Microsoft Security

保護及治理 AI 的四大當務之急

教戰手冊

內容

前言

生成式 AI 正在徹底改變企業創新和成長的方式。根據 Gartner 報告，預計到了 2026 年有 80% 的企業會採用 AI 應用程式，也就不足為奇了。[1] 但是在迅速採用下，隨之而來的是新一波的安全性顧慮。超過 80% 的領導者和安全性專業人員表示，他們最擔心的就是透過這些 AI 系統洩漏敏感性資料。[2]

這不只是一個假設的問題，企業已經面臨這個問題。從無監督地帶進工作場所的影子 AI 工具，到敏感性資料洩漏風險增加，這些挑戰非常真實。此外，還有對監管責任的恐懼感 (超過 62% 的領導者承認他們並不完全了解 AI 法規 [3])，很顯然地，安全性團隊面臨艱鉅的任務。

但是有一種前進的方法。為了協助您的安全性團隊面對這些新的挑戰，我們概述了四個基本步驟：

1 準備您的環境

首先，為了自信地採用 AI，您需要對資料進行分類和標記。但並不止於此，建立健全的身分識別和存取治理是防止資料過度曝光的關鍵。

2 發現風險

接下來，深入洞悉資料流程、誰可以存取資料以及何處有安全性缺口。您這就是透過這種方式制止像過度佈建的使用者未經授權存取和過度共用敏感性資料等問題。

3 保護 AI 應用程式和敏感性資料

一旦 AI 系統上線後，它們需要持續的保護。這意味著防禦資料洩漏和 AI 特定攻擊 (例如提示注入) 等威脅，這些威脅可以操控 AI 操作的方式。

4 控管使用方式

最後，隨著 AI 法規的演變，您的方法也必須跟著變通。控管您的 AI 系統以偵測及處理諸如串通、干擾或濫用敏感內容等問題，對於保持合規性至關重要。

這些當務之急非但在於將風險最小化，也攸關讓您的團隊能夠安全地利用 AI 的力量。讓我們更深入探討如何將這些挑戰轉化為策略性優勢。

第 1 章

準備您的環境

在採用 AI 之前，必須考慮如何收集、儲存、處理及使用貴組織的資料。AI 環境是一個複雜的生態系統，所以以下是準備時要著重的三個重要區域：

資料管理

建立 AI 系統的第一步是保護它所依賴的資料。分類和標記資料以有效地進行管理。對於資安長，這可確保對資料流程的控制、防止資料外洩，以及維持法規合規性。

AI 法規

> 對於所有想要整合生成式 AI 的人來說，等著他們的是一個無法預測的法規環境。如歐盟人工智慧法案 (EU AI Act) 和美國國家標準與技術局人工智慧風險管理架構 (NIST AI RMF) 等新標準預計將有類似於一般資料保護規定 (GDPR) 的影響。

身分識別和存取

一旦擁有資料可見度後，下一步是控制誰可以存取該資料。實施良好的身分識別和存取治理，以減輕未經授權的存取和內部威脅等風險。

AI 開發和部署

在保護資料後，您可以專注於在其間移動資料的 AI 應用程式。確保這些應用程式的安全開發和定期監視，有助於在整個 AI 生命週期中防止漏洞。

第 2 章

發現風險

在 AI 系統的操作階段，確定潛在風險來源至關重要。對三個重要區域的可見度是不可或缺的：在系統間移動的資料、處理該資料的應用程式，以及與之互動的人。

資料風險

發現資料風險牽涉到在敏感性資訊的儲存、處理和傳輸方式中找出漏洞。目標是防止敏感性資料的暴露（經由洩漏或資料外洩）或損毀（經由資料破壞）。

AI 系統中資料風險的例子

> **資料洩漏：**發生在機密資訊暴露於未經授權方時。在 AI 系統中，資料在訓練、處理或產生階段期間可能會洩漏。
>
> **資料外洩：**發生在惡意執行者繞過安全性措施以未經授權存取敏感性資料時。
>
> **資料破壞：**涉及破壞 AI 系統的訓練資料，損害其完整性和扭曲輸出。

應用程式風險

應用程式風險可能來自您的官方 AI 應用程式，以及員工可能擅自使用之未經批准或有風險的 SaaS AI 應用程式。官方應用程式可能有漏洞，例如錯誤設定或未修補的軟體。未經批准的應用程式若未針對安全性進行審查，將帶來風險，因此能夠偵測並封鎖其使用至關重要。

AI 系統中應用程式風險的例子

> 提示注入攻擊透過引入導致系統以非預期方式行事的惡意輸入，來操控 AI 應用程式。例如，使用者可能會輸入設計用來繞過安全性通訊協定的命令，導致 AI 洩露敏感性資料或執行未經授權的動作。

使用者風險

使用者風險與來自使用者和外部攻擊者的人類相關漏洞有關。內部威脅可能會導致惡意或意外洩漏。外部攻擊者可以透過惡意動作利用 AI 系統。偵測不尋常的行為與活動對於及早發現這些風險至關重要。

AI 系統中使用者風險的例子

> **內部威脅：**可存取敏感性 AI 系統的不滿員工試圖變更訓練資料，操控結果以破壞組織的運作。
>
> **外部攻擊：**網路釣魚攻擊以 AI 管理員為攻擊目標，誘騙他們洩露其認證，好讓攻擊者存取和操作敏感性資料或模型。

第 3 章

保護 AI 應用程式和敏感性資料

一旦建立可見度來協助發現風險後，您可以在 AI 系統執行並與真實資料和使用者互動時，將焦點轉移到持續保護上。保護包括保衛每個接觸點上的敏感性資料、根據風險調整安全性措施、控制整個系統的存取，以及快速回應新興威脅。

保護敏感性資料

保護整個生命週期內的敏感性資料。這包括在傳輸和儲存期間加密以確保資料安全、存取控制以確保只有授權的使用者可以與之互動、根據敏感度進行標記將資料分類，以及資料外洩防護 (DLP) 措施以偵測未經授權的共用和移動。

根據風險調整安全性

調整控制來配合不同使用者、裝置和系統所造成的風險等級，藉此維持保護而無須不必要地限制存取或妨礙效率。調適性控制允許對高風險的使用者施行更嚴格的原則，同時對較低風險的情形套用較輕度的控制。

控制存取

確保只有授權的使用者可與 AI 應用程式和敏感性資料互動。當您實施靈活、集中化的原則時，您可以依使用者角色和行為進行調整。存取控制應根據資料敏感度和使用者風險、實施多重要素驗證、限制對高風險使用者的存取，以及定期檢查權限以確保合規性和安全性。

回應威脅

當偵測到風險時，需要快速行動。安全性資訊和事件管理 (SIEM) 工具可協助分析記錄，以找出不尋常的模式，例如可疑的使用者行為或網路活動。這些工具可以觸發自動回應，隔離已入侵的應用程式、撤銷存取權或警告安全性團隊，確保您的 AI 系統即時受到保護。

第 4 章

控管使用方式

一旦 AI 系統受到保護，最後一步是管理使用方式。隨著 AI 法規不斷演變，控管使用方式對於維護合規性及處理違反政策 (如共謀、騷擾或共用不安全的內容) 至關重要。這包括實施法規和行為控制、設定明確的 AI 使用政策、定義保留和修改章程，以及為不斷變化的法規做準備。

隨時掌握監管義務

確保您的組織遵守外部法規，如歐盟 AI 法案和 GDPR，以及內部政策。定期審查與稽核有助於及早發現違規行為，並增進符合道德規範的 AI 使用，使一切正常進行。

法規要求範例

> GDPR[4] 第 5 (1)(e) 條指出，個人資料應「以允許辨識資料主體的形式保存，其時間不應超過處理資料之目的所需的時間。」這要求組織設定保留期間，一旦不再需要個人資料時，就將其立即從資料中刪除。

為法規變更做好準備

隨著 AI 法規不斷演變，維護合規性需要適應性。定期更新政策、處理法規責任，並培訓員工，以確保做好滿足新需求的準備。

設定明確的 AI 使用指南

針對 AI 的適用方式建立明確的標準，確保合乎道德的做法、保護資料處理及尊重隱私權。處理特定治理問題，例如防止幻覺或著作權外洩，並定期審查政策，以保持與不斷變化的業務和法規需求的相關性。

為資料保留和刪除設定明確的指導方針

定義資料應儲存多久以及何時應該移除，以便符合安全性和法規需求。這包括為 AI 互動建立特定時間範圍，例如提示和完成，以確保合規性和保護敏感性資訊。

Microsoft Security

鞏固您的 AI 未來

生成式 AI 為創新和生產力提供了令人難以置信的契機，但也帶來新的風險。為了充分利用 AI，同時保護您的組織安全，您需要一種新方法來管理資料、保護您的系統並保持遵守不斷演變的法規。

您可以按照本教戰手冊中的步驟執行，即準備環境、發現風險、保護 AI 應用程式和敏感性資料以及控管使用方式，您可以確保您的 AI 系統受到安全保護，而且您的組織已準備好迎接未來的挑戰。

Microsoft 安全性提供了工具和支援，可協助您安全地使用 AI、保護敏感性資料，以及負責任地管理 AI 使用。

了解更多關於 Microsoft 全方位的 AI 安全性方法

[1]《Hype Cycle for Generative AI》，Gartner, Inc.，2023 年 9 月 11 日

[2]《First Annual Generative AI Study:Business Rewards vs. Security Risks》，第 8 頁，Information Security Media Group (ISMG)，2024 年 1 月 31 日

[3]《First Annual Generative AI Study:Business Rewards vs. Security Risks》，第 6 頁，Information Security Media Group (ISMG)，2024 年 1 月 31 日

[4]《一般資料保護規定 (GDPR) 第 5 (1)(e) 條》，一般資料保護規定 (GDPR)，歐洲議會和歐盟理事會，2018 年 5 月 25 日