電子書

在企業中應用 AI

部署和執行 AI 應用程式的最佳實務

NUTANIX

簡介

人工智慧（AI）正在以前所未有的方式改變商業世界，幫助企業提高效率、生產力、創新和客戶滿意度。AI 還會給組織帶來新的機會、複雜性、挑戰和風險。在本電子書中，我們將探討 AI 在企業中的重要性、其現狀和趨勢，以及在公司中實施和管理 AI 專案的最佳實務和策略。

目錄

AI 不斷演變的世界

AI 已經在企業中站穩腳跟。這不再是一個超前新奇的概念，而是一個正在改變企業營運方式的現實。AI 已成為許多產業必不可少的組成部分，包括製造業、醫療照護、金融、教育和公共部門等。

多虧了資料、演算法和運算能力方面的進步，AI 得以持續高速發展。傳統 AI 設計要旨在於根據預置規則和資料執行特定任務，用於流程自動化、供應鏈最佳化、自然語言處理、資料分析和網路安全等目的。然而，一種被稱為生成式 AI 的新型 AI 已經橫空出世，企業紛紛從中獲取即時價值。

隨著 OpenAI 的大型語言模型（LLM）GPT-3 在 2022 年底的發布，生成式 AI 開始變得流行起來。它與傳統 AI 的不同之處在於，它在包含數萬億個資料點的龐大資料集上進行訓練，從習得模式中生成新的內容。生成式 AI 也因其模仿一般人類反應的能力而聞名，這與傳統 AI 的特定任務能力極大不同。

LLM 是基礎模型，例如 GPT-3 和 GPT-4、Llama 2、PaM 2 和 Falcon，這些模型已經在大量未標記的資料上進行了訓練。建立這些 LLM 需要相當可觀的投資，這超出了大多數組織所能負擔的範圍。另一種方法是建立一個緊湊的 LLM，這通常涉及使用專屬的組織資料對 LLM 進行微調，以便為各種用例生成特定答案。

儘管傳統 AI 和生成式 AI 大不相同，但它們都深深具有改變當今和未來企業的能力。

55%

來自 McKinsey 全球調查的受訪者表示，他們的組織早在 2023 年初就已採用 AI。

33%

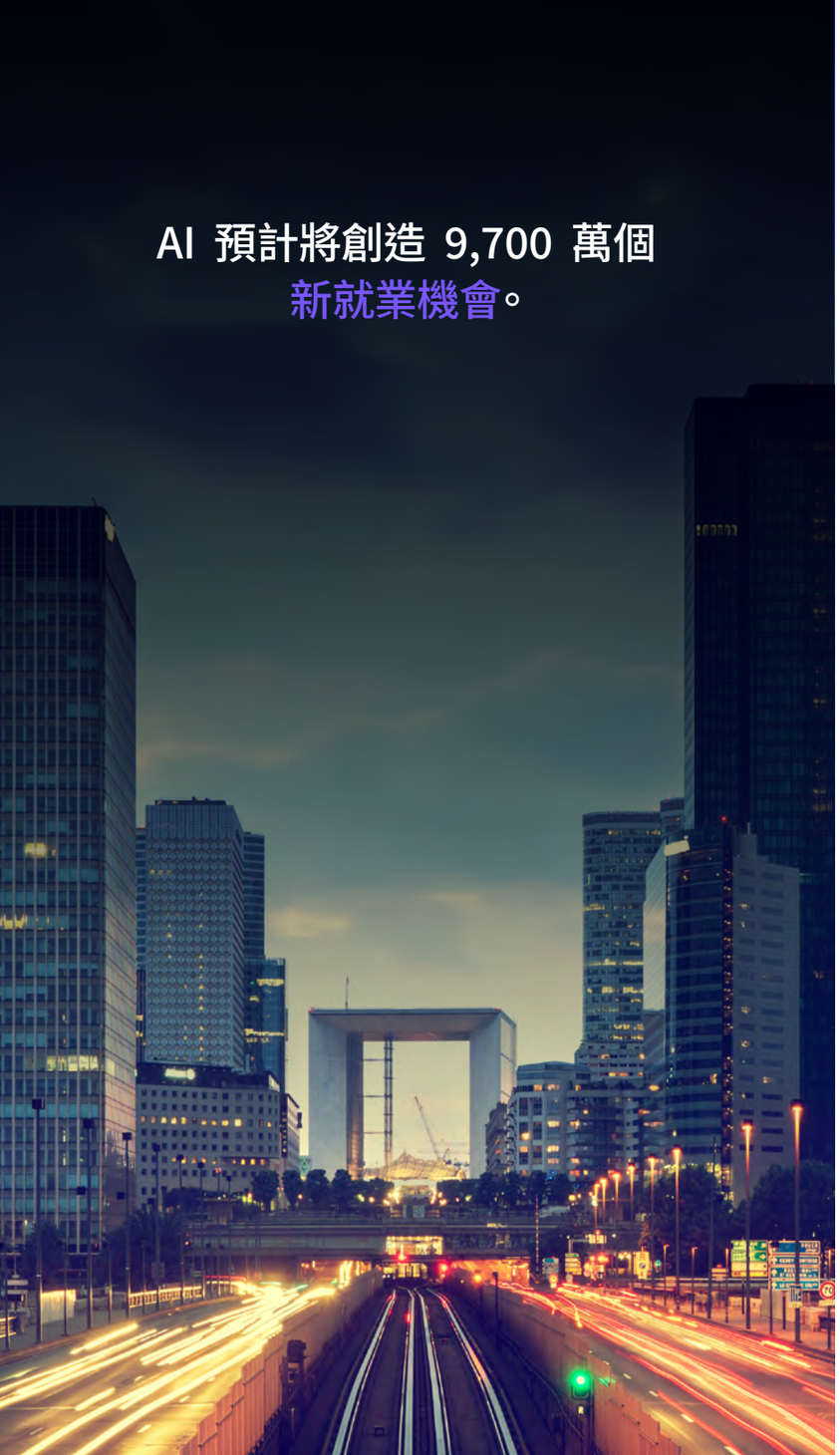
的組織已在至少一項職能中定期使用生成式 AI。

84%

的技術高階主管希望將 IT 基礎架構現代化以支援 AI，資料來源於《2023 年 Nutanix 企業 AI 現狀報告》。

4,070 億美元

2027 年預估的 AI 市場規模，將比 2022 年增加 3,200 億美元。



實施 AI 的挑戰

儘管 AI 會帶來巨大的優勢，但組織在實施和發展 AI 實務時仍面臨關鍵挑戰。為了確保 AI 計畫取得成功，他們需要仔細規畫，讓所有必要的利害相關者參與進來，並持續監控其 AI 應用程式和資料。

AI 模型管理與監控

AI 中的模型管理與監控對於確保 AI 系統的品質、可靠性和效能至關重要。持續對模型進行微調是必要的，以生成及時且相關的結果。監控 AI 模型涉及對生產中所有 AI 模型的指標和回饋進行蒐集、分析和報告。

這些功能構成了組織機器學習作業（MLOps）的核心，這是一套專為部署和維護機器學習模型而設計的實務。每個模型及其相關資料的訓練和微調，都會建立一個新版本的生成式 AI 應用程式。這些模型都必須加以監控和控制，以確保每個版本都具有更高的品質和受到保護，並且像任何其他的企業應用程式一樣，具有完整的資料生命週期管理。

員工技能缺口

由於 AI 在企業中是一個相對新的工具，因此組織在其工作人員中面臨著技能缺口。許多企業缺乏正式的 AI 培訓，這使得再培訓成了當務之急。技能缺口仍然是 AI 採用的最大障礙。

員工也有極大的動力來提升技能：82% 的高階主管認為熟練使用 AI 的員工應該獲得更高的薪酬，74% 的高階主管認為他們應該更頻繁地得到升遷機會，而 72% 的高階主管認為他們的公司應該增加對 AI 學習和發展計畫的投資。

業務與 IT 的結合

與大多數新技術一樣，在資源節省、員工生產力提升，以及增強產品和服務方面，企業方認為 AI 幾乎具有無限的潛力。AI 可以幫助企業做出更快、更好的決策，提升客戶體驗和滿意度，自動執行任務和增強安全性等等。

IT 在規畫、執行和監控企業中的 AI 專案方面肩負著巨大的責任，包括設計和開發 AI 基礎架構、部署和維護 AI 系統、確保 AI 實踐符合隱私權和治理標準等等。處於峰值效率時，最佳的 AI 環境有助於降低複雜性、增強可預測性並完整保護企業的 AI 資產。

雖然業務方和 IT 方的職責迥然不同，但雙方之間的和諧對於 AI 在任何組織中的蓬勃發展都至關重要。

資料一致性

你可以利用組織資料來微調你的模型。一旦你的模型經過訓練和更新，就可以在邊緣進行部署和推理，以削減延遲。高品質的資料使你的模型更加有效率並產出最佳成果。

結構化資料在資料庫中以預定義的格式和結構進行組織，且可搜尋。在 AI 的世界中，我們從參數和嵌入方面對其進行查閱。參數和嵌入會被標示和分類，並輸入到傳統 AI 模型中，而未標示的參數（非結構化資料或原始資料）則輸入到 LLM 中。接下來需要驗證模型，以確定資料品質是否代表準確性的提升。更多的資料並非總是好事；重要的是資料的品質。

資料淨化是從參數和嵌入中偵測和移除不正確記錄的過程，例如錯誤、重複、不一致或不相關的部分。此外還需要對資料進行清除——這是一個非常重要的過程，將個人可識別資訊（PII）和其他敏感資料（包括社會安全號碼、客戶姓名和護照號碼等）從資料集中移除。

成本

AI 實務的成本考量取決於許多因素，例如所需的 AI 模型類型、當前和未來的基礎架構、現有的 AI 員工技能等等。無論你選擇在組織中如何設立 AI 專案，請考慮下列成本領域：

實施：對於那些想要快速起步的企業而言，本地端 LLM 實例和按需付費的雲端服務都是不錯的選擇。

基礎架構：升級可能涉及更新的伺服器、GPU，和其他硬體。

能源：大型 AI 系統可能需要耗費大量電力。

徵才與培訓：提升員工技能和招募相關人才的成本可能很高昂。

資料採集與管理：從感測器和其他方式中獲取原始資料可能是一項昂貴的配置。管理資料包括建構、淨化和維護措施。

訓練和升級 AI 模型：成本可能因模型規模和類型、硬體軟體需求等而有所不同。

效能：監控資料使用情況和其他功能的成本可能很高。使用應用程式效能監控工具會有所幫助。

互操作性問題

操作企業 AI 平台通常會涉及多個雲端，從你的本地端資料中心開始，然後擴展到任意數量的雲端供應商。在雲端之間利用資料、應用程式和服務非常複雜，需要對每個雲端供應商的工具集都有特定的瞭解。供應商服務可以統合私有雲和公有雲，並充分利用每種雲端的優勢，同時提供簡易性和全系統的安全性，以確保在你整個雲端資產中 AI 營運的業務持續性。

AI 使用案例就如想像力一樣無邊無際

跨產業

- 支援和技術知識庫聊天機器人
- 文件自動化
- 詐欺偵查
- 影片推論
- 內容創作

特定產業

- **健康照護：**病患資料挖掘和分析，以識別風險因素
- **金融服務：**偵測洗錢及其他非法活動
- **製造：**將影像識別用於品質檢驗和檢查
- **物流：**預測需求模式以增強供應鏈運作
- **公共部門：**自動化預算編列，以實現更智慧的資源分配

特定的生成式 AI

- 流程自動化
- 程式碼輔助開發
- 審查電子商務應用程式的彙總
- 特定客戶的行銷活動
- 研究與開發中的預測性建模

關鍵 AI 策略

當你開創企業 AI 之路並邁出第一步時，遵循一組基本策略來幫助你形成實踐的目標、範圍和方法非常重要。以下內容將幫助你實現流程、資源和成果的有效管理，確保你的 AI 營運具備最高品質、可歸責性和可靠性。

確定建置和部署的位置

企業喜歡可預測性，特別是當他們在其本地資料中心運行 AI 模型，施以微調以提升效率，然後將其部署在邊緣，反覆地循環這個流程來微調他們的模型時。可預測性源自利用已知資源的熟悉性和可靠性，例如，使用自己的資料，而不是來自雲端服務的資料。此種做法的例子包括：

- 用於零售自助結帳損失預防的影片推論，其中小型 AI 系統運作於自助結帳販賣店的邊緣。
- 在多個地點部署具有人臉辨識功能的安全與監視系統。
- 即時變更的智慧交通號誌燈，可為救護車或緊急救援車協調有效路線。

雲端成本

如果不持續監控，按需付費的雲端服務可能會迅速失去控制。請確保你在這方面不會遇到任何的意外情況。費用可能會增加的一種情況是，當你使用自己的資料來微調雲端中的大型 LLM（與本地端 LLM 不同，後者也利用大型 LLM，但在本地處理資料）。這種做法不僅成本高得令人望而卻步，而且還存在洩漏敏感資料的風險。

延遲

延遲（系統回應請求或執行指令所需的時間）會影響 AI 應用程式的效能和效率。資料的即時或近乎即時處理，以及結果的預期頻率和計時，皆依賴於具備效率、低延遲性的 AI 系統。影響 AI 延遲的因素有很多，包括網路條件、資料源和執行 AI 模型的系統之間的速度、AI 模型的複雜性和大小、AI 系統的架構，以及資料和 AI 系統的位置。GPU 處理能力也扮演其中一角。有時需要快速的 GPU 來實現低延遲；在其他時候，較便宜且較小的 GPU 可以提供類似的延遲，但可以更好地針對成本進行最佳化。

資料重力

隨著 AI 系統的發展，它們利用越來越多的資料，這反過來又需要更大、更廣泛的應用程式、模型和服務來支援它。這個概念被稱為資料重力，它對於 AI 實務在企業層面的發展效率扮演著一個重要的角色。資料重力會影響 AI 系統的效能、安全性和可擴充性，如果管理不當，還會帶來涉及頻寬和資料移動的額外挑戰。當你發展 AI 平台時，請試著預測你的資料將有多大的「質量」，並提前規畫。

靈活的 AI 軟體堆疊

此外，請考慮使用開放原始碼軟體堆疊來部署 AI。為了保持當前和未來 AI 專案的最大靈活性，明智的做法是不要被鎖定在專屬技術中。具明確主張的 AI 堆疊，是開放原始碼軟體和精選服務的套裝解決方案，這讓你能夠在機器學習（ML）框架、MLOps 平台、資料科學平台和 AI 堆疊的其餘部分選擇你自己的技術。

制定嚴格的資料安全、隱私和主權標準

所有組織，不論是區域性組織還是全球性組織，都必須遵守資料法規，以避免嚴厲處罰。AI 加劇了企業的擔憂，因為它可能涉及利用敏感資料，而這些資料在訓練和微調模型時可能會洩漏到外界。

法規合規性有助於確保資訊安全，它能夠要求組織遵守規則，以保護其資產免受威脅行為者影響。負責資料安全和合規性的人員應確保他們熟悉並遵守其營運所在國家的法規要求。

透過嚴格遵守一個或多個安全框架來保護你的AI 模型和資料，包括美國國家標準與技術研究所的網路安全 2（NIST CSF 2）、開放式 Web 應用程式安全專案（OWASP）以及 Gartner® 文章中界定的 AI TRiSM「處理 AI 模型中的信任、風險與安全性問題」。[1]許多組織在其安全實務中將這些框架和其他框架結合使用，所以像是 OWASP 的 10 大 AI 安全威脅清單（重點關注 LLM 應用程式）等，都是你的 AI 安全實務的寶貴資源。

另外，請切記加密保護所有敏感資料的絕對必要性。在 2022 年的一項調查中，55% 的受訪者表示，他們的組織將敏感或機密資料傳輸到雲端，並施以加密技術，或以其他方式使其無法讀取。無論如何，不要讓惡意攻擊者能輕易查看和處理你的資料。

應用 AI 來提升組織效率

AI 已成為眾多產業的變革力量，它具備深刻重述工作未來的潛力。雖然企業渴望實現 AI 的經濟效益，但他們也應該意識到 AI 可提升員工工作成果的獨特能力，而不是取代其工作，並尋找在企業中將這一點付諸實施的方法。AI 可用於自動執行例行任務，使員工能夠專注於更具創造性和生產力的工作。

[1] Gartner 文章。處理 AI 模型中的信任、風險與安全性問題，2023 年 9 月 5 日

AI 助理

一個有趣的例子是 AI 助理，這是一個可幫助員工完成任務和決策過程的交談式介面。AI 助理可協助起草電子郵件、回答特定問題、提供情境感知的支援，並引導員工完成複雜的任務和流程。它們甚至可以執行高度特定的任務，例如資料分析、統一不同的系統以連接來自不同平台的工具和應用程式，以及為開發實務生成程式碼。

AI 維運

AI 維運是利用 AI 技術來維護和改善 IT 基礎架構的過程，包括自動執行效能監控、資料備份和工作負載排程等關鍵任務。透過蒐集和分析來自多個來源的資料，AIOps 可以為你的 AI 應用主動帶來即時分析與回饋。這有助於降低營運成本、縮短問題緩解時間、簡化 AI 營運，並在整個 AI 平台上實現預測性服務管理。

需求預測

AI 可以大幅增強需求預測，也就是估算產品或服務未來需求的過程。它可以幫助提高準確性並減少錯誤；將天氣、事件、假日和促銷等外部因素納入考量；透過預測供需變化來增強敏捷性和回應能力；實現更有效的資料驅動決策；並透過測試和衡量方案的有效性來提升創新。

增進 AI 實踐

要在組織中利用 AI 的優勢，並克服不可避免的挑戰，需要持續著重將新想法和創新應用到你的 AI 實務中。以下是一些可以幫助你逐步發展 AI 營運的關鍵領域。

提升你的員工技能

缺乏具備科學、技術、工程和數學（STEM）技能的員工，是眾所皆知的難處，要找到專精的 AI 員工也並非易事。68% 的高階主管認為，在其員工中存在中至極度情況的 AI 技能缺口。因此，如前所述，最重要的是盡可能地提高現有員工在 AI 知識和技術方面的技能，以幫助彌補該缺口。

這種做法對於在整個企業範圍內建立和鞏固 AI 工作成果方面將大有幫助。你可以從簡單地訓練員工開始，讓他們在 AI 助理和其他 AI 工具中輸入最有效的提示，以生成最佳結果。然後，員工就會更加熟悉 AI，並開始將其視為自己職位的福利。

為了提升這種實務，提示工程作為一種專業的工作角色也越來越受歡迎，其需要領域知識、創造力以及對 AI 模型如何運作的理解。這代表了 AI 在這產業增進勞動力空缺的一種方式。

工作場所中出現的其他專業 AI 角色包括 AI 培訓師、AI 工程師、AI 研究人員、AI 倫理學者、AI 諮詢顧問和 AI 分析師。潛在的新角色可能包括 AI 情緒分析師、在偏差和管制措施方面具備專業知識的 AI 輸入與輸出經理、AI 合規性經理等等。

自動化

自動化是建構 AI 實務的關鍵。例如，讓 AI 盡可能地成為一項自助服務，這對於你組織中的資料科學家和其他 AI 實踐者至關重要。當員工被迫建立工單，並等待 IT 部署裝載有他們所需的程式館和工具的容器或虛擬機（VM）時，專案就會變慢。那麼自動執行這一流程，以便資料科學家可以部署自己的容器和虛擬機，是極為可取的做法。此外，自動化可降低人為錯誤。

理想的資料放置

訓練模型時所使用的資料的位置和存取難易程度，會顯著影響 AI 系統的效能、效率和成本。你的資料與 AI 模型越接近，模型的執行速度就越快，而且你對它的控管力也越大。但更重要的是，請注意不要在公用位置儲存或使用你的 AI 資料，因為這可能會使其暴露於潛在的威脅和洩漏風險。

有必要再次強調的是，在整個組織中減少或消除資料孤島的重要性，還有定期淨化資料，可為你的 AI 營運提供更高效率。

單一控制面板

你可以透過使用單一控制面板來集中管理 AI 平台，從而為你的 AI 營運提供更高的效率。利用通用介面，有助於透過統合基於角色的存取，和由政策驅動的控制項來降低複雜性，例如，讓你能夠使用一套標準化的政策和控制項來管理你的整個 AI 資產。這個控制面板會在你的 AI 平台運作的每個環境中，將 AI 應用程式、工作負載和資料移動性結合起來。

39%

的企業聘請軟體工程師，而 35% 的企業則聘請資料工程師來擔任與 AI 相關的職位（根據 2022 年數據）。

結語

AI 是一項令人著迷且強大的技術，它具有以無窮無盡的方式轉變組織和世界的潛力。從增強創造力和生產力，到解決關鍵業務挑戰和跨產業進行創新，如果在企業中以明智且合乎倫理的方式利用 AI，AI 可以成為一種積極的變革力量。隨著我們來到本書結尾，這裡提供一些最後的想法，當你開始建構 AI 實務時，可對其多加思量。

在你的 AI 實踐中創造目標

你可以從小型 AI 專案開始，並將它們用作提前體驗與歷練的機會，以快速最大化資料和洞見的價值。盡一切努力在你的 AI 道路上，對用例、資料和分析以創意思維來看待，然後建構一個全方位的 AI 實務，可觸及企業的每個部分，讓你的員工具備比以往任何時候都更高的生產力。

你的資料是你通往 AI 成功的窗口

你的組織資料不僅獨一無二，而且具有無限的價值。開始使用 AI 來挖掘它，你將從中獲得更多的價值——以前價值不大的資料會突然變得非常有趣。由於你的資料是 AI 營運的命脈，因此請盡可能使用最強的安全方法來保護它，並用嚴格的治理和合規性措施來加以包覆。以這種方式處理本地端資料，可以讓你更好地控管資料，並有助於降低 AI 營運的風險。

致謝

本書到此結束，感謝撥冗參閱。欲詳細瞭解 Nutanix 能如何協助你快速踏上 AI 之旅，並為你提供安全性、隱私性和控管力，請前往 nutanix.com/ai。

NUTANIX
info@nutanix.com | www.nutanix.com | @nutanix