AI 安全缺口：資安團隊的零信任實施指南

這是適合從業人員的指南，旨在指導如何運用零信任原則，保障 AI 的整個生命週期安全，涵蓋從影子 AI 探索到代理執行時期保護。30/60/90 天實施藍圖和領導層報告的關鍵績效指標。

電子書



©2026 Zscaler, Inc. 保留所有權利。

介紹

資安團隊知道他們面臨 AI 問題，但是大多數人目前還無法回答，究竟有多嚴重。

您目前環境中正在運行哪些人工智慧 (AI) 應用程式？ 您的開發人員透過整合開發環境 (IDE) 外掛程式和命令列介面 (CLI) 工具查詢哪些模型？哪些資料正在透過提示詞進行傳輸？其原始碼、客戶個人識別資訊 (PII) 或內容是否受監管披露規則的約束？哪些代理能夠代表使用者自主行動，及可以存取哪些系統？

AI 在您的環境中實際運作的情況與您能夠看到和管理的範圍之間的缺口，才是風險的真正所在。風險並非存在於理論上的攻擊場景中，而是存在於日常生活中未經批准、出於好意卻未經資安部門批准的 AI 工具的使用中，因為沒有人想到要詢問這些工具的使用情況。影子 AI 和嵌入式 AI 在已批准的軟體即服務 (SaaS) 平台中的 AI 功能是造成這種缺口的主要來源。例如，IT 部門批准用於特定用途的工具，依據預設就內建 AI 助手，現在正在處理內部資料，而沒有任何安全審查。

本指南中的資料和統計來自 *Zscaler* 於 *2026* 年發布的《*ThreatLabz 2026* 年 *AI* 安全報告》，該報告基於對 *9893* 億次資料的分析。*AI/ML 2025* 年 *1* 月至 *12* 月期間，*Zscaler Zero Trust Exchange* 上觀察到的交易。除非另有說明，所有資料均反映該研究期間的情況。

©2026 Zscaler, Inc. 保留所有權利。

零信任提供一種基於原則的應對措施：

不要將任何應用程式、使用者、提示詞或連接器視為本質上值得信任。在每一層都強制執行基於身分識別和語境的存取控制。持續驗證。發生意外時，盡量縮小爆發半徑。挑戰在於定序。在政策落實之前，必須先提高可視性。在自動化功能處於頂層之前，存取控制必須保持穩定。如果建構序列錯誤，則產生的控制措施會阻礙合法工作，或無法阻止實際風險。

83%	18,033 TB	39%	93%
企業年增率AI/ML 交易－如今已達到 基礎設施規模。	資料量將透過 AI 應用程式傳輸 (2025 年統計)，這超過大多數計 畫能夠追蹤的資料量。	AI 交易遭到主動攔截， 治理壓力已經浮 現在檯面上。	年增率的 AI 資料 傳輸 量已超過資 安計畫容量

來源：Zscaler ThreatLabz 2026 年 AI 安全報告 (第 5–6 頁)。

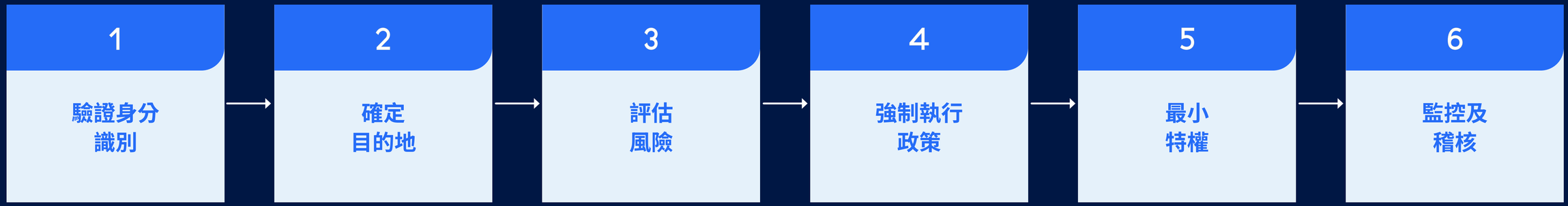
本指南涵蓋的內容

1. 七項零信任原則專門適用於 AI 系統、代理和資料流。
2. 分階段進行 30/60/90 天實施藍圖內建相依性順序
3. 涵蓋新興 AI 風險領域：影子 AI 、提示詞注入、代理系統、模型語境協議 (MCP)/代理之間 (A2A) 的流量、嵌入式 SaaS AI 和 AI 基礎設施態勢
4. 用於向領導層報告 AI 安全計畫進展的關鍵績效指標模型

©2026 Zscaler, Inc. 保留所有權利。

零信任原則套用於 AI

以下七項原則構成實施藍圖中所有事項的決策架構。各自都是用來評估控制措施的決策工具。這種定序方式反映資安團隊在實務上實際應用這些措施的方式。



按序列應用。每個原則都建立在前一個原則的基礎上。

©2026 Zscaler, Inc. 保留所有權利。

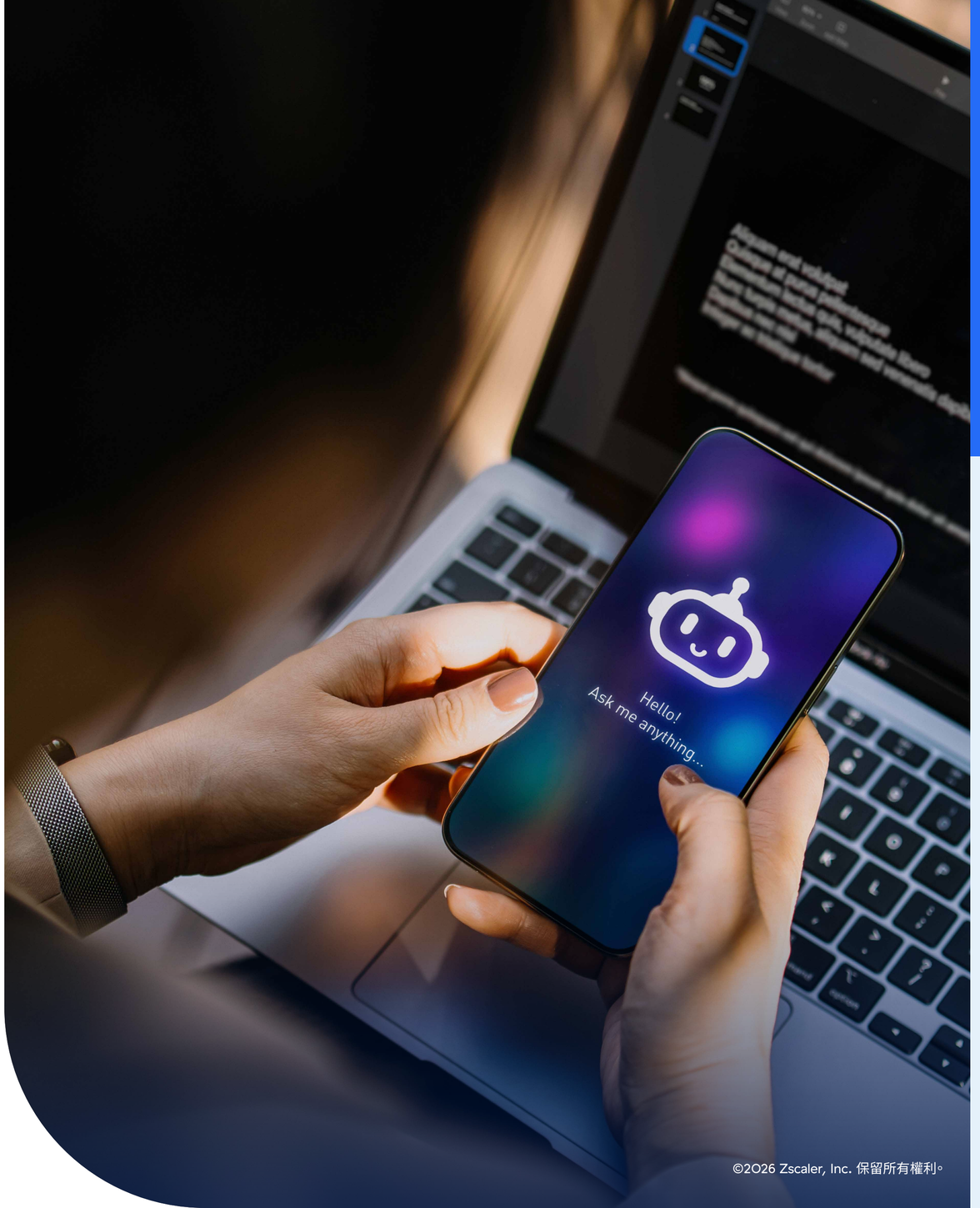
原則 1：
驗證使用者、開發人員和服務的身分

AI 存取權始於身分識別。每個存取生成式 AI (GenAI) SaaS 工具的使用者、每個從 IDE 外掛程式查詢模型端點的開發人員以及每個連線至 AI 基礎架構元件的服務帳戶都應該透過一致的身分識別控制平面進行身分驗證。單一登入 (SSO) 和多因素身分驗證 (MFA) 是基本功能，但細節很重要。

工作團隊和開發人員群體具有不同的風險狀況，需要不同的政策對待。行銷分析師使用 ChatGPT 撰寫內容草稿所帶來的資料外洩風險，與軟體工程師的 IDE 透過連線至 MCP 伺服器並可存取內部儲存庫的 GitHub Copilot 外掛程式路由程式碼補全所帶來的資料外洩風險不同。AI 代理的服務間身分驗證是大多數組織尚未準備好應對的新興邊緣案例。

關鍵控制

- AI SaaS 和 AI 開發人員工具的 SSO 和 MFA 強制執行
- 針對員工使用 GenAI 和開發人員使用 AI 工具（IDE、CLI、MCP 整合）制定不同的身分識別政策
- **AI 代理的服務帳戶控制：**最小特權範圍、短期憑證和稽核日誌
- **裝置態勢作為身分識別訊號：**受管理端點與自帶設備 (BYOD) 使用者享有不同的存取權



©2026 Zscaler, Inc. 保留所有權利。

原則 2：
確定經批准和未經批准 AI 的目的地

並非所有 AI 應用程式都相同，全面封鎖一切既不切實際也不具安全效率。**目標是分層目的地政策：**經批准的工具擁有完全存取權並進行內嵌式檢查；容許的工具在受到監控的情況下存取，並設置使用限制；未經批准或高風險的工具則被攔截或隔離。AI 攻擊面比大多數組織預想的要大得多，隨著新的 AI 功能不斷整合到現有平台中，AI 攻擊面在一年內成長了四倍。

關鍵控制

- 分層目的地政策可依應用程式和使用者群組進行允許、警告、隔離或攔截。
- 動態應用程式探索功能可使新的 AI 工具出現在流量中時使之浮現
- 針對 GenAI SaaS、企業級 SaaS 中的嵌入式 AI 和 AI 開發人員工具制定不同的追蹤政策
- 針對新的 AI 工具請求，制定審批工作流程，促成使用者能夠明確傳遞路徑，而無需繞過各種控制措施。

超過 3,400

不同 AI 應用程式在各企業追蹤到的流量中在一年內增加了 4 倍，並且每月都在增加。

Zscaler ThreatLabz 2026 年 AI 安全報告（第 5 頁）



©2026 Zscaler, Inc. 保留所有權利。

原則 3：
基於語境的決策評估風險

AI 的存取權決策絕不應該是非此即彼。同一使用者存取相同 AI 工具會取決於情境（裝置、資料敏感度和操作）而帶來不同的風險。有效的風險評估需要預先定義敏感資料，包括原始程式碼、個人識別資訊 (PII)、支付卡產業 (PCI) 資料、受保護的健康資訊 (PHI)、未發布的財務業績以及機密併購資訊 (M&A) 詳情。企業流量中出現的主要違規類別可以準確地讓您知道員工提交的內容：姓名和國民身分識別資料、原始碼和醫療記錄。

關鍵控制

- 明確區分原始碼的「紅色資料」類別（PII/PCI/PHI）、秘密和未發布的計畫
- 依使用者群組、部門、裝置態勢和資料敏感度進行風險評分
- 基於語境訊號的動態政策調整，而非僅僅是靜態允許/封鎖規則
- 針對包含文件上傳的 AI 互動和僅包含文字提示詞的 AI 互動，應採取不同的風險處理方法。

超過 4.1億

與 CHATGPT 相關的資料外洩防護 (DLP) 違規行為類型排名：姓名洩漏、SSN、原始碼、醫療資料*Zscaler ThreatLabz 2026 年 AI 安全報告*（第 *6、15* 頁）

Zscaler ThreatLabz 2026 年 AI 安全報告（第 *6、15* 頁）



©2026 Zscaler, Inc. 保留所有權利。

原則 4：
按工作階段強制執行內嵌式政策

日誌會告訴您發生了什麼事。等您找到記錄顯示開發人員向未經批准的模型提交專有原始碼時，資料已經消失。大多數組織透過部署 DLP 來對上傳到 AI 工具的檔案進行檢查，卻不對提示詞文字進行檢查，從而加劇了這個問題。開發人員直接在提示字元視窗中輸入原始程式碼會清除這些控制措施，而不會觸發任何警報。資料傳輸量的成長表明，透過這些不受保護的管道傳輸的資料量非常龐大。

關鍵控制

- 對提示詞、回應和檔案上傳進行內嵌式資料外洩防護 (DLP)，而非僅僅在 URL/類別等級
- 在政策強制執行點可採取允許、警告、編校、封鎖和隔離等的動作
- 利用語境感知偵測技術，對領先的 GenAI 應用程式進行提示詞分類
- 對離題、受限制和競爭敏感的互動進行內容審核
- 瀏覽器隔離機制用於允許但有風險的 AI 工具：無需完全信任即可存取

93%

年增率的企業 AI 資料傳輸成長速度超過了大多數資安計畫能夠追蹤或控管的速度。

Zscaler ThreatLabz 2026 年 AI 安全報告（第 5、14 頁）



©2026 Zscaler, Inc. 保留所有權利。

原則 5：
最大幅度減少接觸和暴險

AI 系統建立的存取路徑通常位於傳統身分識別和網路模型之外。索增強生成 (RAG) 流程可以使模型存取文件語料庫。代理可以連線至內部工具、資料集、向量儲存或 MCP 伺服器。即使這些權限沒有出現在標準的存取權審核中，也和任何使用者的存取權一樣真實且有效。

最大幅度地減少存取和暴險意味著減少 AI 系統可以接觸到的內容，並限制 AI 組件之間的通訊方式。代理、資料管道、RAG 組件和後端資料來源應該只透過明確核准的路徑進行連線。如果沒有最小特權範圍界定和分段，權限過高的代理或遭入侵 AI 組件可能會成為橫向移動路徑。

關鍵控制

- 針對 AI 應用程式、代理、模型連接器、資料集、向量儲存和 AI 基礎設施組件進行基於角色的存取範圍控制
- 定期審查並減少對敏感資料集和 AI 連線系統的過度特權存取
- AI 代理的最小權限集：定義 AI 代理可以存取的內容，限制其他所有內容
- AI 服務、代理平台、資料管道、RAG 組件和後端資料來源的分段
- 限制 AI 組件之間的東西向存取，僅允許通過明確定義的路徑
- 對 MCP 和 A2A 流量進行協定感知檢查和日誌記錄

最大幅度地減小每個 AI 存取路徑的爆發半徑

將每個 AI 連接器、代理權限、RAG 管道、向量儲存、MCP 伺服器和組件之間的連線都視為暴險點。確定存取權範圍，對其通訊區域進行分段，檢查流量，並記錄活動。



©2026 Zscaler, Inc. 保留所有權利。

原則 6：
持續監控和可稽核性

持續驗證是一種運作模式，而非達到某個里程碑後便可拋諸腦後的事項。對於 AI 系統而言，這意味著記錄提示詞和回應、針對政策違規進行即時警示、提供使用情況儀表板，以及無需在五種不同工具間手動關聯日誌的調查工作流程。**這背後的監管壓力不容忽視：**包括《歐盟人工智慧法案》(EU AI Act)、美國國家標準與技術研究院人工智慧風險管理框架 (NIST AI RMF) 在內的新興 AI 治理框架，均要求具備可證明且有文件記錄的控制措施。

關鍵控制

- 內嵌式 AI 互動的提示詞和回應日誌記錄
- 按使用者、部門和 AI 應用程式劃分的使用情況儀表板，並疊加風險訊號
- 針對 AI 相關政策違規和異常情況的警報和調查工作流程
- 符合稽核要求的報告：正在使用哪些 AI、套用了哪些控制措施以及發生了哪些違規行為
- 治理態勢與 NIST AI RMF、歐盟 AI 法案及其他適用框架的對應



©2026 Zscaler, Inc. 保留所有權利。

1	2	3	4	5
反應型	**可見的**	**受控管**	**自動化**	**連續的**
無清查臨時攔截	已知的 AI 足跡基準政策	以角色為基礎的政策治理工作流程	自動化指導持續合規	紅隊演練 + 執行時期護欄

專案成熟度 →

第 1–30 天
建立可視性

您無法保障您看不見的 AI 安全。前 30 天著重於建立足夠的清查和基準控制，以便做出明智的決策，而不是全面封鎖

第 1–30 天 建立可視性	第 31–60 天 建置政策成熟度	第 61–90 天 自動化和驗證
• 部署 AI 應用程式探索 • 定義紅色資料類別 • 基準存取政策 • 首個領導力儀表板	• 基於角色的存取政策 • 治理工作流程已上線 • AI–SPM 推出 • 私有 AI 保護	• 自動化使用者指導 • 持續合規 • 紅隊演練節奏 • 閉環護欄

30/60/90 天實施藍圖概覽



©2026 Zscaler, Inc. 保留所有權利。

30 天實施清單

- **建立定期的 AI 工具審查節奏：**探索在流量中發現的新工具、收到的異常請求以及做出的政策變更
- 建構首個面對領導層的 AI 安全儀表板：AI 應用程式足跡（經批准與影子）、按類別劃分的 DLP 事件、政策措施與趨勢線
- 針對最常見的 AI 相關資安事件編制事件處理手冊：提示詞訊息中包含敏感資料、存取被封鎖的目的地、上傳大檔案到 AI 工具等
- **集中管理 AI 安全日誌：**在一個儀表板視圖中顯示使用者活動、應用程式使用情況、 提示詞/回應事件、政策措施和違規行為

實施

- **利用前兩週的遙測資料審查並調諧資料外洩防護 (DLP) 政策：**減少誤報，提高已確認資料暴險模式的偵測能力
- **啟用內容審核功能，以確保內容符合規範：**包括離題互動、受限制主話題、競爭敏感的內容以及不當輸出
- **在允許但有風險的層級中，為 AI 工具增添瀏覽器隔離：**使用者可以存取該工具，但無法上傳、下載和處理資料。複製/貼上受限於瀏覽器層
- **將資料外洩防護 (DLP) 覆蓋範圍擴展至完整的紅色資料範圍：**原始程式碼（包括其他語言和模式）、個人識別資訊 (PII)（超出常見格式）、支付卡產業資料安全標準(PCI) 持卡人資料、個人健康資訊 (PHI) 和機密商業內容

擴展資料保護

- **發佈 AI 可接受使用情況的交通燈指南：**綠色（經批准）、黃色（限制使用）、紅色（已封鎖，並說明原因）。公開例外申請流程。
- 啟用提示詞的內嵌式資料外洩防護功能，首先針對紅色資料類別啟用高置信度偵測規則：原始碼模式、憑證格式、PII 結構描述
- **限制未經批准工具上的高風險操作：**上傳檔案、大型複製/貼上區塊，以及將 AI 輸出下載到非受管理儲存裝置
- **強制執行目的地政策：**對於明顯未經批准或高風險的 AI 目的地，按使用者群組和應用程式發出警告或封鎖。

套用基準控制

- **建立 AI 風險基準清查：**哪些應用程式正在存取哪些資料類型，以及哪些使用者群體是 AI 使用量最大的使用者。
- **明確定義紅色資料類別：**原始碼、個人識別資訊 (PII)、支付卡產業資料 (PCI DSS)、個人健康資訊 (PHI)、機密資訊和憑證、未公開的財務資料、機密併購或策略計畫

定義哪些內容需要保護

- 啟用高流量 AI 應用程式的提示詞可視性，以便瞭解哪些資料正在透過提示詞和回應進行傳輸
- 繪製哪些 AI 工具與內部系統整合，以及存取哪些資料類型
- **使影子 AI 浮現：**從未正式批准但積極使用的工具
- 部署 AI 應用程式探索，以清查 GenAI SaaS、企業平台中的嵌入式 AI 以及整個環境中正在使用的 AI 開發人員工具

需要考慮的權衡

先採取警告政策而不是硬性封鎖，可以減少摩擦，並在 DLP 規則開始攔截合法工作之前發現誤報。代價是，即使在預警階段，資料暴險事件仍然可能發生。權衡一下您對紅色資料風險的承受能力，如果原始程式碼外洩是一個重大風險，則對程式碼模式偵測採取嚴格限制措施，即使經歷早期出現誤報的痛苦也是值得的。

30 天交付成果

- AI 應用程式清查及風險等級分類（允許、容忍、禁止）
- 初始存取政策適用於風險最高的 AI 應用程式和使用者群組
- 針對紅色資料類別啟用及時 DLP，初始調諧已完成
- 首個具備 AI 足跡和風險指標的領導儀表板
- 已發布 AI 可接受使用指南和例外申請流程

第 31–60 天
提升政策成熟度

第 31–60 天在可視性和初始控制的基礎上，建立成熟的政策模型、有效的治理工作流程和 AI 基礎設施清查。



©2026 Zscaler, Inc. 保留所有權利。

60 天實施檢查清單

完善存取政策模型

- **制定部門專屬的 AI 存取政策：**人力資源、財務、工程和銷售部門對 AI 工具的需求和資料暴露風險各不相同
- **分層條件政策控制：**受管理裝置與自帶裝置 (BYOD) 存取層級、地理存取訊號、裝置態勢風險評分
- **將 AI 應用程式的存取權正式劃分為三個層級：**經批准層級（完全存取權，可進行內嵌式檢查）、容忍層級（受監控的存取權，使用受到限制）、禁止層級（被封鎖，但有例外路徑）。
- 審查並終止最初 30 天內提出的例外申請，使其符合目前更成熟的政策模型。

建立治理流程

- **將 AI 治理要求映射到內部政策和適用框架：** NIST AI 風險管理框架、歐盟 AI 法案、HIPAA（若涉及PHI）
- **規範 AI 工具審核流程：**提交、安全審查、風險分類、有條件核准和定期重新驗證
- 為與 AI 應用程式相關的資料類型建立標準敏感度標籤，並為每個標籤制定明確的處理規則
- **明確 AI 安全事件的上報路徑：**由誰負責回應，什麼情況構成需要通報高階主管的事件，以及需要保留哪些證據。

清查 AI 基礎設施

- 清查 AI 模型、代理、服務、資料集、向量儲存和管道，並建置包含從資料集到執行時期使用溯源資訊的人工智慧物料清單 (AI-BOM)
- **執行 AI-SPM 評估：**識別 AI 基礎架構元件中的組態錯誤、權限過大、敏感資料暴險和漏洞。
- **優先處理高風險 AI-SPM 補救：**權限過高、敏感資料暴險以及不受限制的 AI 組件存取路徑。
- **驗證對主要 AI 平台的覆蓋範圍：**雲端託管模型、企業部署的代理平台、MCP 伺服器整合和非託管 AI 基礎設施

保護私有 AI 應用程式

- **為私有 AI 應用程式的控制功能增添提示詞注入檢查：**精心設計的對抗性輸入旨在覆寫模型指令或提取敏感資訊
- **為 RAG 管線實施資料投毒監控：**偵測對文件語料庫的未經授權修改，這些修改可能會影響模型輸出
- **啟用針對具有異常偵測功能的私有 AI 系統的提示詞/回應日誌記錄：**異常查詢模式、非預期輸出內容和超出範圍的互動
- 為高敏感度私有 AI 應用程式建立模型互動審查工作流程

©2026 Zscaler, Inc. 保留所有權利。

60 天交付成果

- 使用已上線的條件強制執行，完善以角色為基礎的存取政策模型
- AI 治理工作流程運作：工具審核、敏感度標記、升級路徑定義
- AI-SPM 可視性基準已完成，並附有優先排序的補救待辦項目
- AI-BOM 涵蓋主要的 AI 基礎設施組件
- 針對優先順序最高的內部 AI 應用程式實施私有 AI 保護措施
- 已定義合規報告節奏，首份報告已出爐

第 61–90 天
自動化和驗證

第 61–90 天使計畫能夠自我維持：自動化強制執行、持續合規性監控以及透過紅隊演練進行對抗性驗證。



©2026 Zscaler, Inc. 保留所有權利。

90 天實施清單

自動強制執法

- **針對違反政策的行為實施自動化使用者指導：**在上報到封鎖之前，先根據具體情境發出警告並提供指導
- **整合重複違規行為的 ITSM 或 SOAR 工單系統：**對於重複觸發相同控制措施的使用者，自動建立調查工單
- **針對高風險行為組態自動隔離和隔離觸發：**提交大量敏感資料、存取被禁止的 AI 目的地、異常提示詞模式
- **建立回饋機制：**長期追蹤政策違規趨勢，辨識出需要調整政策而非強制執行失效的行為類別

鎖定合規報告

- **組態 AI 合規性態勢的持續監控：**哪些控制措施處於啟用中狀態，哪些資產受到監控，以及存在哪些覆蓋盲點
- **將控制措施對應到適用的框架：** NIST AI RMF（包括用於生成式 AI 風險的 NIST AI 600–1）、歐盟 AI 法案義務、GDPR 以及HIPAA（若涉及PHI）
- **建構符合稽核要求的合併報表：**AI 應用程式使用摘要、及時採取 DLP 措施及結果、已授多和已過期的政策例外、AI–SPM 發現補救狀態
- **建立每季一次的 AI 安全計畫審查節奏：**審查政策執行、覆蓋範圍缺口、新的風險領域以及藍圖調整。

紅隊演練和強化

- **為內部建構的 AI 應用程式建立 AI 紅隊演練節奏：**預先建立的攻擊探測庫以及針對特定應用場景的客製化測試案例
- **執行涵蓋所有 AI 攻擊面的測試：**提示詞注入、越獄、語境投毒、個人識別資訊擷取、原始碼外洩、離題行為和模型基準回歸
- **為生產環境 AI 系統部署執行時期保護政策：**偵測提示詞注入、資料外洩（個人識別資訊和原始碼）、越獄嘗試和內容審核違規行為
- **將紅隊演練結果自動轉化為執行時期護欄：**測試與生產環境執行之間的缺口是大多數 AI 應用程式安全計畫失敗的原因

16分鐘	72%	100%
是 AI 紅隊演練測試中首次出現關鍵失敗的中位數時間 這是在超過 25 個環境、超過 22.2 萬次對抗性攻擊中的執行結果	企業在首次測試中就暴露了關鍵漏洞。 即使在成熟的環境中，關鍵風險也從一開始就存在。	AI 系統經測試均至少有一個嚴重漏洞。 沒有任何 AI 系統是絕對安全的。持續測試必須是強制性。

©2026 Zscaler, Inc. 保留所有權利。

驗證架構一致性

- **比對所有 AI 存取路徑與六項零信任原則：**驗證每項原則是否套用於員工、開發人員和 AI 基礎架構的存取介面。
- **辨識並彌補交會處的政策缺口：**這是強制執行層級結束、另一層級開始但強制執行涵蓋不一致所在處
- 透過統一的政策模型使強制執行標準化，涵蓋身分驗證、目的地控制、風險評估、內嵌式強制執行、存取和暴險最小化、以及監控。

90 天交付成果

- 自動化違規回應和使用者指導工作流程已投入運營
- 持續的合規性監控已上線，並記載框架一致性
- 供領導階層和稽核部門使用的治理合併報表
- 已建立 AI 紅隊演練節奏，並啟用執行時期保護政策
- 將紅隊演練結果閉環轉化為生產環境護欄
- 計畫記分卡和季度藍圖，以進行持續治理



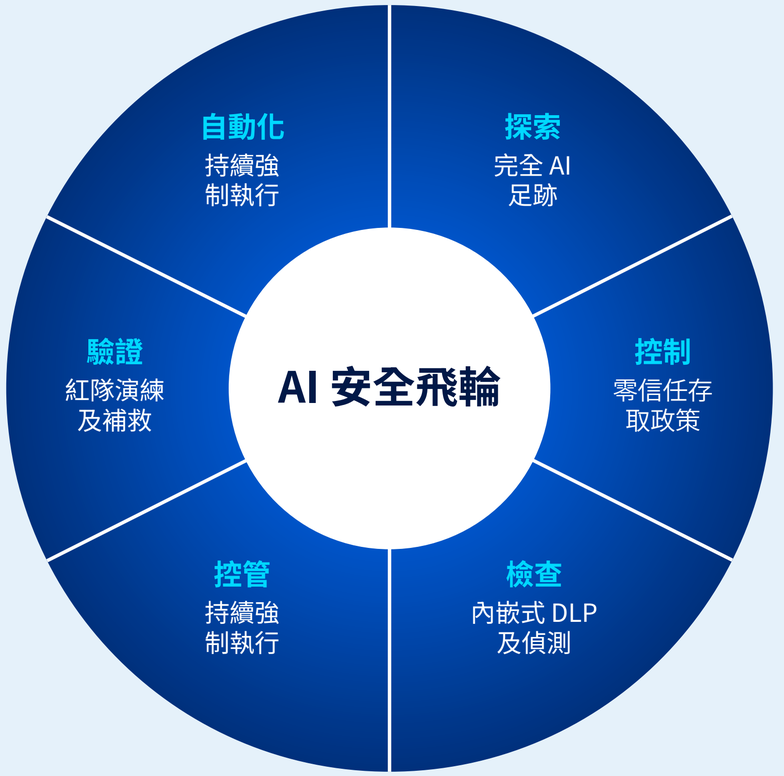
©2026 Zscaler, Inc. 保留所有權利。

Zscaler 如何協助安全採用 AI

Zero Trust Exchange™ 平台旨在控管企業級 AI 並保障安全。這擴展同一平台的功能，該平台已經為全球數千家企業控管使用者、工作負載和網路存取，而不是對舊型安全架構進行改造。

飛輪的每個階段都對應一組特定的平台功能。下表顯示 Zscaler 在每一層提供的功能。

探索 瞭解您的 AI 足跡	控制 對 AI 強制執行零信任存取	保護 強化和守護 AI 應用程式
AI 資產管理 Zscaler 可以存取您環境中的每個 AI 應用程式、模型、代理、MCP 伺服器和嵌入式 SaaS AI 功能，包括 IT 從未核准的工具。	**AI Access Security** 對 AI SaaS、企業平台中的嵌入式 AI 和 AI 開發人員環境強制執行零信任存取控制，並在每一層進行內嵌式檢查。	**AI 紅隊演練 及 AI 護欄** 在真實的對抗條件下測試私有建置的 AI 應用程式，然後自動將測試結果轉換為執行時期保護政策。
• 在所有使用者群體中探索影子 AI 及使用可視性 • **AI 物料清單 (AI-BOM)：**從資料集到模型、代理和執行時期的溯源 • **AI-SPM：**持續評估組態錯誤、權限過大和敏感資料暴險等問題 • 對企業部署和 SaaS 嵌入式代理進行 AI 代理偵測	• 依使用者群組、應用程式和資料類型執行精細的允許、警告、隔離和封鎖政策 • 對所有 AI 流量的提示詞、回應和文件上傳中實施內嵌式 DLP • 用於對 AI 工具進行非受管理和自帶設備 (BYOD) 存取的零信任瀏覽器 • 為領先的 GenAI 應用程式提供提示詞擷取和分類功能	• **自動化、持續性紅隊演練：**提示詞注入、越獄、語境投毒、資料外洩、行為漂移 • 提示詞強化和模型基準測試，並根據 NIST AI RMF、歐盟 AI 法案和 OWASP LLM Top 10 進行治理映射 • **執行時期護欄：**越獄偵測器 PII/原始碼外洩、內容審核和離題行為 • **閉環政策產生：**紅隊演練結果自動轉化為生產環境偵測器。



平台優勢

大多數資安廠商只解決了一部分的 AI 安全問題。Zscaler 在單一平台上涵蓋完整的 AI 生命週期：隨時隨地探索 AI，使用零信任政策控制誰可以存取哪些 AI，並從建置到執行時期保護 AI 應用程式和基礎設施。紅隊演練結果與生產執行時期護欄之間的閉環是任何單一解決方案都無法複製的能力。

©2026 Zscaler, Inc. 保留所有權利。

常見問題集

「零信任 AI」在實務上意味著什麼？

零信任 AI 意味著將零信任的核心原則（不假設任何隱式信任，持續驗證）套用於您環境中的每一次 AI 互動。實際上，任何 AI 應用程式、使用者、裝置、提示詞或連接器都不是本質上值得信任的。存取權的授予是基於已驗證的身分識別、裝置態勢、資料敏感性和工作階段情境。內嵌式套用控制措施。每一次存取決策都會被記錄並可稽核。

在最初的30 天內，我們應該採取哪些措施來降低影子 AI 的風險？

首先要進行探索。您無法管理您看不見的 AI，而 ThreatLabz 追蹤了企業流量中超過 3400 個 AI 應用程式，這一數字在一年內成長了四倍，這主要是由於 IT 部門已核准的工具中隨附嵌入式 AI 功能。一旦掌握了這些應用程式，就對風險最高的應用程式進行分類，並設定基準目的地政策。為風險最高的資料類別增添及時 DLP 功能。發布透明的可接受使用指南和例外申請流程，確保使用者能夠合規地展開工作，而不是有動機繞過控制措施。

AI存取控制和內嵌式DLP如何保護提示詞和上傳內容？

存取控制決定了誰可以在什麼條件下從哪些設備和位置存取哪些 AI 工具。內嵌資料外洩防護 (DLP) 在內容層運行，檢查使用者實際提交的內容、提示詞文字、檔案上傳等。copy/paste 即時互動。這兩種控制措施相輔相成：存取控制限制誰能存取哪些 AI ，而資料外洩防護 (DLP) 控制哪些資料透過這些允許的通道傳輸。任何一方的漏洞都會造成資料洩露，這就是為什麼即使組織部署了 DLP 來控制上傳操作，但如果忽略了及時檢查，仍然會面臨嚴重的資料外洩風險。

什麼是 AI-SPM，對採用 AI 有何重要意義？

AI-SPM 是 AI 基礎設施組態、權限和暴露風險的持續評估，相當於雲端基礎設施領域對 AI 的 CSPM。AI-SPM 能夠探索 AI 資產，根據安全最佳做法評估其組態，識別過度權限和敏感資料暴險，並追蹤安全態勢隨時間的變化。AI-SPM 至關重要，因為 AI 基礎設施中存在大量組態錯誤，而大多數從 AI 實驗過渡到部署的組織對此知之甚少。

哪些關鍵績效指標（KPI）能夠最有效地向領導層證明 AI 風險正在降低，同時又不影響生產力？

最有效的框架是將風險降低指標與採用率指標結合，使領導階層能夠同時看到這兩個趨勢。影子 AI 應用程式數量的減少與經批准 AI 採用率的提高，清楚地表明治理機制正在發揮作用。及時採取資料外洩防護措施（阻止、警告或允許分發）體現了執法力度的平衡。例外請求的處理時間和 AI 工具的審核週期則反映治理機制的回應速度。

ZSCALER AI SECURITY

- Zero Trust Exchange 平台
- AI 資產管理
- AI Access Security
- AI 紅隊演練
- AI 護欄

zscaler.com/ai-security



©2026 Zscaler, Inc. 保留所有權利。

zscaler™ | 快速應對。 持續守護安全。

關於 Zscaler

Zscaler（納斯達克股票代碼：ZS）是零信任安全領域的先驅和全球領導者。 全球最大的企業、關鍵基礎設施組織和政府機構都依賴 Zscaler 來保護用戶、分支機構、應用程式、數據和設備，並加速數位轉型計畫。Zscaler Zero Trust Exchange™ 平台分佈於全球 160 多個資料中心，結合先進的人工智慧技術，每天可抵禦數十億次網路威脅和策略違規行為，並透過降低成本和複雜性，為現代企業釋放生產力提升潛力。

© 2026 Zscaler, Inc. 保留所有權利。Zscaler™ 及其它商標，均列於 **zscaler.com/legal/trademarks** 是 (i) 在美國註冊的商標或服務標誌，或是 (ii) Zscaler, Inc.在美國和/或其它國家的商標或服務標誌。其它所有商標，均為其各自所有者的財產。

+1 408.533.0288 **Zscaler, Inc. (HQ) • 120 Holger Way • San Jose, CA 95134** **zscaler.com**