星球永續健康線上直播

智慧數位資安 (4)

釣魚攻擊智慧生命週期治理

2026 年 4 月 22 日

當前的戰爭與衝突，不僅是傳統軍事對抗的延伸，更反映出人類社會數位化發展進入以「數位威脅」新階段。跨國家、跨組織乃至個體之間的信任機制，正面臨前所未有的衝擊與重構，使數位資安不再僅屬技術層面的議題，更成為攸關整體社會運作與制度穩定關鍵基礎。以人工智慧驅動的釣魚攻擊為代表的資安威脅已由過去大規模、低精準度的攻擊模式，轉變為高度個人化且具情境合理性的精準滲透。本週將探討人工智慧資安生命週期治理，以及 AI 釣魚攻擊生命週期評估。

健康科學新知

美國-伊朗談判僵持:「危局僵持」

中東衝突已由傳統軍事對抗的框架，演變為同時牽涉核擴散管控、能源通道主權、區域安全秩序與大國外交博弈的複合性危機。今年 2 月美以對伊朗發動聯合軍事行動，迄今已在多個層次上深刻重塑中東地緣政治，並對全球能源市場與國際外交秩序產生廣泛衝擊。在核談判層面，美伊雙方的根本分歧至今仍未縮小。美國要求伊朗接受長達二十年的核活動全面暫停，涵蓋停止鈾濃縮與處置現有濃縮鈾庫存。伊朗則僅願考慮短期、有限度的限制措施，堅持和平使用核能屬於不可讓渡的主權權利，並明確以解除制裁及解凍海外資產作為談判前提。雙方在核權利的定義範疇、制裁交換的具體機制以及互信建立的優先順序上存在根本性分歧，使得即便談判管道重新開啟，達成全面協議的可能性依然極為有限。荷莫茲海峽問題則構成這場衝突另一個關鍵的地緣支點。伊朗藉由對海峽的實質控制形成對外施壓的戰略籌碼，美國的反制封鎖行動進一步激化雙方對峙，導致國際油價在衝突高峰期一度突破每桶百美元，並在全球範圍內推升通膨壓力。由於全球約 20%的石油與液化天然氣出口須經此咽喉要道，這場戰爭早已不僅是區域衝突，而是直接牽動全球經濟穩定的系統性風險。能源供應衝擊所引發的外溢效應，也反向形

成美國國內的政治壓力，迫使川普政府在維持軍事強硬姿態與保留外交談判空間之間持續尋求平衡。在大國外交層面，巴基斯坦藉由伊斯蘭馬巴德的多輪接觸及軍方高層訪問德黑蘭，逐漸在美伊之間扮演溝通管道的角色。然而，巴基斯坦的斡旋能量有其結構性侷限為訊息傳遞者，而非擁有雙方高度互信主導調停方。其外交能動性背後，既涉及中國在幕後發揮的隱性影響力，也折射出巴基斯坦自身藉此提升國際地緣政治能見度的戰略盤算。因此，巴基斯坦雖能協助促成初步接觸、傳遞談判訊號，卻尚不具備單獨左右停火進程或和平框架走向的實質能力。綜合以上分析，當前局勢最可能走向的結果，並非一份能夠終結戰爭的全面和平協議，而是某種有限、暫時性降溫安排。伊朗於 2026 年 4 月 19 日宣布重新封鎖荷莫茲海峽，伊斯蘭革命衛隊海軍明確警告，任何試圖進入海峽的船隻將被視為與敵人合作並遭到攻擊。同日已有多艘商船與油輪在海峽附近遭到炮擊，印度亦就兩艘懸掛本國旗幟的船隻中彈事件向伊朗提出嚴正交涉。荷姆茲海峽航道成為這場複合危機核心節點。

美國調停以色列-黎巴嫩衝突:「停火未安」

在美國調停下，以色列與黎巴嫩之間目前已出現停火安排，這使長期受戰火影響的居民得以陸續返回家園，也讓這場危機暫時從高強度軍事對抗轉入停火後的重建與政治協商階段。不過，居民返鄉所見並不是恢復正常生活的起點，而是滿目瘡痍的現實：南貝魯特及黎巴嫩南部多地可見被炸毀的建築、焦黑的外牆、散落的瓦礫、受損的商店與無法居住的住宅。許多人回到社區後，只能短暫查看房屋、取回少量物品，隨後再度離開，顯示停火雖然降低了即時軍事危險，卻未真正解除人民的生活危機。當前黎巴嫩危機不能只被理解為單純的以黎衝突，而必須同時從軍事、人道、政治與外交多個層面加以把握。以色列在黎巴嫩南部邊境村莊的大規模爆破與夷平，造成的並不只是軍事目標的摧毀，更包括居民家園、地方經濟、社區網絡、家庭記憶與離散社群精神依附的深刻破壞。戰爭的影響生活世界瓦解。停火讓人們可以返回，但返回後面對的卻是失序與失落，這使「回家」本身成為一種帶有創傷色彩的經驗。外交方面上周由美國主持的華盛

頓會談，象徵以色列與黎巴嫩三十多年來重新出現直接接觸，也顯示美國正試圖以調停者角色推動局勢降溫。然而，這一進展仍然十分脆弱。以色列與美國傾向將和平理解為削弱真主黨、解除其武裝，並重新塑造黎巴嫩南部的安全秩序；黎巴嫩政府則更重視停火的延續、人道壓力的減輕、居民返鄉與國家主權的恢復。雙方對和平的理解並不一致，因此停火雖已出現，真正的政治解決卻仍相當遙遠。其中最難突破的兩大問題，仍是真主黨解除武裝與黎巴嫩主權安排。真主黨的武裝力量不只是軍事工具，更是其政治正當性、組織存在與支持者安全感的重要基礎，因此不可能輕易放棄。黎巴嫩政府雖主張由國家軍隊壟斷武力，但受限於宗派分裂、軍隊能力不足，以及真主黨在社會與政治中的深厚基礎，實際上缺乏強制解除其武裝的能力。於是，談判形成根本矛盾：以色列真正交戰的對象是真主黨，卻只能與黎巴嫩政府協商，但黎巴嫩政府無法完全控制真主黨，停火協議穩定性仍需要觀察。

伊美衝突持續影響石油供給:「能源挑戰」

中東戰爭已使全球能源市場進入高度失衡狀態，問題核心不只是油價上升，而是荷姆茲海峽受阻、美國封鎖伊朗港口、原油供應短缺與運輸成本飆升，共同造成全球石油供應鏈重組。表面上，布蘭特與西德州原油期貨價格仍在每桶 100 美元上下，似乎尚未全面失控，但實體原油現貨、到岸成本與煉油利潤已承受遠比期貨價格更沉重的壓力。期貨市場呈現的價格訊號，並未充分反映當前危機的真實嚴重程度。當前危機最重要的特徵，是供應與運輸短缺先發生，再反過來壓縮需求。全球原油與凝析油產量損失估計已達每日 1,300 萬桶，約占全球供應的 15%，規模極為罕見。與疫情時期由需求崩跌導致產量下滑不同，這一次是供應受阻、運輸受限、成本上升，迫使消費與生產活動被動收縮。這也使「需求破壞」成為未來市場調整的重要機制。荷姆茲海峽承載全球約五分之一石油供應關鍵航道，在戰爭與封鎖之下通行嚴重受阻。即使市場一度因美伊可能重啟談判而出現樂觀情緒，油價短暫回穩，但實際海峽流量仍遠低於正常水準，顯示供應風險並未真正解除。換句話說，金融市場的平靜，並不等於實體能源市場已恢復正常。

文稿也說明，美國對伊朗港口的封鎖，使中東戰事進一步外溢為全球能源與大國競爭問題。中國批評美國此舉危險，因為這不僅破壞脆弱停火，也威脅中國自身能源進口與海運安全。亞洲則因高度依賴荷姆茲海峽石油供應，開始大舉搶購大西洋市場原油，進一步推高歐洲與美國的現貨價格。這顯示中東戰爭雖發生於區域，但其能源後果已迅速擴散至全球。印度是危機中最脆弱的國家之一。由於高度依賴進口，且同時面臨伊朗原油再度受阻與俄羅斯原油豁免到期，印度承受了雙重供應擠壓。其石油儲備有限，宏觀經濟指標也已受到衝擊，反映能源安全問題已從進口與煉油層面，擴大為經濟成長與外交自主空間的壓力來源。這也顯示，在大國制裁與區域戰爭交錯下，中等強權的政策操作空間正被明顯壓縮。

另一方面俄羅斯則因美國短期放寬制裁而在 2026 年 3 月暫時恢復部分石油出口收入。這說明美國能源政策本身也陷入矛盾：一方面希望透過制裁維持地緣政治壓力，另一方面又不得不顧及全球市場穩定與油價風險。俄羅斯收入的反彈並非穩定復甦，而是政策暫時鬆動下的短期結果。若戰爭無法迅速結束，當前看似尚可維持的局面，很可能只是更大調整前的過渡階段。

美國封鎖重擊古巴:「能源反噬」

美國對古巴形成事實上的石油封鎖後，古巴的醫療系統與科學研究都遭到嚴重衝擊。長時間停電、燃料短缺與交通癱瘓，使醫院難以維持正常運作，也讓許多研究計畫被迫中止或延後。文章以哈瓦那的醫院與生技研究機構為例，呈現這場危機如何從能源問題擴大成公共衛生與科研體系的全面壓力。在醫療方面，醫師必須在一天長達 20 小時以上的停電中勉強維持照護，甚至在手術進行時也可能突然失去電力。醫院不只設備運作受影響，也逐漸面臨物資不足。文章提到，孕產婦照護、重症手術與基本公共衛生都受到威脅，外部觀察者也注意到營養不良、垃圾無法清運與生活環境惡化等問題，顯示危機已深入一般民眾的日常生活。在科研與生技產業方面，古巴原本就因美國長年制裁而難以取得設備、試劑與國際合作資源，如今又因燃料危機、航班取消與資金來源減少而雪上加霜。儘管古巴曾發展出肺癌免疫療法與 COVID-19 疫苗等重要成果，許多臨床

試驗仍被迫暫停，國際合作也逐漸萎縮。

古巴科學界仍努力撐住基本運作，例如安裝太陽能板、使用備用發電機、讓研究人員在家工作，或以電動腳踏車協助通勤。某些重要的臨床試驗仍持續進行，像是阿茲海默症與癌症相關藥物的研究。文章最後傳達的重點是，古巴科學家最重要任務是保存現有能力與人才等待情勢好轉。

環境化學物質誘發 腸道菌群失調與糖尿病:「微菌失衡」

農藥可能透過擾亂腸道微生物群，進一步影響代謝、免疫、神經功能，甚至與非肥胖型第二型糖尿病有關。文章以印度農工長期接觸農藥後罹患糖尿病與腸胃不適的案例開場，指出這類現象讓研究者開始懷疑，農藥的健康效應可能不只限於急性中毒、神經毒性或致癌性，還可能牽涉到較隱微但長期的腸道生態變化。近期科學雜誌撰文全球農藥使用量持續上升，而新研究工具讓科學家得以更細緻地觀察農藥對腸道菌相的影響。印度研究團隊發現，農村地區雖然糖尿病盛行率低於都市，但仍有相當比例患者，且與肥胖、高膽固醇等典型危險因子無明顯關聯，因此推測環境化學物質可能參與其中。後續小鼠研究顯示，常用殺蟲劑 chlorpyrifos 會改變腸道菌群組成，使有益菌下降、潛在有害菌增加，同時造成高血糖與糖尿病表現，而且這些小鼠並未變胖，顯示其病理途徑可能不同於一般生活型態相關的糖尿病。文章也指出，農藥的影響不只是改變細菌種類，還會改變微生物產生的代謝物。這些代謝物與腸道屏障、發炎反應、免疫調節及腸腦軸訊號密切相關，因此農藥暴露理論上可能進一步影響情緒、行為與其他全身性功能。某些人體研究也發現，一般人尿液中普遍可檢出農藥殘留，且殘留濃度較高者，腸道微生物組成與代謝型態也有所不同。不過目前仍缺乏足夠堅實的人體因果證據。因為飲食、生活方式、基因與多重化學暴露都會影響腸道微生物群，所以要證明農藥就是造成特定疾病的直接原因並不容易。研究者認為，未來需要更多介入性研究，例如降低飲食中的農藥暴露後觀察腸道菌相是否改變，才能更清楚確認其作用機制。現階段最實際的建議仍是盡量減少農藥暴露，但對許多農業工作者而言，這往往不容易做到。

AI 創新強化與資安隱患:「利害並生」

人工智慧正快速進入醫藥與金融等關鍵產業，並同時展現創新潛力與風險挑戰。在醫藥領域，Novo Nordisk 與 OpenAI 的合作，顯示 AI 已被用於加速藥物研發、提升資料分析能力，並延伸至製造與營運層面，成為企業全面數位轉型的重要工具。同時，其他藥廠亦積極投入 AI 藥物開發，反映 AI 已成為產業競爭核心。然而，在金融與資安領域，Anthropic 的新模型展現出能發現大量系統漏洞的能力，引發包括 Bank of England 在內的監管機構高度警覺，並啟動跨部門風險評估。此顯示 AI 技術同時具有強化安全與放大威脅的雙重特性。人工智慧一方面推動醫療創新與效率提升，另一方面也對金融穩定與國家安全構成潛在衝擊，凸顯其已成為影響產業發展與監管體系的關鍵基礎技術。

AI 訓練記憶雙面刃:「記憶成患」

最新研究指出，大型語言模型在訓練過程中可能出現「記憶化」現象，不僅學習語言規則，還可能逐字記住並輸出訓練資料，帶來版權與隱私外洩風險。研究團隊開發開源工具 Hubble，系統性分析不同資料在模型中的保留情形，發現越晚加入訓練的資料越容易被記住，凸顯模型效能與資訊安全之間的張力。未來若能發展「去學習」技術，有望讓模型選擇性遺忘特定內容，兼顧應用價值與風險控管。

生成式 AI 醫學影像挑戰:「真偽難辨」

生成式人工智慧已能產生高度擬真的醫學影像，卻也帶來真偽難辨的挑戰。最新研究指出，放射科醫師在未提示下僅約四成能察覺 AI 影像，整體判別準確率約七成五，連大型模型表現亦有限。此現象恐影響臨床決策、研究可信度與法律判讀。專家建議透過訓練辨識技巧、導入數位浮水印與監管機制，以提升影像來源透明度，在促進技術應用同時降低潛在風險。

虛假引文影響科學文獻 「虛證成災」

大型語言模型（LLM）廣泛應用於學術寫作，卻引發「幻覺引用」問題。最新分析指出，AI 可能生成不存在的文獻或錯誤引用，且已滲入正式出版物。研究顯示，2025

年約 2%至 6%的論文含有潛在虛假引用，高於過去水準，甚至推估影響逾十萬篇文獻。此現象不僅干擾研究查證，也可能誤導科學發展。期刊與出版商已開始加強審查與撤稿機制，但專家強調，研究者仍需對引用內容負最終責任，以維護學術可信度。

AI 資安生命週期治理

電影《終極警探 4.0：網路恐攻》引入現代網路攻擊的運作模式。相較於過去以實體對抗為主的動作情節，攻擊場域轉向數位空間，呈現關鍵基礎設施遭受網路滲透的風險。劇情設定於美國國慶連假前夕，一個匿名組織以「弱點測試」為名，招募具備高階技術能力的駭客，要求其撰寫可穿透政府核心系統的程式碼。受測試的對象涵蓋多個關鍵部門，包括金融體系、聯邦調查機構以及交通與通訊等基礎建設系統。然而，這項所謂的測試實際上是一場精心設計的攻擊行動。當駭客完成程式並提交後，隨即遭到預先植入裝置的遠端引爆而喪生，顯示攻擊者在技術利用之外，同時進行資訊封鎖與人員清除，以確保攻擊計畫不被外洩。隨著事件發展，聯邦調查機構發現其系統已遭植入後門，關鍵作業網路受到干擾甚至無法正常運作。同時，攻擊者透過製造實體威脅迫使人員撤離，使防禦能力進一步削弱。此一策略不僅體現網路攻擊與實體行動的結合，也反映出「分散注意力」與「削弱應變能力」的複合式攻擊手法。在取得駭客所撰寫的攻擊程式後，主謀得以進一步滲透金融、交通與其他關鍵系統，形成對國家基礎設施的全面控制能力。主謀 Gabriel 鎖定駭客 Matt 的 IP 位址後，派出行動小組進行滅口行動，企圖在攻擊全面展開前排除潛在變數。在此同時，聯邦調查機構亦針對全國既有監管名單中的駭客展開追查行動，並派員前往接觸 Matt，以釐清事件全貌。於公寓突襲過程中，Matt 遭遇武裝攻擊，所幸在關鍵時刻成功脫離現場，並與前來調查的執法人員共同撤離。隨著事件升級，Matt 透過其駭客網絡建立外部通訊管道，並與其技術夥伴合作，逐步拼湊出整體攻擊架構，確認此次行動並非單一入侵事件，而是一項高度組織化的系統性攻擊計畫。進一步分析顯示，Gabriel 所啟動的核心行動為所謂的「Fire Sale」計畫。此一概念象徵對國家關鍵基礎設施的全面性癱瘓，其目標並非單點破壞，而是透過分階段攻擊造成系統性崩解。在第一階段，攻擊鎖定交通網路與運輸系統，包括道路、航空與鐵路，

使整體運輸體系同步失效，造成大規模混亂。第二階段則轉向金融體系，入侵華爾街與全國財經機構，導致股市劇烈波動，ATM 與銀行系統全面停擺，使經濟活動陷入癱瘓。第三階段則進一步延伸至電力系統，透過能源基礎設施的失效，完成對整體社會運作的全面瓦解。此一攻擊模式顯示，現代網路攻擊已不再侷限於資料竊取或單一系統入侵，而是朝向跨系統、跨領域的整合性攻擊發展，其影響範圍可迅速擴散至整體社會運作層級。最終，在多方合作下，透過技術分析與系統重建，逐步恢復關鍵系統運作，並揭露整體攻擊架構，成功阻止危機進一步擴大。

在智慧數位資安的面向中，可以清楚看到「攻防同源」核心概念。相較於過去資安多以防護為主，現在的情境已轉變為攻擊與防禦來自同一技術體系、彼此交織並持續演化的關係。一方面是「AI 強化資安」（AI for Cybersecurity），透過人工智慧協助資安防護，例如資安營運（SOC）自動化、異常偵測以及威脅情報的自動分析，提升整體防禦能力。但另一方面，同一套 AI 技術也成為攻擊的目標，也就是「資安保護 AI」（Cybersecurity for AI）。例如攻擊者可能透過訓練資料汙染、提示詞攻擊、注入式攻擊，甚至模型竊取等方式，直接對 AI 系統進行滲透與操控。因此，AI 同時既是防禦工具，也是攻擊標的。在這樣的架構下，資安已不再只是單向的防護問題，而是需要在「防禦能力」與「攻擊預測」之間，同時進行整合與治理的動態系統。

在智慧時代的數位資產防護中，防護目標已不再侷限於單一系統，而是需要從多個面向進行整體性的規劃與治理。首先，在「資料（Data）」層面，人工智慧高度依賴大量資料進行訓練、驗證與即時輸入，因此資料本身即成為最關鍵的資產之一。然而，資料來源在蒐集與處理過程中，可能面臨惡意竄改、外洩或污染等風險，進而影響後續模型的學習結果。其次，在「模型（Model）」層面，無論是基礎模型、微調模型，或是大型語言模型，其權重參數與 embedding 結構，都可能成為攻擊目標。例如參數被竄改、後門植入，或模型行為被操控，皆可能導致整體系統產生偏誤甚至錯誤決策。第三，在「系統（System）」層面，包含 API、推論伺服器、GPU 運算環境以及向量資料庫等基礎架構。若系統層遭受攻擊，例如 API 濫用或權限配置錯誤，將直接影響模型的運作與服務

穩定性。在「知識（Knowledge）」層面，現代 AI 系統透過 prompt templates 與 RAG（檢索增強生成）機制建構知識體系。然而，一旦前端資料、模型或系統受到污染，最終產出的知識內容也可能隨之失真，進而影響決策品質。在「身分（Identity）」層面，使用者帳號、API Key 以及各類授權機制，一旦遭到竊取或濫用，將可能導致系統被全面入侵，甚至擴散至金融、醫療等高度敏感領域，對個人與組織造成重大風險。最後，在「信任（Trust）」層面，這是整體數位資產防護的核心。AI 模型的輸出是否可信、決策是否具可解釋性，以及是否能維持品牌與系統的信譽，皆直接影響其在現實世界中的應用價值。因此，當數位資安由傳統的靜態防護，進一步擴展至涵蓋資料、模型、系統、知識、身分與信任的動態治理架構時，真正的挑戰已不僅是技術問題，而是如何確保整體系統在面對攻擊與不確定性時，仍能維持其可靠性與可被信任。

在智慧產品的資安生命週期規劃中，若從技術流程來看，整個 AI 系統從資料蒐集到模型運作的每一個階段，皆可能成為攻擊的切入點。首先，在「資料蒐集與標註（Collection）」階段，主要風險來自資料來源不明與惡意樣本混入，也就是所謂的資料汙染（data poisoning）。特別是在醫療等高價值資料領域，若資料在蒐集初期即遭污染，將直接影響後續模型訓練的正確性。接著，在「模型訓練（Training）」階段，可能出現後門植入（backdoor insertion）的風險。攻擊者若在訓練過程中植入隱藏機制，將使模型在特定條件下產生錯誤輸出，進而影響整體決策。在「模型評估（Evaluation）」階段，則可能面臨測試資料污染與評估指標失真的問題。例如，若評估資料混入偽造或生成式影像，將使模型表現被高估，甚至掩蓋潛在錯誤，導致後續應用風險提升。進一步，在「系統部署（Deployment）」階段，常見風險包括權限錯配與 API 濫用。一旦系統接口遭到不當存取或利用，將可能造成資料外洩或系統功能被惡意操控。在「智慧推論執行（Inference）」階段，則需面對資料外流與規避攻擊（evasion）。攻擊者可透過操控輸入資料，誘導模型產生錯誤判斷，甚至繞過既有防禦機制。最後，在「維運與退場（Ops & Retirement）」階段，仍存在模型漂移（model drift）與版本污染的風險。若未持續監控與更新，模型可能因環境變化或資料偏移而逐漸失準，甚至被惡意利用。

在數位資安領域中，傳統攻擊仍然廣泛存在。例如勒索軟體與惡意軟體，會直接癱瘓系統或竊取資料；釣魚訊息則透過社交工程手法，誘使使用者洩漏帳號密碼；SQL 資料隱碼攻擊與阻斷服務攻擊（DDoS），則針對系統漏洞與服務可用性進行破壞。此外，也包括中間人攻擊（Man-in-the-Middle）、零日漏洞攻擊（Zero-day exploit）、DNS 隧道攻擊以及跨網站指令碼攻擊（XSS），這些技術性攻擊手法，主要針對通訊、系統與應用層進行滲透與操控。而社交工程則從人為層面切入，成為許多攻擊成功的關鍵入口。在這些傳統攻擊基礎之上，人工智慧的導入進一步放大了攻擊的規模與精準度。現代攻擊不僅可以自動化生成釣魚內容，還可能延伸至模型層級的攻擊，例如模型竊取、訓練資料汙染、評估過程干擾，甚至在系統部署階段透過 AI 進行自動化滲透。更進一步，生成式 AI 技術的發展，使攻擊者能夠產生高度擬真的內容，例如影像、文字或身分資訊，增加辨識難度，並強化社交工程攻擊的效果。這類「反生成式 AI」的應用，已成為新興資安風險的重要來源。

在資安風險中，醫療體系是當前最具代表性的高風險場域之一。透過前述案例可以理解，醫療體系的發展不僅關係到公共健康，同時也可能成為未來金融詐欺與資料外洩的重要來源。首先，醫療體系具有「高價值資產」的特性。病患醫療紀錄（PHI）資料密度極高，在暗網市場上的單筆售價可達數百至上千美元，遠高於一般信用卡資料，且具有高度變現價值。因此，一旦資料外洩，其影響不僅限於個人隱私，更可能擴散至金融與保險等相關領域。其次，在「營運脆弱性」方面，醫療服務全年無休且無法中斷，當面臨系統加密或勒索攻擊時，組織往往傾向支付贖金以維持運作，進一步提高其成為攻擊目標的機率。第三，在「認知資源耗竭」的情境下，醫護與相關人員長時間工作且承受高度壓力，難以對每一封郵件進行充分判讀，使其更容易成為釣魚攻擊的受害者。例如帶有時效壓力或認證要求的郵件，往往利用緊迫性誘導使用者採取行動。最後，在「情報暴露風險」方面，醫療人員因專業認證更新頻繁且人員流動率高，其職涯資訊常公開於如 LinkedIn 等平台，成為攻擊者進行自動化蒐集與精準鎖定的重要來源。

在醫療體系中所呈現的脆弱性，若與人才數位資料平台（如 LinkedIn）上的社交情

報結合，將形成一個高度風險的交集區域。一方面，醫療產業本身具備高合規壓力（如 HIPAA）、全年無休的營運特性，以及高度敏感的病患醫療資料（PHI），使其成為高價值攻擊目標。另一方面，在 LinkedIn 等平台上，醫療相關人員因職涯發展與專業需求，會頻繁更新其職務資訊、專業經歷與資安認證。這些公開資訊，使攻擊者能夠自動化蒐集並建立高精準度的人才輪廓。當這兩個面向產生交集時，人工智慧即可辨識出「高風險目標族群」，例如近期取得資安認證、任職於醫療機構且具高權限的員工。這些人員原本是組織中的專業人才，卻在數位環境中轉化為高價值的攻擊目標。此外，由於醫療人員流動率高、職務與技能持續更新，其公開資訊呈現高度動態特性，使攻擊者能夠持續調整攻擊策略，提升攻擊成功率。因此，可以理解為何當前資安防護愈發困難。傳統以弱點掃描與防毒軟體為主的「靜態防護」模式，已無法有效應對這類結合 AI 與社交情報的精準攻擊。在智慧時代下，資安已轉變為一個「動態對抗」的過程，不僅需要即時偵測風險，更必須持續評估系統、模型與決策的可信度，才能在不斷演化的威脅環境中維持安全與穩定。

AI 釣魚攻擊生命週期評估

2024 年由美國衛生及公眾服務部發出警訊的 AI 驅動魚叉式釣魚攻擊事件中，駭客透過 LinkedIn 蒐集醫療機構員工資訊，打造高度客製化釣魚郵件，誘使點擊驗證連結竊取帳密。由於醫療產業持有高度敏感病患資料並受 HIPAA 規範，成為高價值攻擊目標。此次點擊率達 17%，恐導致帳密外洩、橫向移動與資料竊取，進一步影響醫療服務運作與病患隱私。專家警示，AI 工具正大幅提升社交工程攻擊成功率，醫療機構須強化資安防護與人員警覺。AI 正加速網路攻擊流程精準化。最新分析指出，駭客透過 AI 分四階段發動魚叉式攻擊：先大規模蒐集 LinkedIn 資料鎖定目標，再冒用專業認證包裝釣魚內容，接著在關鍵時段精準投遞，最後誘導受害者進入偽造登入頁竊取憑證。AI 可自動辨識潛在高價值目標並生成高度客製化郵件，大幅提升攻擊成功率。專家警告企業須強化員工資安意識與驗證機制，以防範日益精密的 AI 社交工程威脅。

AI 正重塑網路釣魚攻擊手法。駭客可利用 AI 工具生成高度個人化郵件，冒充權威

機構並結合真實情境，提升可信度。同時利用時間偏誤與損失規避等心理機制，透過期限壓力誘使受害者快速點擊連結或輸入帳密。此類攻擊不僅內容精緻，且能大量自動化投放，大幅提高成功率。專家提醒，面對看似合理的驗證通知或緊急訊息，應保持警覺並多重確認，以降低受騙風險。本案例說明多層防禦機制在實務中仍可能遭到繞過。攻擊者鎖定中型醫療機構中具高存取權限、且近期取得資安認證的員工，利用 LinkedIn 公開資訊進行精準釣魚。郵件內容專業、語氣自然，成功避開郵件過濾機制；同時透過新註冊且無惡意紀錄的網域與視覺相似字（typosquatting）規避信譽檢查。在人員層面，因缺乏警覺而點擊連結;組織層面,分散投遞策略亦降低群體通報效果。最終導致約 17% 的點擊率，顯示單一防線不足，需強化跨層整合與行為偵測能力。

攻擊者取得高權限帳號後，得以橫向移動至核心系統與病患資料庫，竊取高價值醫療資料並於暗網流通，造成隱私與法規風險。同時可部署勒索軟體，加密關鍵系統，直接影響醫療服務運作，導致檢查延誤、急診壅塞甚至手術中斷。在長期面，攻擊者可能建立後門並維持持續存取，進行間諜活動或竊取研發成果。此類事件凸顯醫療體系在身分驗證與權限控管上的關鍵脆弱性，需強化零信任架構與持續監控機制。醫療機構需先辨識自身系統暴露面，包括 AI 自動化監測工具、企業節點與員工於 LinkedIn 等公開資料，避免敏感資訊被過度揭露。同時，應預測可能成為高風險攻擊目標的單位與人員，並提前進行資料去識別化與權限控管。另一方面，也需辨識 AI 工具本身的風險，如外部 AI 爬蟲蒐集資料、公開資訊被濫用，以及內部 AI 系統可能存在的資安漏洞。透過整合分析攻擊者如何運用 AI 進行滲透，建立全面性的防禦策略，以降低精準攻擊與資料外洩風險。

面對用 AI 攻擊 AI 的新型釣魚事件防護策略中一方面部署進階 AI 釣魚辨識與監測工具（如 ChatSpamDetector），並強化對相似拼字、語意變體與社交工程模式的偵測；另一方面在系統端建立嚴格的輸出安全護欄，避免內部 AI 被惡意誘導生成可用於釣魚的模板，同時限制模型仿真真人機構或個人語氣的能力，以降低被濫用的風險。在偵測階段，重點從「規則比對」走向「情境與語意異常分析」。攻擊者可用分散投遞與微調

變體躲避監控，因此需要監控 AI 模型的異常查詢模式與操控行為，辨識是否有人透過反覆試探繞過防護閾值；同時也要偵測系統處理外部輸入時的預期運算峰值與訊號異常，讓偵測機制能跟上更細微、更即時的威脅型態。

一旦偵測到異常存取，回應策略強調「自動化智慧應變」與「AI 安全事件處理」並行。可實行零信任架構的動態權限撤銷，並在可疑行為出現時即刻啟用更強的多因素驗證（如 MFA/FIDO2），同時阻斷疑似橫向移動至病患資料庫路徑同時避免節點中斷造成醫療健康服務影響。在事件處理上，需隔離遭污染或被劫持的模型運作環境，必要時中斷 AI 系統與核心業務系統的 API 介接，並保留互動日誌與生成紀錄供鑑識追查。復原階段採「系統＋模型」雙軌復原，目標是在最短時間恢復關鍵服務並降低再受攻擊機率。建議先快速重建受損的關鍵醫療系統，確認備份資料未被加密或遭竄改後再執行安全還原，同步修補入侵點並重新確保醫療資料的安全與可信度；對 AI 模型則需清洗遭污染的訓練資料與快取儲存，重新校準與部署正確且完成驗證的模型，並提升防護基準，讓復原後的模型具備更強的抗干擾能力，維持整體運作穩定。

以上內容將在 **2026 年 4 月 22 日（三）10:00 am** 以線上直播方式與媒體朋友、全球民眾及專業人士共享。歡迎各位舊雨新知透過星球永續健康網站專頁觀賞直播！

- **星球永續健康網站網頁連結: https://www.realscience.top/7**
- **Youtube 影片連結: https://reurl.cc/o7br93**
- **漢聲廣播電台連結: https://reurl.cc/nojdev**
- **不只是科技：https://reurl.cc/A6EXxZ**

講者：

陳秀熙教授/英國劍橋大學博士、許辰陽醫師、陳立昇教授、嚴明芳教授、林庭瑀博士

聯絡人：

林庭瑀博士　電話: (02)33668033　　E-mail:happy82526@gmail.com

劉秋燕 電話: (02)33668033 E-mail: r11847030@ntu.edu.tw