

## 技術詞彙表

本技術詞彙表包含本文件中使用的若干技術詞彙的解釋。因此，該等詞彙及其涵義可能與標準的業界涵義或用法不同。

|            |   |                                                                                                                                                      |
|------------|---|------------------------------------------------------------------------------------------------------------------------------------------------------|
| 「AI」       | 指 | 人工智能，為計算機科學的一個領域，專注於通過機器模擬人類智能                                                                                                                       |
| 「算力積木」     | 指 | 專為高效執行人工智能任務所需核心數學運算而設計的硬件單元。該等組塊針對矩陣乘法、卷積、注意力機制及其他在深度學習模型中佔主導地位的張量運算進行了優化                                                                           |
| 「AI推理」     | 指 | 使用預先訓練的模型(其參數已自訓練得出)執行預定任務的過程，透過其學習到的模式或知識處理新的未知輸入數據，並產生輸出數據，例如預測、分類或決定                                                                              |
| 「AI推理芯片」   | 指 | 為執行具高算力效能、最低延遲及優化用電的經訓練人工智能模型而設計的專門集成電路，專注於從新的未知輸入數據中得出輸出數據，例如預測、分類或決定                                                                               |
| 「AI模型」     | 指 | 可將非結構數據作為輸入數據，並透過其「智力」(即感知世界、轉錄及組織資料、增強或生成內容或作出決定的能力)將其轉化為資料輸出的數學算法                                                                                  |
| 「AI訓練」     | 指 | 將機器學習模型暴露於標記或結構化的資料中，教導其辨識模式或進行預測的過程。在訓練過程中，模型會使用梯度下降等算法反覆調整其內部參數(例如權重、偏置)，以最小化其預測輸出與實際「正確」輸出之間的差異(損失)。這個階段需要大量的計算資源(例如GPU/TPU)和大型數據集，並產生一個經訓練作推理的模型 |
| 「算法」       | 指 | 基於執行一系列指定動作以解決問題的程序或公式，尤指通過電腦                                                                                                                        |
| 「算法芯片化」    | 指 | 一種協同設計方法，可直接將計算算法映射至定製芯片架構，針對特定工作負載進行硬件優化                                                                                                            |
| 「ASIP-LSF」 | 指 | 一種專用的處理器架構，將專用指令集處理器(ASIP)與優化後的載入存儲框架(LSF)相結合，旨在通過針對目標應用程序定製指令集和記憶體訪問模式來加速特定領域的工作負載                                                                  |

## 技術詞彙表

|                                                |   |                                                                                                                                                                                 |
|------------------------------------------------|---|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 「大數據」                                          | 指 | 超出傳統數據管理工具處理能力的極為龐大且複雜的數據集。大數據分析範疇的重點為檢查該等大型數據集，以發現可為各領域的決策及戰略規劃提供數據的模式、相關性及見解                                                                                                  |
| 「藍牙」                                           | 指 | 一種短距離無線技術標準，用於固定及移動設備之間的短距離數據交換                                                                                                                                                 |
| 「chip-to-chip mesh torus互連」或「C2C mesh torus互連」 | 指 | 一種用於多芯片系統的高性能、可擴展網絡拓撲結構，其中處理單元以環形或網格配置連接                                                                                                                                        |
| 「chiplet」                                      | 指 | 一種小型模組化集成電路，當中封裝了特定功能，設計用於與其他chiplet整合於單一封裝中，以形成完整的系統級芯片(SoC)，並可將多個chiplet整合到一個封裝中，從而擴展所得芯片的性能                                                                                  |
| 「雲」                                            | 指 | 一個由群集於互聯網／內聯網的遠程服務器組成的網絡，用於存儲、管理及處理數據，以取代本地服務器或個人電腦                                                                                                                             |
| 「雲計算」                                          | 指 | 將電腦數據及程序存儲在多台可通過互聯網存取的服務器的做法                                                                                                                                                    |
| 「集群」                                           | 指 | 一組互連的計算機或服務器，作為統一系統共同運作，以提高性能、可靠性或可擴展性                                                                                                                                          |
| 「CNN」                                          | 指 | 一種專為處理網格結構化數據而設計的深度學習架構，通過卷積操作自動學習分層空間特徵                                                                                                                                        |
| 「算力」                                           | 指 | 一台計算機執行一項操作的能力                                                                                                                                                                  |
| 「CPU」                                          | 指 | 中央處理器，為複雜的電子電路組件，用於執行機器的操作系統及應用程序，為通用計算的核心                                                                                                                                      |
| 「深度學習」                                         | 指 | 人工智能及機器學習的分支，模擬人類大腦等生物神經系統的運作，並利用多層神經網絡以在物體探測及識別、語音識別以及自然語言處理等作業中達致高精確度。有別於傳統機器學習技術，深度學習能對如圖像、視頻或文字等數據自動進行表徵學習，而無需引入手動編碼規則或人類領域的知識。其極具靈活彈性的架構使其能夠直接通過原始數據學習，如獲提供若干數據，更可提升其預測精準度 |
| 「晶粒」                                           | 指 | 從晶片上切割下來未進行封裝的單顆芯片                                                                                                                                                              |

## 技術詞彙表

|                                    |   |                                                                                                                    |
|------------------------------------|---|--------------------------------------------------------------------------------------------------------------------|
| 「die-to-die chiplet」或「D2D chiplet」 | 指 | 一種高帶寬、低延遲的通信技術，可使多個chiplet作為單一集成系統運作                                                                               |
| 「邊緣」                               | 指 | 一個分佈式計算範式，其中數據透過接近數據來源的閘道匯集及由本機電腦系統處理，以提供即時服務。由於不需要雲端傳輸，此架構可滿足重要的產業需求，包括即時作業、應用程式智能化以及保障安全性／私隱。此架構一般部署於終端裝置及雲端平台之間 |
| 「邊緣計算」                             | 指 | 使計算及數據存儲更接近需要提高反應時間及節約帶寬的位置的分佈式計算模式                                                                                |
| 「GPU」                              | 指 | 圖形處理器，一種專門的電子電路，旨在快速操控及改變記憶，以加快圖像的創建                                                                               |
| 「自主可控」                             | 指 | 完全在組織內部利用自身資源、專業知識和基礎設施開發、構建或創造                                                                                    |
| 「集成電路」                             | 指 | 於半導體基板上裝配的電子電路                                                                                                     |
| 「IoT」                              | 指 | 物聯網，指嵌入電子設備、軟件、傳感器及網絡連接的物理裝置網絡，使該等物件能夠收集及交換數據，從而使該等裝置能夠互相通訊及與用戶通訊                                                  |
| 「ISA」                              | 指 | 指令集架構，一套標準化的指令、操作和規則，用於定義處理器與軟件的互動方式。它作為軟硬件接口，管理程序如何控制處理器、存取內存及執行運算                                                |
| 「KOL」                              | 指 | 關鍵意見領袖                                                                                                             |
| 「LLM」                              | 指 | 基於深度學習架構的先進人工智能模型，通過使用海量文本數據集訓練以理解、生成和處理類似人類的語言                                                                    |
| 「LSTM」                             | 指 | 一種專門用於建模序列數據中長期依賴問題的循環神經網絡(RNN)                                                                                    |
| 「近存超融架構」                           | 指 | 一種先進的運算範例，通過將運算單元與高帶寬存儲系統物理整合，從根本上優化數據處理效率                                                                         |
| 「NPU」                              | 指 | 神經網絡處理器，一種專門通過預測AI模型、生成式AI模型及推理AI模型操作加速機器學習算法的微處理器                                                                 |
| 「算子」                               | 指 | 用於對原始數據進行預處理的預建函數                                                                                                  |

## 技術詞彙表

|         |   |                                                                                                       |
|---------|---|-------------------------------------------------------------------------------------------------------|
| 「研發」    | 指 | 研究與開發                                                                                                 |
| 「RNN」   | 指 | 一種用於處理序列數據的神經網絡，通過隱藏狀態保持對先前輸入的「記憶」                                                                    |
| 「系統級芯片」 | 指 | 於單一芯片內整合功能上多樣化的子模組的集成電路，形成專為特定應用場景定製的完整系統。系統級芯片通常整合多種異質元件，並將通用處理器、硬體式編碼單元、基頻處理器及其他專用功能區塊等元件整合至統一的硅芯片中 |
| 「流片」    | 指 | 集成電路設計於量產(即設計被送至代工廠裝配於硅晶圓時)前的最後階段                                                                     |
| 「TOPS」  | 指 | 每秒萬億次運算，為計算單位，代表處理器每秒可進行一萬億( $10^{12}$ )次操作，常用於量化AI加速器、NPU或其他深度學習硬件的推理性能                              |
| 「終端」    | 指 | 進行本地化數據處理、提供實時回應及與用戶或環境實體互動，同時將複雜工作自邊緣節點或雲端平台卸載的硬件裝置                                                  |