駛計程車而設計的人工智慧車用超級電腦，它可以裝在比一片車牌還小的電路板上，取代現今許多自動駕駛車輛所使用，大小足以裝滿後車廂的運算裝置。

Pegasus 平台採用兩具 Xavier SoC 及兩個下一代 NVIDIA GPU。客戶將在 2018 年中拿到擁有每秒 320 兆次處理能力的首批 Pegasus 樣品。

目前已有超過 25 家公司使用 NVIDIA 技術來開發全自動駕駛計程車，而 Pegasus 將為推動量產助一臂之力。

### INTEL

由於 X86 架構的高功耗，英特爾幾乎與整個行動時代無緣，若不是 PC 和伺服器領域的統治地位，英特爾或許早已舉步維艱。隨著汽車技術的發展，智慧汽車產業爆發了對汽車級高性能處理器的巨大需求，這巨大的市場當然是英特爾不願再錯過的。

Intel 為強化 AI 晶片競爭能力，英特爾在 2015 年花費了高達 167 億美元收購了那斯達克的 FPGA 晶片公司 Altera，這是 Intel 歷史上最大的收購，希望藉助 FPGA 與輝達在機器學習和智慧駕駛領域一較高下。

2016 年 8 月 Intel 又以 3.5 億美元收購深度學習公司 AI 新創業者 Nervana，這家新創公司提供的深度學習演算法在人工智慧和智慧駕駛領域都有優異的效果，並在 2017 年 10 月 17 日宣布推出「Nervana 神經網路處理器」，搶占 AI 晶片市場。

接著在 2016 年 9 月再以 4 億美元收購視覺晶片公司 Movidius，歷經將近一年的研發，於 2017 年 8 月底宣佈推出了全新的 Myriad X 視覺處理器 ( Vision Processing Unit ; VPU )，這是全球第一個配備神經運算引擎 ( Neural Compute Engine ) 的系統單晶片 ( SoC )，可用於加速產品端的深度學習推理，未來可應用在無人機、機器人、智慧攝影機、虛擬實境與擴增實境等產品。

2017 年 3 月，INTEL 再以 153 億美元收購以色列先進駕駛輔助系統 ( ADAS ) 的企業 Mobileye，補全了英特爾在自動駕駛領域只有硬件沒有方案的缺失。目前 Intel 與 Mobileye 共同開發平台將以兩組 Mobileye EyeQ5 晶片搭配 Intel Atom C3xx4 處理器以及 5G，並且藉由 Mobileye 軟體系統實現以精簡成本達成 Level 4-Level 5 的全自動駕駛技術，與其他的自動駕駛巨頭公司進行競爭。

## **NXP 恩智浦**

恩智浦為全球最大的汽車電子及人工智慧物聯網晶片公司，逾 60 年的領先經驗及技術，不斷推動人工智慧、物聯網、自動駕駛及端點到端點安全市場的持續創新。

恩智浦(NXP)半導體認為機器智慧的進展，將使得無人車(driverless car) 在未來完全普及。相較於其他競爭對手在汽車電子市場的布局，

恩智浦在 ADAS 上的競爭優勢尤其明顯。原因是 ADAS 的產品需要比較精密的計算能力，而恩智浦布局較早。加上在微控制器 (MCU) 上面有其技術上的優勢，因此，相對競爭對手有著明顯的差異化表現。另外，恩智浦也正在規劃能搭配自駕車 Level 4 駕駛系統運作的鏡頭影像處理器，比起目前多數車廠大幅度使用其他廠商的自駕車 Level 3 系統的影像處理器要有更好的效能。

除了在技術上的優勢之外，NXP 不但生產晶片，更可以讓客戶自行修正從底層到應用層的各项參數，這樣的功能相對更為開放。

## **IBM**

IBM 從 2008 年開始研究能模擬人類大腦的晶片項目，然後在 2014 年推出第一個 TrueNorth 晶片。TrueNorth 是 IBM 參與 DARPA 的研究項目 SyNapse 的成果。

強大的運算能力和市場需求持續推動認知架構發展，IBM 早已朝該方向落實協同創新與開放生態圈之策略。IBM 於 2013 年與 Google 等共同成立 OpenPOWER 基金會，開放 IBM POWER 晶片技術，使得 AI、機器學習所需的認知架構運算效能大幅提升；為加速開源領域的開放協作，IBM 亦透過 Docker、MongoDB、Redis 等開源技術社群，讓開發者體驗 Power 支援的 Linux 開發環境。

2016 年 IBM 在年度 Hot Chips 技術大會上介紹 Power 9 處理器，這款突破性晶片主要面向大型企業與超級計算機場景，負責處理分析、大數據、機器學習以及其它各類高強度任務。經過一年的研發，在 2017 年 12 月 13 日，IBM 在北京正式發表了新一代 Power 9 伺服器晶片架構。該架構針對目前熱門的 AI 運算領域優化，並期望作為新一代通用平台，協助 GPU 等計算架構解開來自 X86 平台的層層限制，發揮真正的計算性能。作為 IBM 長久以來的合作夥伴，NVIDIA 也首發了

NVLINK 2.0 規格，並在 Power 9 平台上展現出了遠超過 X86 平台的性能表現。

## AMD

處理器大廠美商超微 ( AMD ) 在 2018 年 CES 大展中所舉辦展前技術日 ( AMD Tech Day )，執行長蘇姿丰 ( Lisa Su ) 宣布將於 2018 年第 2 季之後推出 12 奈米第二代 Ryzen 桌上型處理器，以及 7 奈米 Vega 繪圖晶片，並將結合處理器及繪圖晶片的運算效能，迎接大數據分析及機器學習的人工智慧新時代。

2017 年 9 月特斯拉公司已和超微 ( AMD ) 合作，將開發自家的自動駕駛車人工智慧 ( AI ) 晶片。

## Google

Google 花了十年打造伺服器中心，處理每日數十億次的網路搜尋需求。在 2016 年 5 月 Google I/O 大會，Google 首次公布了自行研發專屬晶片 TPU (Tensor Processing Units)，加快機器學習腳步，表示 Google 資料中心已使用了一年，並宣稱 TPU 性能優於 CPU 與 GPU。

2017 年 Google I/O 大會，Google 宣布正式推出第二代 TPU 處理器，第二代 TPU 處理器加深人工智慧在學習和推理的能力，據 Google 內部測試，第二代 TPU 晶片針對機器學習的訓練速度比目前市場上的 GPU 節省一半時間。新的 TPU 包括了 4 個晶片，每秒可處理 180 兆次浮點運算。Google 還找到一種方法，使用新的電腦網路將 64 個 TPU 組合到一起，升級為所謂的 TPU Pods，可提供大約 11,500 兆次浮點運算能力。

## ARM/Softbank

全球 IC 矽智財領導廠商安謀國際 ( ARM ) 在 2017 年 5 月發表三款針對人工智慧 ( AI ) 應用的處理器架構，並已授權十家廠商，預計 2018 年推出相關應用晶片，搶攻 AI 商機。

AI 應用的處理器平台 ARMDYNAMIQ，包括二顆處理器 Cortex A75、Cortex A55，以及一顆繪圖處理器 Mali-G72，全都瞄準人工智慧新商機與高階行動運算。這也是安謀首次針對 AI 發展推出技術平台，及運算功能更強大的處理器，並且重新定位大小核的架構，可以針對不同應用，提供不同大小核設計架構。

由於機器學習及人工智慧 ( AI ) 是市場當紅話題，未來會成為智慧型手機及智慧電視的必備功能。2018 年 3 月安謀 ( ARM ) 再度宣布全新 Mali 多媒體 IP 套件，包含 Mali-G52 與 Mali-G31 繪圖處理器、Mali-D51 顯示處理器、Mali-V52 視訊處理器等矽智財，將高效能繪圖及顯示運算延伸至行動裝置與數位電視市場。

以 ARM 此次推出的高效能 Mali-G52 來說，能處理更複雜的圖像，讓受功耗與頻寬限制的主流行動裝置系統得以運行更多機器學習功能，並同時具備高出 30% 的效能密度，提升 15% 能源效率，減少裝置的耗電與發熱，在運行像 AR 這類極為耗電的遊戲時使用時間能撐得更久。機器學習效能比前一代產品高 3.6 倍，確保所有層級裝置都能支援次世代的機器學習使用情境。

Mali-G31 則是 ARM 首款採用創新 Bifrost 架構的超高效率繪圖處理器，是 ARM 旗下產品線中，支援 OpenGL ES 3.2 與最新世代 Vulkan API 的最小處理器，比前一代 Bifrost 產品 Mali-G51 體積小 20%，但效能密度卻高出 20%，不僅節省晶片面積，同時更提供絕佳能源效率。

## Apple

身為手機巨頭，蘋果首款人工智慧晶片是 A11 Bionic。蘋果 2017 年秋季新品發表會中，A11 Bionic 隨 iPhone X 一起亮相。A11 除了在 CPU、GPU、ISP 方面提升，還首次搭載機器學習的硬體專用「神經網路引擎」( Neural Engine )，採用雙核設計，每秒運算次數最高可達 6,000 億次，相當於 0.6TFlops，以幫助

加速人工智慧任務，即專門針對 Face ID、Animoji 和 AR 等應用程式的 ASIC。

## Qualcomm 高通

在 2018 年世界行動通訊大會( MWC )上，晶片大廠高通宣布人工智慧 ( AI ) 引擎 ( Qualcomm Artificial Intelligence Engine ) 啟動，打造終端裝置使用者之 AI 體驗。這款 AI 引擎將支援驍龍 Snapdragon 高階 8 系列 845、835、820，甚至以及中階 660 行動平台，打造終端裝置 AI 使用體驗。

高通 AI 引擎的主要軟體核心元件包括：神經處理引擎 ( NPE ) 軟體架構，支援首次公佈於 Google 的 Android Oreo 作業系統的 Android 神經網路 API，以及 Hexagon 神經網路程式庫大幅加速 AI 應用，如卷積、池化與活化等功能。

在 AI 引擎支援下，Snapdragon 核心硬體架構高通 Hexagon™ Vector 處理器與高通 Adren GPU 與高通 Kryo CPU 能夠快速且有效於終端裝置執行 AI 應用程式，此異質運算方式提供開發人員及 OEM 廠商在智慧型手機及其他終端裝置的 AI 使用者經驗上進行優化。

## Broadcom 博通

網通晶片大廠博通 ( Broadcom ) 推出已獲矽認證 ( silicon-proven ) 的 7 奈米矽智財 ( IP )，將以特殊應用晶片 ( ASIC ) 搶攻當紅的人工智慧 ( AI )、5G 及高寬頻網路等市場。

博通為客戶打造的 7 奈米 ASIC 去年底完成設計定案 ( tape-out )，博通也說明將把 7 奈米 ASIC 晶圓代工及 CoWoS 封裝訂單交由台積電負責。

博通的 7 奈米 ASIC 除了鎖定深度學習等 AI 應用，以及搶進高效能運算 ( HPC ) 晶片市場外，也將採用台積電 CoWoS 先進封裝技術將 HBM Gen2/Gen3 等記憶體直接整合在晶片當中。至於高速 SerDes 矽智財則有助於打造

高速交換器或路由器應用晶片，以及完成支援 5G 高速及寬頻網路的系統單晶片。博通表示，為主要前期客戶打造的 7 奈米 ASIC 在去年底已完成設計定案。

## AI 晶片新創公司崛起

除了榜單內提到的企業，Compass Intelligence 還在報告中提到了值得關注的 AI 晶片創業公司。



另外，根據 Github 上一張列有全球 AI 晶片廠商的資訊圖，包括大公司的產品，也包括新創公司，而新創家數則是不斷增加中。



當市場將目光放在 NVIDIA 押寶圖形處理器(GPU)在人工智慧(AI)上的應用潛能之際，新創公司試圖在優化推論和機器訓練方面，打造速度更快、更節能，更適合搭載在下一代人工智慧裝置的晶片。圖形處理器不再是傳統計算架構，而是成為處理人工智慧所需的快速運算晶片，而新創公司認為他們可以做得更好，即使尚未出貨，募資金額仍不斷增加。

過去幾年來，深度神經網路的興起，讓創業投資(VC)的資金重新回到了半導體領域。根據 CB Insights 調查，2017 年風險投資者在晶

片新創公司的投資額超過 15 億美元，幾乎是 2 年前投資金額的 2 倍。

位於加州的 Graphcore、Cerebras、Wave Computing 以及 2 家位於北京的新創地平線機器人 (Horizon Robotics) 和寒武紀科技 (Cambricon) 都募得超過 1 億美元資金。

已秘密開發晶片超過 10 年的加州 KnuEdge 累計募資超過 4,700 萬美元。美國德州的 Mythic 累計獲投超過 5,500 萬美元。另一家北京新創公司深鑿科技 (DeePhi) 已募得 4 千萬美元。來自上海燿知電子科技有限公司 (Think Force)，也獲得將近 7,000 萬美元投資。

### 值得矚目的 AI 晶片新創

#### Graphcore



- 2016 年成立
- 總部：英國布里斯托
- 技術：智慧處理晶片 (IPU)
- 產品：機器學習加速器，比現有最快的系統快 10-100 倍，將用於無人駕駛汽車和雲端運算，預計在 2018 年量產銷售。
- 獲投 1.1 億美元 (A,B,C 輪)
- 投資方：三星、Bosch、Dell 等公司創投、DeepMind 創辦人
- 入選 CB Insights 2018 AI 100 排行榜。

#### Wave Computing

- 成立於 2010.12
- 總部：美國加州

- 核心技術：機器學習數據流處理架構



DPU (Dataflow Processing Unit)，宣稱 DPU 加速神經網路訓練上能夠超越 GPU 1000x

- 最新動向：Wave Computing® 在 2018 年 5 月宣佈與包括中國百度、谷歌、史丹福大學、哈佛大學、加州大學柏克萊分校、明尼蘇達大學、多倫多大學在內的多家企業和大學聯合發佈了一套用於測量和提高機器學習軟硬體性能的國際基準 MLPerf。
- 獲投 1.17 億美元 (A,B,C,D 輪)

#### 寒武紀



- 2016 年成立
- 總部：中國北京
- 前身：中國科學院計算機研究所衍生公司
- 產品：2017 年推出商用深度學習專用的 Cambricon-1A 處理器，應用於手機、安防設備、可穿戴設備等終端。2018 年 5 月再發布第三代 IP 產品 Cambricon 1M 和最新一代雲端 AI 晶片 MLU100 和板卡產品。MLU100 採用寒武紀最新的 MLUv01 架構和 TSMC 16nm 的先進工藝，平衡模式下的等效理論峰值速度達每秒 128 萬億次定點運算，高性能模式下的等效理論峰值速度更可達每秒 166.4 萬億次定點運算。
- 獲投 1.0 億美元 (A 輪)

- 投資方：包括阿里巴巴創投、聯想創投、國科投資等
- 入選 CB Insights 2018 AI 100 排行榜。

### 地平線機器人



**Horizon Robotics**  
地平线机器人技术

- 成立於 2015 年·由前百度深度學習研究院) 創始人余凱創辦(地平線機器人)
- 總部：中國北京
- 產品：深度神經網絡處理晶片 NPU (Neural Processing Unit) · 2017 年底推出「征程 1.0 處理器」(自駕車)和「旭日 1.0 處理器」(智慧攝影·應用智慧城市與商業)
- 獲投 1 億美元 (A+輪)
- 投資方：Intel Capital 等

### Cerebras



- 2016 年成立
- 總部：加州洛杉磯
- 產品：深度學習專用的 ASIC 晶片
- 獲投獲投 1.12 億美元(A,B,C 輪)
- 入選 CB Insights 2018 AI 100

### 燭知電子科技有限公司 (Think Force)



- 成立於 2017

- 總部：中國上海
- 核心技術：深度學習智慧晶片
- 產品：智能晶片、人工智慧硬體平台·提供一站式行業應用解決方案。
- 獲投 4.5 億人民幣(約 7,000 萬美元) (A 輪)
- 投資方：紅杉資本等

### Mythic



- 2012.5 成立
- 總部：美國德州奧斯丁
- 產品：深度學習的神經網絡晶片
- 獲投 5,520 萬美元(SEED,Debt Financing, A,B)
- 投資方：Lux Capital, Softbank venture 等

### KnuEdge



- 2005 年成立
- 總部：美國加州聖地亞哥
- 產品：KnuPath 神經網路晶片
- 獲投 4,700 萬美元
- 投資方：Hasso Plattner Ventures 等

### 深鑒科技 DeePhi Tech



- 2016 年成立

- 總部：中國北京
- 方案：神經網路深度壓縮技術和 DPU 平台的解決方案。
- 晶片：「聽濤」系列 SoC ( 2018Q3 )，聚焦安防領域
- 獲投 4 千萬美元(A 輪)
- 投資方：Xilinx、聯發科、清華控股等公司與 VC

### 追趕美國！中國即將宣布 3,000 億人民幣半導體基金

中國近年來積極發展自己的半導體產業，以降低對外國科技的依賴，這方面的努力在美國政府以國家安全為由，反對大陸企業購併美國晶片公司後，更顯得緊迫。

中國 2014 年成立的晶片發展基金，已成為美國近年對大陸科技政策抱怨的重點。美國貿易代表署 3 月 22 日公布的貿易行為報告中指出，大陸 2014 年的基金由政府部門和國營企業掌控，顯示「中國政府高度介入成立基金，以達到國家戰略目標」。

中國在經歷美國因中興通訊違約而祭出禁止美國公司向中興供應芯片 7 年的制裁後，在 5 月 4 日結束的美中貿易談判，觸及半導體產業，中方要求美國放寬對晶片的出口限制，並調整對中興通訊的出口禁令。

根據華爾街日報引述熟知內情人士說法，中國大陸「很快」就會宣布成立 3,000 億人民幣的新基金，用於提升大陸設計和製造先進微處理器和圖形處理器，以縮小與美國和其他競爭對手的科技差距，此舉勢必會使美中貿易緊張關係再升高。

人工智慧晶片可能讓原來的晶片市場重新洗牌，也是主要科技大國必爭的領域。過去晶片一直是該領域大廠所掌控，但 AI 晶片新革命中，也看到 IT 科技巨頭跨入，研發自己

所需的處理器外，新創業跨入這場戰場中，究竟誰能勝出，後勢值得關注。

---

### 參考來源

Compass Intelligence、NVIDIA、智東西、科技新報、科技產業資訊室、DigiTime、工商時報、經濟日報

網址：

1. <https://goo.gl/WsngWc>
2. <https://goo.gl/GBv54N>
3. <https://goo.gl/CXrQGD>
4. <https://goo.gl/5geF3D>
5. <https://goo.gl/UUjPMu>
6. <https://goo.gl/gKBen1>
7. <https://goo.gl/u2QUjo>