星球永續健康線上直播

星球健康週新知 &

專題: 智慧數位資安 (4)

釣魚攻擊智慧生命週期治理

2026-04-22

CHE團隊：
陳秀熙教授、許辰陽醫師、陳立昇教授、嚴明芳教授、林庭瑀博士、
劉秋燕、羅崧瑋、林家妤、陳虹彣、邱士紘、尤翊庭、王斌俞

資訊連結:
https://www.realscience.top/7

星球永續健康線上直播

https://www.realscience.top/4

Youtube影片連結: https://reurl.cc/gWjyOp

漢聲廣播星球永續健康:
https://audio.voh.com.tw/TW/Playback/ugC_Playback.aspx?PID=323&D=20240615

新聞稿連結: https://reurl.cc/no93dn

本週大綱

- 健康科學新知 (2026 / W16)
- AI資安生命週期治理
- AI釣魚攻擊生命週期評估

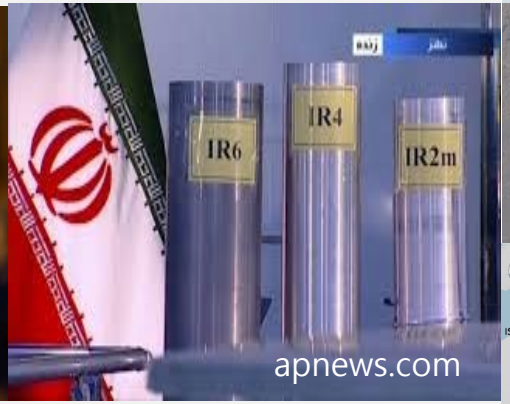
健康科學新知

2026 / W16

美國-伊朗談判僵持:「危局僵持」

美方聚焦核禁令年限，
雙方擬以海峽解封換解凍資產與暫停濃縮鈾

伊朗視濃縮鈾權為尊嚴紅線，
要求撤銷制裁，與美方持續僵持



美軍封鎖伊朗致油價震盪，川普面臨《戰爭權力法》屆期壓力，面臨國會監督與制裁

巴基斯坦在大國支持下發揮地理窗口功能，
成為美伊兩國衝突之協調者

伊朗宣布再度關閉荷姆茲海峽，並警告任何靠近船隻將遭攻擊，已有多艘商船中彈受損

美國調停以色列-黎巴嫩衝突:「停火未安」

真主黨反對這場談判，也不接受被逼繳械，讓談判前景更不明朗

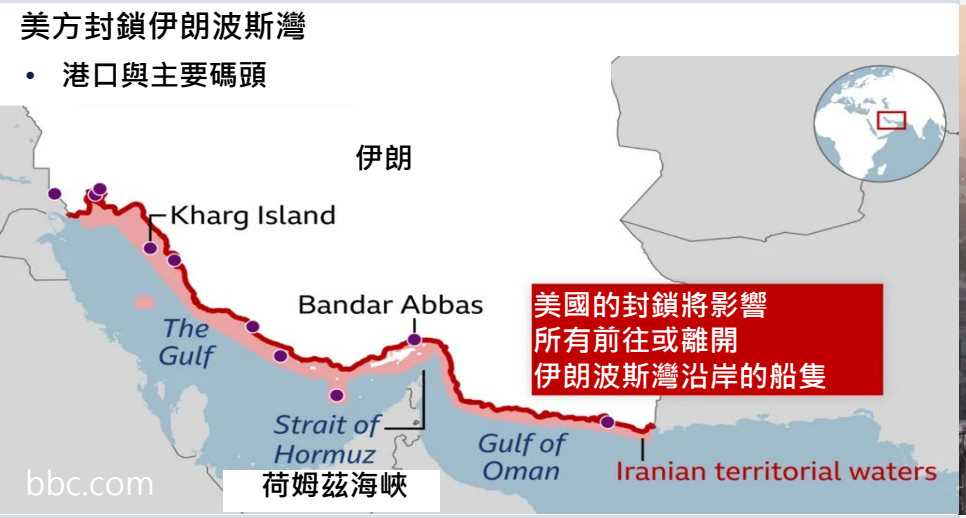
以色列持續爆破村莊，黎巴嫩居民失去家園，返鄉希望也越來越遠

以色列主打解除武裝，黎巴嫩主張優先停火，這讓談判目標一開始就存在明顯落差

在美方調停下以色列-伊朗停火協議於4月17日生效，居民逐步返回戰火後家園

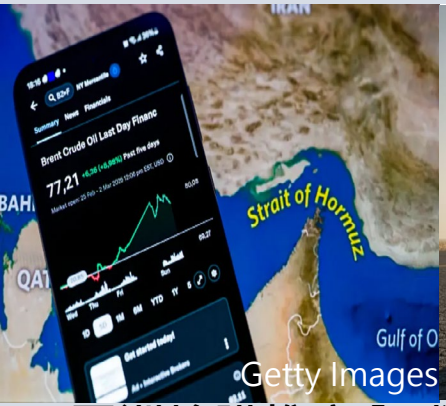
伊美衝突持續影響石油供給:「能源挑戰」

中國批美對伊封鎖是「危險且不負責」指其破壞停火協議，使跨國政經陷入新動盪



美伊重啟談判點燃和平希望，油價下跌有效緩解成本壓力，注入強勁復甦動能

美封鎖海峽且俄油豁免屆期，印度能源進口受限，對經濟與能源穩定構成挑戰



美國核發30天臨時豁免令，使俄石油收入大幅成長，穩定全球能源供應

亞洲搶購推高全球油價，產量損失規模類比疫情，供應短缺衝擊終端需求

美國封鎖重擊古巴:「能源反噬」

Richard Stone, *Science*, 2026

危機來源：石油供應中斷惡化經濟民生

- 美國封鎖古巴航運造成石油運輸中斷，古巴燃料供應快速惡化
- 委內瑞拉曾是古巴主要供油來源，馬杜洛遭美方拘捕後，輸往古巴石油幾乎中斷，能源危機快速惡化
- 能源經濟衝擊造成民生、交通全面失序，醫療與科學研究高度仰賴能源供應受到嚴重影響

直接衝擊：醫療與研究被迫停擺

- 全國多地每日停電長達 20 小時以上
- 醫院手術可能在停電中斷，需依賴發電機維持搶救
- 公共衛生條件惡化，孕婦以及重症照護風險升高
- 臨床試驗被迫暫停，實驗室運作大幅受限

生技韌性：古巴生技在困境中續行

- 古巴生技自製藥物、疫苗與試劑，維持基礎醫療服務
- 合作、供應、資金受阻下，部分研究仍持續推進
- 科學家靠太陽能、備援與遠距工作，努力維持研究現場

環境化學物質誘發腸道菌群失調與糖尿病:「微菌失衡」

研究背景

Sanket Jain, *Science*, 2026

- 非典型病例：
  印度鄉村地區，許多農民都非肥胖且從事耕耘等高體力活動工作卻被診斷出第 2 型糖尿病
- 腸道菌群假說：
  越多研究顯示，暴露於農藥會干擾腸道微生物生態，進而影響代謝與免疫系統

臨床意義

✓ 因發病機制不同，若糖尿病由環境化學物質誘導，可能需要不同臨床護理策略

✓ 農藥不僅改變代謝，還可能干擾「腸-腦軸」→ 引發**憂鬱行為或免疫失調**

研究方法

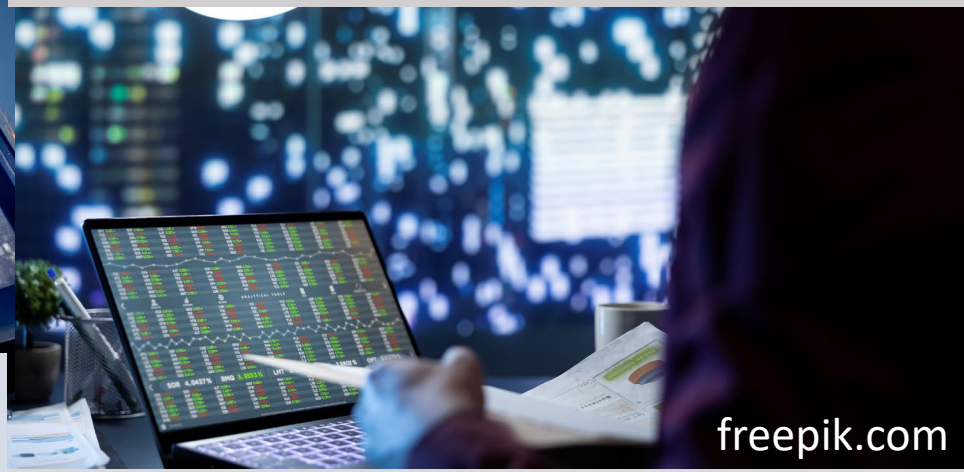
1. 動物模型模擬：
   模擬當地飲食殘留農藥數據，對小鼠進行長期暴露實驗
2. 體外菌株篩選：
   將代表性人類腸道菌暴露於各式農藥，觀察其代謝小分子產量變化
3. 人類樣本監測：
   追蹤受試者尿液中農藥殘留，並與其糞便樣本中微生物代謝進行關聯分析

AI創新強化與資安隱患:「利害並生」

諾和諾德宣布與 OpenAI 合作，將人工智慧導入藥物開發流程

OpenAI將協助分析大量資料、找出潛在藥物，縮短從研發到病人可用治療的時間

Mythos 找出大量高風險金融統漏洞使英國監管單位與金融機構擔憂不當使用資安威脅



Anthropic警告Mythos可在作業系統與瀏覽器中找出嚴重漏洞可能造成資安威脅
模型上線後將只提供特定單位避免不當使用

AI訓練記憶雙面刃:「記憶成患」

Peter Hall, *Science*, 2026

- 大型語言模型訓練提供大量書籍、文章與網頁RWD資料有助於模型進化學習
- 模型互動生成新內容同時也受訓練時看過的RWD影響
- 此特徵可能牽涉版權爭議，也可能外洩地址、信用卡等敏感資訊

模型測試設計

- 研究者開發開源工具 Hubble，專門研究 AI 的記憶化現象
- 團隊取得大量算力，訓練 20 多個測試模型進行比較

資訊保留風險

- 越晚放進訓練流程的資訊，越容易被模型牢記並重現
- 訓練後段雖能強化重要知識保留，同時提高敏感資料被記住的風險

後續挑戰與意義

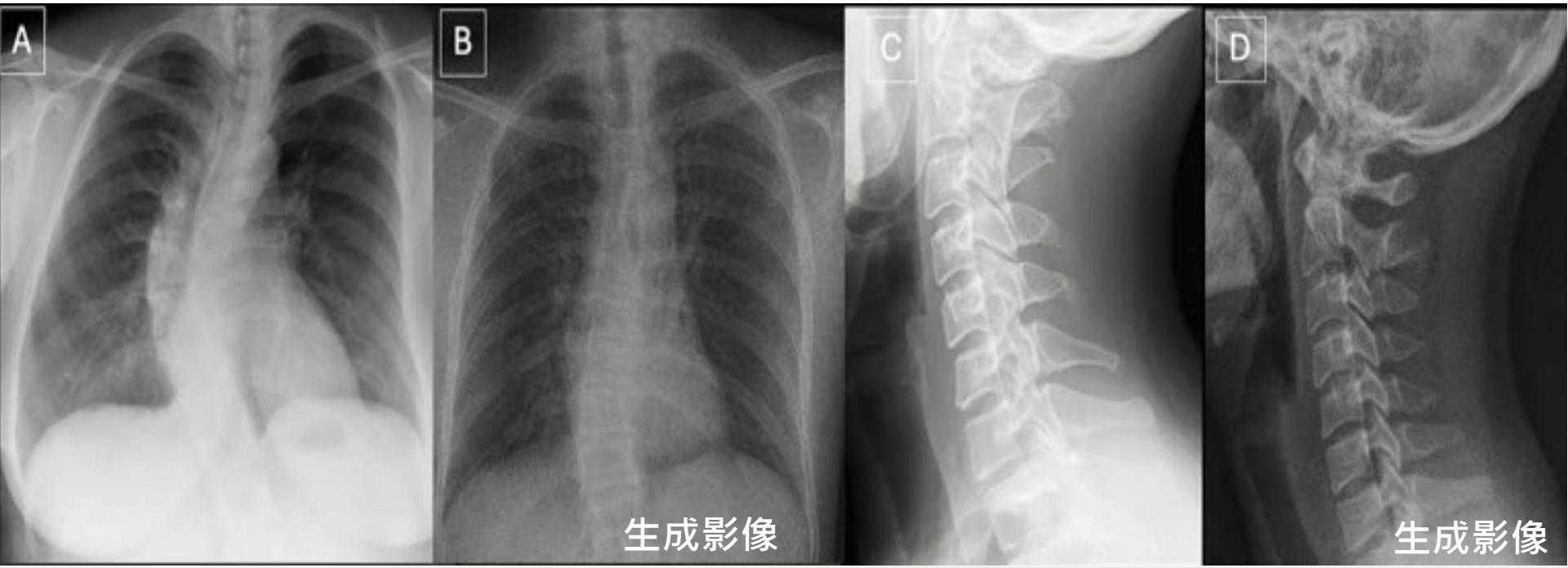
- 敏感資料若提早放入訓練，或許可降低被記住機率
- Hubble 未來可用來測試 AI 是否能真正做到針對特定資料強制遺忘 避免版權與隱私資料疑慮

生成式AI醫學影像挑戰「真偽難辨」

Nature 652 , 20 (2026)

生成式人工智慧可產生高擬真醫療影像
專業醫師也難以分辨真偽 新技術考驗醫療決策與研究

以下兩組X光片，猜猜看哪邊是真實影像、哪邊是AI生成影像呢?



- 放射科醫師對AI生成X光影像辨識準確率僅約七成五
- 虛擬生成醫學影像可能影響臨床決策與醫療研究
- 訓練與限制AI及數位浮水印等監管可提升辨識能力確保應用安全

虛假引文影響科學文獻「虛證成災」

參考文獻引用錯誤

Miryam Naddaf and Elizabeth Quill, *Nature*, 2026

- 越來越多研究人員使用大型語言模型(LLM)輔助文獻檢索、撰寫論文和格式化參考文獻。然而有時這些模型會產生不存在的參考文獻

虛假引文疑慮研究結果

- 《*自然*》雜誌與 Grounded AI 合作，分析2025年超過 4000 篇出版物，涵蓋五大出版商(愛思唯爾 (Elsevier)、Sage、施普林格·自然 (Springer Nature)、泰勒弗朗西斯 (Taylor & Francis) 和威立 (Wiley))，發現有 65 篇發表文獻至少包含一個無效參考文獻
- 分析結果顯示2025年有2.6%的論文至少包含一條潛在的虛假引用，高於2024年的約0.3%

科學雜誌虛假文獻應對措施

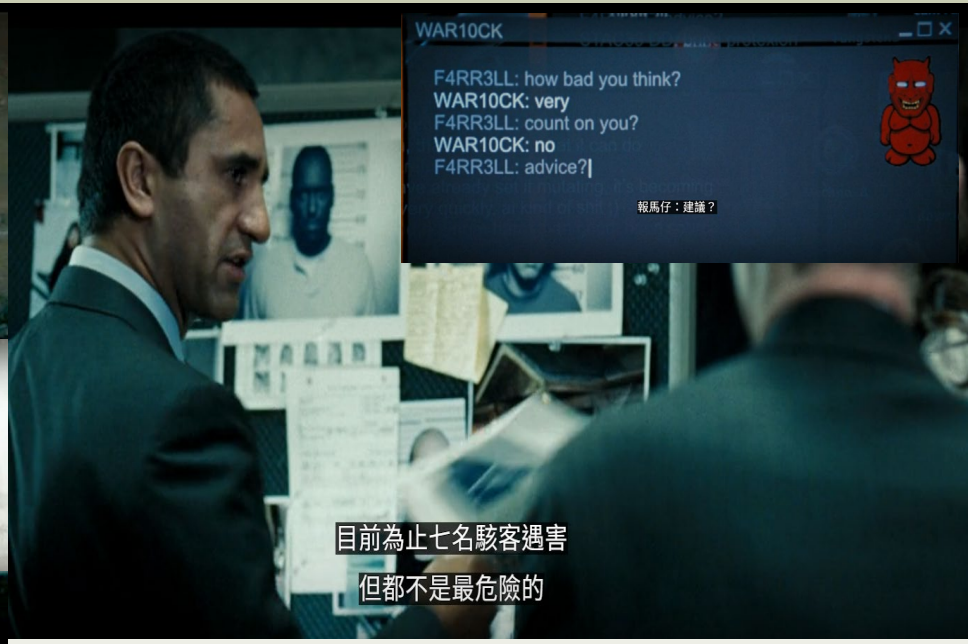
- Wiley 雜誌發言人表示，一些小問題可能會向作者提出，要求其澄清
- Springer Nature雜誌表示一旦發現稿件中包含虛假引用將撤回稿件。

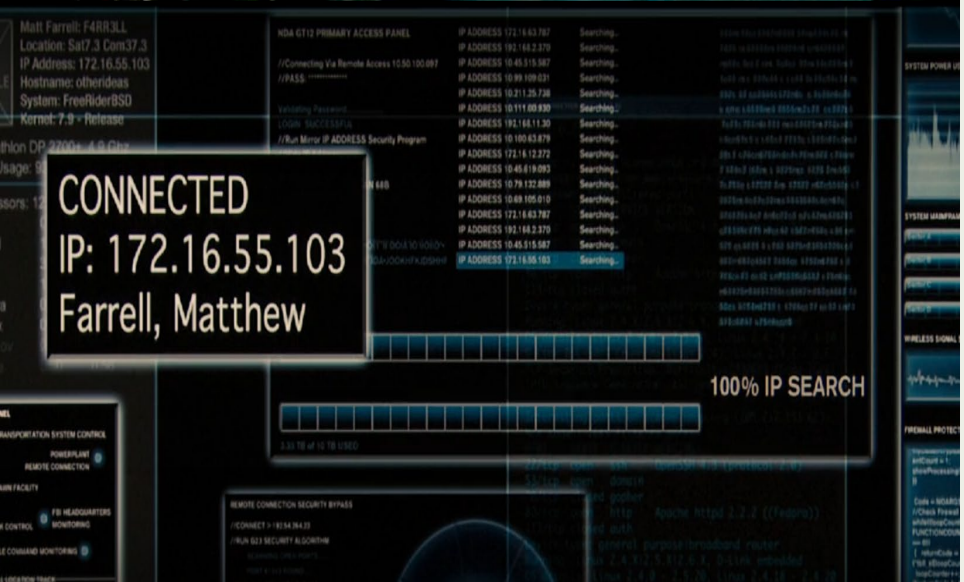
AI資安生命週期治理

終極警探4.0: 網路恐攻

- 美國國慶連假前夕匿名駭客以弱點測試名義受僱寫下可穿透政府核心系統的程式碼
- 程式碼送出後寫碼者便被植入電腦爆裂裝置遠端引爆喪生
- FBI 網安部門偵測到攻擊緊急搜尋駭客欲取的攻擊原始碼

滅口追殺: 公寓突襲與聯邦追查





- 主謀者Gabriel鎖定駭客Matt IP位址，派出殺手小隊上門清除
- 麥克連臨危以身體直覺擊退突擊手，與 Matt 從烈焰公寓狼狽脫逃。
- Matt 透過駭客好友 Warlock 建立外圍通訊，從制度縫隙拼湊陰謀全貌

全面瓦解: Fire Sale 三階段癱瘓

- Gabriel 啟動 Fire Sale，第一階段駭入交通網路陸空鐵運輸系統同步癱瘓
- 第二階段鎖定華爾街與全國財經機構，股市暴跌、ATM 與銀行全面停擺，電力系統為全面癱瘓第三階段

智慧數位資安面向

AI 強化資安
AI for Cybersecurity

AI 輔助資安防護
資安營運(SOC)自動化
異常偵測
威脅情報自動分析

資安保護AI
Cybersecurity for AI

AI 成為攻擊目標
訓練資料汙染
提示詞攻擊
注入式攻擊、模型竊取

AI同時是防禦工具與攻擊標的

融合AI<-> Cybersecurity
防護與攻擊預測兩面

智慧時代數位資產防護目標

資料 (Data)

訓練、驗證、測試與即時輸入的敏感資料

模型 (Model)

基礎模型、微調模型、權重參數與 Embedding 邏輯

系統 (System)

API、Inference server、GPU 與向量資料庫

知識 (Knowledge)

Prompt templates、RAG 文件庫與商業決策邏輯

身分 (Identity)

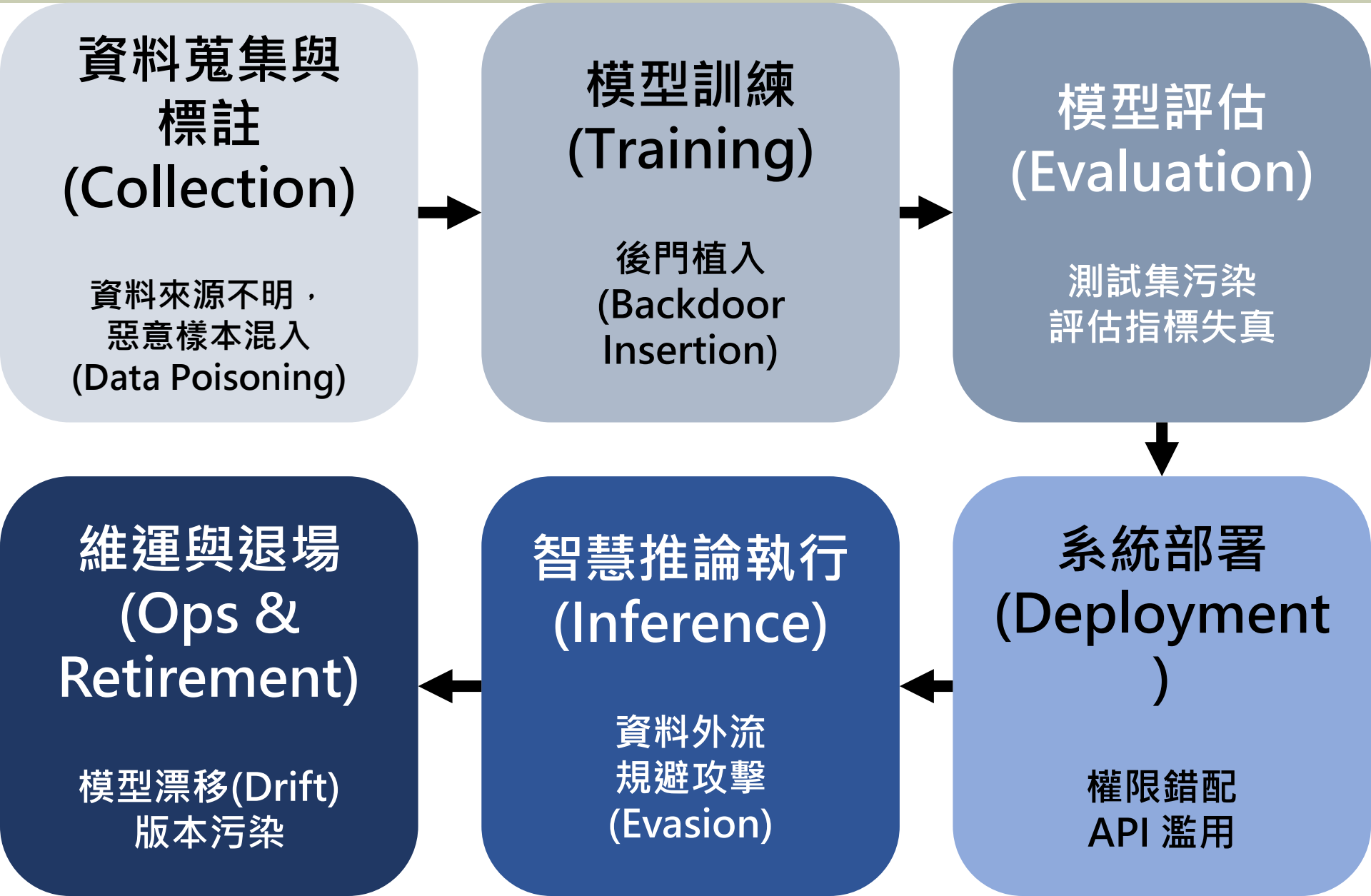
使用者帳號、API Key 與 Tool calling 授權

信任 (Trust)

模型輸出可信度、決策可解釋性與品牌商譽

AI時代數位資安由傳統靜態安全架構擴展至動態模型、知識邏輯與決策信任為核心全方位數位資產保護

智慧產品資安生命週期規劃



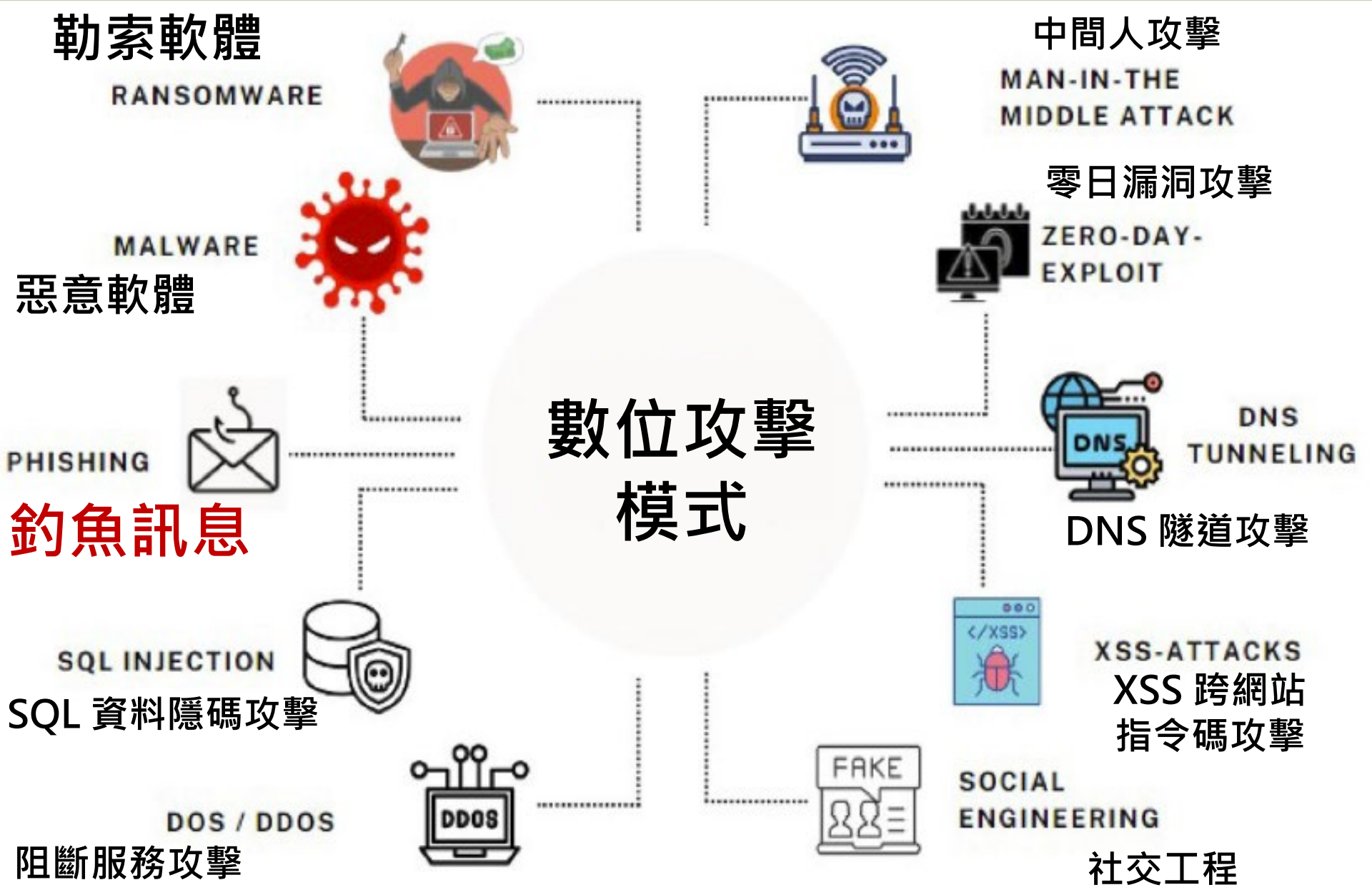
AI資安生命週期治理

Kaur et al., 2023

週期	AI for Cybersecurity	Cybersecurity for AI
Identify	找出系統風險	找出AI本身風險
Protect	用AI防禦攻擊	保護AI模型
Detect	偵測威脅	偵測AI被攻擊
Respond	自動回應	AI安全事件處理
Recover	系統復原	AI模型復原

數位資安常見攻擊型態

Swetha et al., 2025



醫療體系面臨之 AI 精準釣魚威脅

Clabby, 2024

高價值資產

病患醫療紀錄（PHI）密度極高。暗網單筆售價高達 $250–$1,000，遠高於信用卡資料，且具極高變現價值

營運脆弱性

醫療服務全年無休且無法中斷
面對系統加密威脅
組織支付勒索軟體贖金之妥協率極高

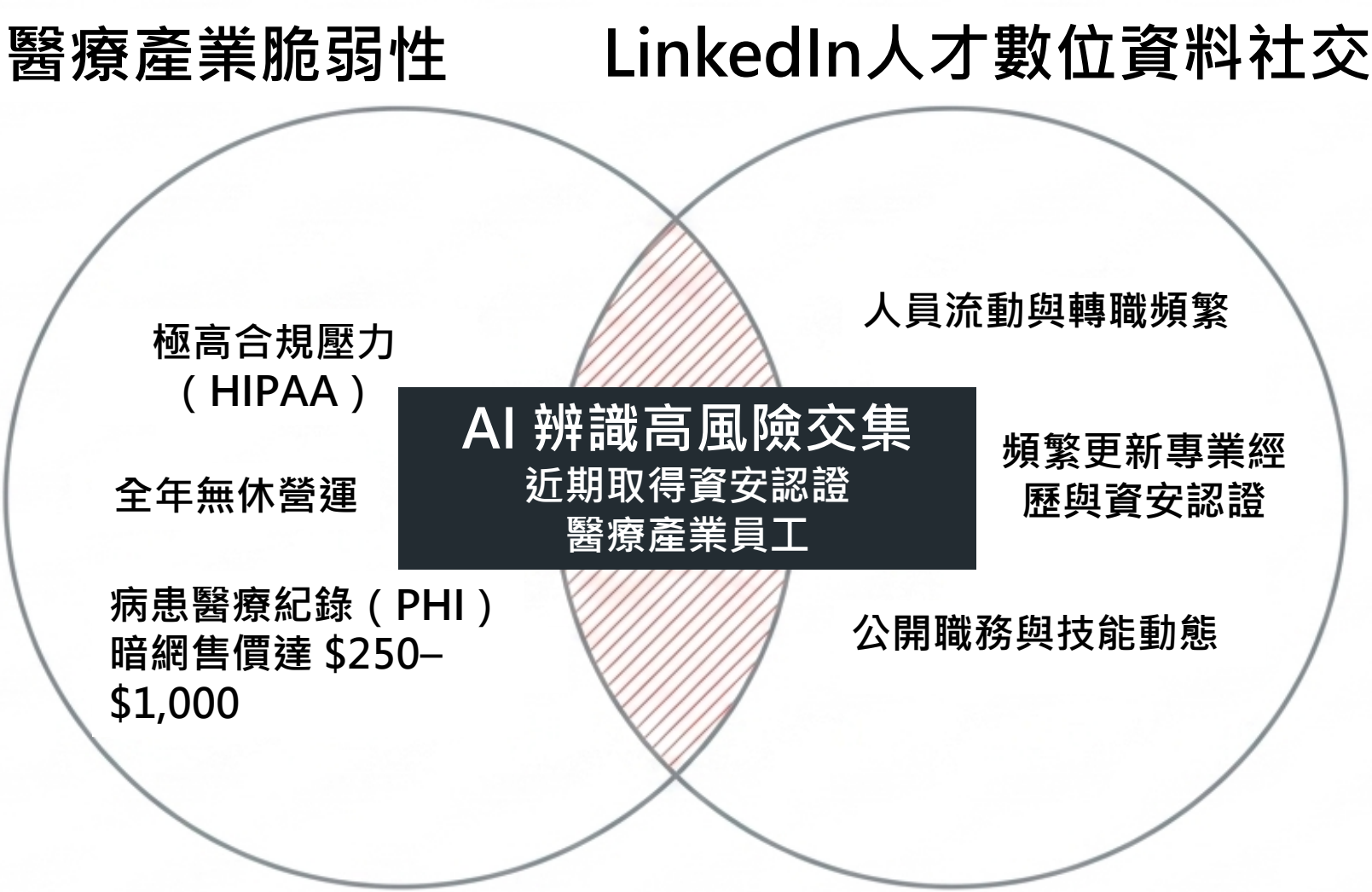
認知資源耗竭

醫護與相關人員工時長、壓力大
難以對單一郵件投入充分投入關注
導致誤入陷阱

情報暴露風險

專業認證更新頻繁，且人員流動率高
LinkedIn 成為攻擊者自動化爬取
職涯動態之高風險情報來源

醫療與 LinkedIn 情報危險交集



公開認證與職務動態，讓近期取得資安認證的員工，從專業人才變成 AI 眼中的高價值目標

AI釣魚攻擊
生命週期評估

AI 驅動魚叉式釣魚攻擊事件

醫療產業持有高敏感病患資料，且受 HIPAA 等法規嚴格規範，因此成為高價值攻擊目標。

美國衛生及公眾服務部(HHS)警示：醫療產業為 AI 驅動釣魚與社交工程高風險目標

無附件、信件內容破綻少、寄件者無黑名單紀錄，呈現低特徵攻擊樣貌

事件核心

事件時間：2024年
攻擊方式：LinkedIn 情報蒐集
目標：取得資安認證醫療機構員工
目的：誘使點擊驗證信連結取得帳密

影響規模

點擊率：17%
風險來源：醫療資料敏感度高、受 HIPAA 規範，病患資料、計費系統、研究資料皆可能受影響

結果損失

- 影響病患隱私、醫療服務中斷與觸發後續勒索攻擊風險
- 點擊後導致帳密外洩、橫向移動與資料竊取

AI 工具自動蒐集個人資訊產生高度客製化

釣魚郵件提高社交工程攻擊成功率 醫療機構為高風險目標

AI工具攻擊目標精準鎖定流程

Clabby, 2024

階段一 情報蒐集

機制：
AI 爬蟲大規模抓取LinkedIn檔案

AI 應用：
自動辨識30–90天內取得資安認醫療機構工作人員為目標

階段二 武器化

機制：
選擇冒充 ISC2、CompTIA 等認證實體

AI 應用：
批次客製釣魚信，整合員工真實姓名、特定認證名稱與專業術語

階段三 精準投遞

機制：
於員工取得認證後 7–21 天內，選擇週二至週四上午投遞

AI 應用：
生成多個語意等價但措辭不同之變體，並透過多通道分散投遞以規避群體通報

階段四 漏洞利用

機制：
導向偽冒登入頁

AI 應用：
成功竊取具備高權限之企業網域憑證，完成滲透

智慧釣魚攻擊生成與心理操控機制

Clabby, 2024

AI 個人化釣魚信件生成

寄件人：ISC² Certification Services <verify@isc2-certverify[.]org>
主旨：[Action Required] 您的 CISSP 認證待驗證 — Certificate ID #74921

親愛的 [員工姓名]：

恭喜您於近期通過 CISSP 認證！根據新實施的數位防偽政策，考量您任職於 [Insert Healthcare Org Name] 關鍵基礎設施，您需於 14 日內完成身分驗證。若未完成，認證將暫時停用。

權威冒充與情境合理性
偽裝官方網域與術語
符合受害者對認證機構通訊之心理預期

近期偏誤：
結合目標剛取得認證之記憶，將釣魚內容合理化為後續標準流程

損失緊迫威脅：
設定 14 日期限，強化失去認證之心理壓力，促使受害者採取行動

多層防禦機制失效分析

Clabby, 2024; Qi et al., 2024

受害組織與規模

中型醫療組織

攻擊載體與情報源

偽冒專業認證機構之釣魚信

（利用 LinkedIn 公開檔案）

目標群體特徵

鎖定高存取權限員工

(共通點：近期剛取得資安認證)

攻擊成效

17% 釣魚攻擊連結點擊率

防禦層一：電子郵件過濾

新註冊網域無黑名單、無惡意附件，無垃圾郵件特徵

防禦層二：網域信譽檢查

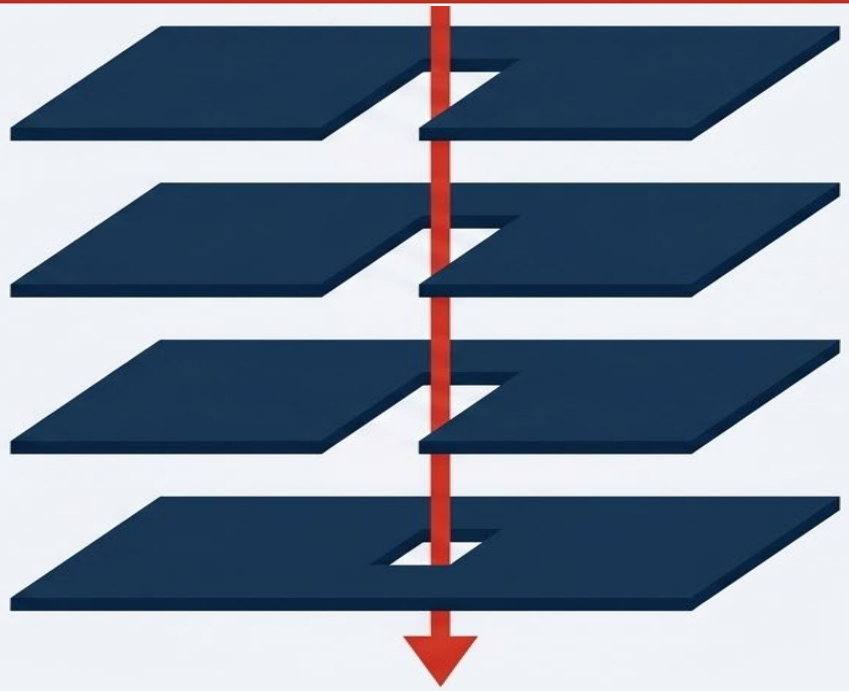
視覺相似性攻擊（typosquatting）規避演算法

防禦層三：員工資安訓練

信件無拼字錯誤、語法自然，並利用情境降低警覺

防禦層四：群體通報機制

批次分散投遞策略，無法觸發群體警示



憑證外洩攻擊擴散與健康照護衝擊

Clabby, 2024; Qi et al., 2024

資安人員
憑證外洩

Compromised
Security
Credential

醫療資料外洩

橫向移動至病患資料庫
竊取高價值
病患健康資訊於暗網流通

勒索軟體部署

利用高權限部署惡意軟體並
加密關鍵醫療系統導致
檢查、急診、手術停擺

持續潛伏

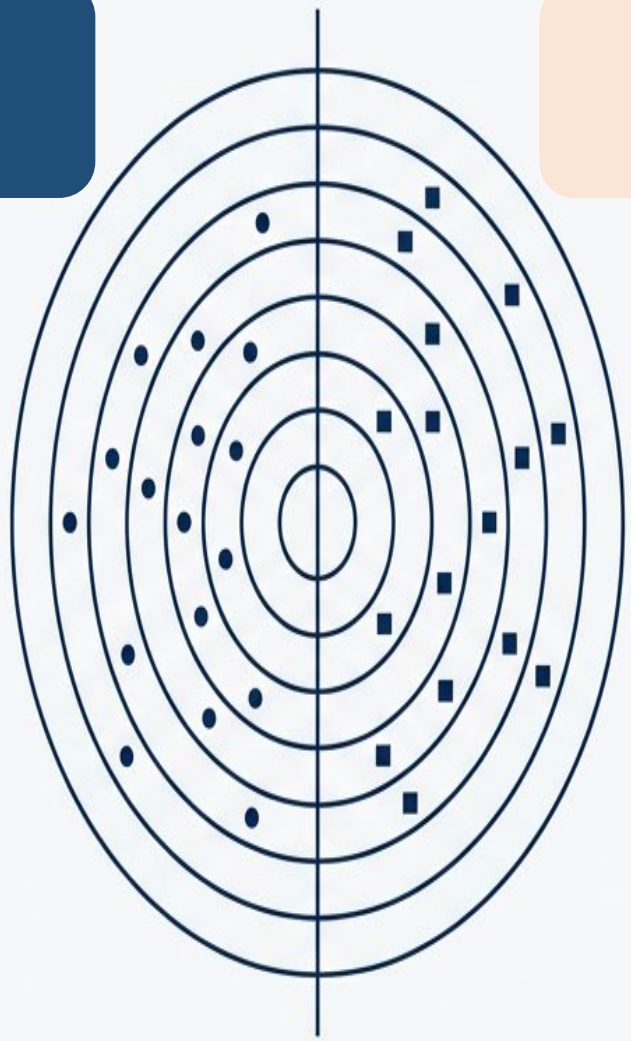
建立後門與持久化存取
若涉及新藥或疫苗研發
可能轉為國家級間諜活動

AI風險與目標識別: Identify

Clabby, 2024; Qi et al., 2024

辨識醫療資訊系統風險

- 部署 AI自動化曝險監測工具
- 持續掃描企業節點與員工 LinkedIn 公開資料
- 預測高風險目標並提前隱藏關鍵資訊



辨識AI工具風險

- 識別外部AI爬蟲
- 公開資料威脅防範
- 盤點內部 AI 工具是否具備資料外洩或遭惡意利用之漏洞
- 分析攻擊者如何運用 AI入侵

公開資料可被 AI工具快速蒐集與分析，

攻擊者得以精準鎖定高風險目標，系統與 AI 風險需同步辨識防範

AI 系統防禦與模型保護策略: Protect

Clabby, 2024; Qi et al., 2024

用AI 禦攻擊

保護AI模型

- 部署進階AI釣魚辨識監測工具 (如 ChatSpamDetector)
- 強化對相似拼寫網域之檢測機制
- 針對認證機構冒充等高社交工程風險類別建立動態黑名單

- 建立嚴格輸出安全護欄
- 防止內部 AI 系統被惡意誘導生成釣魚樣板
- 限制 AI 模型模仿特定真實機構或個人語氣的能力

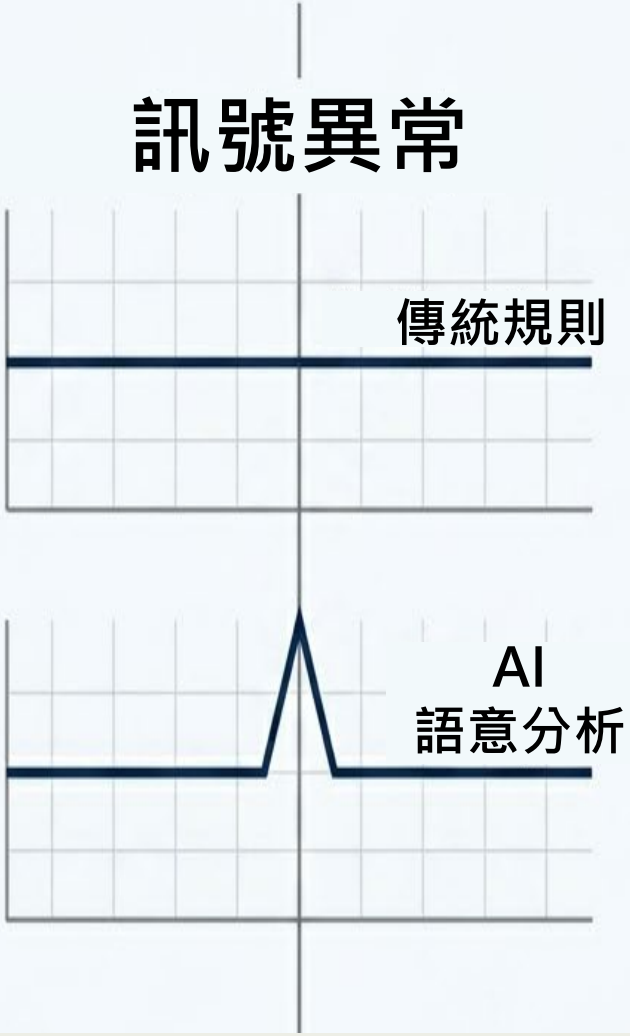
AI資安攻擊可假冒官方語氣並規避傳統郵件過濾，需結合 AI 防禦與模型保護機制以動態郵件過濾與安全護欄，降低釣魚攻擊風險

智慧釣魚異常偵測: Detect

Clabby, 202

偵測威脅

- 從規則比對轉向情境不合理性分析
- 偵測極端精微的語義異常與緊迫性操弄(如驗證期限)
- 識別分散釣魚訊息投遞策略微小變體



偵測AI被攻擊

- 監控針對 AI 模型異常查詢模式與探測行為
- 識別攻擊者是否利用迭代探索防禦閥值
- 偵測 AI 系統處理外部輸入時非預期運算峰值

攻擊者以分散投遞與微調變體規避監控，傳統規則難以偵測，
智慧資安可導入語意分析與行為監控機制防範

自動化回應與安全事件處理: Response

Clabby, 2024; Qi et al., 2024

自動智慧應變

AI安全事件處理

- 實行零信任架構與**動態權限撤銷**
- 偵測到**異常存取**即自動強制觸發多因素驗證（MFA/FIDO2）
- **即時切斷**失竊憑證至病患資料庫（PHI）的橫向移動路徑。

- 隔離遭污染或被劫持的 AI 模型運作環境
- 中斷 AI 系統與其他核心業務系統的 API 介接
- 保留 AI 互動日誌與生成紀錄，供鑑識團隊分析濫用軌跡

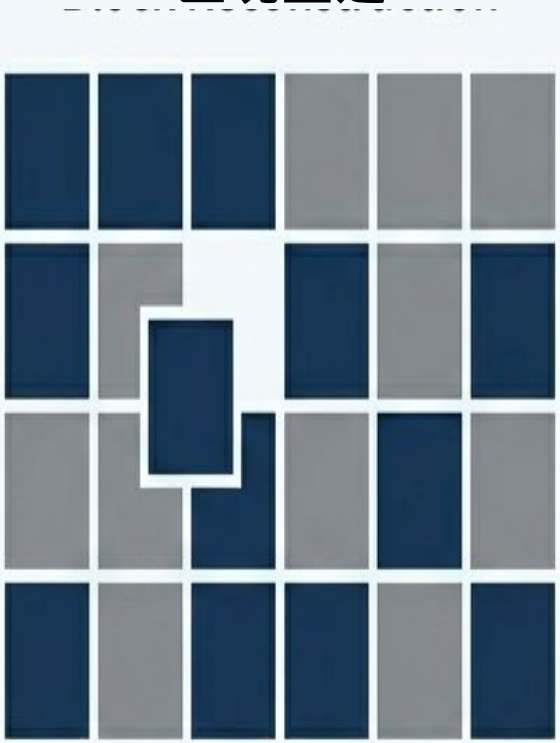
建立自動化智慧應變與AI事件處理機制即時阻斷異常存取
隔離受影響模型保留入侵紀錄，強化整體資安應變能力

AI雙軌復原韌性強化: Recovery

Clabby, 2024; Qi et al., 2024

系統復原

- 快速重建受損的關鍵醫療系統
- 確保備份資料未受勒索軟體加密，執行安全還原
- 修補被入侵系統，重新確保醫療資料的安全與可信度



AI模型復原

- 清洗遭惡意污染的訓練資料與快取儲存
- 重新校準並部署正確且驗證後AI模型
- 提升防護基準，確保復原後 AI 模型具備強化抗干擾能力

透過系統與模型雙軌復原，快速恢復關鍵服務

清理污染資料並強化防護能力，確保整體運作安全穩定

星球永續健康
線上直播


林庭瑀
博士


陳秀熙
教授

國立台灣大學


梅少文 主持人


侯信恩主持人


楊心怡製作人


林家妤


陳虹彣

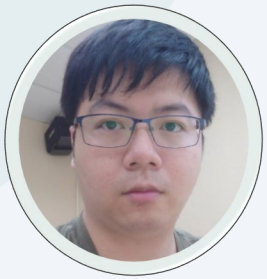

許辰陽
醫師


尤翊庭


王斌俞


邱士紘


劉秋燕


羅崧瑋


嚴明芳
教授


陳立昇
教授

台北醫學大學