Adobe

AI 的轉折點

如何在組織中
負責任地採用 AI

目錄

點擊此功能表圖示
即可返回本頁面

I. 簡介：負責任地採用 AI 已成為如今的當務之急

隨著 AI 成為推動全產業轉型的催化劑，企業高管正面臨前所未有的壓力，需要迅速創新、應對競爭威脅並提升營運效率。然而，在企業內部競相採用 AI 也引發了新的風險。如果缺乏審慎的監督，這種快速的 AI 實作可能會導致法規失誤、營運中斷以及長期的聲譽損害。平衡速度驅動與責任要求不再是權衡取捨的問題，而是戰略上的當務之急。

治理 AI 並管理其風險往往是一項艱鉅的任務。新出台的指引、框架與政策，加上不斷變化的國際、聯邦和地方法規，這些複雜性往往會讓組織不知從何入手，從一開始就為利害關係人帶來了挑戰。Adobe 憑藉負責任創新的經驗，基於 AI 道德原則（問責性、責任感和透明度），讓我們對如何克服這些挑戰有深入的分析。

我們的經驗表明，雖然通往負責任的 AI 創新之路看似艱辛，但只要運用合適的工具、策略和思維方式，成功指日可待。

組織在制定 AI 策略時所面臨的核心決策之一，就是要確定是建立、購買還是自訂 AI 解決方案，或是三者兼具。以下方法專門針對希望購買 AI 解決方案的組織，目的是為希望從外部尋求 AI 解決方案的組織在現有的價值與商業實務基礎上建置 — 與當前的利害關係人接軌。本框架以獨立研究為基礎，並參考 AI 治理方面的專家訪談，提供可行的發展路徑，協助組織評估其當前狀況，並提供最佳實務，以便在整個企業範圍內落實負責任的 AI 原則。它包括以下實用步驟：建立員工生成式 AI 使用指引，透過詳盡的問卷調查評估供應商，以及更新 AI 治理流程來跟上不斷演進的環境。

無論您處於 AI 歷程的哪個階段 — 無論是評估組織的 AI 就緒程度，還是完善現有策略 — 本框架都提供了一套行之有效的方法，將人類智慧與尖端的 AI 治理相結合，負責任地進行擴展。透過遵循這一路線圖，組織可以有效地評估、試行、採用和監控 AI 解決方案，建立一個堅韌的基礎，促進信任，降低風險，並推動持續的業務價值。

II. 框架概觀：打造一個可擴充、合乎道德的 AI 未來

成功的生成式 AI 實作需要的不僅僅是一份行動清單 — 它需要一個戰略分層方法，每一個階段都建立在上一個階段的基礎上，為永續創新和合乎道德的 AI 實務奠定基礎。本框架可作為一系列環環相扣的搭建組塊，旨在整合每個階段負責任的 AI 實務 — 從評估組織準備程度到有效擴充與持續監控 AI 系統。

本框架並非將 AI 的採用視為流程驅動的練習，而是專注於建立與組織需求協調演進的系統。它強調人類監督與先進 AI 技術之間的重要平衡，確保組織能夠發揮 AI 的潛力，同時與道德、法規和營運目標保持一致。

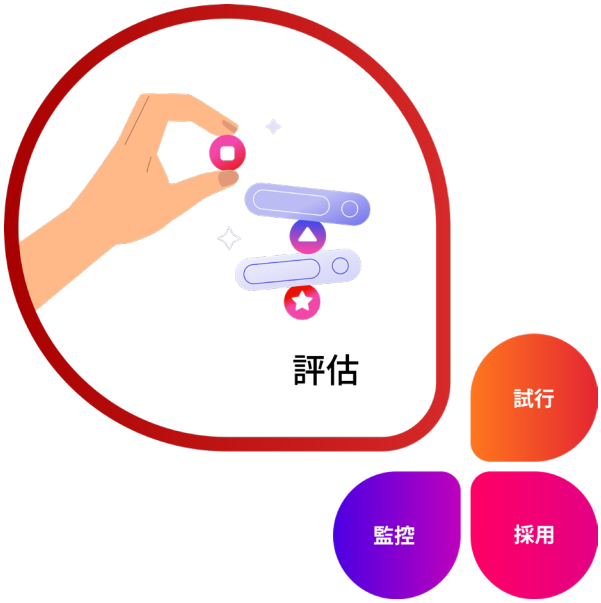
本框架中的每個階段（準備評估、負責任的試行、擴大採用範圍和持續監控）都支援長期成功，作為整合支柱在每個步驟中互相加強。透過在每個階段融入負責任的 AI 實務，企業可以在促進信任、透明度和問責性的同時，駕馭 AI 採用的複雜性。

以專業知識為基礎，以研究為依據

Adobe 與一家獨立研究公司合作，調查生成式 AI 的採用情況，收集了來自不同產業的 200 多位 IT、組織與合規領導者的見解。這項研究強調 AI 採用的當前實務、挑戰與成功策略。此外，Adobe 還對產業專家進行了深入訪談，並審核了全球標準，包括歐盟 AI 法案、NIST AI 風險管理框架、新加坡的 AI Verify、IEEE 標準 7000 以及 ISO 42001。這些努力確保該框架適用於各行各業及各種規模的組織，無論 AI 採用進度如何。

1. 評估：組織準備程度並選擇負責任建置的 AI 技術

負責任地採用 AI 的歷程從領導者開始。**評估階段**可為決策者提供所需的工具、數據及分析，以評估 AI 如何融入組織的戰略優先事項。這一階段可讓跨職能領導者檢視組織的技術基礎設施、治理框架和 AI 素養，從而協助確定整體準備程度。



1.1 評估組織準備程度

儘管許多組織已經開始採用 AI，但在接受調查的組織中，只有 21% 的組織已完全制定出負責任的 AI 優先事項，78% 的組織仍在進行中或處於規劃階段，突出表明了對準備方法的明顯需求。IT、合規、風險管理與策略領導者對於建置負責任的 AI 基礎至關重要。這首先需要全面檢視組織的治理框架和 AI 素養，以找出可能影響 AI 採用的缺口。

組織需要採取**整體方法**來評估 AI 準備程度，將**自上而下的領導計劃**與日常接觸 AI 的員工提供的**自下而上的意見回饋**結合起來。

準備行動：

進行全面的準備審核 — 評估組織的技術基礎設施、治理標準、AI 相關政策、負責任的創新框架，以及合規實務，以找出優勢和需要改進的領域 — 確保符合策略目標和負責任地採用 AI 的需求。

協作找出並解決主要差距 — 記錄安全性、隱私、法律、合規與透明度標準方面的其他 AI 政策需求，同時讓跨職能團隊（包括 IT、法律、合規及業務部門）參與進來，優先確定可操作的後續步驟。

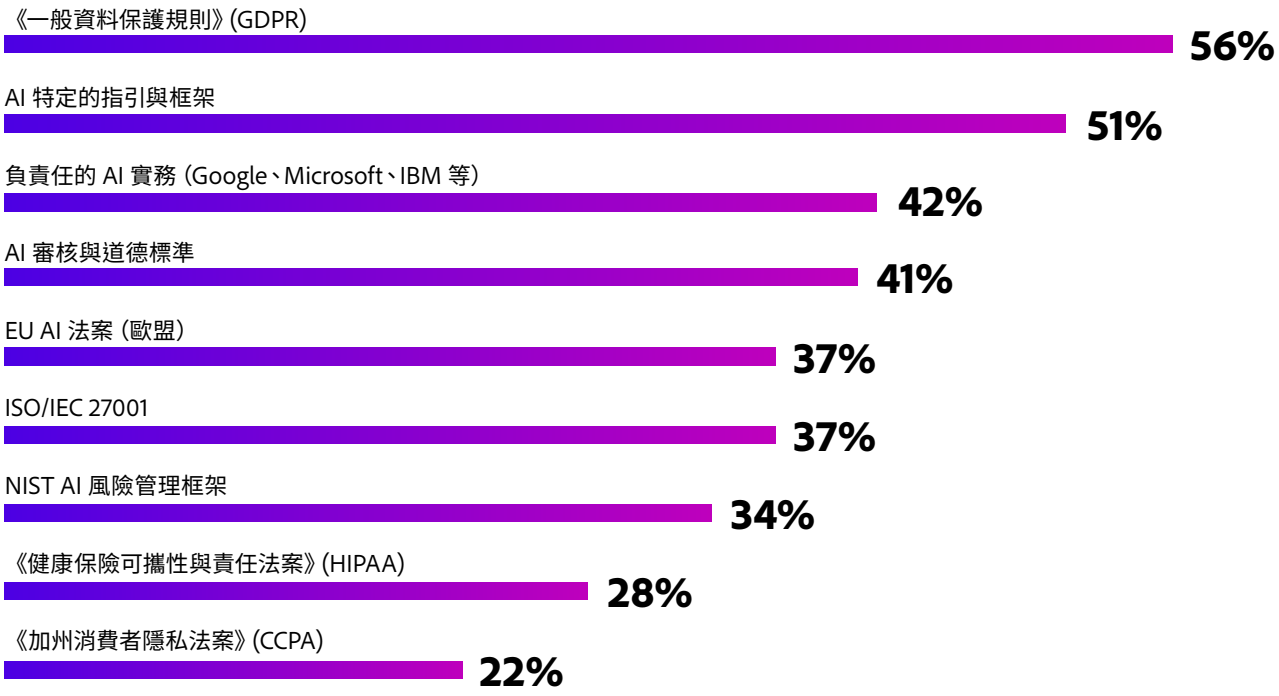
建立並授權治理團隊 — 指派團隊監督 AI 治理，確保符合內部負責任的 AI 標準與外部法規框架 — 賦予這些團隊權力與資源，主動管理風險並適應不斷變化的需求。

1.2 選擇負責任建置的 AI 技術

首先，全面檢視貴公司現有的治理標準。這些標準可能已經包含**隱私、安全性、可及性**以及**法律考量**等關鍵領域。**《一般資料保護規則》(GDPR)** 和 **AI 特定框架**等全球**基準**是許多組織維持合規與風險監督的組成部分。此外，**地區性政策**及**產業特定標準**（例如 **AI 審核**與**責任標準**）也應納入治理標準。

AI 技術所需的安全性與隱私標準或認證

在受訪者總數中所佔的百分比，依降序排列。



一旦組織勾勒出負責任的 AI 期望和治理框架，接下來便是為負責任建置的 AI 技術確立選擇標準。這些標準應整合現有的標準，並重點關注與生成式 AI 相關的獨特要素，如來源的透明度、輸出的準確性、訓練數據授權、偏見緩解以及文化本地化。

根據研究結果，組織在評估生成式 AI 技術時使用的重要標準包括：

1. **訓練數據評估 (72%)**
2. **AI 使用披露 (63%)**
3. **危害緩解 (60%)**
4. **來源透明度 (55%)**
5. **偏見緩解 (50%)**

這些因素確保所選擇的 AI 技術同時滿足**業務需求**與**道德責任**，支援組織的長期成功。

組織應制定針對性的選擇標準，使 AI 解決方案符合戰略業務目標和負責任的 AI 原則。這些標準強調：

透明度 確保 AI 流程是可解釋和可追蹤的。	**文化本地化** 調整 AI 系統以尊重不同的文化與地區背景。
準確性 維持高標準的數據真實性和預測可靠性。	**偏見緩解** 主動減少偏見，支持公平公正的 AI 結果。

記錄評估和選擇過程的每個階段，加強適應性與問責性，建立一個可以隨著 AI 進步和法規變化而演進的靈活治理模型。以下是評估階段需要採取的步驟摘要：

評估

步驟 1：評估組織準備程度

- 定義並傳達公司有關負責任使用技術（包括 AI）的標準。
- CIO 和/或跨公司委員會將審核當前系統與業務流程，找出從負責任的 AI 採用受益最多的領域。
- 匯總來自內部業務與職能領導者對負責任的 AI 採用可考慮的其他使用案例的意見。

步驟 2：選擇負責任建置的 AI 技術

- 針對 AI 的考量，審核現有的治理標準，包括隱私、安全性、可及性以及法律。
- 透過重點關注透明度、準確性、偏見、文化本地化及合規，制定選擇標準，整合先前建立的、滿足負責任 AI 期望的標準。
- 評估並選擇最符合既定標準和業務需求的 AI 技術，記錄決策過程。

2. 試行：識別並試行高影響力的使用案例。

試行階段是 AI 實驗與實際運作之間的橋梁。這一階段使主要利害關係人能夠評估該技術的效能，特別是其如何與業務目標和負責任的 AI 目標保持一致。它不僅僅測試技術可行性，而是著重於使主要領導者和利害關係人能夠以有意義的方式直接與該技術接觸。這關係到讓人們能夠運用 AI，根據其可擴展性做出明智的決策，並確保其符合道德、營運和法規標準。

試行讓組織有機會在情境中對 AI 系統進行壓力測試，協助他們瞭解哪些地方可能需要責任評估和透明度文件，以及新功能相對於預期的表現。透過記錄分析和收集可行的學習成果，組織可以制定一個負責任地擴展 AI 的路線圖，建立一個可支援即時和長期目標的基礎。

2.1 確定並準備優先使用案例

發展一個有說服力的 AI 商業案例包括讓主要利害關係人、前線員工參與其中，從而提供關於 AI 潛力的全面視角。透過讓那些將直接與技術互動的人員提早參與，組織可以找出 AI 能帶來切實利益的高影響力使用案例，例如在行銷內容創作、編碼、工作流程自動化和數據管理領域。

具體化 — 重點關注過程，而非角色：與其圍繞特定角色（例如「開發人員的 AI」）來構建使用案例，不如重點關注 AI 可以簡化和改進的流程，例如「AI 輔助編碼以自動化例行程式碼審核與錯誤檢測」。

建立可度量的使用與成本節省指標：儘管 ROI 很重要，但 AI 試行還應強調更廣泛的回報，例如生產力、上市速度、員工滿意度及增強的客戶體驗 — 這些指標通常稱為「體驗回報」。

提升影響力超越短期利益：應將 AI 計劃定位為長期轉型的驅動因素。使用案例不僅應滿足當前的營運需求，還應符合數位化轉型或競爭差異化等戰略目標。

2.2 根據業務和負責任的 AI 標準試行

透過雙重視角（業務表現與負責任的 AI 標準）評估試行，確保 AI 計劃同時符合營運目標與負責任的 AI 基準。超過半數 (54%) 接受調查的組織已為其優先使用案例確立了可接受的風險水平。組織應該系統性地記錄這些評估，擷取學習成果，為未來的 AI 專案提供指引。這種結構化的方法為可擴展的 AI 實作奠定了堅實的基礎。

試行時的行動：

設定業務與負責任的 AI 基準：確定營運目標（例如生產力、成本節省）和負責任的 AI 指標（例如透明度、公平性）。

確立風險臨界值：設定風險參數，並建立持續評估的框架，以有效管理並減輕 AI 相關的風險。

擷取並分享學習成果：制定標準化流程，記錄試行成果，以支援透明度並指導未來的擴展工作。

試行

步驟 1：確定優先使用案例

- 從現有的業務優先使用案例中，找出 2-3 個 AI 道德與責任非常重要的試點。
- 針對這些使用案例，確立指標和臨界值，以追蹤業務與負責任的 AI 表現。

步驟 2：根據業務和負責任的 AI 標準試行

- 執行試行，並根據需要進行額外的技術、業務及責任驗證與測試。
- 根據預先定義的業務與負責任 AI 預期的指標和臨界值，評估試行成果，並記錄學習成果納入未來的評估與測試方法。
- 根據試行成果和分析，推進至採購/採用階段。

---

3. 採用：在組織內負責任地整合 AI

採用階段標誌著從試行到全組織整合的過渡。這一階段的重點是從實驗性應用過渡到全面運作的 AI 系統 — 在負責任地部署 AI 的同時，將從試行中汲取的經驗融入現實的實務。

在這一階段，您的員工會主動掌握 AI 在其現有工作流程中的角色。根據他們在試行階段的實際經驗，員工有能力推動 AI 的採用。

3.1 訓練並賦能組織

有效擴展 AI 需要一支知識淵博的員工隊伍，他們既要瞭解 AI 的功能，也要瞭解使用 AI 時所帶來的道德責任。量身打造的訓練計畫應能協助各角色和各部門的員工善用 AI 工具。許多組織 (89%) 都意識到訓練的重要性，其中近三分之二的組織將負責任的 AI 指引納入訓練中。訓練應該將技術能力與問責性、透明度和法規合規等原則相結合。

訓練與賦能時的行動：

使訓練與治理保持一致：在訓練材料中納入負責任的 AI 指引，確保員工瞭解合規、風險管理及透明度要求。

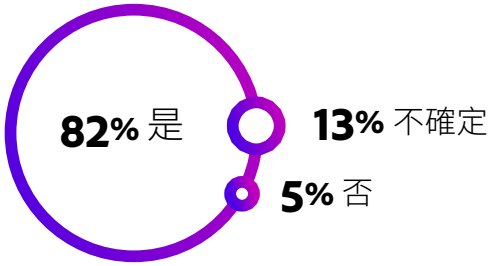
針對角色自訂訓練：針對特定職能的需求制定量身打造的訓練模組，包括業務與負責任的 AI 最佳實務。

3.2 部署時牢記責任

大規模採用 AI 需要建立一個確保負責任使用的治理框架。各組織應將其 AI 計畫與現有的治理政策保持一致，同時持續改進政策，以滿足不斷變化的法規、營運和負責任的 AI 標準。

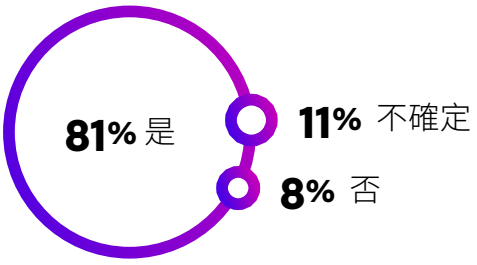
計劃在標準中納入負責任的 AI 考量以便廣泛部署

在受訪者總數中所佔的百分比，依降序排列。



將負責任的 AI 考量納入技術治理工作中

在受訪者總數中所佔的百分比，依降序排列。



負責任的部署行動：

培養問責文化：鼓勵團隊瞭解 AI 對於營運工作流程和利害關係人信任的廣泛影響，並在每個層級灌輸責任感。

持續演進訓練：隨著 AI 治理框架的演進，更新訓練計劃，以反映新的最佳實務與法規變更。

採用

步驟 1：訓練並賦能組織

- 根據負責任的 AI 原則，針對如何以及何時使用 AI，制定使用案例特定的指引。
- 在解決方案推出時，為相關團隊提供包含最佳實務的全面訓練。
- 在整個組織內慶祝並分享成功案例。

步驟 2：部署時牢記責任

- 針對每項技術，編纂核心採用要求（例如，業務影響、整合難易度及風險緩解）。
- 與業務領導者緊密合作，根據需要協調取捨。
- 將關鍵 AI 與責任考量融入現有的治理框架（例如，存取、控制、角色）。

4. 監控：持續監督與改進

隨著 AI 系統進入全面部署階段，持續監控與改進變得至關重要。**監控階段**強調即時追蹤、嚴格的效能審核，以及主動的風險管理方法，以確保 AI 系統持續有效、合規並符合組織目標。透過融入負責任的 AI 指標並建立結構化的審核流程，組織可以適應不斷變化的法規需求和新出現的風險，同時培養持續的信任和營運價值。

4.1 根據業務和負責任的 AI 基準監控效能

一流的技術成果監控結合了自動化效能追蹤與人類專業知識。儘管許多組織已經採用即時監控工具來評估 AI 系統效能，但加入人工監督後，這些工具的有效性會顯著提升。人工團隊更有能力分析數據、發現風險，並針對必要的調整做出明智的決策。

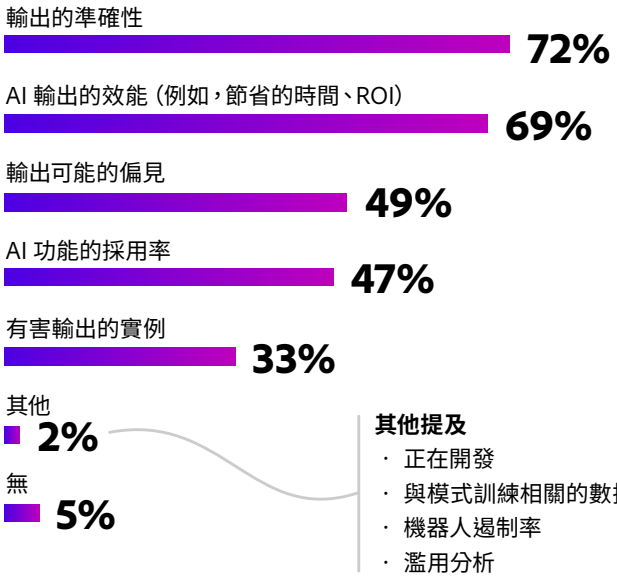
雖然 69% 的組織使用即時監控工具，但當這些工具與人工判斷相結合時，效果會顯著提高。許多組織將技術指標視為優先考量，72% 的組織著重於準確性，69% 的組織著重於 ROI，然而負責任的擴展還需要關注道德層面的問題。融入人類監督能確保透明度和可預測性，從而在內部和外部建立信任。監控可進一步主動檢測偏見，49% 的組織會追蹤這項指標，33% 的組織會監控有害的輸出。如果沒有持續、主動的監控，AI 系統可能會損害誠信與信任度。透過

完善效能監控以應對技術與道德風險，公司能夠保護其品牌、建立使用者信心，並為負責任地擴展 AI 打下堅實基礎。

這種技術與人員間的協作可讓組織及早發現潛在風險，例如數據不準確、新出現的偏見或合規失誤。

追蹤技術效能與有效性的 AI 特定考量

在受訪者總數中所佔的百分比



其他提及

- 正在開發
- 與模式訓練相關的數據隱私
- 機器人遏制率
- 濫用分析

近 1/3 的組織

沒有足夠的員工支援技術效能指標和業務成果的持續改進

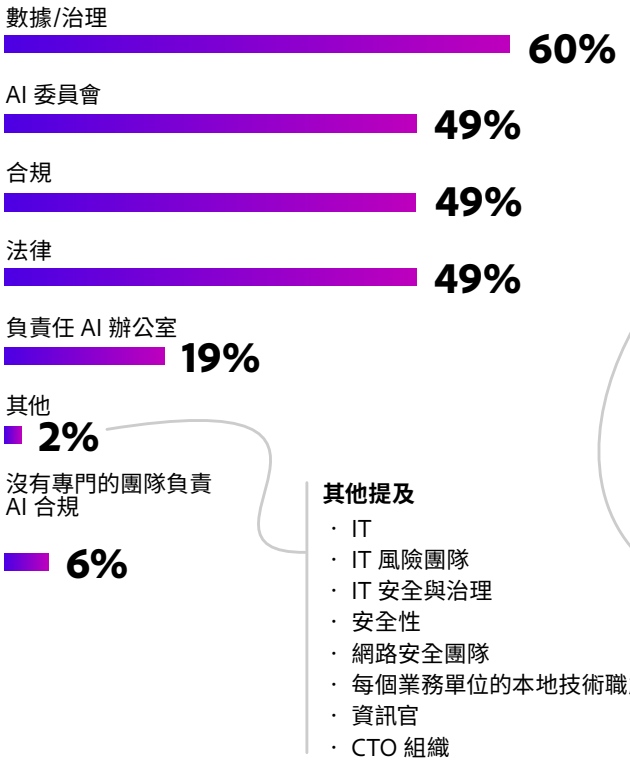
4.2 持續風險管理

AI 的風險管理是一個持續的過程，應隨著 AI 系統的發展而不斷演進。建立結構化、跨職能的 AI 風險管理方法，使組織能夠主動應對業務風險與聲譽風險。定期審核應涉及整個組織的利害關係人，包括數據科學家、業務領導者及法律/合規官員，確保對技術效能與負責任的 AI 目標進行全面的評估。

AI 風險管理是一個持續的過程，會隨著技術進步而不斷演進。60% 的組織涉及數據和治理團隊，49% 的組織包括 AI 委員會、合規和法律團隊 — 強調了跨職能協作的必要性。這種主動的方法使組織能夠與內部價值和外部期望保持一致。其「原因」是為了建立能夠適應法規變化的適應力。68% 的組織強調在風險管理中納入負責任的 AI，因此全面的文件和持續的風險評估至關重要。

追蹤技術效能與有效性的 AI 特定考量

在受訪者總數中所佔的百分比

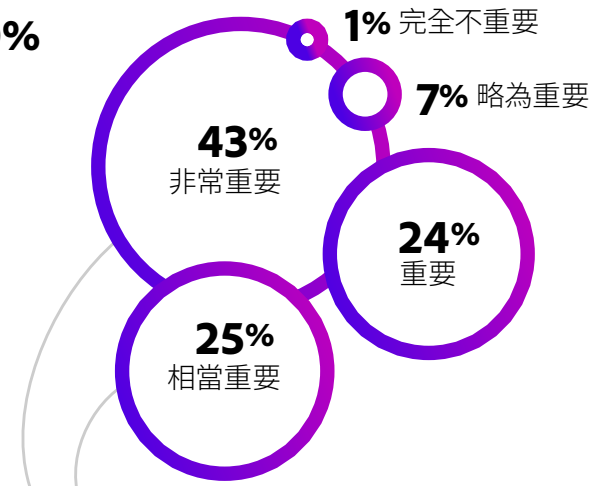


其他提及

- IT
- IT 風險團隊
- IT 安全與治理
- 安全性
- 網路安全團隊
- 每個業務單位的本地技術職能
- 資訊官
- CTO 組織

負責任且合乎道德的 AI 使用考量對確保合規的團隊的重要性

表示有指定團隊專注於監控 AI 法規和標準的企業所佔百分比



68%
非常重要
+
相當重要

透過建立嚴格的效能指標和培養持續風險管理的文化，組織可以確保 AI 部署始終與業務目標和負責任的 AI 優先事項保持一致。主動監控，結合跨職能的審核與詳盡的文件，使組織能夠充滿信心和責任感地引領 AI 驅動的轉型。

風險管理的主要動作：

建立跨職能風險審核：建立由數據科學家、合規官員和法律專家參與的定期風險評估，以根據即時數據識別新出現的風險。

一致地追蹤與報告：持續收集員工和最終使用者的意見回饋，以發現可用性問題、偏見或意外行為。

監控

步驟 1：監控效能

- 定義並追蹤較長期的 AI 效能指標，包括業務和責任目標。
- 建立持續的審核並討論結果，以持續提高業務效能，同時保護組織的責任原則。

步驟 2：部署時牢記責任

- 指定並授權角色以追蹤不斷演進的 AI 法規和標準（例如，新加坡的 AI-Verify、美國國會提案等），並確保公司標準相應更新。
- 制定一個持續識別和降低與使用 AI 相關風險的流程。
- 更新有關組織如何遵守公司標準的文件。

ADOBE 案例研究

內部使用生成式 AI

在 Adobe，我們將生成式 AI 視為一項變革性技術，它能夠提升人類創造力，而非取代人類創造力。我們鼓勵內部以負責任的態度探索生成式 AI 技術，並與我們自己的 AI 道德原則的問責性、責任感和透明度保持一致。

2023 年 6 月，Adobe 成立了一個由 CIO 和 CHRO 讚助的跨職能內部工作組，旨在幫助員工在 Adobe 內部以安全、負責任和靈活的方式探索和使用生成式 AI。該工作組與全公司的領導者和主題專家合作，透過瞭解生成式 AI 使用的假設範圍、建立適當的指引以及簡化實驗，著重於為基層員工實驗提供深思熟慮的方法。這項計畫已經形成了四個基於角色的工作組，代表整個 Adobe 中不同的生成式 AI 使用案例，並建立了接收流程、生成式 AI 風險容忍框架以及使用案例審核藍圖，其中考慮了不斷演進的道德、安全、隱私及其他法律問題。此外，我們還提供基於特定使用案例的核准生成式 AI 工具和模型清單，以及員工使用生成式 AI 的指引。2024 年 3 月推出的供應商生成式 AI 指引包括有關使用生成式 AI 及所選產品功能的訓練課程。

此項計劃的實作幫助簡化了流程，實現了更快速的實驗，並在可能的情況下進行規模化應用，能夠評估全公司的生成式 AI 現狀。Adobe 會繼續在企業內部促進分享學習成果與分析，創造集體探索的協作生態系統。隨著生成式 AI 在自有產品中的擴展、生成式 AI 技術與模型的普及，以及法律與法規指引的演進，本計畫也在不斷演進。實驗審核受到監控 — 團隊正在開發核准後的實驗追蹤，包括投入生產規模的實驗。

III. 將最佳實務融入組織中

為了負責任地採用、監控和最佳化 AI 系統，組織應將重點放在幾個營運領域：提供全面的員工指引、嚴格評估供應商，以及確立強大的 AI 治理工具。這樣，公司不僅能確保其 AI 計劃符合不斷變化的法規標準，還能基於現有的治理和風險管理工作，並與培養信任、透明度和問責性的實務保持一致。

本節概述將這些最佳實務融入日常營運的實用步驟。

1. 員工使用指引：

針對組織的特定需求和風險量身打造 AI 使用指引，對於確保負責任的部署至關重要。這些指引應能協助員工掌握法規標準和治理協定，讓 AI 技術與數據安全性、透明度和問責性的承諾保持一致。

1.1 數據敏感度：

明確規定何時數據處理應在本機或嚴格的存取控制下進行，以防止未經授權的存取；這也意味著避免使用可能生成或操縱敏感輸出的提示。此指引可保護專屬資訊，並維持對數據隱私法規的合規。

1.2 AI 使用的透明度：

隨時公開 AI 的參與情況，例如當 AI 用於建立內部文件、客戶介面或外部溝通時。這一實務可以培養問責性，並維持對 AI 所生成內容真實性與可靠性的信任，從而維護公司的品牌和聲譽。

1.3 帳戶管理政策：

針對需要註冊帳戶的生成式 AI 工具，建立明確的使用政策，包括規定是否可以使用組織的電子郵件帳戶、明確規定哪些工具經核准用於業務目的，以及不鼓勵將個人帳戶用於工作內容。這可防止未經授權的使用，並幫助與組織更廣泛的資訊安全計畫保持一致。

透過根據組織背景量身打造這些指引，員工可以自信且負責任地使用生成式 AI 工具，助力打造一個創新與誠信並重的環境。

2. 供應商評估：範例問題

評估 AI 供應商需要提出資訊豐富的問題，並瞭解您要尋求的答案，以確保他們的系統符合負責任的 AI、法律和法規標準。以下問題旨在提供基準評估，使組織能夠對潛在合作夥伴關係做出明智的決策，並降低與 AI 採用相關的風險。

主題	供應商問題	推理	Adobe 範例
數據來源與用途	「在開發與訓練 AI 系統時使用了哪些特定類型的數據？」	深入瞭解用於訓練 AI 模型的數據來源、性質和範圍。此問題可確保供應商的數據實務符合購買者負責任的 AI 標準，並符合法律要求。	**Firefly：**Adobe 不會將企業使用者內容（包括 Firefly 輸入與輸出）納入用於訓練 Firefly 基礎模型的數據集中。
智慧財產權合規	「是否使用了任何可能具有著作權、智慧財產權或授權限制的數據集？」	這項詢問可確認所有數據來源合法，並且是透過核准的機制獲得，以避免潛在的法律爭議。	**Firefly：**客戶使用 Adobe 認證登入後，可隨時在 stock.adobe.com/Dashboard/LicenseHistory 檢視授權歷史資訊。
訓練數據和邏輯透明度	「您能否詳細解釋開發 AI 系統時所套用的訓練數據和邏輯？」	這些方面的透明度可辨識潛在的偏見，並提供對模型推理過程的理解，這對於評估其可靠性與公平性至關重要。	**AEP AI 助理：**Adobe 不會使用任何客戶數據來訓練或微調 Azure OpenAI 服務。
輸出清晰度	「您能否提供 AI 系統輸出的簡單語言描述？」	確保輸出能讓非技術審核人員理解，從而實現有效的決策與負責任的使用。	**Firefly**：Adobe 自動為某些 Firefly 生成的資產生成內容證書，以幫助提供透明度，表明該資產是使用生成式 AI 建立的。
人類監督	「如果人工審核是 AI 系統的一部分，那麼人類參與的程度和性質為何？」	瞭解自動化流程與人工判斷之間的平衡，可評估系統的運作動態，並找出可能需要人工干預以維護品質與責任標準的領域。	**Acrobat AI 助理：**Adobe 嚴格限制可存取此資訊的人員，僅限於直接參與 Adobe 生成式 AI 服務開發的少數受過訓練的 Adobe 員工。
公平性與偏見評估	「AI 系統是如何評估偏見的，這些評估的結果是什麼？」	這個問題展示了供應商對公平 AI 實務的承諾，以及他們檢測和減緩可能對不同人口群體不成比例地產生影響的偏見的方法。	**Acrobat AI 助理：**Adobe 團隊進行測試，以減少我們的生成式 AI 產品可能產生的偏見和有害結果。請參閱為企業建置的生成式 AI 解決方案簡介。
風險緩解	「是否對潛在有害輸出進行過評估，以及採取了哪些措施來減輕這些風險？」	評估供應商在識別和解決潛在負面結果方面的主動措施，展示其對傷害的補救。這意味著 AI 系統經過測試，可確保安全和負責任地運作。	**AEP AI 助理：**Adobe 使用內部開發的內容過濾器：(a) 判斷 AEP 中 AI 助理的輸入（提示）是否符合 Adobe 的生成式 AI 使用者指引，以及 (b) 過濾掉任何違反這些指引的生成回應（例如，仇恨言論和粗俗語言）。

透過提出這些有針對性的問題，組織可以評估 AI 供應商，並做出符合其 AI 責任標準與營運優先事項的明智選擇。此方法可管理風險，並確保與 AI 供應商的關係建立在透明、合規與負責任的創新基礎上。

3. AI 治理工具

實作強大的 AI 治理，確保 AI 系統的開發、部署和監控方式符合組織價值與法規標準。目前已有許多法規標準，例如歐盟 AI 法案以及新加坡的 AI Verify 等框架。在美國，公司應遵循各州的全面隱私法律和 NIST AI 風險管理框架，因為其很可能是未來法規的基礎。

以下治理工具可協助組織管理 AI 風險，並增強透明度、問責性和安全性：

AI 庫存	建立 AI 系統清單，根據風險概況和戰略優先事項對其進行分類，作為集中的存儲庫。	您是否已經記錄了 AI 使用案例並將相關風險分類？
意見回饋機制	建立強大的意見回饋通道，以擷取 AI 開發團隊以外的人員（例如最終使用者、客戶或公眾）的見解。	有哪些意見回饋通道可用來擷取最終使用者、客戶或公眾的見解？
系統限制文件	記錄 AI 系統限制，包括 AI 模型的知識缺口資訊，以及能夠可靠使用其輸出的情境。	您是否已經記錄了您的 AI 使用案例的已知或預期限制？
內容來源	追蹤並驗證 AI 相關數據的來源、歷史和修改，包括追蹤訓練數據來源、使用的演算法和轉換。	您如何追蹤 AI 相關數據的來源、歷史和修改，包括從建立到最終使用的數據來源和轉換？
AI 測試與紅隊測試	評估風險，如訓練數據的無意暴露、易受逆向工程影響以及與模型擷取相關的風險。	您已經採取了哪些測試協定？這些協定如何處理特定的 AI 使用案例風險？
安全軟體開發	AI 系統應整合至組織的安全軟體開發生命週期，遵守編碼與部署的既定最佳實務。	您的 AI 使用案例如何整合現有的安全軟體開發協定？
訓練	訓練應涵蓋相關政策、程序與合規要求，讓利害關係人掌握有效管理 AI 風險並依照組織標準行事的知識。	您是否參加了 AI 治理與風險管理訓練？

將這些最佳實務融入現有流程，確保以負責任的方式採用、管理和最佳化 AI 技術。透過整合這些活動，例如提供詳盡的員工指引、嚴格評估供應商，以及確立全面的 AI 治理工具，組織可建立符合治理標準與法規要求的彈性實務基線，使其不僅能夠應對當前挑戰，還能負責任地擴展 AI 以取得未來的成功。

IV. 負責任的實作建置負責任的創新

要在公司中最大限度地發揮 AI 的潛力，需要採購負責任的技術、確立明確的使用指引、開發專門的訓練，以及部署強大的治理。這種方法可推動業務價值，確保 AI 計畫符合法規預期，並維持負責任的實作標準，在整個組織中融入負責任的 AI 文化。

一致的監督和適應讓 AI 計畫不會偏離正軌。透過規定效能指標、進行定期評估，以及主動管理風險，組織可以領先於法規變化，並維持 AI 專案的完整性。

這一框架可讓組織在負責任的 AI 環境中扮演領導角色，與 AI 一起邁向未來。此方法著重於影響、整合與誠信，為永續創新與持久成功鋪平了道路。