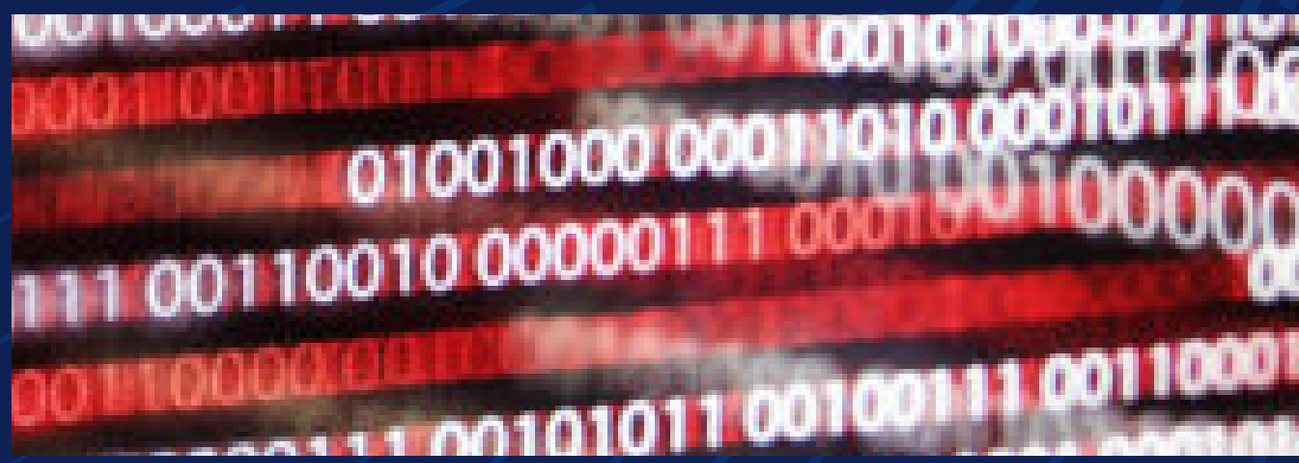
破除網路安全誤解：

破解 AI 安全迷思



AI 正在改變各個產業，但許多組織在保護 AI 方面陷入迷思，認為這相當複雜，不過實際上並非如此。事實是什麼？保護 AI 系統不必完全從零開始，現行網路安全性原則就能有效應對 AI 帶來的獨特難題。

Dell Technologies 瞭解 AI 背後的架構，能助您調整現有解決方案來適應這個新架構。讓我們破解最常見 AI 安全迷思並揭露事實，與您共同有效保護系統。

迷思 1：「AI 系統過於複雜而無法保護。」

事實：AI 確實會產生新的網路安全風險，例如提示注入、資料操縱和敏感資訊洩露等等。此外，代理式 AI 系統還會擴大攻擊面，因為攻擊者可能利用這類系統操控結果或提升權限。

話雖如此，組織仍必須找出這些漏洞並採取措施，來防止 AI 系統遭受傳統和 AI 特有威脅。其實，組織可以管理這些風險，並保護 AI 模型。務必瞭解一個重點：AI 系統需要大量資料做為輸入，且會產生大量資料做為輸出。因此您應採取以資料保護為主的關鍵安全策略，同時包括：

- 零信任原則：例如身分管理、角色型存取和持續驗證。
- 定期滲透測試和漏洞管理：找出漏洞。
- 記錄和稽核：驗證資料輸入和輸出。

迷思 2：「我現有的工具都無法保護 AI。」

事實：保護 AI 安全不是從零開始，而是更聰明地運用既有工具。大多數現有網路安全工具都能調整，以有效保護 AI 系統。雖然 AI 性質獨特，但這項技術本身只是推動業務的眾多工作負載之一。身分管理、網路分段及監控、端點和資料保護等基本網路安全實務，仍是保護 AI 環境的必要做法。關鍵在於調整這些做法來應對特定 AI 難題，例如保護訓練資料和演算法，以及降低對抗性輸入等風險。

強大的防禦始於妥善維護網路衛生，例如系統修補、存取控制和漏洞管理。重要的是調整這些實務來處理 AI 特定風險。整合以 AI 為主的策略與現行安全防護方法，再搭配適當工具，就能有效管理 AI 安全。

不過還有一個重點：更新硬體可在對抗網路攻擊方面發揮關鍵作用。舉例來說，採用新型 AI PC 即可針對「端點」這個主要攻擊向量，設下強大的第一道防線。隨著 Windows 10 支援結束，過時的 PC 會引發風險。此外，Windows 11須有可信平台模組 (TPM) 版本 2.0。這是有助於加密、安全開機和防範韌體攻擊的安全晶片，但許多舊款 PC 完全沒有 TPM，或只支援舊版 TPM。Dell 提供的安全商用 AI PC，均內建這些增強功能。

AI 基礎結構 (例如伺服器和儲存空間) 也是如此。Dell AI Factory 不僅採用針對 AI 安全最佳化的硬體，還內建眾多安全功能，包括安全供應鏈、資料不變性、隔離和加密等等。

迷思 3：「AI 安全性不過是指保護資料。」

事實：AI 安全性不只是基本的資料保護，而是保護整個 AI 生態系統，包括模型、API、輸出、系統和裝置。當 AI 越深入整合關鍵應用程式，這項技術遭到誤用或利用的風險越高。若無完善的安全防護措施，不肖人士可能會竄改 AI 模型來輸出有害或誤導性內容，或未經授權利用 API 存取敏感系統，AI 系統則可能在輸出內容中不慎暴露私人或機密資訊。

因此，必須採用多層次做法，才能全方位保護 AI 安全。這包括保護模型不受對抗性攻擊，避免不肖人士藉此操縱輸入資料來欺騙 AI 系統；使用強大的身分驗證方法保護 API，防止未經授權的使用情形；以及**持續監控輸出內容**，找出可能代表攻擊或故障的異常或可疑模式。有效保護 AI 安全，不只是確保 AI 系統的完整性和可靠性，還須降低惡意使用或產生意外結果的風險，這樣才能建立使用者和利害關係人的信任。

迷思 4：「AI 不需要人類監督。」

事實：治理和人為監督至關重要，可確保 AI 系統以可預測且符合倫理與人類價值觀的方式運作。其中進階 AI 系統，尤其是具有自主決策能力的代理式 AI，甚至帶來獨特的難題，需要完善的措施加以防護。沒有適當的監督，這些系統可能偏離預期目標，或展現可能構成風險的意外行為。

要解決這個問題，就必須建立明確界限、導入分層控管機制，並確保人類持續參與關鍵決策過程。此外，定期稽核、確保 AI 作業公開透明及徹底測試，可進一步強化問責制並提升信任，進而防止誤用並促進負責任的 AI 技術部署。

強化 AI 安全性的最佳實務

要填補 AI 特有的安全性缺口，組織須採用主動的策略性方法。
以下是保護 AI 系統的 10 個最佳實務：

分層安全性架構：
使用分段、防火牆和強大的身分驗證，保護基礎結構、軟體和資料的每一層。

監控與驗證 AI 輸出內容：
透過異常偵測、日誌記錄和警示來監控 AI 輸出內容的異常模式或有害行為。

保護供應鏈：
實施強大的供應商管理計畫。稽核供應商和第三方元件、驗證完整性，並使用經簽署的程式碼來防止 AI 開發生命週期中的漏洞。

制定復原計畫：
定期備份資料並測試災難回復計畫，以將停機時間降至最低，並確保發生入侵事件時能快速復原。

保護訓練資料和模型：
監控資料完整性並套用完善的驗證工具，來防範資料中毒、對抗性輸入和其他威脅。

導入強大加密機制：
使用強大的演算法來加密靜態和傳輸中的敏感資料，並以安全的方式管理及定期輪換加密金鑰。

加強存取控制：
實施最小權限原則、導入角色型存取控制 (RBAC)、定期輪換認證並稽核權限，以防止未經授權的存取。

定期執行安全性稽核和滲透測試：
經常評估系統漏洞及執行滲透測試，藉此預先找出風險，以防攻擊者利用。

保護 API：
使用功能強大的驗證協定 (如 OAuth 2.0)、強制執行 HTTPS 加密，並定期更新 API 以消除潛在漏洞。

訓練員工掌握 AI 安全性最佳實務：
定期訓練團隊掌握安全開發和威脅識別方法，以及維護強大的安全性實務，以防止入侵。

Dell 的價值主張：實用的 AI 安全性解決方案。

AI 安全性其實並不像表面上那樣複雜。事實是什麼？保護 AI 與保護現有工作負載並無差別，關鍵在於瞭解架構並套用合適的策略。這正是 Dell Technologies 能提供協助的地方。

Dell 將協助您更容易保護 AI 安全。我們會運用您目前的解決方案，緊密整合到以 AI 為主的架構，因此您無須全面翻新基礎架構，就能應對提示注入、API 濫用和對抗性攻擊等挑戰。

Dell 具備豐富專業知識，為您破除 AI 安全性相關迷思，證明保護 AI 確實可行。無論您是剛展開 AI 之旅，還是想增強防禦能力，我們都會協助您抱持自信、有效保護您的投資和系統，建立韌性數位未來。讓我們共同以更簡單的方式保護 AI 安全。

能夠協助的 Dell 產品和解決方案

精選 Dell 解決方案	說明
Dell AI Factory	Dell AI Factory 透過安全供應鏈保護 AI 工作負載，確保從開發到部署作業，都在值得信任的基礎結構上執行。這項解決方案具備資料不變性、隔離和加密等功能，可保護敏感模型和資料集、抵禦網路威脅，並在資料導向的動態環境中，以高效率流暢執行可擴充的 AI 作業。
網路韌性	PowerProtect 具備不變性和隔離等多項進階功能，可保護 AI 工作負載、確保資料完整性並防範網路威脅。這項解決方案提供端對端加密和異常偵測，同時能快速復原資料，將停機時間降至最低。
Dell Trusted Workspace (端點安全性)	一套內建和選用附加功能組合，專為保護商用 AI PC 和其中執行的 AI 工作負載而設計。根據安全供應鏈實務建構而成，內建功能包括 SafeBIOS 和具備 TPM 的 SafeID。選用附加功能包括安全元件驗證、具備 ControlVault 的 SafeID，以及合作夥伴軟體 CrowdStrike 和 Absolute，可最大化工作空間安全性。
AI 安全性諮詢服務	一套服務組合，可協助開發並實施全方位 AI 安全性策略。內容包括諮詢服務、AI vCISO 和資料安全性規劃。
受管理 AI 安全營運服務	實現跨堆疊深度可視性以快速偵測和回應威脅。功能包括 Managed Detection and Response、受管理 AI 防護、AI 滲透測試，以及事件回應與復原服務。
安全性軟體整合	設計、安裝及設定安全性工具，來保護存取管理、應用程式、網路、雲端等等。

造訪 **dell.com/cybersecuritymonth**，瞭解如何解決當今一些主要的網路安全挑戰

DELL Technologies