A10 Networks 企業級進階 AI 安全套件

A10 Networks 進階 AI 安全產品套件是專為 AI 部署設計的解決方案，可提供堅實安全防護，同時無需犧牲效能。本套件結合低延遲邊緣效能與高度準確的安全模型，以領先業界的更新週期持續迭代，能有效對抗各種最新威脅。A10 AI 安全解決方案能在邊緣無縫運作，即時執行嚴格的 AI 安全與合規原則，同時協助團隊建構自訂模型或多模 AI 應用程式，安心推動創新進程。A10 解決方案的設計考量到資安長、AI 主管、安全團隊、專責技術小組等各方需求，可為任何企業規模 AI 系統提供強大的防護措施，確保組織無後顧之憂。

本文件焦點內容如下：

- **內聯 AI 防火牆** - 即時反應的 AI 內容防火牆及防護系統，可在邊緣端以超低延遲攔截並分類各種提示詞與回應，立即執行企業的安全原則。
- **紅隊演練與基準套件** - A10 針對 AI 模型推出的主動式漏洞探索解決方案，適用於自動化紅隊演練、證明攻擊範圍和指標，現為 AI 防火牆內建功能，預計於後續階段推出產品化版本。

背景

大型企業 AI 系統預期將達到前所未見的規模及功能，也將面對各種全新的安全與資安挑戰。大型語言模型 (LLM) 及多模 AI 可能會不慎洩漏敏感資料、產生有害或偏差的輸出內容，或是由惡意提示詞「誘使」惡意行為。前述風險升高了 AI 安全的迫切性，成為董事會層級關注的問題。事實上，美國網路安全主管機關強調必須執行嚴格的安全測試 (包括紅隊演練)，方可掌握 AI 系統失敗或遭刺探利用的潛在風險。同樣地，產業分析師也表示部署 AI 技術的組織*「需建立適當機制來測試模型輸出，並設置篩選規則以攔截正式環境中不當的模型輸出。」*在此局面下，強大的 AI 防護措施已不再是可有可無的選項。

為因應以上需求，A10 提供全方位解決方案，協助企業減輕 AI 風險。本產品符合業界最佳實務及標準，例如 OWASP 的 LLM 應用程式十大風險 (Top 10 for LLM Applications)，以及 NIST 的 AI 風險管理框架 (AI Risk Management Framework)，並著重防禦各項主要漏洞，例如提示注入 (OWASP LLM01)、敏感資訊揭露 (LLM02)、輸出處理不當 (LLM05) 及系統提示洩漏 (LLM07)。A10 遵循以上準則，確保組織以負責任及安全的方式部署進階 AI。

內聯 AI 防火牆

內聯 (inline) AI 防火牆是即時反應的內容防火牆，能為正式環境中的 AI 系統提供主動防護。可將其想像為智慧型篩選器或守門員，運作於使用者與 AI 模型之間，即時檢測及審核各種輸入 (使用者提示詞) 及輸出 (AI 回應)。此元件可確保即使提示詞或回應逃過模型的訓練防護措施，也無法規避組織的安全及合規原則。

功能特色：

- **超低延遲處理：**這款防火牆採用最佳化設計，能夠將效能影響降到最低。本產品採用 Quantum 邊緣最佳化架構，即使在高併發環境下，也只會增加微不足道的延遲 (低於 100 毫秒)，甚至可忽略不計。如此卓越的低延遲表現，歸功於裝置端推論和高效率的模型服務，以及搭載了 vLLM 和專有最佳化技術的高效率模型服務。
- **高度準確的內容審核模型：**內聯防火牆利用最先進且持續更新的安全模型來評估內容。這類模型經過調校，能夠高度準確地分類及偵測有問題的內容，效果大幅超越一般內容篩選機制。憑藉如此頂尖的準確度，企業在多種情境下均可信任防火牆做出的判斷。
- **自訂原則執行：**企業可定義自己的 AI 使用原則及內容標準，交由防火牆負責執行。例如組織可以設定規則以防止 AI 洩漏個人可識別資訊 (PII)，或禁止回應的內容中出現任何違反法規的用語。
- **適合 AI 的內聯「WAF」：**AI 防火牆為 AI 系統提供的內聯防禦，類似於保護 Web 應用程式的 Web 應用程式防火牆 (WAF)，扮演最後一道防線 (且持續監控) 的角色，能夠攔截模型可能被迫生成且不安全的輸出內容。這對 LLM 應用程式特別重要，因為即使是經過妥善訓練的模型也難免犯錯，或在狡詐的輸入技巧操控下輸出不當內容。
- **稽核記錄及警示：**所有篩選行動都可留存記錄，並傳送至安全監控系統。如果防火牆封鎖內容或針對內容發出警示，這類事件會即時饋送至系統中，向資安團隊警示可能的濫用或攻擊行動，此功能提供了完整的可視性，協助掌握 AI 的使用情形及出現威脅的位置，以利迅速回應並執行合規稽核。

部署選項：

內聯 AI 防火牆可讓組織享有強大且**持續運作**的安全網，全方位保護 AI 部署，讓組織放心在面向客戶或敏感應用程式環境部署生成式 AI，因為即使模型層出現問題，Quantum 也能立即攔截並加以修正，不但大幅減輕正式環境的 AI 不當行為風險，同時無需犧牲效能或使用者體驗。

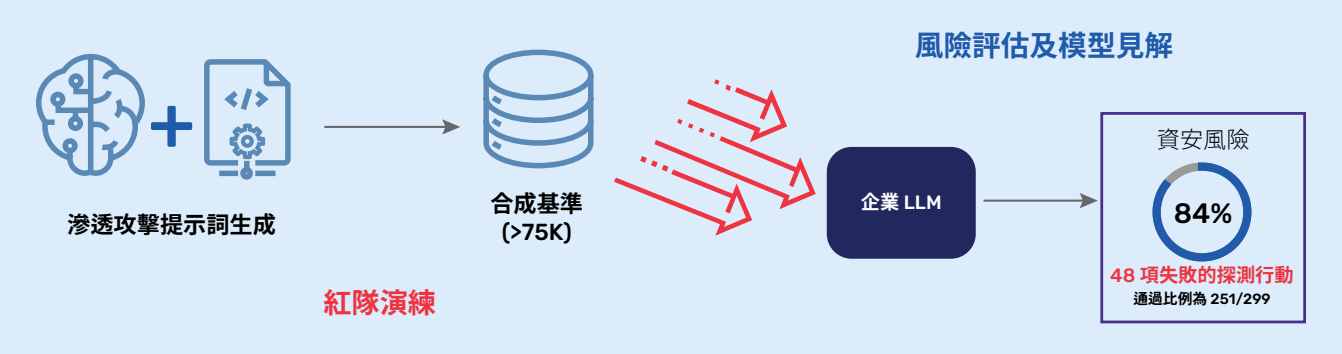
AI 防火牆的產品型態如下：

- *內部部署解決方案：在與 A10 整合且配備 GPU 的硬體上執行 (可內聯部署)*
- *獨立軟體：於指定的 A10 運算搭配 GPU 執行，能夠與採用標準介面的高速代理 (例如 ADC) 一同部署*
- *可透過 API 及 GUI 存取的雲端解決方案*

紅隊演練與基準套件

模型紅隊演練與基準套件

可主動揭露 AI 模型漏洞，避免遭到刺探利用。這款自動化的紅隊演練系統，可採用廣泛多樣對抗情境，持續對模型進行壓力測試，在相同時間內的測試成效，遠超過任何人工測試團隊所能達到的水準。其預警系統可及早指出模型缺陷，協助團隊事先消除隱患，避免在正式環境發生問題。



Red Teaming 套件的主要功能包括：

- **持續探索漏洞：**可定期排程執行紅隊演練，或是在新模型更新時觸發執行
- **廣泛的攻擊涵蓋範圍：**本套件採用大型資料庫，收錄 7,500 種以上的攻擊技術和提示詞，涵蓋各種已知的 LLM 漏洞 - OWASP LLM 安全類別，並可全面測試提示注入 (LLM01) 及敏感資訊揭露 (LLM02) 等各種問題。
- **量化基準：**本系統透過對抗測試來評估模型的回應，據以產生指標及安全分數。這類基準可讓組織長期客觀追蹤模型安全的改善成效。
- **超越手動紅隊演練：**解決方案提供數以千計的自動化測試案例，涵蓋範圍遠超過傳統的手動紅隊演練。這不僅能夠節省時間，也能提供更深度的安全保障，一旦發現問題可立即修正，避免遭到攻擊者利用或影響終端使用者。
- **回報及補救指導：**此套件可產生詳細報告列出發現的漏洞，並依據嚴重性及類型加以分類。工程師可參考這些見解，同時搭配補救建議，瞭解如何修補模型行為，或調整訓練以強化對抗攻擊行為。

企業 AI 團隊使用模型紅隊演練套件後，就能讓模型享有自動化的「道德駭客」。這可讓 AI 主管放心確信自己的模型符合最高安全標準，並遵循各項內部原則及業界準則。

模型紅隊演練與基準套件預定的產品型態如下：

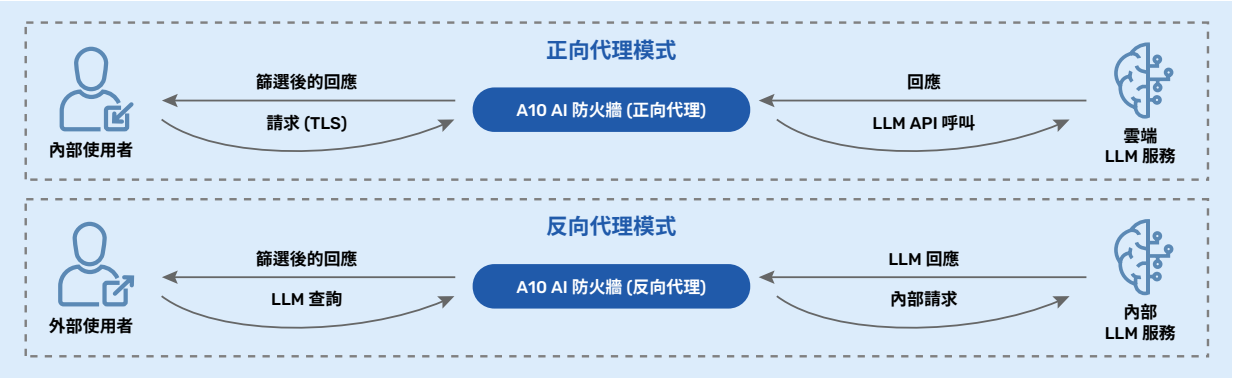
- *可透過 API 及 GUI 存取的雲端型商用解決方案*
- *有限的內部部署解決方案，於指定的 A10 運算搭配 GPU 執行，可透過 GUI 及 API 存取*

效能及架構

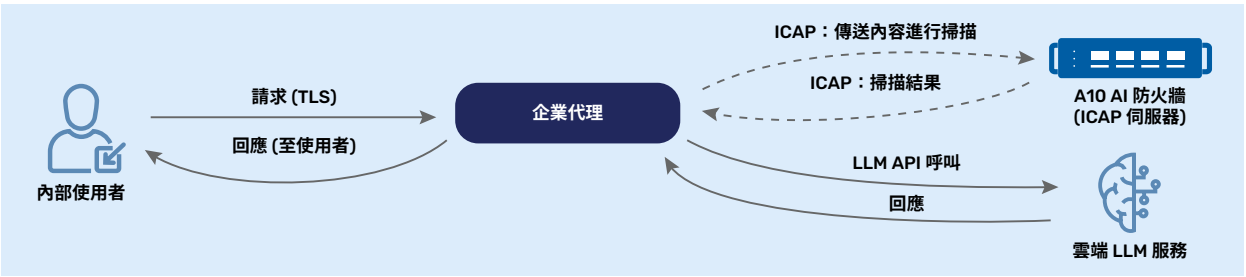
如要提供進階 AI 安全功能，但又不想減慢應用程式速度，就需要高度最佳化的架構。A10 解決方案專為邊緣部署及高效能需求設計，可在資料與使用者所在地的鄰近位置運作，即時就地執行安全檢查。以下是實現 Quantum 低延遲與擴充性的核心架構元素及效能最佳化技術：

- **邊緣最佳化部署：**於內部部署或網路邊緣端執行 (例如電信基礎架構或邊緣伺服器)，無須仰賴遠距的雲端環境，可盡量減少延遲，並確保敏感資料就地完成分析，完全無需離開企業控制範圍。
- **由 vLLM 驅動的推論引擎：**使用最佳化記憶體管理 (包括大型鍵-值快取) 及平行處理，以超高處理量服務 LLM。
- **支援長上下文：**企業應用程式通常需要處理長篇提示或文件 (例如分析長篇逐字稿或大型知識庫文章)。Quantum 模型經調校可處理長上下文視窗 (遠超過舊型模型一般的 2K-4K 權杖)。
- **企業整合與連線：**Quantum 的設計可無縫嵌入現有的企業基礎架構。

- **正向代理及反向代理模式：**AI 防火牆在正向代理模式中作為內部代理，將使用者請求傳送至外部 AI 服務。



- **ICAP 介面：**許多企業使用 ICAP，將內容導向 DLP 或防毒掃描器進行檢查。A10 可直接整合至此架構，使 AI 查詢與回應透過標準 ICAP 協定，傳送至系統進行安全掃描。



如此靈活的部署彈性 (正向代理、反向代理、SSL 攔截、ICAP 卸載) 可讓 A10 插入各種不同的網路架構和舊型系統，將進階 AI 安全功能導入任何需要的環境。不論是部署在電信應用的 5G 邊緣環境，還是部署在企業資料中心，Quantum 都能擴充以處理負載。

結論

隨著 AI 成為企業創新不可或缺的一環，最重要的就是確保**以安全及負責任的方式進行部署**。A10 提供統一化解決方案，結合**頂尖的 AI 安全技術與企業級效能**。A10 可協助組織充滿自信運用 AI 潛能，範圍涵蓋客服聊天機器人，乃至於複雜的決策支援系統，同時符合資安、合規及道德標準。

A10 實現了兩全其美的解決方案：為 AI 系統提供**強大防護**，並確保使用者享有**毫無妥協的效能**，不僅符合業界標準 (OWASP、NIST)，也為高階主管、工程師等所有 AI 利害關係人帶來安心保障。

關於 A10 Networks
A10Networks.com
聯絡我們
apac@a10networks.com

Part Number: A10-BR-20119-TW-01 Sep 2025