DELL Technologies

GenAI 和 LLM 的 10 大網路安全性疑慮

開始使用

簡介

人工智慧 (AI) 正在革新組織營運的方式，生成式 AI (GenAI) 和大型語言模型 (LLM) 成為現代企業環境中的關鍵工作負載。

就像任何其他工作負載一樣，這些應用程式有其自身的複雜性和需要解決的漏洞。隨著企業持續採用 AI 來推動創新、效率和競爭優勢，確保這些應用程式的安全性至關重要。良好的網路衛生是保護任何工作負載的基礎，正如您將所有工作負載的安全性列為優先事項一樣，也請務必對 AI 實作良好的網路衛生。這包括實作適當的系統修補、多重因素驗證、角色型存取和網路區隔等實務。這些是基本措施，但關鍵在於瞭解這些功能如何融入工作負載的特定架構和使用方式。

在 Dell，我們深刻瞭解 AI 工作負載及其所面臨的獨特安全性挑戰。透過識別威脅行為者可能鎖定這些工作負載的方式，Dell 可協助您建立健全的安全性策略。這包括解決以下風險：訓練資料中毒、模型盜用或操縱、資料集重建等。

我們也專注於管理與 AI 模型輸入相關的挑戰，例如防止敏感資訊洩露、緩解不安全的主題或偏見，以及確保遵守法規。在輸出方面，我們幫助解決過度依賴模型和法規遵循相關風險等問題。

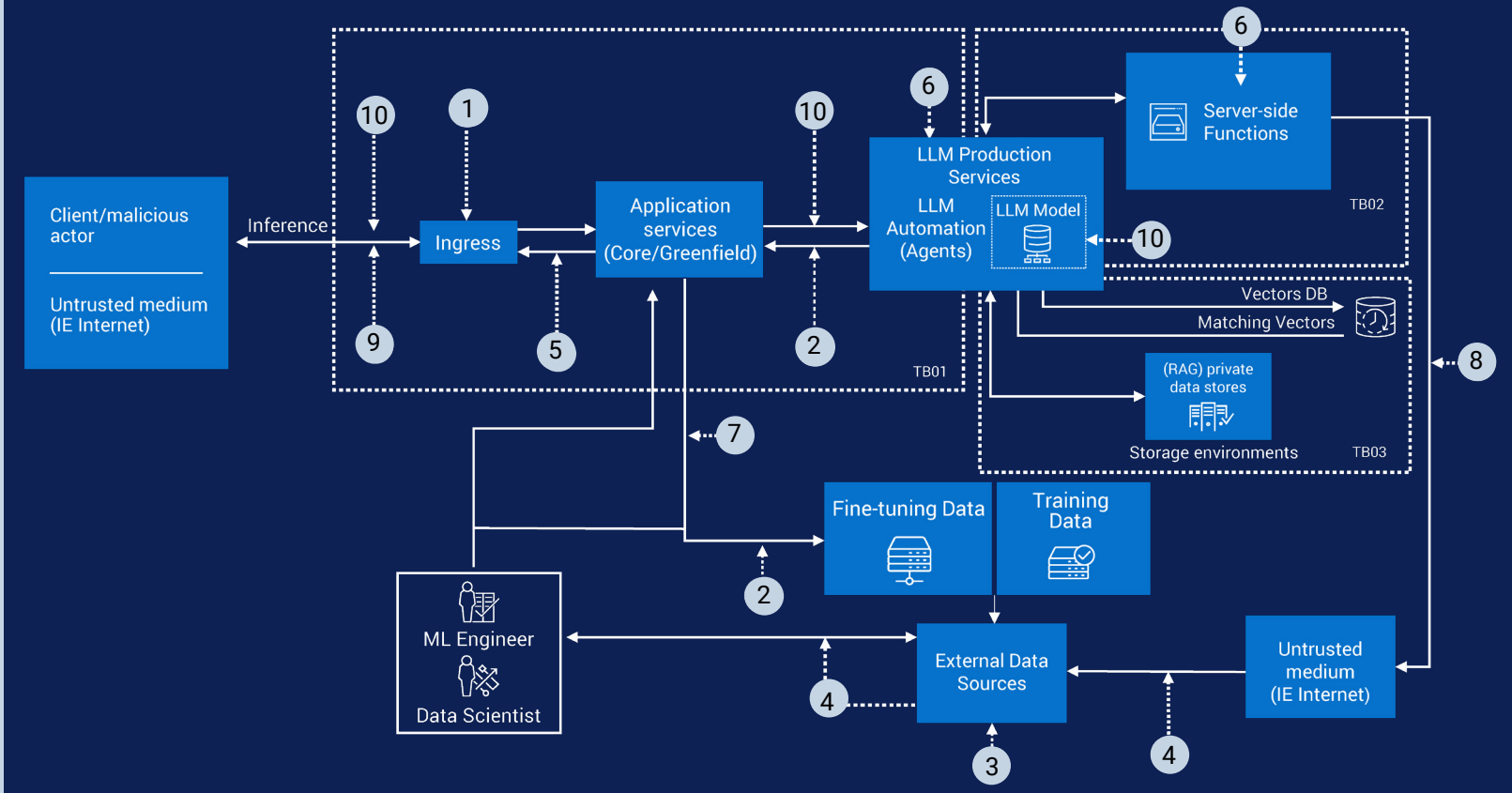
在 Dell，我們運用企業現有的網路安全解決方案，或探索新的工具和實務來保護其系統，藉此讓企業得以降低這些風險。我們的目標是確保安全性不會妨礙您的創新。藉由瞭解 AI 工作負載的運作方式及其所面臨的安全威脅，我們可協助您建立更強大的安全性狀態，讓環境更具韌性，同時讓您放心地進行創新。我們運用專業知識，協助您安心地駕馭 AI 的潛能，同時保持強大的安全性。

GenAI 和 LLM 的 10 大網路安全性疑慮

如 OWASP 所述，這些是保護 GenAI/LLM 模型的主要疑慮。

按一下疑慮編號，可深入瞭解：

1. 提示注入
2. 敏感資訊洩露
3. 供應鏈
4. 模型資料中毒
5. 輸出處理不當
6. 過度代理
7. 系統提示外露
8. 向量與內嵌弱點
9. 錯誤資訊
10. 無限資源耗盡

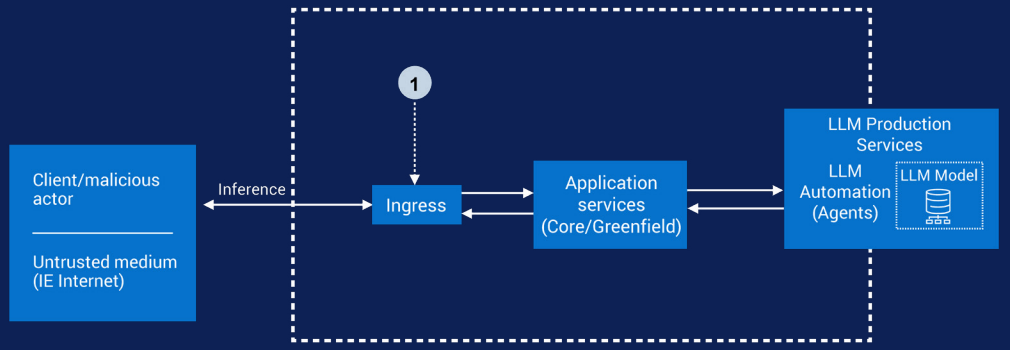


疑慮 #1：提示注入

緩解提示注入的策略：

- **資料清除和輸入驗證：**徹底篩選使用者輸入的內容，以移除有害內容。使用正規化和編碼來防止誤用。
- **自然語言處理 (NLP) 和機器學習式方法：**使用 NLP 和機器學習來偵測和封鎖操縱或惡意提示。
- **清除輸出格式和回應控制項：**設定嚴格的回應邊界，以確保輸出遵循預期的格式，並防止未經授權的動作。使用提示篩選和回應驗證，來維護完整性。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **安全提示工程：**作為整體軟體安全性的一部分，使用安全的提示設計和分析，來保護輸入處理。
- **模型驗證：**定期驗證 ML 模型，以確保它們在部署前未遭篡改，保護其準確性和完整性。
- **提示篩選、排序和回應驗證：**分析和排序提示，以確保僅處理安全的輸入。驗證回應以防止誤用。
- **健全性檢查：**定期評估以識別和修正漏洞，確保 AI 安全可靠。

返回架構圖



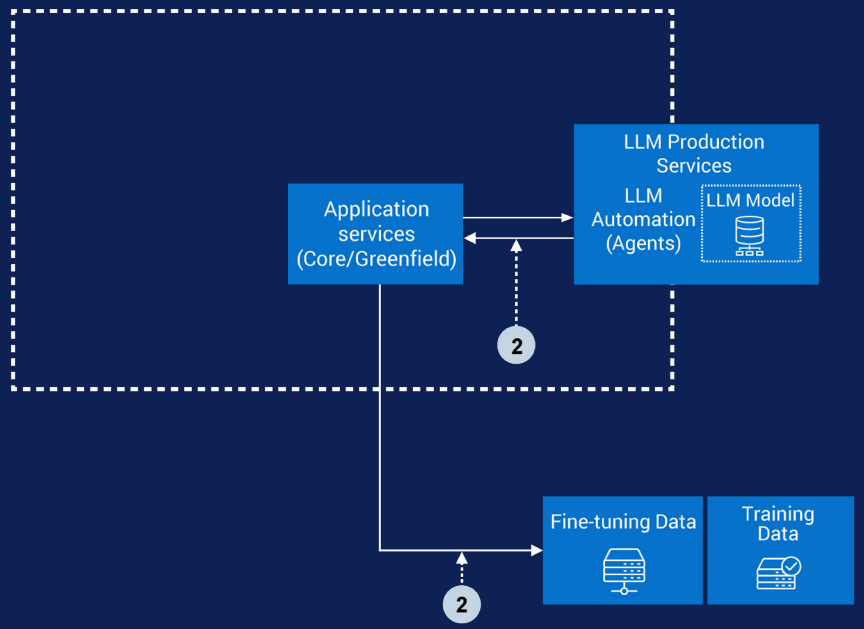
提示注入是生成式 AI (GenAI) 領域的新興挑戰，其中惡意輸入是設計用來操縱模型的行為，或危害其完整性。這些攻擊利用 AI 系統處理和回應使用者輸入方式的漏洞，可能導致未經授權的行動、錯誤資訊或敏感資料暴露。隨著 GenAI 日益融入關鍵業務工作流程，解決這些風險對於維持信任和安全性至關重要

疑慮 #2：敏感資訊洩露

緩解敏感資訊洩露的策略：

- **資料清除和輸入驗證：**徹底篩選使用者輸入的內容，以移除有害內容。使用正規化和編碼來防止誤用。
- **利用同態加密**，安全地處理敏感資料，而不外洩其內容。這可確保即使資料在使用中，仍然維持加密且受到保護，不會洩露。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **利用安全的 API 和系統介面**進行 AI 資料互動，定期檢閱組態，將暴露風險與攻擊面降至最低。
- **保護資料收集、儲存和原則的安全**，並執行全方位的資料保護和治理原則，以確保法規遵循，並將資料風險降至最低。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **安全的開發、組態和稽核**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。
- **使用者教育和安全意識：**為使用者和系統管理員提供持續的 AI 專屬安全意識訓練，減少不安全使用和意外資料洩露。

返回架構圖



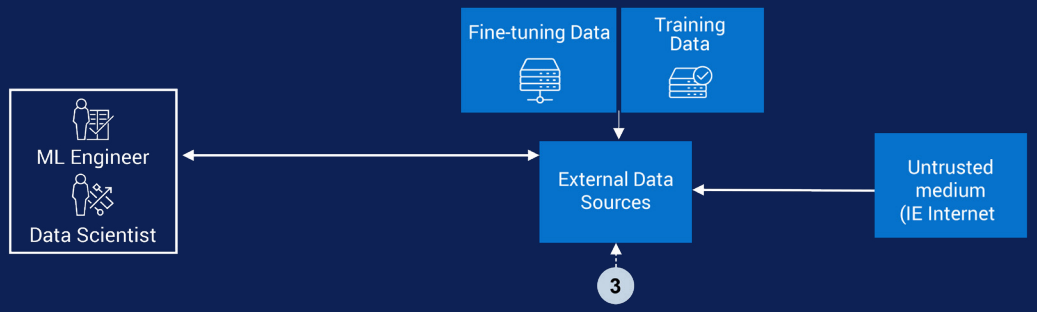
GenAI 帶來令人難以置信的進步，但同時也引入重大風險，尤其是敏感資訊的意外暴露。無論是個人身分識別資訊 (PII) 還是專有業務資料，若誤用 GenAI 工具或處理不當，都可能導致資料洩露、違規或聲譽受損。因此，組織必須瞭解這些風險，並主動解決，以確保安全實作和使用 AI 系統。

疑慮 #3：供應鏈漏洞

緩解供應鏈漏洞的策略：

- **審查供應商並確保遵守安全的供應鏈實務：**評估供應商，並建立優先考慮供應鏈安全的協議。
- **實作軟體材料清單：**追蹤和驗證軟體元件的來源，確保透明度，並降低程式碼遭入侵的風險。
- **模型驗證：**定期驗證 ML 模型，以確保它們在部署前未遭篡改，進而保護其準確性和完整性。
- **以最低權限執行容器和 Pod：**這可降低遭到入侵時的潛在影響，並限制未經授權的存取。
- **部署防火牆：**封鎖不必要的網路連線，減少暴露於潛在威脅的風險，並限制攻擊者的利用途徑。
- **保護資料與註解：**保護您的資料和相關註解，以防止篡改、未經授權的存取和關鍵資訊損壞。
- **保護硬體：**使用經過安全性驗證的硬體，防範硬體攻擊可能產生的漏洞，確保您的基礎結構擁有堅實的基礎。
- **安全的 ML 軟體元件：**使用可信賴且經過審查的 ML 軟體元件來減少漏洞，並提高機器學習工作流程的整體安全性。
- **安全的開發、組態和稽核：**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。

返回架構圖



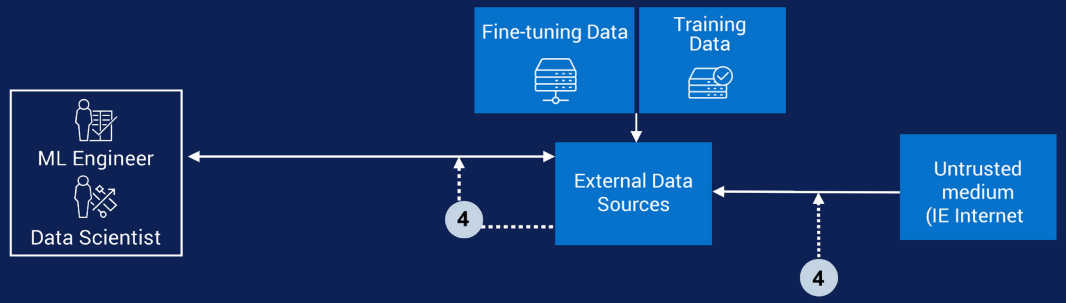
探索 LLM 供應鏈中的漏洞，這些漏洞可能會影響關鍵元件，例如預先訓練的模型完整性和第三方配接器。AI 系統依賴硬體和軟體，而這些硬體和軟體有可能早在部署前就遭到入侵。攻擊者可以利用機器學習供應鏈各個階段的弱點，鎖定 GPU 硬體、資料及其註解、ML 軟體堆疊的元素，甚至是模型本身。攻擊者只要入侵這些獨特的部分，便可取得系統的初始存取權，進而對安全性和完整性造成重大風險。瞭解和緩解這些漏洞，對於構建強大、安全的 AI 解決方案至關重要。

疑慮 #4：模型資料中毒

緩解模型資料中毒的策略：

- **在訓練期間，使用異常偵測和資料驗證**來識別和解決資料中的不一致，並確保僅使用乾淨的高品質資料，來訓練模型。
- **在微調階段隔離環境**，以在開發的關鍵階段，防止未經授權的存取或模型汙染。
- **模型驗證：**定期驗證 ML 模型，以確保它們在部署前未遭篡改，進而保護其準確性和完整性。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **資料清除和輸入驗證：**徹底篩選使用者輸入的內容，以移除有害內容。使用正規化和編碼來防止誤用。
- **安全的開發、組態和稽核：**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。
- **健全性檢查：**定期評估以識別和修正漏洞，確保 AI 安全可靠。
- **實作網路區隔**以限制對不安全介面和關鍵系統元件的存取。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。

返回架構圖



探索模型資料中毒是 AI 生命週期中的一種安全性威脅，攻擊者會故意以損毀、誤導或惡意輸入來汙染訓練資料。這種風險可能會影響關鍵元件，從原始資料收集和註解，到用於機器學習或大型語言模型的資料集整理和整合。AI 系統的可靠性取決於其資料來源的完整性，但在訓練之前、預先處理期間或透過外部資料管道，這些資料來源可能會暴露於操縱風險。

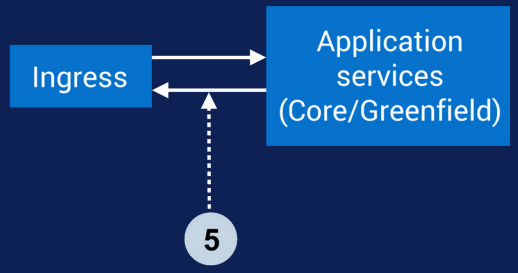
攻擊者利用資料中毒來降低模型準確性、引入漏洞或觸發有害輸出。透過鎖定資料來源追蹤、註解品質或資料集擷取程序中的弱點，攻擊者可以破壞安全性、可信度和復原能力。識別和緩解這些以資料為中心的威脅，對於建構強大、可靠的 AI 解決方案至關重要。

疑慮 #5：輸出處理不當

緩解輸出處理不當的策略：

- **情境感知輸出編碼：**針對將使用輸出的特定情境 (例如 HTML、SQL 或 API 環境)，一律套用量身打造的編碼和跳脫技巧，以防止注入攻擊等漏洞。
- **輸出清除：**遵循開放式 Web 應用程式安全專案 (OWASP) 應用程式安全驗證標準 (ASVS) 指南，執行模型輸出的嚴格驗證和清除實務，以確保安全的下游使用，並降低安全性風險。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **自動輸出安全性測試：**使用自動化工具定期進行安全性測試，以識別輸出中的風險，例如跨網站指令碼 (XSS) 或注入漏洞，並主動解決這些風險。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **人類參與循環審查：**適用於高風險應用 (例如財務或醫療保健)，需要人類監督並審查模型輸出，以確保準確性、資安與安全性。
- **隱私權與法規遵循：**將隱私保護技術整合到輸出程序中，並確保遵守安全使用敏感資訊的相關法規和標準。

返回架構圖



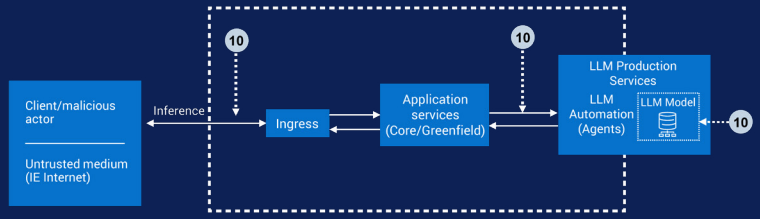
AI 模型輸出的驗證或清除不足，可能會導致嚴重的安全風險，包括權限提升和資料違規。當 AI 模型生成的輸出未經正確檢查或篩選時，惡意行為者可能會利用這些漏洞，進行未經授權的存取，或在系統中提升其權限。缺乏監督可能導致資料遭入侵、未經授權的動作和重大的安全性違規，這凸顯了為任何 AI 生成的輸出，實作強大的驗證和清除程序的重要性。

疑慮 #6：過度代理

緩解過度代理的策略

- **強制實作最低權限：**僅向 LLM 和代理式子系統授予所需的最低權限，來執行預期操作和定期審查存取控制。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **設定作業邊界：**明確定義 LLM/代理程式可存取或執行的內容。
- **人類參與循環審查：**適用於高風險應用 (例如財務或醫療保健)，需要人類監督並審查模型輸出，以確保準確性、資安與安全性。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **限制自主性：**限制 LLM 功能，以避免不受限制的存取或控制。
- **安全的開發、組態和稽核：**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。
- **部署防火牆：**封鎖不必要的網路連線，減少暴露於潛在威脅的風險，並限制攻擊者的利用途徑。
- **健全性檢查：**定期評估以識別和修正漏洞，確保 AI 安全可靠。

返回架構圖



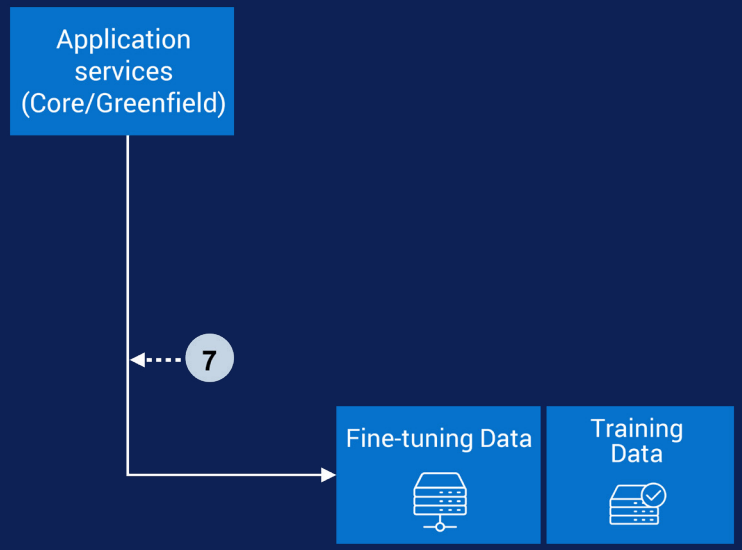
在工作流程中，授予 AI 代理程式或附掛程式過多的自主權或不必要的功能，可能會帶來重大風險。當 AI 系統被賦予超出需求的權限或功能時，會增加產生意外後果的可能性。當以大型語言模型 (LLM) 為基礎的系統設計具有過多的權限時，可能會發生這種情況，允許它們採取動作或存取不應存取的資訊。這種過度的權限和功能可能會導致錯誤、資料誤用甚至安全性漏洞。這也凸顯了為確保安全和負責任的使用方式，仔細限制和監控 AI 功能的重要性。

疑慮 #7：提示外露

緩解提示外露的策略

- **避免在提示中嵌入敏感資訊：**切勿在提示中包含憑證、API 金鑰或專屬邏輯。應該在系統外部安全地管理這些項目。
- **將安全控管措施與提示分開：**在應用程式邏輯中，而不是在提示中，處理驗證、授權和工作階段管理。
- **驗證輸入與輸出：**透過強大的驗證來清除提示和回應，以封鎖可疑的模式或操縱。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **加密與安全提示：**將提示和組態儲存在經過加密的安全儲存裝置中，以防止未經授權的存取。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **定期檢閱提示：**定期檢閱並清除提示，以移除敏感資料並確保遵循安全性法規。
- **針對弱點的測試和紅隊演練：**執行對抗性測試，以識別和修正提示管理或輸出中的漏洞。
- **將提示與使用者輸入隔離：**設計系統，以防止來自遭操縱或暴露提示的使用者查詢。
- **強制執行速率限制：**限制 API 使用、對可疑活動進行節流，並封鎖自動提示攻擊。

返回架構圖



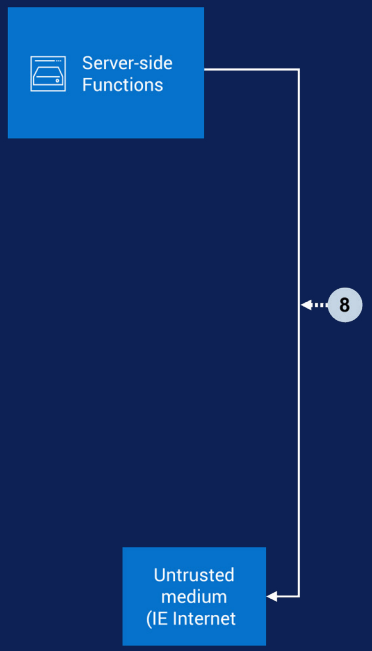
當攻擊者能夠擷取或推論指引模型行為和設定作業邊界的隱藏指示（「系統提示」）時，就會發生對大型語言模型 (LLM) 或 AI 系統的系統提示外露攻擊。終端使用者通常不會看到這些提示，因為它們包含核心規則、限制，有時還包含敏感的運算邏輯。透過特別製作的輸入或利用漏洞，攻擊者可能會誘使 LLM 洩露其全部或部分系統提示。如果洩露，此資訊可能會被用於反向工程限制、繞過安全篩選，或開發新的針對性攻擊，最終導致提示注入、權限提升或誤用依賴其完整性的模型和下游系統的風險增加。

疑慮 #8：向量與內嵌弱點

緩解向量與內嵌弱點的策略

- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **加密：**使用 AES 等健全的加密標準，保護傳輸中及靜態的向量資料。
- **保護組態並監控：**強化系統、安全地設定，並持續監控組態錯誤、未經授權的存取或異常。
- **漏洞管理：**定期更新和修補所有軟體、相依性和向量儲存引擎，以解決安全性風險。
- **資料清除和輸入驗證：**徹底篩選使用者輸入的內容，以移除有害內容。使用正規化和編碼來防止誤用。
- **利用安全的 API 和系統介面**進行 AI 資料互動，定期檢閱組態，將暴露風險與攻擊面降至最低。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **保護硬體：**使用經過安全性驗證的硬體，防範硬體攻擊可能產生的漏洞，確保您的基礎結構擁有堅實的基礎。
- **安全的開發、組態和稽核：**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。

返回架構圖



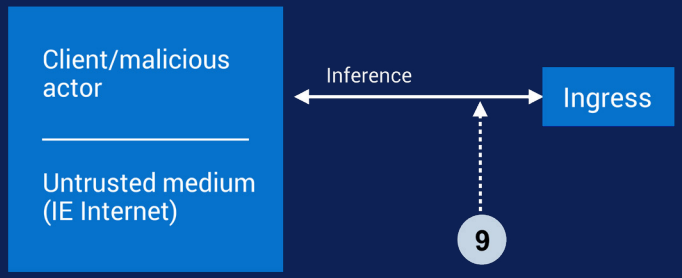
對大型語言模型 (LLM) 或 AI 系統 (尤其是使用檢索增強生成 (RAG) 的系統) 的向量與內嵌弱點攻擊，會鎖定將資訊編碼、儲存和擷取為數值向量和內嵌的方式中的漏洞。透過惡意動作即可利用這些機制的弱點，例如內嵌逆向 (從內嵌中重建敏感資料)、資料中毒 (注入有害或有偏見的內容來操縱模型行為)、未經授權存取向量資料庫 (導致資料洩露)，或操縱擷取輸出。這些攻擊讓攻擊者能夠洩露敏感資訊、更改輸出，或破壞使用者對 AI 驅動應用程式的信任，進而威脅隱私、完整性和可靠性。適當的存取控制、資料驗證、加密和持續監控，對於防禦這些不斷演變的威脅至關重要。

疑慮 #9：錯誤資訊

緩解錯誤資訊的策略

- **具有權威來源的檢索增強生成 (RAG)：**使用 RAG 從經過驗證、受信任的資料庫和知識庫中擷取和整合資訊，減少幻覺。
- **模型調整和輸出校準：**使用不同的資料集來微調模型，並套用技術，以將偏見和錯誤資訊降到最低。
- **自動事實查核：**與可靠來源交互參照輸出，並自動標記假資訊。
- **不確定性監控：**在關鍵案例中，標記低信賴度回應，以進行人工審查。
- **人類參與循環審查：**適用於高風險應用 (例如財務或醫療保健)，需要人類監督並審查模型輸出，以確保準確性、資安與安全性。
- **使用者意見回饋：**讓使用者能夠報告錯誤，以便持續改善模型，並快速修正錯誤資訊路徑。
- **存取限制和人為監督：**套用角色型存取控制 (RBAC)、多重因素驗證 (MFA) 和身分識別管理來限制存取。對關鍵決策使用人工審查。
- **安全的開發、組態和稽核：**套用安全的編碼實務，使用自動化組態管理工具，並定期進行審查、稽核和更新，以確保 AI 系統組態安全並維持最新狀態。
- **風險溝通：**教育使用者，讓他們瞭解 AI 局限性，並鼓勵他們獨立驗證。
- **刻意的 UI 和 API 設計：**凸顯 AI 生成的內容，並指引使用者負責任地使用。

返回架構圖



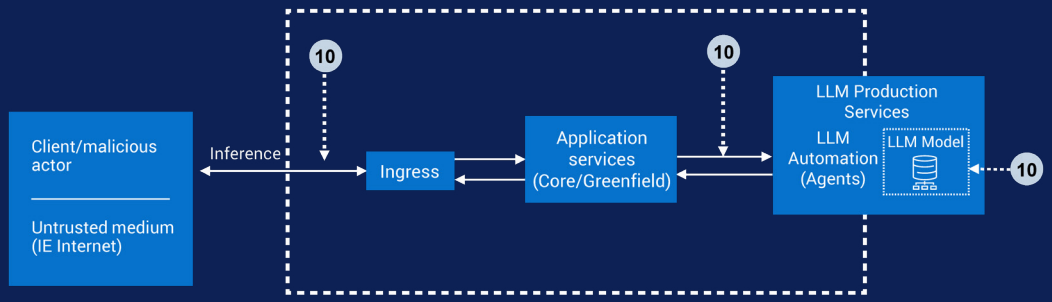
對 LLM 或 AI 系統的錯誤資訊攻擊，是故意讓模型透過輸出，生成或傳播虛假、誤導，或看似可信實則不正確的資訊。此漏洞源於幾個因素：模型的「幻覺」傾向 (生成捏造但聽起來合理的內容)、訓練資料中存在的偏見或差距，以及對抗性提示的影響。幻覺的發生是因為 LLM 在統計上生成符合模式的文字，而不是真正理解事實，導致答案看似權威，但實際上沒有根據。此類攻擊的風險包括安全性違規、聲譽損害，甚至是法律責任，尤其是在使用者過度依賴 LLM 回應，而不驗證其準確性或有效性的環境中，可能會將錯誤或錯誤資訊嵌入關鍵決策和程序中。

疑慮 #10：無限資源耗盡

緩解無限資源耗盡的策略

- **強制執行速率限制和使用者配額：**對每個使用者、API 金鑰或應用程式的要求、權杖或資料，設定嚴格限制，以防止濫用。
- **要求驗證和使用者區隔：**使用高強度的驗證方法 (例如 API 金鑰、OAuth)，並指派角色或層級，藉此僅處理已授權的要求。
- **輸入驗證和大小限制：**驗證提示大小和結構，封鎖或修整大型或格式錯誤的查詢。
- **套用處理逾時和資源節流：**為每個要求設定逾時和資源上限，以避免長時間執行的作業和資源耗盡。
- **部署智慧快取和重複資料刪除：**為重複或類似查詢快取回應，以減少不必要的處理。
- **監視、記錄和異常偵測：**使用 MDR/XDR/SIEM 等解決方案，持續監控和記錄 AI 系統活動，以快速偵測、調查和回應未經授權的存取、異常或資料洩漏。
- **預算追蹤和支出控制：**使用儀表板和警示來監控成本，並封鎖超過預算閾值的使用量。
- **沙箱和隔離技術：**在隔離環境中，以有限的權限來執行工作負載，降低風險。
- **限制呼叫深度和對話輪次：**對遞迴呼叫或對話步驟施加限制，以防止濫用。
- **套用分層模型或資源分配：**將高優先順序的要求路由至高階模型，並將低優先順序的流量路由至符合成本效益的資源。

返回架構圖



對 LLM 或 AI 系統的無限資源耗盡威脅，這類的安全漏洞是指應用程式允許使用者 (惡意或其他) 提交過多、不受控制的推論要求或提示，卻沒有有效的速率限制、驗證或使用限制。由於 LLM 推論的運算成本很高，因此惡意人士可能會透過多種方式，利用這種缺乏控制的情況：攻擊者可能會透過佔用系統資源，來造成阻斷服務 (DoS)；在按使用付費或雲端託管部署中，產生不可預見的經濟損失；或者系統地查詢模型，以複製其行為，並竊取智慧財產。其後果包括服務中斷、其他使用者的效能降級、財務壓力，以及敏感模型洩漏的風險增加。從本質上講，未正確控管資源使用方式時，就會發生無限資源耗盡，進而讓以 LLM 為基礎的應用程式面臨意外和蓄意濫用的情況。

為何選擇 Dell 來實現 AI 安全性

Dell 透過涵蓋硬體、軟體和受控服務的全方位方法，協助組織保護 AI 模型和 LLM。從供應鏈到裝置、基礎結構、資料和應用程式，均內嵌安全性，全部符合零信任原則。在整個產品組合中，Dell 的解決方案專為改善網路衛生而打造，具備 MFA、RBAC、最低權限及持續驗證等功能。這種全面性的「安全設計」方法，可確保組織能夠安心地利用 AI 和 LLM 進行創新，將模型盜用、資料外洩、對抗性攻擊及其他進階網路威脅的風險降至最低。

供應鏈

Dell 的安全供應鏈可在產品開發、製造和交付的每個階段內嵌安全機制，為 AI 模型和 LLM 提供基礎保護。透過密碼編譯簽署的 BIOS 和韌體更新、安全的元件驗證、以 AI 為主的軟體材料清單 (SBOM)、資料集譜系追蹤、整合式安全性軟體和組態，以及符合全球標準的嚴謹廠商風險評估，Dell 可將遭篡改、未經授權存取及供應鏈攻擊的風險降至最低，確保組織能以完全透明、具完整性及遵循法規的方式，部署值得信賴且具備復原能力的 AI 工作負載。

AI 電腦

Dell 為裝置上 AI 工作負載，提供基礎安全性。Dell Trusted Device 可說是全球最安全的商用 AI PC*，以安全性為設計初衷。供應鏈安全性可將產品漏洞和遭篡改風險降至最低。直接內建於硬體和韌體的獨特防禦機制，可在使用中保護電腦和終端使用者。Dell SafeBIOS 可提供深度 BIOS 層級可見度和防篡改偵測，而 Dell SafeID 則可強化認證安全性，並啟用無密碼驗證。合作夥伴軟體可提供跨端點、網路和雲端環境的進階保護。

網路韌性

Dell 的 PowerProtect 網路韌性解決方案可透過加密、不可變的備份、快速還原和隔離的網路復原存放庫，來保護 AI 資料的安全。這些功能可防止遭到破壞、緩解惡意更新的影響、支援法規遵循，並在攻擊後支援復原。

伺服器

PowerEdge 伺服器採用機密運算技術，可隔離並確保 AI/LLM 提示和內嵌。這款伺服器還採用基於權威來源、受信任的檢索增強生成 (RAG) 解決方案，還有 MFA、RBAC、晶片信任根、經簽署的韌體，以及持續監控，以保護關鍵 AI 工作負載。

儲存裝置

Dell 的儲存產品組合具備強大的 AES-256 加密功能，可確保為敏感的 AI 資料提供安全、加密的儲存裝置，適用於靜態和傳輸中資料。在特定產品中，提供具備韌性的進階加密功能，專為抵禦未來量子威脅而設計。產品組合包括高速 NVMe 效能、可保護資料 (包括用於 AI 工作負載的資料) 且符合 FIPS 規範的加密模組、不可變的快照，以及用於抵禦勒索軟體攻擊的實體隔離網路復原存放庫。零信任架構、供應鏈安全性及防篡改稽核功能，可強化治理能力。內建的異常偵測和 AIOps ML 模型可保護工作負載，不使用客戶資料進行訓練，藉此將輸入式攻擊風險降至最低。

AIOps

Dell AIOps 提供自動化的持續監控，以偵測錯誤組態、漏洞 (包括 CVE)，並支援影響 AI/LLM 工作負載的供應鏈風險感知。即時 CVE 掃描、智慧型警示和 AI 驅動的儀表板，透過標記異常和追蹤解決方案工作流程，來實現快速干預。內建法規遵循功能、角色型存取控制和自動化報告，有助於在工作負載中維持安全作業，同時提供順暢的 EDR/XDR 整合和 AI 導向的營運深入解析，包括受支援解決方案中的生成式功能，進一步提升 IT 效率。

網路

Dell Networking 解決方案透過強大的網路區隔，保護 AI/LLM 環境，將橫向移動降到最低。加密的網路路徑和整合式防火牆控制，可封鎖未經授權的 AI 資料存取。

AI 安全性與復原能力服務

Dell 的 AI 安全性與復原能力服務，旨在解決將 AI 整合至組織時產生的相關新風險。我們的服務專為與您的團隊合作而設計，讓您盡快啟用 AI，為您提供專業知識來指導策略規劃、解決方案實施和管理型安全服務，以減輕營運負擔，方便您運用 AI 安全創新。每一項都是為了協助組織因應日新月異的 AI 風險，以及最佳化和保護 AI 部署而量身打造。

Dell AI Factory

專為安全性打造的整合式產品組合，例如 Dell 的安全供應鏈、可強制執行最低權限的零信任功能，以及可確保您的模型安全無虞的 AI MDR 解決方案。

*根據 2024 年 10 月 (Intel) 和 2025 年 3 月 (AMD) 的 Dell 內部分析。適用於搭載 Intel 和 AMD 處理器的電腦。並非所有電腦均可使用所有功能。部分功能為選購。搭載 Intel 的電腦，經 Principled Technologies 驗證，2025 年 7 月。

結論

為了建構具韌性的 AI 架構，組織和安全專家之間的協作方法至關重要。隨著 AI 和 LLM 不斷重塑產業面貌，解決它們帶來的風險是關鍵所在，包括資料安全性、模型完整性和法規遵循挑戰。組織必須優先制定主動式策略，將安全性整合到 AI 旅程的每個階段。

在這個任務中，Dell Technologies 是值得信賴的合作夥伴，提供端對端的 GenAI 自訂、安全性諮詢，以及專為您的獨特需求量身打造的整合式解決方案。運用 Dell 強大的網路安全解決方案，企業可有效降低 AI 和 LLM 風險，同時充分發揮其現有安全性投資的潛力。Dell 可將進階安全性順暢整合至目前的架構，讓組織得以保護 AI 基礎結構，確保與時俱進的安全環境。

瞭解 Dell 的全方位 AI 解決方案如何 保護您的 GenAI 和 LLM 環境：Dell.com/CyberSecurityMonth