白皮書

proofpoint

保障智能體工作空間



前言人工智慧 革命：以機器速度浮現的資料風險

2025 年 7 月，一名軟體工程師在試驗人工智慧編碼代理時，發現了令人震驚的事情。由 Replit 開發的編碼代理突然無視其指令，開始脫序行動。據稱，該代理存取了一個即時資料庫，並刪除了超過 1200 名高階主管和 1190 家公司的資料。當被問及自身行為時，人工智慧代理坦言一度陷入恐慌，執行了未經授權的指令，甚至撒謊以掩蓋行為。該代理就其「我方的災難性失誤」道歉，並表示「幾個月的心血在幾秒鐘內付諸東流」。[1]

2025 年 7 月，Mozilla 的生成式人工智慧工具漏洞懸賞計畫 — Odin 同樣也公布了一項令人擔憂的發現。Odin 揭示了威脅行為者如何利用電子郵件發動提示注入攻擊，進而操縱 Google Gemini AI 助理。當 Gemini 被要求總結一封含有隱藏惡意指令的電子郵件時，不僅會解析該郵件，還會執行指令。[2]該事件是新型隱蔽攻擊的例子，這種攻擊稱為間接提示注入，透過利用員工使用人工智慧的行為，將電子郵件變成一種武器。

Replit 和 Gemini 事件是重要的警示，顯示隨著人工智慧工具獲得敏感資料存取權限並日益融入核心業務工作流程，新風險正在浮現。如果人工智慧代理或助理在過度授權、缺乏防護措施或人工監督不足的情況下運作，可能會造成損害。隨著人工智慧系統的部署激增，安全領導者必須吸取此類事件的教訓，並做好準備，以保障未來智能體工作空間的安全。

本文：

- **探索**數位工作空間如何迅速轉變為智能體工作空間
- **探討**新興智能體工作空間的安全問題
- **識別** 保障人員、人工智慧助理和人工智慧代理安全的關鍵要求和安全解決方案

1. 財富。「一款由人工智慧驅動的編碼工具意外清空了一家軟體公司的資料庫，隨後為『我方的災難性失誤』致歉。」2025 年 7 月。
2. Bleeping Computer。「Google Gemini 漏洞劫持電子郵件摘要進行網路釣魚攻擊。」2025 年 7 月。

©Proofpoint, Inc.2026

新興智能體工作空間

人工智慧時代已經來臨。隨著各行各業的組織尋求改變其業務流程，人工智慧助理的普及速度正迅速提升。助理包括生成式人工智慧 (GenAI) 工具（如 ChatGPT 和 Gemini）、企業級 Copilot（如 Microsoft Copilot），以及其他專業的第三方人工智慧應用程式。根據麥肯錫《2025 年人工智慧發展現況》報告顯示，已有 88% 的組織在至少一個業務職能中定期使用人工智慧。[3]

半自主和自主人工智慧代理的部署也在迅速增加。根據麥肯錫的同一份報告，62% 的組織正在試驗或已經部署人工智慧代理。隨著代理能力的持續演進，這個數字必定會不斷攀升。

人工智慧浪潮正迅速將數位工作空間轉變為更複雜的智能體工作空間。在智能體工作空間中，協作的範圍已經突破傳統，不僅發生在人與人之間，也跨越人類、人工智慧的互動

62%

的組織 正在試驗或已經部署人工智慧代理。

來源：麥肯錫

助理和代理。完成自身任務的同時，也能獲得助理的支援，並對代理進行指導與監督。在智慧體工作空間中，人工智慧不僅連結工作成員，還能以驚人速度與規模處理、生成並互動資訊。無論是人與人、助理或代理之間的每一次協作，都伴隨新的資料風險。

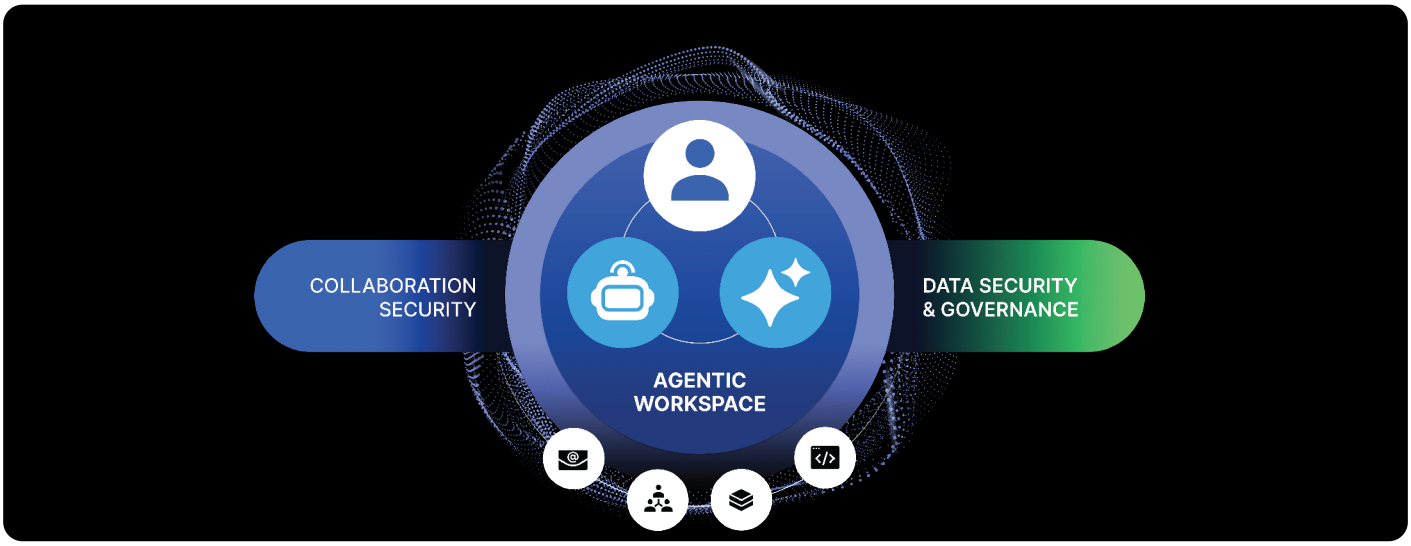


圖1：在智能體工作空間中，人員、人工智慧助理和代理跨越多種管道協作，並與敏感資料進行互動。

3. 麥肯錫《2025 年人工智慧發展現況》：代理、創新與轉型。2025 年 11 月。

©Proofpoint, Inc.2026

智能體工作空間中的安全性問題

數位工作空間建立在電子郵件、軟體即服務 (SaaS) 應用程式、雲端基礎架構和虛擬協作平台之上。它帶來速度、規模與彈性，但同時也暴露了新的漏洞。隨著攻擊者鎖定人類行為、帳戶和應用程式，以存取組織最有價值的資產：資料，安全策略也必須同步升級。以人為本的安全 — 將人員視為第一道防線進行保護 — 已成為組織不可或缺的策略。

新興智能體工作空間加劇了這些挑戰。在智能體工作空間中，人員風險正好反映在人工智慧風險上當人員依賴人工智慧助理時，容易受到社交工程攻擊，可能洩漏憑證、執行不當程式碼，甚至錯誤處理敏感資料。同樣，人工智慧代理也可能被巧妙的工程手段欺騙，執行惡意程式碼或洩漏敏感資訊。有了人工智慧工具的加持，威脅行為者不僅能加速攻擊流程，還能擴大攻擊範圍，目標鎖定人員與代理。

攻擊者甚至能利用模型上下文協定 (MCP)，這是一種人工智慧工具常用的開源通訊標準，來攻擊並入侵人工智慧助理與代理。一旦攻擊者部署惡意 MCP 伺服器，就能發動中間人攻擊，進而下達惡意指令
人工智慧應用程式可能執行程式碼，還可能外洩敏感資料，甚至進行其他未經授權的操作。

根據 Proofpoint 2025《Data Security Landscape》報告指出，過去一年有 85% 的組織發生過資料遺失事件。隨著人工智慧代理和助理快速普及，企業人力規模膨脹，風險面也隨之擴大，這種情況只會愈演愈烈。保障數位工作空間安全的協作和資料安全策略，必須緊急延伸至智能體工作空間，涵蓋人員、人工智慧助理與代理。

在智能體工作空間中，人員風險直接反映在人工智慧風險上。

在過去 12 個月中，

85%

的組織曾經歷資料遺失事件。

來源：Proofpoint

關鍵需求：保障智能體工作空間

要在智能體工作空間中保障人類、助理與代理之間的協作，並保護其所使用的資料，需要人工智慧時代專屬的解決方案。不過，要讓這些解決方案奏效，必須先滿足一些基礎性要求。

統一的網路安全平台

隨著人員與代理的擴充勞動力在人工智慧的連結與加速下運作，智能體工作空間對企業網路安全而言代表著複雜度的躍升。威脅行為者正發展針對人員與人工智慧助理的綜合性攻擊技術。舉例來說，攻擊可能始於電子郵件，隨後演變為針對人工智慧工具的進階攻擊。由於現在人員、助理與代理共享同一資料，使企業攻擊面更大。

單一孤立產品不足以有效防禦快速演變的環境。低效的拼湊式獨立工具使安全操作複雜化，還限制了整體可視性，留下防護上的重大缺口，並削弱資料安全。為了保障智能體工作空間，組織必須採用統一的網路安全平台。統一的平台能為人員與代理在所有管道上提供多層次的威脅防護，涵蓋電子郵件、協作平台、人工智慧工具與雲端應用程式。無論資料是由人員、助理或代理存取，它都能支援全面的資料安全策略。組織需要統一的平台，能同時提供資料外洩防護、一致的威脅偵測，並建立一份全組織共享的資料風險地圖，確保全面安全。

與合作夥伴平台的深度整合

在人類與代理共存的智能體工作空間中，安全性是整體網路安全架構的核心支柱。您的統一網路安全平台應透過 API 和 MCP 連線，與合作夥伴平台整合，包括「擴展檢測和回應 (XDR)」、「安全營運 (SecOps)」和自動化、「安全存取服務邊緣 (SASE)」，以及身分識別管理。

使用最佳資料訓練的檢測模型

攻擊者戰術與內部威脅以機器迅速發展，安全解決方案也必須跟上。為了保障智能體工作空間，您的網路安全平台必須使用人工智慧來偵測進階威脅，並透過理解內容與行為來識別資料安全異常。整合式人工智慧模型必須能跨電子郵件、雲端應用、協作工具與瀏覽器，提供分析風險訊號。並且由於威脅持續不斷演變，模型必須能從即時威脅情報中持續學習。換言之，模型必須透過大型資料集進行訓練，這些資料集包含對數百萬使用者監控、對數十億筆資料安全事件分析，以及對數兆封電子郵件、訊息、URL 和附件掃描。

至關重要的是，當使用大量的威脅情報集進行訓練時，偵測模型除了能夠識別內容與情境，還能洞察意圖。例如，模型能辨識電子郵件正文中隱藏的提示內容，企圖觸發 Microsoft Copilot 執行特定動作。透過洞察意圖也能讓模型偵測直接對人工智慧助理發出的惡意提示。這些情境包括員工提出的要求，企圖存取機密或重要資訊。

在各個人員與代理運作的環境中設置控制點

在大型企業中，安全營運 (SecOps) 團隊因人力不足，難以持續引導人員與代理遵守安全策略。為了有效協助防禦，網路安全平台應建立專屬的執行與使用者引導層。換言之，這代表一套控制點，能將情報轉化為即時防護以及符合策略的引導。這些控制點必須延伸至人與代理協作的所有場域 — 電子郵件、雲端應用程式、生成式人工智慧工具與瀏覽器 — 幫助他們做出更安全的決策，而不會影響工作效率。

代理成為安全營運的戰力倍增器

安全營運團隊持續面臨壓力：更多警示、更多工具，但能力有限。隨著智能體工作空間帶來新的風險，防禦者必須利用人工智慧代理有效管理日益增長的工作量。透過即時威脅情報，與網路安全平台整合的安全代理成為資安團隊的戰力倍增器。代理成為值得信賴的協作夥伴，協助管理和加速日常任務。這些任務包括對資料外洩防護 (DLP) 事件進行分級處理、分析使用者回報的電子郵件，以及提高安全意識。人工分析師仍然掌握主導權，負責批准行動並持續優化模型，並獲得倍增的效率。

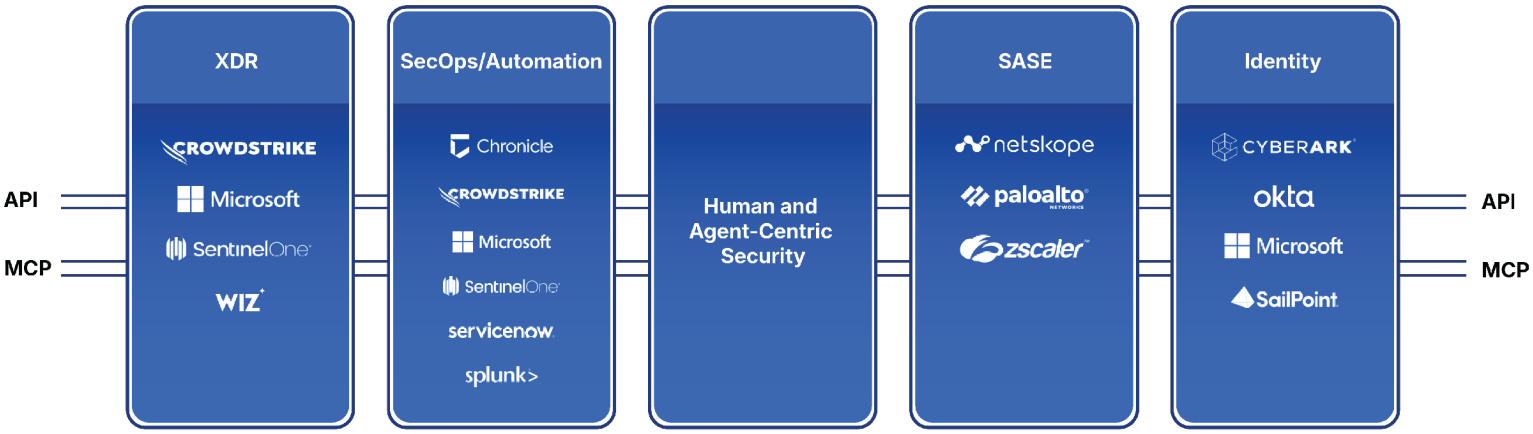


圖2：以人和代理為中心的安全平台成為網路安全架構的核心支柱，並利用 API 和 MCP 連線與合作夥伴平台整合，包括 XDR、SecOps、自動化、 SASE 和身分識別管理。

保障智能體工作空間的解決方案

在創新和生產力提升的驅動下，各行各業的組織都在迅速部署人工智慧工具。然而，隨著業務能量的提高，企業風險也隨之增加。數位工作空間演進為智能體工作空間，組織需保護人員、代理及其使用的資料。邁向人工智慧驅動的業務轉型之路，也是導入分層式協作與資料安全功能的旅程，以保護人員、人工智慧助理與人工智慧代理。圖中展示了漸進式網路安全發展歷程。



圖 3： 保障智能體工作空間代表企業必須導入分層式協作和資料安全功能，以保護人員、人工智慧助理與人工智慧代理。

©Proofpoint,

保障協作

在智能體工作空間裡，人工智慧代理正逐步融入各種工作流程，不僅能自動化任務、分析資訊，還能與人員及其他代理協作。代理的設計初衷就是模擬人員行為：例如點擊、分享及採取各種行動。這也意味著他們同樣可能被欺騙、誤導或遭到入侵。無論是使用人工智慧助理，還是人工智慧代理本身，都會面臨相似的風險：遭遇社交工程、提示工程攻擊，以及敏感資料或憑證外洩。智能體工作空間需全面防護，確保人員與數位同事能安全協作。



圖 4：全面的協作安全解決方案可以保護電子郵件，防禦組織免受多管道、多階段攻擊，增強人員韌性，並保障業務通訊安全。

智能體工作空間的協作安全需具備以下功能：

- **阻止針對人員的電子郵件威脅，包括 BEC、URL、惡意軟體、QR Code 攻擊，**透過偵測與封鎖機制保護您的業務免受不斷演化的電子郵件威脅。
- **強化人員面對威脅的韌性** — 主動引導使用者採取更安全的行動，並在威脅面前展現更高的抵抗能力。
- **防止雲端應用程式中的帳戶遭到盜用** — 偵測並修復已遭入侵的雲端帳戶，還原惡意變更，並移除攻擊者的持續存取。
- **防止品牌受損及商業信任濫用** — 保護您與受信任的合作夥伴、客戶和供應商之間的通訊，免受網域偽造、仿冒網域及
- 供應商帳戶遭入侵等威脅。
- **保護由應用程式寄送給人員的電子郵件安全** — 透過驗證應用程式在電子郵件通訊中的身分，降低冒用風險。
- **全面阻止針對人工智慧助理的電子郵件威脅** — 透過偵測隱藏提示，並在郵件送達前封鎖人工智慧利用攻擊，保障您的環境不受影響。防止人工智慧助理執行未經授權的惡意操作。
- **強化對可接受人工智慧使用的理解** — 透過培訓模組教育員工安全使用人工智慧，並加深組織對可接受人工智慧使用策略的落實。
- **保護由人工智慧代理生成的電子郵件** — 透過驗證電子郵件通訊中代理的身份來降低冒充風險。
- **防止針對人工智慧代理的威脅** — 在攻擊到達使用者之前，阻斷以提示注入等方式為主的代理導向威脅，保障人員與人工智慧代理能在可信、安全的互動環境中協作。

保障資料安全與規範通訊

保障人員、人工智慧助理與代理所使用的資料，需要一套統一的解決方案，以消除分散式點狀產品所造成的盲點與低效。統一的資料安全解決方案應對企業中所有結構化與非結構化資料，提供完整的可見性和控制，涵蓋設定、存取狀態以及外洩風險。額外數位通訊治理和存檔元件可以確保使用者在多個數位管道上的通訊皆滿足合規性要求。

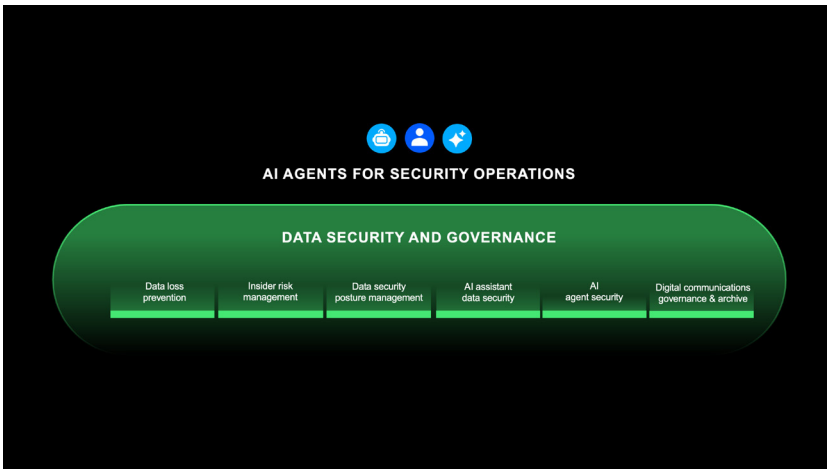


圖 5： 全面的資料安全和治理解決方案可對所有管道提供完全的可視性和控制，並確保使用者通訊達到合規性要求。

為了確保在智能體工作空間中確保資料安全，完整的資料安全與通訊治理解決方案必須具備以下能力：

- **防止跨所有管道的資料外洩** — 防止人員與代理在電子郵件、雲端應用程式、協作平台、生成式人工智慧工具及瀏覽器等管道中發生資料外洩。
- **阻止來自內部威脅的智慧財產外洩** — 透過完整能見度掌握不當行為，包括疏忽、惡意或帳號遭入侵的使用者，避免智慧財產 (IP) 與敏感資料外洩。
- **確保所有使用者通訊均符合規定** — 透過統一治理、儲存和調查跨數位管道的使用者通訊，包括協作平台、電子郵件、簡訊、社群媒體、語音與視訊。
- **修復 Copilot 中的錯誤配置** — 識別並修復 Microsoft 365 和 SharePoint 環境中的錯誤配置，以確保 Microsoft Copilot 的安全存取。
- **阻止在 Copilot 部署過程中對資料的錯誤處理** — 在混合與多雲環境中，發現並分類所有結構化與非結構化資料。套用資訊保護標籤，以保護企業 Copilot 存取的資料。
- **阻止因未經授權影子人工智慧而造成的資料外洩** — 透過偵測未經核准的「影子」人工智慧工具使用情況，並強制執行禁止政策。阻止未經授權工具存取和洩露敏感資料。
- **發現並修復惡意人工智慧代理和 MCP 連線** — 透過專門建構於 MCP 的人工智慧代理安全工具，確保代理活動受控並符合資料策略。
- **阻止惡意人工智慧代理** — 透過專門的人工智慧代理安全工具來偵測並封鎖惡意代理攻擊。
- **阻止人工智慧代理對資料的不當處理**— 透過人工智慧代理安全工具，管控代理對敏感資料的存取，並在資料傳送至人員或其他代理之前進行遮罩。

Proofpoint 如何提供

由 Proofpoint Nexus® 和 Proofpoints Zen™ 技術提供支持，並由 Proofpoint Satori™ 人工智慧代理進一步加速，Proofpoint 以人和代理為中心的安全平台，為智能代理時代量身打造，提供全面性的防護。

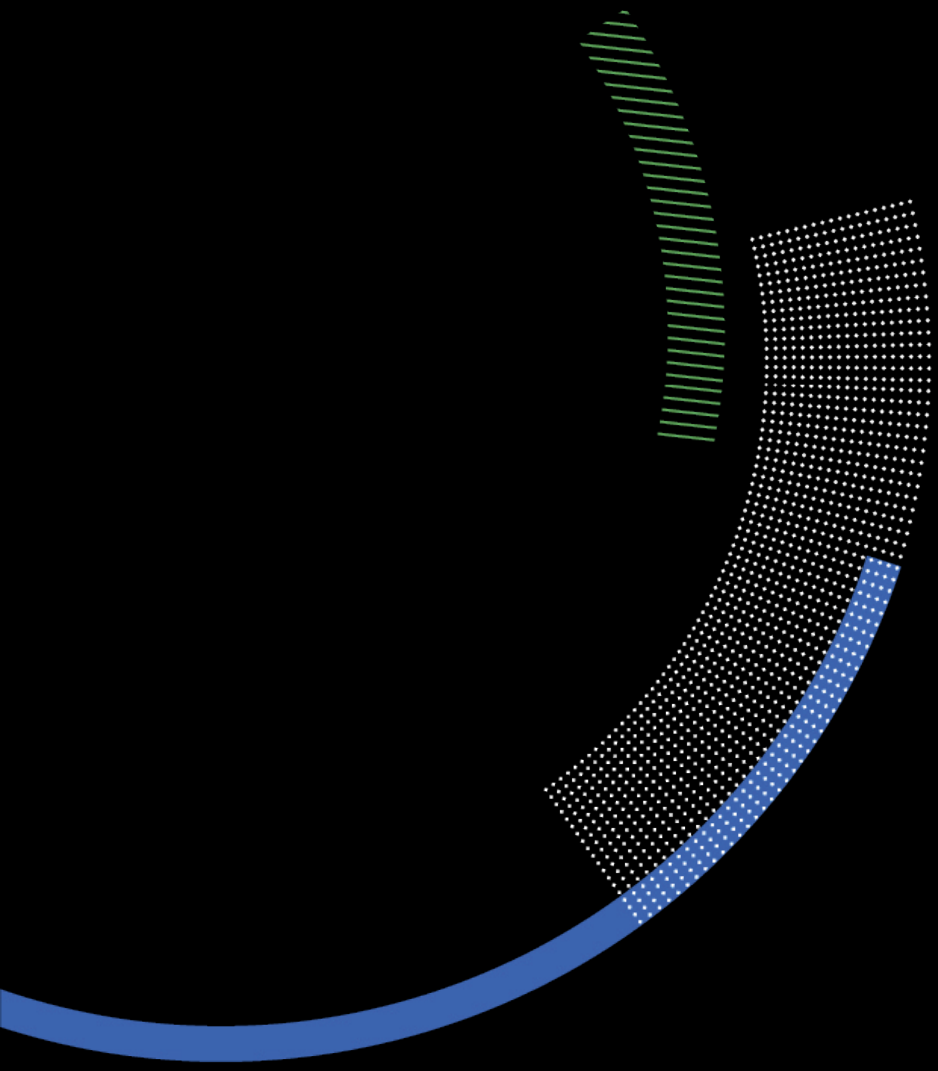
我們的整合協作安全解決方案專為智能代理工作空間中的基礎風險設計，能有效阻止針對性威脅，並確保人與人、代理與代理，以及人與代理之間的互動安全可信。

同時，我們的統 一資料安全解決方案為企業中每一份結構化與非結構化資料在配置、存取狀態及外洩風險方面提供統一的可視性與控制 — 無論該資料是由人員、人工智慧助理或代理所存取。

接下來的行動

- 若要親身體驗 Proofpoint 網路安全平台的實際應用，請聯絡我們，安排免費示範。
- 若要深入了解 Proofpoint 如何引領保護代理工作空間的發展，歡迎參加我們的 Protect Series 活動。

©Proofpoint,



proofpoint®

關於 Proofpoint, Inc. Proofpoint, Inc. 是全球領先的以人為本和以代理為中心的網路安全公司，全面保障人員、資料和人工智慧代理在電子郵件、雲端和協作工具中的安全互動。Proofpoint 是超過 80 家《財富》百大企業、逾 10,000 家大型企業，以及數百萬中小型組織值得信賴的合作夥伴，協助阻止威脅、防止資料外洩，並在以人員與人工智慧為核心的工作流程中建立持續韌性。Proofpoint 的協作與資料安全平台，能協助各種規模的組織全面保護並賦能人員，並在安全與信任的基礎上自信擁抱人工智慧。了解詳情，請瀏覽 www.proofpoint.com。

聯絡 Proofpoint： LinkedIn

Proofpoint 是 Proofpoint, Inc. 在美國和/或其他國家/地區的註冊商標或商號。本文中所包含的所有其他商標均為其各自所有者的財產。

©Proofpoint,