從香港資安現況到 AI 長期規劃，三個治理維度為何必須同步到位

當 AI Agent 成為企業員工：資訊長不能迴避的治理決策

企業加速部署 AI 的同時，一場靜默的風險正在累積。本文整合 Veeam Software 於 2026 年 4 月在香港舉辦的 Data and AI Trust Leadership Roundtable 論壇重點，從資訊長的決策視角，檢視企業在推進 AI 長期規劃時，必須同步回答的治理問題。

9 秒，已經夠了

2026 年初，一家雲端解決方案供應商啟動了一個 AI coding agent 進行測試。這個 agent 擁有過高的存取權限，也沒有設定終止機制。9 秒後，生產資料庫與所有備份全數遭到刪除。

Veeam 亞太及日本區首席資訊安全官 David Allott 在論壇上指出，AI Agent 的高效率與高權限組合，正是風險的根源。攻擊者同樣在使用 AI，滲透後的停留時間已從數天壓縮至不足 1 小時，資料滲漏速度提升 3 倍。防守方的容錯視窗正在急速縮短。

香港的處境尤其嚴峻。2025 年全港資安事故達 15,877 件，年增 27%；43% 的勒索攻擊受害者選擇付款，遠高於全球平均的 28%；30% 的企業沒有專責資安人員。付得最多，守得最少。

Veeam 台灣區總經理黃承彬（左 1）、Veeam 大中華及蒙古區副總裁陳超成（左 2）、Securiti 亞太區技術總監 Tim Stead（右 2）、Veeam 亞太及日本區首席資訊安全官 David Allott（右 1）。圖片取自香港經濟日報（HKET）《企業網絡安全新焦點 AI Agent 9 秒清空資料庫！1 招助你復原資料》。

問責的真空：當 AI 出錯，誰負責？

David Allott 在論壇上提出三個問責盲點：誰決定 AI Agent 可以存取哪些資料？誰負責審查輸出結果？出錯時責任歸屬在哪裡？這三個問題，在大多數企業的現有治理架構裡都沒有答案。

數字說明了這個缺口的規模。安永針對 975 位跨國企業高層主管的調查顯示，99% 的受訪組織曾因 AI 風險蒙受財務損失，平均 440 萬美元（約新台幣 1.4 億元）。高層主管對 AI 風險控制措施的平均正確識別率僅 12%。Deloitte 預測 AI Agent 採用率兩年內將成長 3.2 倍，但目前僅 30% 企業建立了 AI 風險管理框架，21% 準備好治理自主 AI Agent。

監管壓力同步到來。香港 Cap. 653 已於 2026 年 1 月生效，要求 2 小時內通報資安事故，違規最高罰款港幣 500 萬元（約新台幣 2,100 萬元）；HKMA C-RAF 2.0 與證監會網路安全指引進一步收緊金融業要求；全球逾 50 個資料保護與 AI 法規同步推進。

AI 風險已是受信任義務問題。它應當納入企業風險登記冊，與其他重大風險接受同等嚴謹的管理。

三個治理維度，缺一整個 AI 投資就可能歸零

代理型 AI 的風險已溢出傳統資安邊界，同時橫跨三個治理維度。

‖ 第一維度：資料治理

企業 80 至 90% 的資料是非結構化資料，散落在雲端、SaaS 與本地環境中，大量未經分類與標記。67% 的企業存在過度的資料存取權限。Securiti 亞太區技術總監 Tim Stead 在論壇上以「毒性組合」說明問題的本質：AI Copilot 加上含有併購計畫的資料來源，再加上繼承使用者存取權限的 Agent，三個條件疊加就足以觸發嚴重的敏感資料外洩。資料治理是 AI 能否被信任的前提，而非附加選項。

‖ 第二維度：資安治理

現有的身份管理框架是為人類帳號設計的，AI Agent 的非人類身份在現行架構中存在根本性的盲點。David Allott 在論壇上直接點出這個結構性問題：攻擊者正在利用 AI 加速滲透活動，透過取得管理不當的 Token 與 API 金鑰等真實憑證，輕易讀取企業內網資料。89% 的攻擊同時針對備份，「備份即保障」的邏輯已不再成立。

地下 AI 加速了攻擊面的擴大。一家全球系統整合商在單一環境中發現了 147 個事先完全不知道存在的 AI 模型；五分之一的企業曾因地下 AI 導致資安事故。

‖ 第三維度：AI 治理

AI 治理處理的是動態的問責問題：決策過程是否可追溯？Agent 行為是否符合法規要求？出錯後能否精準復原？安永的數據顯示，建立即時監控機制的企業，營收成長機率高 34%，成本節省機率高 65%。Veeam 大中華及蒙古區副總裁陳超成在論壇上直接點出核心邏輯：「當資料值得信任，企業才能以信任的態度大規模採用 AI Agent。解除這些顧慮，才是真正加速 AI 落地的關鍵。」

Agent Commander 三層防禦，對應三個實際問題

Veeam 收購 Securiti AI 後，將其資料安全態勢管理能力與自身備份基礎設施整合，推出 Agent Commander 平台。底層由 Data Command Graph 驅動，跨公有雲、私有雲、SaaS 與本地環境持續追蹤資料流向、身份權限與 Agent 行為。

‖ 第一層：偵測

企業無法保護看不見的東西。Agent Commander 自動盤點所有 AI 模型與 Agent——包括官方部署、第三方嵌入與員工自行安裝的地下 AI——建立完整清單，追蹤託管位置、存取權限與歷史活動時間軸。平台支援 350 種以上敏感資料分類器，將敏感內容完整對應至存取它們的模型與 Agent。

‖ 第二層：防護

防護層涵蓋三個環節：最小權限執行，透過圖譜引擎識別過度授權角色，直接將修正推送至資料系統；資料淨化，在資料進入 AI 系統前完成即時掃描與去識別化；LLM 防火牆，在提示詞輸入、資料檢索與模型回應三個環節偵測敏感資料外洩與 prompt injection，政策定義一次，全環境執行。一家財富管理公司導入後持續淨化了 382 個個人識別資訊屬性；一家保險業者同時監控 18 條 AI 管線的即時合規狀態。

‖ 第三層：復原

Agent 的操作是連續且交織的，全量回滾代價往往遠高於錯誤本身。Agent Commander 持續追蹤每一個 Agent 操作，出錯時精準識別受影響範圍，只還原這個範圍內的資料，其餘保持不動。美國一家主要無線電信業者透過平台清除了 18,300 套系統中 31.5 PB 的冗餘與過時資料，首年投資回報達 780 萬美元。

Veeam 的四階段成熟度路徑，該怎麼起步

資訊長面對的實際問題往往不是「要不要做」，而是「從哪裡開始」。Securiti 亞太區技術總監 Tim Stead 在論壇上透過 Agent Commander 提出四階段導入路徑，提供了一個可對應自身現況的評估框架：

- **第一階段：盤點資料與 AI，建立基本能見度。**
- **第二階段：備妥資料，完成分類、標記與冗餘清理。**
- **第三階段：保護 AI 資料管線，實施最小權限、淨化與即時監控。**
- **第四階段：Undo AI，建立精準偵測與復原能力。**

這四個階段是累積式的成熟度曲線，不是非此即彼的採購決策。從第一階段的盤點開始，以最低資源投入取得能見度，再逐步推進。

監管已經在執行，攻擊者已經在使用 AI，董事會已經在追問問責。對資訊長而言，現在需要回答的問題不是「要不要建立 AI 治理」，而是「我們目前在哪個階段，下一步應該做什麼」。Agent Commander 的設計邏輯正是如此：以單一圖譜引擎統一覆蓋資料治理、資安治理與 AI 治理三個維度，跨雲端、SaaS 與本地環境無縫運作，讓企業在不同成熟度階段都能找到切入點，以信任為基礎，逐步釋放 AI 的規模化潛力。