



AI 驅動的電信業：保障行業未來與盈利能力

利用 AI 推動創新、重新定義業務並提升電信行業的收益。

目錄

簡介	1
電信業 AI 的主要應用案例	4
電信業的 AI：提升營運效率的主要應用案例	4
電信領域的 AI：通過新服務和擴充服務推動增長的主要應用案例	5
應對複雜性：AI 採用與實施挑戰	5
Supermicro 與 NVIDIA – 幫助電信公司採用人工智慧來優化營運並釋放策略成長	6
提升營運效率的主要應用案例	7
1. 重新構思聯絡中心以提升客戶體驗	7
2. 簡化網路營運	8
3. 業務洞察的預測分析	8
電信領域的 AI：通過新服務和擴充服務推動增長的關鍵應用案例	9
1. 主權 AI 工廠的 AI 基礎設施	9
2. 透過邊緣 AI 將 5G 營利化及 6G 研究	10
Supermicro 與 NVIDIA：一系列解決方案	11
向前邁進	13
欲了解更多資訊	13

簡介

AI：點燃電信業的下一次進化

電信行業的各組織（電信公司）正處於關鍵的轉折點。電信公司在塑造全球數位生態系統、推動經濟增長和連接全球人們方面發揮著關鍵作用。作為有潛力數位化整個國家並與各行各業合作的行業，電信公司擁有獨特的能力來賦能其他領域並推動廣泛的數位轉型。通過利用尖端的 AI 技術，電信公司可以提升營運效率、提供卓越的客戶體驗，並在日益動態的市場中保持競爭力。

“電信行業正站在 AI、連接性、安全性和可持續性數位化世界的交叉路口。它推動了 AI 革命，並 [肩負著保護關鍵基礎設施和保障我們世界的責任。](#)”¹

¹ [“Why telecommunications is a lynchpin between cybersecurity and AI for good”](#), World Economic Forum, 2024



未來，電信公司的成功法則將是利用新的預測式和生成式 AI 在每個階段提取價值。根據 BCG 的說法，“[最具創新性的電信公司已經開始利用生成式AI的力量](#)。”² 透過利用網路、客戶和營運產生的大量資料，電信公司可以使用先進的分析和 AI 來提取有價值的洞察、提升網路效能、改善客戶體驗並優化營運效率。

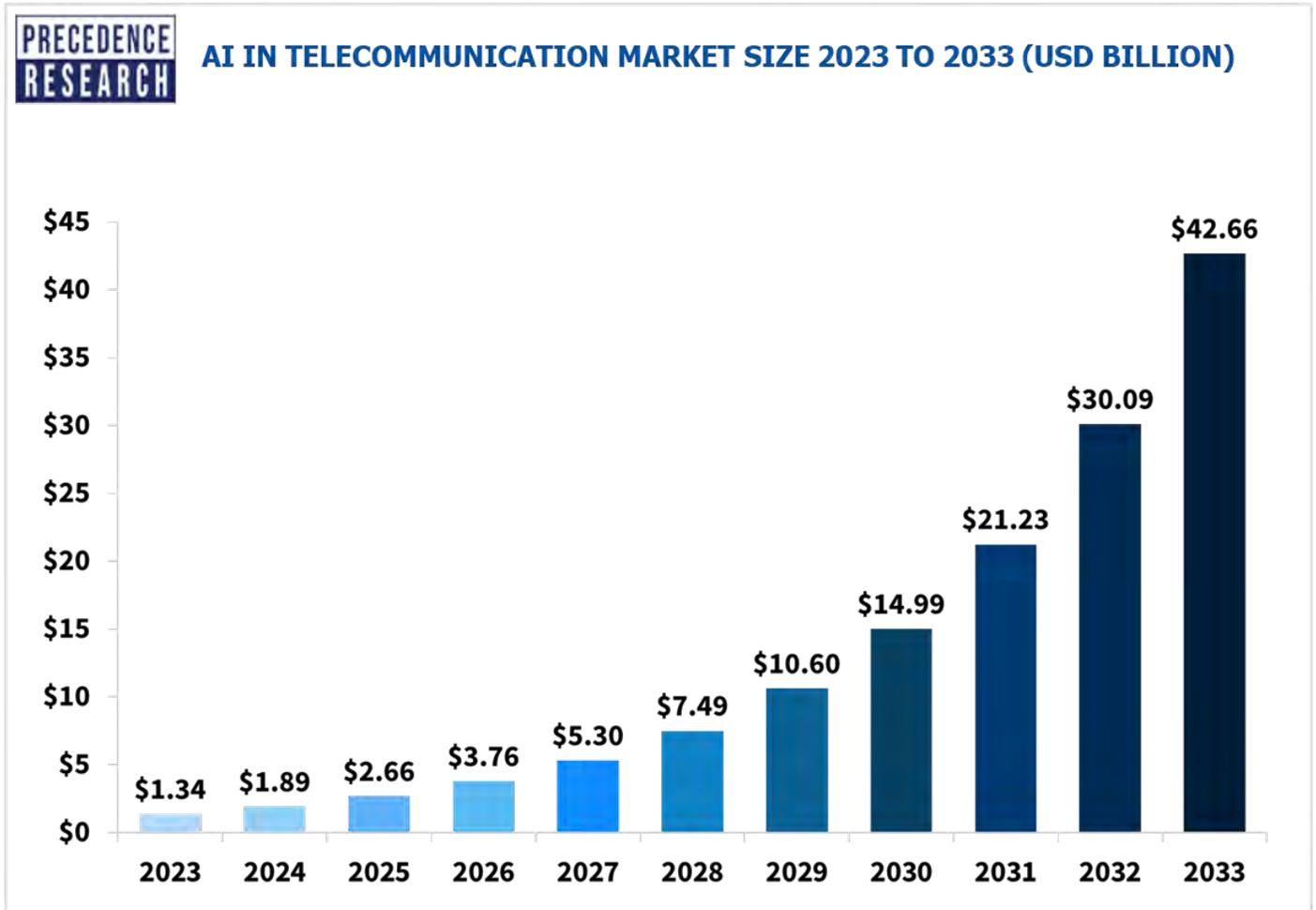


圖 1- [2023至2033年電信業的 AI 的市場規模 \(十億美元\)](#)³

改變數位經濟：AI 如何推動電信 B2B2X 合作夥伴關係的創新

成功建立自身 AI 基礎設施並有效利用 AI 的電信公司，處於理想的位置，可以向他們已經服務的客戶提供 AI 服務和基礎設施，從而開啟新的收入來源，並成為區域 AI 供應商。他們可以向醫療保健、金融、製造和零售等行業提供關鍵基礎設施和資料驅動的洞察，促進協生生態系統的發展。作為 B2B2X 合作夥伴，電信公司不僅推動數位轉型，還可以定位自己為跨行業增長和發展的重要推動者。

² [New Formula for Success | 2024 Telco Value Creators | BCG](#)

³ [Precedence Research, "AI in Telecommunication Market Size, Share, and Trends 2024 to 2033," 2023](#)

這些服務的一些例子包括具備視線檢測功能的高端視訊會議服務、提供即時翻譯的 3D 音頻客服聊天機器人，以及提升網路安全或支援網路邊緣基礎設施即服務的解決方案。

電信營運商在擴展資料中心方面具有得天獨厚的優勢，能夠通過加速運算基礎設施來支持國家，提供主權 AI 平台，促進國家和經濟發展。這種新型資料中心，被稱為 AI 工廠，能夠提升地方政府、企業和初創公司的 AI 能力。電信領導者需要繼續投資並創新他們的 AI 產品，構建其能力，最終目標是成為 AI 工廠的領導者。每一項新的投資都是對先前投資的延續，並代表著潛在的盈利新產品，使電信公司能夠參與行業整體增長。

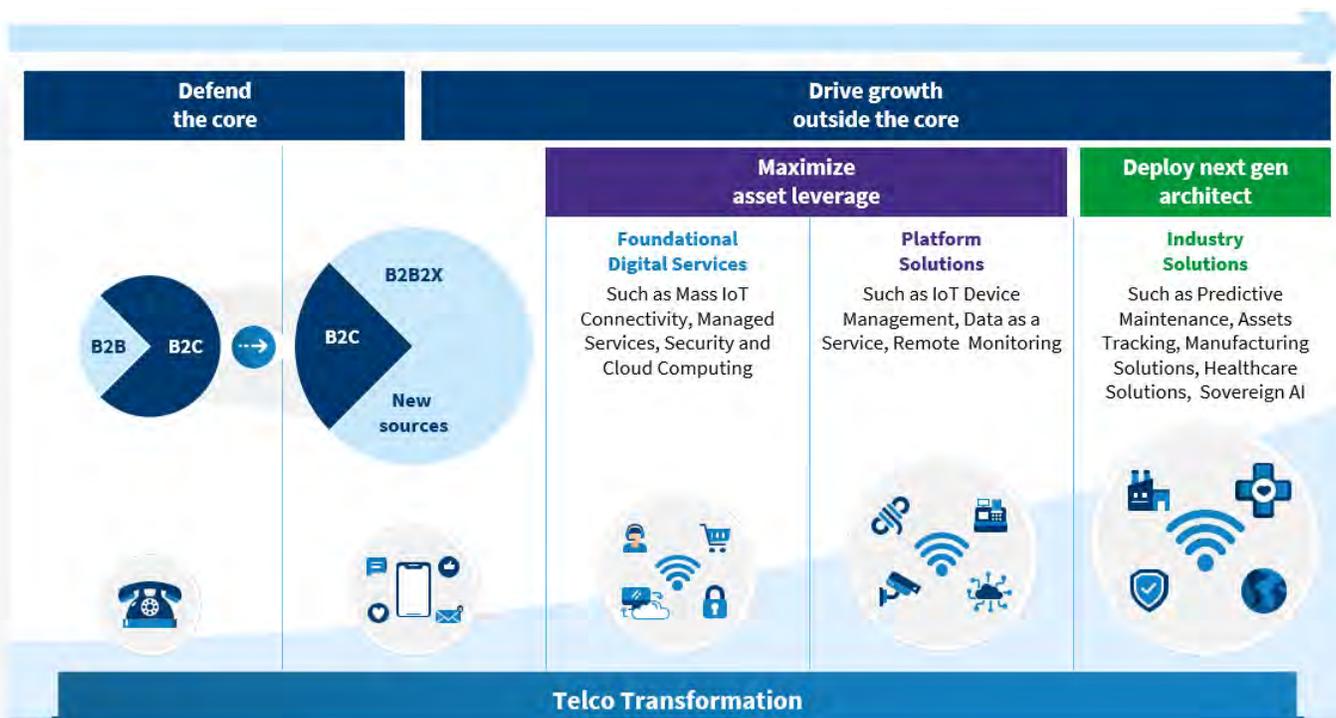


圖 2- 持續投資推動電信轉型

電信業中最常見的 AI 趨勢

1. 電信公司迅速採用 AI

根據 NVIDIA 最近的一項調查，[90% 的電信高管](#)⁴ 表示，他們的組織目前正從實驗階段到全面部署階段參與 AI。該行業的 AI 領導者已經在客戶滿意度提升和成本降低方面取得了顯著的成效。根據麥肯錫的調查，表現最好的五分之一公司在過去五年的收入複合年增長率是同行的 2.1 倍，並且[總股東回報率是同行的 2.5 倍](#)⁵

⁴ [State of AI in Telecommunications. NVIDIA, 2024.](#)

⁵ [The AI-native telco: Radical transformation to thrive in turbulent times. McKinsey, 2023.](#)

自 2020 年以來，使用電信網路來連接員工、改善醫療、通訊和基本服務的組織激增，這一切都得益於高速連接。由 AI 驅動的機遇正在為該行業提供急需的動力，推動行業發展。

2. 客戶體驗仍然是 AI 投資的關鍵驅動力

電信公司正在利用預測式和生成式 AI 來滿足各種業務需求，而最大 AI 機會在於提升客戶體驗。電信公司通過虛擬助理、豐富的體驗和推薦來改善客戶互動，同時利用預測性資料分析來提升服務體驗。此外，AI 驅動的客服工具能高效地處理日常查詢，縮短等待時間，確保迅速解決問題。對於 B2B 客戶，電信公司可以通過超越傳統連接的平台，成為企業客戶構建 AI 基礎設施的重要合作夥伴。

3. 電信公司正在利用 AI 增強網路規劃和營運

AI 正在革新電信公司管理其網路的方式。AI 使電信公司能以創新且強大的方式與網路互動，例如回應關鍵問題、在指定時間內識別事件並推薦有效解決方案。這能透過快速解決問題來改善客戶的服務體驗。例如，預測式 AI 解決方案可以在網路問題影響客戶之前檢測並自動修復問題，從而有效提前處理服務查詢並降低客戶流失率。

這些只是 AI 開始展現巨大潛力的眾多策略領域中的一部分。該技術的使用情況可以通過一些被多家大型電信組織廣泛採用的常見案例來最好地描述。在這份白皮書中，我們將重點介紹五個最常見的應用案例，這些應用案例最有可能對電信行業產生重大影響。

電信行業中的 AI 關鍵應用案例

AI 擁有改變組織內部營運和創造全新產品組合的潛力，關鍵應用案例涵蓋了這兩個領域。

電信行業中的 AI：提升營運效率的關鍵應用案例

1. 重新構思聯絡中心以提升客戶體驗

AI 驅動的聊天機器人可以改變聯絡中心的運作方式，以提升客戶體驗。聊天機器人處理第一線查詢並向人工代表提供資料驅動的解決方案，從而縮短通話時間、提高客戶滿意度並增加生產力。

2. 簡化的網路營運

電信公司可以利用 AI 來簡化昂貴的網路維護工作。使用 AI 可以節省營運團隊的時間，提高效率，同時縮短故障排除的解決時間並提升效能。電信公司可以建立自我修復的網路，能夠在問題影響客戶之前識別並自動修復問題，無需人工干預，並通過使用 AI 來優化頻譜分配和使用，以提升現有客戶的服務質量。

3. 預測式分析以獲取業務洞察

利用 AI 獲得對客戶使用模式和生活方式的更深入洞察，從而在核心產品之上提供量身定制的高端產品。提升客戶互動和滿意度，同時改善利潤率，提供超越連接的服務。通過為其他企業提供有關其自身客戶行為和偏好的寶貴洞察，擴展預測式分析能力至 B2B 客戶。

電信行業中的 AI：推動增長的新服務和擴充服務的關鍵應用案例

1. AI 工廠：國家主權 AI 基礎設施

AI 工廠正在改變主權 AI 基礎設施的開發和部署。AI 工廠使電信公司能夠創建符合資料主權和隱私規範的量身定制的 AI 解決方案。這種方法確保對資料和 AI 模型的完全控制，幫助促進創新，同時滿足國家利益和法規合規要求。實施 AI 工廠以支持主權 AI 基礎設施標誌著向更智慧、更安全、更高效的電信未來邁出了重要一步。

2. 透過邊緣 AI 將 5G 營利化及 6G 研究

電信公司可以利用 AI 生態系統來提供創新的 5G 應用，包括用於自主購物的智慧視訊分析、減少城市擁堵的智慧交通監控，以及沉浸式 5G 擴增實境 (XR) 體驗。此外，邊緣運算在 6G 網路中可以發揮重要作用，為該行業創造顯著的增長機會，例如即時資料分析和在網路邊緣部署 AI 時的智慧決策。電信公司可以利用 AI 來加速 6G 研究，通過優化網路設計、增強信號處理和實現即時資料分析來推動進展。此外，AI 正在提升創新應用的開發，如智慧邊緣運算和沉浸式擴增實境，推動下一代連接的邊界。AI 使電信公司能夠超越研究，通過邊緣的 5G 營利化創造當前的收入機會。

應對複雜性：AI 採用與實施挑戰

與許多行業一樣，完全擁抱 AI 和 AI 驅動的策略在電信行業中仍然具有挑戰性。電信組織在嘗試實施預測式和生成式 AI 解決方案時面臨的一些常見障礙包括：

資料隱私與合規性：確保資料隱私和法規合規性對於處理敏感客戶訊息，並在不同司法管轄區營運的電信公司來說至關重要，有時也是具有挑戰性的。

- 電信資料通常包含有關客戶的敏感資訊，包括個人識別碼、通信模式和位置資料，如果處理不當，可能會危害客戶隱私。
- 在多個地區營運的電信公司難以建立一個穩健且可稽核的 AI 系統，以滿足其產品使用所在每個司法管轄區的資料隱私規範。
- 電信服務供應商在利用 AI 進行資料分析和洞察時，面臨建立穩健安全措施和合規框架的挑戰。

電信公司必須為其客戶創建可靠、安全的 AI 基礎設施，以確保資料主權和合規性。各個司法管轄區的法規各不相同，需要一個安全的平台來建構 AI 解決方案，為電信公司提供值得信賴的基礎。

勞動力挑戰：在 NVIDIA 2023 年的一項調查中，電信公司高層認為，尋找具備必要 [技能的合適熟練勞動力是 AI \(34%\) 面臨的首要挑戰，具體來說是生成式 AI \(55%\) 的首要挑戰](#)。⁶

- 如果沒有精通 AI 的熟練員工，組織將難以成功部署 AI 解決方案並實現其價值。從大型科技公司到小型企業，各種組織都在競相僱用經驗豐富的 AI 開發人員。專業人員必須熟悉現有的基礎設施，才能成功將 AI 解決方案整合到其營運中。
- 員工經常抵制採用需要顯著提升技能和靈活適應快速變化的系統和流程的新技術。

⁶[State of AI in Telco 2024 Report. NVIDIA, 2024.](#)

組織需要具備 AI 就緒的基礎設施，該基礎設施易於部署，能夠良好地融入現有工作流程，並且使用 IT 團隊熟悉的工具來操作和管理。

可擴展性和效能: 電信公司在努力管理海量資料並滿足高客戶期望的同時，保持基礎設施效能的持續擴展可能會帶來相當大的挑戰。

- 隨著客戶需求的增長，電信組織在處理海量資料並通過廣泛的資料中心和網路基礎設施同時支援數百萬用戶方面面臨困難。
- 客戶期望獲得高效能和良好支援，因此任何效能下降或服務中斷都可能對他們的體驗產生負面影響，導致不滿、潛在的客戶流失，以及服務供應商的聲譽受損。
- 高延遲和回應緩慢可能會延遲 AI 回應，導致錯失銷售機會、客戶沮喪以及營運效率低下。

電信環境中的 AI 解決方案必須具備高度的可擴展性和效能，特別是在使用的高峰期。為了快速創新並滿足客戶期望，組織需要專門構建的高效能 GPU 伺服器支援的基礎設施，配備高效能網路，能夠快速且安全地擴展和部署 AI 解決方案。

舊有系統和基礎設施: 過時的技術和分散的資料礙了有效的 AI 部署和可持續增長。

- 舊有系統通常與現代 AI 技術不相容，使整合變得複雜且昂貴。相容性問題可能會減緩 AI 解決方案的部署速度並限制其效能。
- 舊有基礎設施可能會創造出分散的資料環境，導致資料孤島。這使得綜合和全面分析資料變得困難，從而阻礙 AI 生成準確洞察和預測的能力。
- 舊有系統可能無法處理 AI 應用程式對資料處理和儲存的增長需求，限制了 AI 計劃的可擴展性，並阻礙了電信公司充分發揮 AI 能力。

電信公司需要現代化的基礎設施，以簡化孤立資料系統的整合，實現可持續增長。

Supermicro 與 NVIDIA – 幫助電信公司採用 AI 來優化營運並釋放策略成長

Supermicro 基礎設施配合 NVIDIA AI Enterprise 解決方案，為行業提供了針對電信優化的 AI 解決方案，幫助組織改善客戶體驗、簡化複雜的網路營運、通過資料科學提取業務洞察，同時開啟積極影響業績的新途徑。

以下是 Supermicro 與 NVIDIA 為您的組織提供支持的電信行業 AI 關鍵應用案例，以及一些最近已經採用和實施 AI 解決方案的客戶。

提高營運效率的關鍵應用案例

1. 重新構想聯絡中心以增強顧客體驗

生成式 AI 聊天機器人是客戶服務的第一線互動，回答有關服務中斷和帳單變更的問題，並讓人工代表處理更緊急和複雜的查詢。聊天機器人簡化了客戶體驗，根據客戶服務軟體公司 TIDIO 的報告，[62% 的受訪客戶表示他們更願意使用客服機器人，而不是等待人工代表](#)。⁷

一旦請求升級到聯絡中心的人工代表，AI 驅動的解決方案會回顧之前的客戶通話，完成情感分析，並為人工代表創建一份解決方案清單以供考慮。這可以幫助聯絡中心的人工代表獲得有關客戶流失的可能性，和成功解決先前通話問題的洞察，例如臨時降費等。聯絡中心人工代表可以從 AI 驅動的全通路策略中獲得來自之前客戶互動的額外背景訊息，不再只是跟進客戶投訴。

智慧虛擬助理 AI 工作流程

智慧虛擬助理 AI 工作流程加速了端到端智慧虛擬助理解決方案的建置與部署。此工作流程包含使用 NVIDIA NeMo™ 進行 ASR 和 TTS 訓練，以及使用基於 NVIDIA Triton™ 推論伺服器建置的 NVIDIA Riva 進行推論。

為其他行業提供動力

電信公司還可以將 AI 驅動的聯絡中心解決方案作為產品或服務提供給其他行業，幫助企業提升其客戶服務營運。通過提供跨越多個接觸點（包括實體店、網站和行動設備）的整合客戶服務體驗，電信公司使企業能夠採用全通路策略，彌合線上和離線互動之間的差距。這不僅提高了客戶滿意度，還增加了營運效率，使 AI 驅動的聯絡中心解決方案成為對 B2B 客戶具有吸引力的產品。

Supermicro + NVIDIA

Supermicro 的 GPU 優化系統與 NVIDIA 的 AI Enterprise 解決方案相結合，幫助電信組織創建智慧聊天機器人、輔助工具和虛擬助理。這些工具可以存取並使用重要資料，以提高效率並提供競爭優勢。這是通過一種叫做檢索增強生成（RAG）的方法實現的，它將大型語言模型（LLMs）與公司的知識庫連接，解鎖了許多新的可能性。

- 在搭載 NVIDIA H100 NVL GPU 的 Supermicro [SYS-221HE-FTNR\(D\)](#) 上實現 2335 LLM 吞吐量（每 GPU 每秒處理的總 token 數），或者要獲得更高的吞吐量，可以使用搭載 NVIDIA GH200 的 [Supermicro's ARS-111GLNHR](#)。
- NVIDIA NeMo™ 和 NVIDIA NIM™ 有助於加速企業級生成式 AI 模型的開發和部署。
- NVIDIA NeMo™ 是一個端到端的雲原生框架，用於策劃資料、訓練和客製化基礎模型，以及大規模運行推論。它支持文本到文本、文本到影像、文本到 3D 模型，以及影像到影像生成，加速了企業級生成式 AI 模型的開發和部署。
- NVIDIA NIM™ 透過將模型作為微服務提供，促進了生成式 AI 模型的可擴展和高效部署，使應用程式可以透過 API 調用請求推論。
- NVIDIA Merlin™ 是一個開源推薦系統，基於機器學習和深度學習算法，能夠加速從資料處理和訓練到部署 GPU 加速推薦系統的整個流程。

⁷ [“The Future of Chatbots...” TIDIO, 2024.](#)

- 對於最先進的對話 AI，NVIDIA® Riva 提供了加速資料管道、開發高品質虛擬助理應用的工具、即時轉錄和聊天機器人。

2. 簡化網路營運

管理網路可能是電信組織營運部門中最耗時和勞力的工作之一。僅網路維護的現場和服務營運方面，就佔了大多數電信組織營運預算的 60% 到 70%⁸。維護團隊不僅需要不斷更新和修復老化的基礎設施，還需要安裝新的升級，當網路因如天氣事件等不可預見的情況受損時，他們還必須進行緊急修復。

AI、物聯網 (IoT) 和邊緣加速運算正在提升網路效能和效率。AI 被證明是電信公司的強大夥伴，通過幫助識別潛在的網路故障、推薦解決方案、優化現場服務，以及持續監控異常並自動化回應流程，以實現最小的停機時間，應對營運中的障礙。借助 AI 驅動的系統，組織可以實現網路的自我配置、自我優化和自我修復，無需人工干預，同時優化資源管理。

在過去幾年中，電信公司一直致力於建立 5G 基礎設施和擴展覆蓋範圍。5G 操作對於維持高品質和穩定的服務至關重要，而 AI 可以在幾個關鍵領域提供幫助。AI 可以提升動態網路管理，幫助創建自我優化的網路。這些網路可以根據即時條件進行調整，優化效能，確保即使在高峰使用期間也能提供高品質的服務。AI 執行的預測分析可以預測網路資源需求，從而更好地規劃和分配資源，防止擁堵並確保最佳效能。通過智慧路由，AI 可以優先處理特定應用程式和服務的路由，以減少擁堵並提升整體服務品質。

Supermicro + NVIDIA

Supermicro 的 AI 就緒電信基礎設施與 NVIDIA 的 AI Enterprise 解決方案相結合，能夠幫助電信公司創建 AI 驅動的解決方案，構建軟體定義和加速的 5G 基礎設施，並將連接智能帶到邊緣的智能設備上。Supermicro 的 SYS-221HE-FTNR(D) IOT SuperServer 與 NVIDIA 的 H100 NVL 相結合，支持電信微資料中心和網路功能虛擬化，而 Supermicro 的 GPU SuperServers ARS-111GL-NHR、SYS-421GE-TNHR2-LCC 和 SYS-821GE-TNHR 則支援高效能運算、生成式 AI 以及 AI 訓練和推論。

Supermicro 的高效能 AI 開發工作站可以利用強大、專用且安全的平台，極大地提升您的生成式 AI 開發過程，幫助您創建和訓練模型。使用 NVIDIA NeMo™ 構建一個全面的生成式 AI 系統，配備大型語言模型 (LLMs)、多模態、視覺和語音 AI，並通過 NVIDIA TensorRT™ 的即時批次處理來實現低延遲和高效能的推論，使生成式 AI 解決方案的部署更加簡單。

3. 預測分析以獲取業務洞察

在競爭激烈的電信行業中，任何額外的優勢來吸引和保留客戶都能提升利潤率，並為組織提供重要的競爭優勢。利用先進的分析和機器學習，電信公司可以提取有價值的洞察，以改善網路效能、客戶體驗和營運效率。僅通過先進分析，電信公司就可以減少雲端成本，用於如客戶流失預測、網路設備預測維護、進階安全、詐欺檢測等場景。前瞻性的電信領導者正在利用這一機會提供新服務。在 B2C 情境中，這可以轉化為對電信客戶的下一階段洞察，以識別追加銷售機會、建立品牌忠誠度和提供升級的服務。

⁸ [How AI is helping revolutionize telco service operations | McKinsey](#)

為其他行業提供動力

在 B2B 環境中，電信公司可以利用資料幫助其他企業優化其服務。例如，他們可以預測需求激增或停機時間，讓營運商有足夠的時間和資源迅速而有效地解決問題。電信公司可以將預測分析作為服務提供，讓其他行業利用 AI 的力量來改善自身營運。通過將這些複雜的工具和洞察擴展到 B2B 客戶，電信公司不僅能夠多元化收入來源，還能強化其作為創新領導者的地位，推動各個行業的商業成功。

Supernano + NVIDIA

Supernano 的高效能基礎設施與 NVIDIA 的生成式 AI 軟體套件相結合，能夠加快 AI 和機器學習專案的結果產出時間，同時提高成本效益。

- 選擇 Supernano 的 GPU SuperServers ARS-111GL-NHR、SYS-421GE-TNHR2-LCC 和 SYS-821GE-TNHR 來支持高效能運算、生成式 AI 以及 AI 訓練和推論。使用 Supernano 的 SYS-221HE-FTNR(D) IOT SuperServer 結合 NVIDIA 的 H100 NVL 來支援 5G 邊緣和核心運算。
- NVIDIA RAPIDS™ 軟體程式庫套件有助於加速機器學習處理。

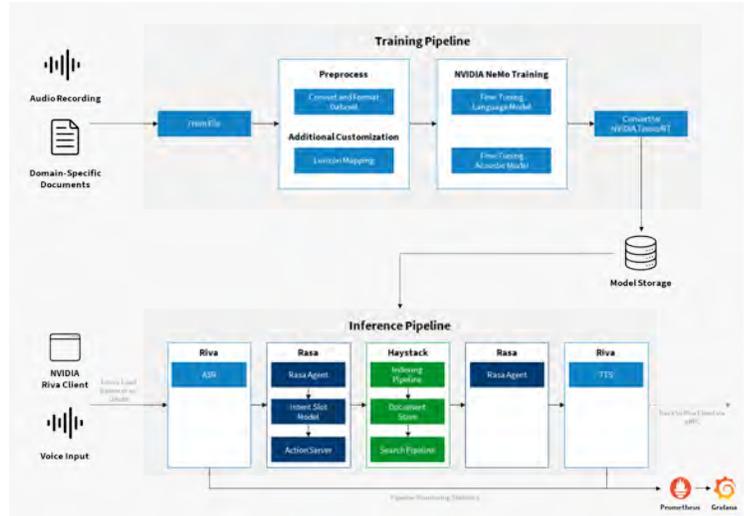


圖 4- AI 訓練和推論管道

電信業的 AI：透過新服務和擴充服務推動成長的關鍵應用案例

1. 主權 AI 工廠的 AI 基礎設施

電信組織已經是政府、企業和初創公司值得信賴的合作夥伴，這些機構都依賴電信資料中心提供安全的高速連接。通過擴展資料中心，加入加速運算基礎設施來支持 AI 模型，電信公司可以將現有資料中心轉變為 AI 工廠。AI 工廠專為 AI 工作流程設計，特別是大型語言模型 (LLM) 的基礎預訓練、微調、檢索增強生成 (RAG) 和推論，這些都是計算密集型的任務。隨著 AI 採用的增長，並且大多數應用開始將生成式 AI 作為新的用戶界面，AI 基礎設施將在各個地區變得越來越必要。各國將迅速意識到 AI 對於國家發展、經濟增長、創造就業和促進創新的重要性。電信服務提供商可以利用他們在資本管理、分散式基礎設施以及資料中心空間和電力接入方面的專業知識。這使他們能夠向國家生態系統 (包括政府、學術界、企業和初創公司) 提供基礎的 AI 工廠能力。Supernano 和 NVIDIA 可以透過 [閱讀這份白皮書](#) 來幫助組織設計和創建他們的 AI 工廠。

創收機會

電信公司不必僅限於提供 AI 基礎設施作為服務。他們還可以開發和部署自己量身定制的 AI 解決方案，以解決各行各業客戶的關鍵應用場景和挑戰。這種方法可以開啟新的收入來源，並在其地區內提供全新的客戶體驗。

通過向其他行業提供 AI 作為服務，電信組織可以進入新市場並多樣化收入來源，為如醫療、金融和製造等領域提供 AI 驅動的洞察和自動化。這些行業中的一些關鍵 AI 工廠應用案例包括製造業的預測性維護、個人化的醫療建議、金融領域的即時詐騙檢測，以及零售行業的高級客服聊天機器人。這種跨行業的應用不僅最大化了對 AI 基礎設施的投資回報，也使電信營運商在更廣泛的 AI 生態系統中樹立了領導者的形象，推動創新和超越傳統市場界限的增長。

Supermicro + NVIDIA

電信營運商可以利用其分散式區域資料中心（RDC）作為加速運算基礎設施，來構建、微調和部署主權 AI 模型。Supermicro 與 NVIDIA 提供了最廣泛的 NVIDIA 認證系統供選擇，為組織建立自身的 AI 能力提供最佳效能和效率，並最終將這些服務提供給其他人。這些解決方案利用 NVIDIA 參考架構在 Supermicro AI 優化系統上，確保最快的部署速度、支援性和效能。

- 為了達到最佳的資料中心效能，可以利用 Supermicro SuperServer SYS-421GE-TNHR2-LCC 或 SYS-821GE-TNHR 搭配 NVIDIA HGX H100/H200，實現高達 15.8k BF16/FP16 TFLOPS 和 31.6k INT8 TOPS 的資料中心叢集運算能力。
- NVIDIA NeMo™ 和 NVIDIA NIM™ 有助於加速企業級 GenAI 模型的開發和部署。
- NVIDIA NeMo™ 是一個端到端的雲原生框架，用於整理資料、訓練和客製化基礎模型，以及大規模運行推論。它支持文本到文本、文本到影像、文本到 3D 模型，以及影像到影像的生成，並加速企業級 GenAI 模型的開發和部署。

2. 透過邊緣 AI 將 5G 營利化及 6G 研究

儘管 5G 技術已經在全球範圍內部署，以滿足不斷增長的行動資料服務需求，但 6G 網路代表了無線服務的下一步，通過整合 AI 和 ML 來創建以前無法在 5G 網路中實現的新應用和案例。通過與電信公司合作，在邊緣部署 5G，許多行業如交通、醫療保健、物流、製造、機器人技術、智慧城市和零售等組織可以視覺化和優化其實體環境，並將運算能力擴展到新的地方，變得更加連接和高效。電信服務供應商通過提供企業和 B2B 服務、管理服務、消費者服務等來實現 5G 邊緣的營利化。

AI 對於 6G 的研究和開發也至關重要，因為 AI 驅動的模擬和建模加速了新技術和應用案例的測試。組織現在可以加快 6G 技術的發展，這些技術將來會連接數兆個設備，為一個超智慧的世界奠定基礎，這個世界由自動駕駛車輛、智慧空間、各種擴增實境和沉浸式教育體驗以及協作機器人支援。

由 AI 驅動的 6G 研究正在為電信公司開啟令人興奮的新收入來源，這些收入來自於包括超可靠低延遲通信和沉浸式 AR/VR 體驗等前沿服務。借助 AI 驅動的網路優化和預測性維護，營運成本大幅降低，為投資創新技術和盈利新服務鋪平道路。

Supermicro + NVIDIA

Supermicro 的伺服器與 NVIDIA 6G 研究雲端平台共同支援從邊緣到核心再到雲端的下一代網路，幫助電信公司快速推進研究。業界領先的研究人員可以利用 6G 研究雲端平台的所有元素來推進他們的工作。

- Supermicro SYS-E403-13E-FRN2T 搭配 NVIDIA L40s、Supermicro SYS-111E-F(D)WTR 搭配 NVIDIA L40、Supermicro SYS-221HE-FTNR(D) 搭配 NVIDIA H100 NVL，以及 Supermicro ARS-111GL-NHR 搭配 NVIDIA GH200 都是促進 5G 優化和前瞻性 6G 研究的有效選擇。
- NVIDIA Aerial Omniverse Digital Twin for 6G：一個參考應用程式和開發者樣本，能夠進行完整 6G 系統的物理準確模擬，從單個塔台到城市規模。它包含了軟體定義的無線接入網（RAN）和用戶設備模擬器，以及逼真的地形和物體屬性。使用 Omniverse Aerial Digital Twin，研究人員可以基於現場特定資料模擬和構建基站演算法，並即時訓練模型以提高傳輸效率。
- NVIDIA Aerial CUDA-Accelerated RAN：一個軟體定義的全 RAN 堆疊，為研究人員提供了顯著的靈活性，以即時客製化、編程和測試 6G 網路。
- NVIDIA Sionna Neural Radio Framework：一個框架，提供與 PyTorch 和 TensorFlow 等流行框架的無縫整合，利用 NVIDIA GPU 生成和捕捉資料，並在大規模上訓練 AI 和機器學習模型。這還包括 NVIDIA Sionna，這是基於 AI/ML 的無線模擬的領先鏈路層研究工具。

Supermicro 和 NVIDIA：多樣化的解決方案

在任何 AI 實施的核心，都是支持系統的強大而一致的解決方案架構。Supermicro 與 NVIDIA 的合作方法正是如此，提供了整合 CPU、GPU、優化記憶體、網路和 NVIDIA AI Enterprise 軟體平台的全方位堆疊解決方案，並在 Supermicro 平台的高可靠性基礎設施中協調運作。

從大規模訓練到智慧邊緣推論，Supermicro 的 AI 優化解決方案簡化並加速了 AI 部署。它們幫助電信部門為 AI 工作負載提供最佳效能和可擴展性，同時優化成本並最大限度地減少對環境的影響。

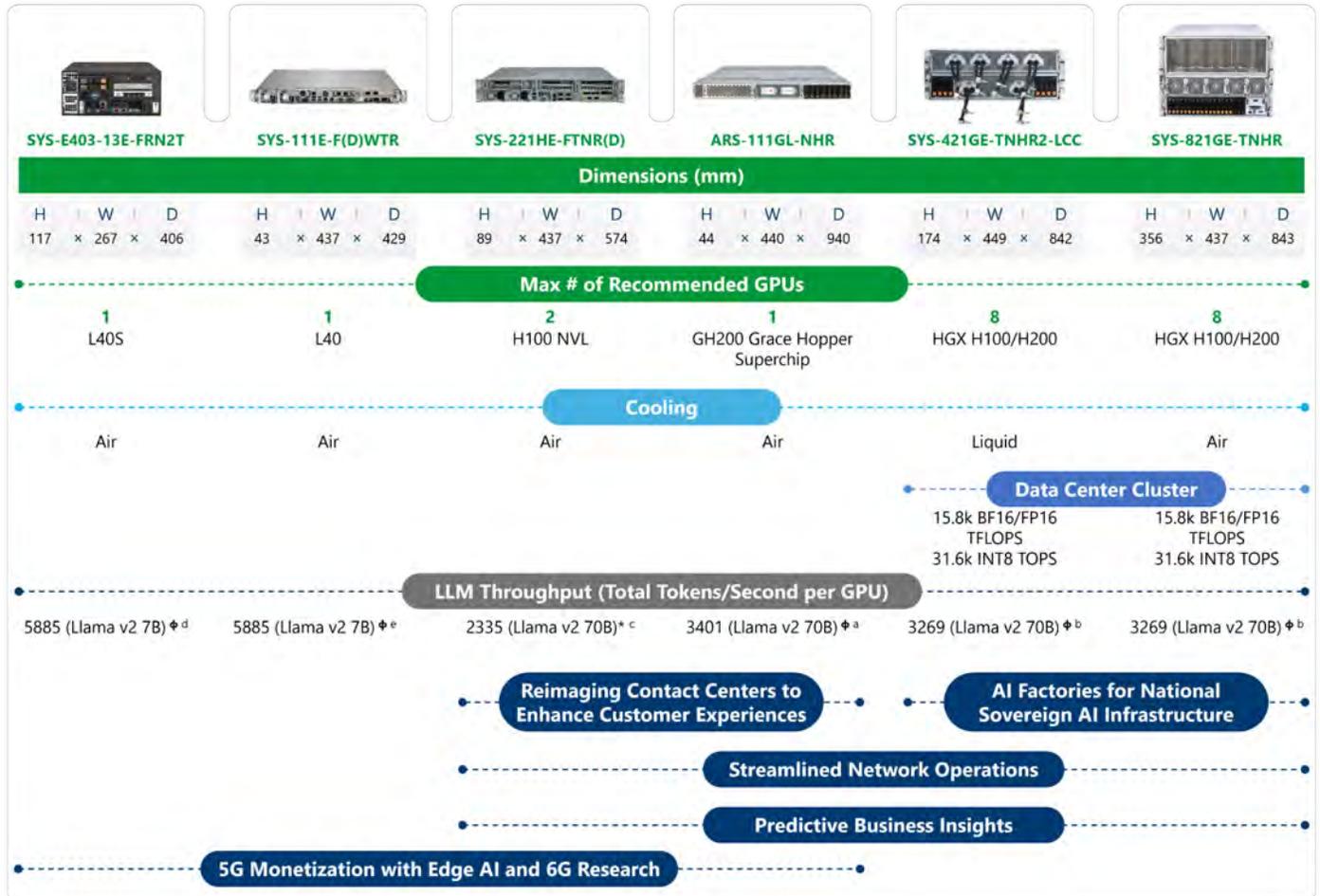
Supermicro 和 NVIDIA 可以幫助電信客戶提供各種 AI 工作負載優化的解決方案，同時使他們成為依賴電信公司滿足營運和連接需求的其他行業的強大盟友。雖然組織可以根據其獨特的環境和需求選擇多種系統，但我們在下頁的產品表中提供了一組簡短的建議。

為您的 AI 應用選擇最佳的 Supermicro 系統

Supermicro 提供靈活的 AI 就緒基礎設施解決方案，確保電信營運商在實施 AI 解決方案時能夠隨需求擴展。無論需求如何，Supermicro 都有相應的解決方案來擴展記憶體、處理能力和儲存，以滿足各種情境。

Supermicro 和 NVIDIA 在引導組織選擇適合其特定 AI 應用的系統方面表現出色。這項支援涉及考慮 AI 模型的大小、系統相容性以及具體的應用案例需求。無論是重新構思聯絡中心以提升客戶體驗，還是預測分析以獲取商業洞察，Supermicro 的邊緣平台產品組合可涵蓋多種形式因素，包括緊湊型盒式、1U 和 2U、壁掛式、機架式，甚至無風扇型號。當這些解決方案與 NVIDIA 強大的 AI 和加速運算平台結合時，它們能夠有效地滿足這些多樣化需求。

Supermicro 和 NVIDIA 解決方案的多功能性是它們在電信及其他行業中廣泛應用的關鍵。這些系統能夠適應各種計算需求。



†: Input Length = 128, Output Length = 128, Batch Size = 256
[‡]: Input Length = 128, Output Length = 128, Batch Size = max
^a: TP = 1, with 1x GH100
^b: TP = 2, with 2x H100
^c: TP = 1, with 1x H100 NVL
^d: TP = 1, with 1x L40S
^e: TP = 1, with 1x L40
 TP = Tensor Parallelism

圖 5-最常用於 AI 應用的 Supermicro 解決方案

安全性、儲存和網路是此架構的基礎要素，確保資料的完整性與傳輸不會受到損害。NVIDIA 的高效能網路平台，加上像 Arccus 等電信生態系合作夥伴，將軟體定義、硬體加速的電信級路由、交換與遙測功能帶入每台 Supermicro 伺服器，如圖 6 所示。這種強大的後端被封裝在 Supermicro 的硬體中，該硬體以其可靠性著稱，設計旨在滿足各種環境的需求。最終結果是一個可擴展的解決方案架構，使電信公司能夠充分發揮 AI 的潛力。該架構如圖 6 所示。

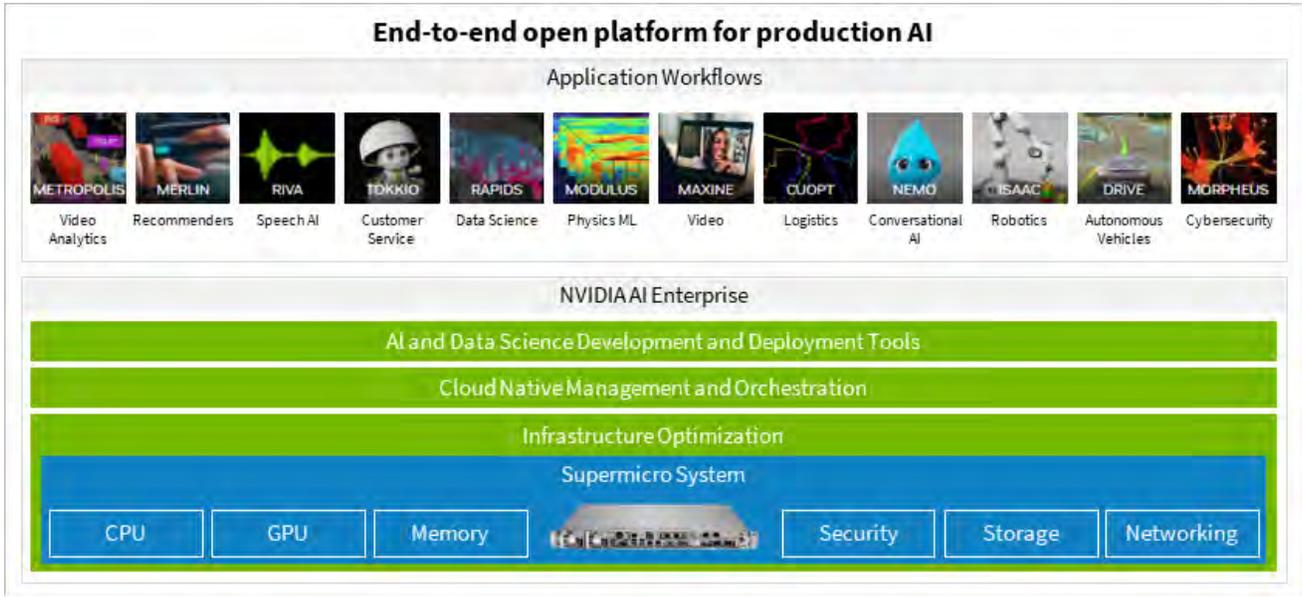


圖 6 - 全方位解決方案

向前邁進

AI 的探索和採用已經在電信產業中展開，成功將 AI 融入工作流程的企業將能創造最大的價值。對於想要利用 AI 的組織來說，Supermicro 和 NVIDIA 提供了一條明確的發展道路。他們的結合專業知識和多樣化解決方案為各種 AI 計劃奠定了堅實的基礎。這種協同作用降低了風險，加快了投入生產的速度。成功的實施不僅能帶來更愉悅的客戶體驗、增加收入，還能提升在各種產業應用中的安全性。

欲了解更多資訊

要了解有關我們的 AI 解決方案的更多信息，請訪問 <https://www.supermicro.com/en/solutions/ai-deep-learning>.

SUPERMICRO

身為高效能、高效率伺服器技術和創新的全球領導者，我們為資料中心、雲端運算、企業 IT、大數據、HPC 和嵌入式市場開發並提供端到端綠色運算解決方案。我們的 Building Block Solutions® 方法使我們能夠提供廣泛的 SKU，並使我們能夠根據您的要求建置和提供應用程式最佳化的解決方案。請參閱 www.supermicro.com。

NVIDIA

自 1993 年成立以來，NVIDIA（納斯達克股票代碼：NVDA）一直是加速運算領域的先驅。該公司於 1999 年發明的 GPU 激發了 PC 遊戲市場的成長，重新定義了電腦圖形學，點燃了現代人工智慧時代，並推動了整個市場的工業數位化。NVIDIA 現在是一家全端運算公司，提供正在重塑產業的資料中心規模產品。更多資訊請造訪 <https://nvidianews.nvidia.com>。