

2026 資產管理金融新知暨法遵系列研討會

「金融業迎向 AI 時代的挑戰與因應策略」

活動實錄



隨著生成式 AI 快速發展，人工智慧已從提升作業效率的工具，發展為具備自主規劃與執行能力的新型技術，為金融業帶來創新應用的機會，但同時衍生出新型態資安風險。金融機構除須持續強化既有資安防護外，更需從治理層面建立完整的風險管理机制，以提升整體資安韌性與風險應變能力。

本研討會特就 AI 時代下新型資安威脅、金融業治理挑戰及實務因應策略等面向，邀請主管機關與專家分享最新趨勢與實務經驗，與會者來自政府機關、投信投顧與金融周邊單位之中高階主管代表，共計 85 位先進踴躍參與。

「金融業迎向 AI 時代的挑戰與因應策略」議程

日期：9 月 4 日（五） 14:00-17:00

地點：證基會西門會議廳

時間	主題	講席
14:00-14:10	主辦單位致詞	張麗真 總經理 證券暨期貨市場發展基金會
14:10-15:30	專題演講 1 迎接 AI 時代：打造金融業資安 防線與實務策略	黃思皓 教授 陽明交通大學資訊管理與財務金融學系
15:30-15:40	意見交流	
15:40-16:00	中場休息	
16:00-16:50	專題演講 2 金融業 AI 治理、資安韌性與監 理新趨勢	林裕泰 處長 金管會資訊服務處
16:50-17:00	意見交流	

主辦單位首長致詞

證券暨期貨市場發展基金會 張麗真總經理

張總經理首先感謝所有與會者的參與。近年生成式 AI 的快速演進，提供金融業在效率提升與創新應用上的巨大潛力，同時也帶來前所未有的資安挑戰。今年 4 月，美國 AI 新創公司 Anthropic 發布 AI 模型 Claude Mythos，此 AI 模型具備自動化偵測系統漏洞與自主建構攻擊程式的能力，大幅提升攻擊速度與複雜性，使傳統以天或週為單位的防禦機制逐漸難以因應。對金融業而言，這不只是技術問題，更是攸關市場信任與營運的一大挑戰。



金管會於今年 7 月 13 日發布「金融業應對 AI 攻擊之挑戰與策略建議」，呼籲金融機構須正視 AI 驅動的資安威脅，強化既有防護措施，並從治理層面建構完整風險管理機制，包括：供應鏈管理、持續監控、事件應變與韌性演練等面向，期望金融機構能在快速變化的環境中保持穩健，並在導入 AI 應用的同時，兼顧資安防線的建構與法令遵循的落實。

本次研討會爰安排兩項重要議題，首先由陽明交通大學黃思皓教授，分享「迎接 AI 時代：打造金融業資安防線與實務策略」；接著是金管會林裕泰處長，分享「金融業 AI 治理、資安韌性與監理新趨勢」，俾業者了解 AI 時代下新型資安威脅、金融業治理挑戰及實務因應策略等面向的重要新知。

專題演講 1：

迎接 AI 時代：打造金融業資安防線與實務策略

陽明交通大學資訊管理與財務金融學系 黃思皓教授

隨著人工智慧技術快速滲透金融產業的核心業務，金融業的資訊安全風險樣貌正

經歷根本性的轉變。黃教授指出，傳統資安威脅依然存在，但生成式 AI、代理型 AI 與深度學習等新興技術的普及，為攻擊者提供更具規模化與自動化的攻擊手段。

一、傳統金融常見的五大資安威脅

(一)老舊核心系統 (Legacy Systems)

金融機構的核心帳務系統多半依賴建置已久的架構，長年累積的技術債使漏洞修補難度極高，難以跟上現代資安標準的要求。



(二)資金與結算系統攻擊

包括偽造 SWIFT 跨國電匯指令、ATM 邏輯吐鈔攻擊 (Jackpotting) 及鎖定交易資料庫發動的勒索軟體攻擊，這類攻擊直接衝擊金融機構的資金安全與營運穩定。

(三)社交工程與帳號盜用

包括針對高階主管的商業郵件詐騙 (Business Email Compromise，簡稱 BEC)，以及利用外洩密碼對網路銀行進行撞庫攻擊 (Credential Stuffing/ATO)。

(四)第三方與供應鏈風險

隨著金融機構日益仰賴委外維運廠商、清算平台及 Open API 介接，這些外部連結點也成為攻擊者入侵的跳板。

(五)內部威脅與權限濫用

當特權帳號缺乏細微度的職務分離控管 (Segregation of Duties, 簡稱 SoD) 時，可能導致內部人員洩漏個資或篡改交易紀錄。

黃教授指出，這些金融業長期以來面對的既有資安挑戰，並未隨 AI 技術興起而消失，反而往往與新型攻擊手法相互交織，形成更複雜的風險態樣。

二、AI 時代下新興的資安威脅

(一)大型語言模型 (LLM)

LLM 係基於深度學習並在大量文本數據上進行自監督預訓練的語言模型，具備理解、生成、總結、翻譯及複雜推理等自然語言處理能力。當企業將 LLM 整合至客戶服務或財務報表分析系統時，攻擊者可透過間接提示注入(Indirect Prompt Injection) 操縱模型權限，進而竊取機密財務記錄、客戶個資或內部交易策略。

(二)對抗性攻擊 (Adversarial Attack)

在推論階段，攻擊者向輸入數據加入微小且難以察覺的特定擾動或惡意越獄提示詞，誘使已訓練完成的模型產生分類錯誤或輸出非預期結果。黃教授以詐欺交易為例，洗錢或盜刷集團會在交易特徵（如轉帳金額、頻率、地理位置標籤）中刻意加入特定的微擾動特徵，誘導金融機構的反洗錢（AML）或即時詐欺偵測 AI 系統，將高風險的非法交易誤判為正常合法交易。

(三)多模態生成式 AI (Multimodal Generative AI)

多模態生成式 AI 能同時處理、理解並生成多種數據類型（文字、影像、語音、影片），透過統一的語意空間實現跨模態轉換與生成。在實務上，犯罪集團利用此技術即時生成高階主管的 Deepfake 影音與語音，對財務人員發動進階商務電子郵件詐騙（BEC），促成非法巨額轉帳；此外，多模態生成式 AI 也能生成合成人臉或動態視訊，藉此繞過數位銀行 eKYC 遠端活體辨識與開戶驗證機制。

(四)代理型 AI / 智慧體 (Agentic AI)

代理型 AI 具備自主性、目標導向推理、工具調用與環境互動能力，能結合規劃、記憶與 API 執行器自主完成複雜的多步驟任務。當代理型 AI 被賦予調用內部銀行 API、自動執行外匯或證券交易、審批信貸權限時，若遭遇目標劫持（Goal Hijacking）或非預期的自我迭代漏洞，攻擊者便可能誘導代理執行非授權的資金轉移、異常高槓桿下單或大量惡意授信，造成系統性風險與重大財務損失。

(五)投毒攻擊 (Poisoning Attack)

攻擊者可能向訓練數據集注入惡意、污染或標籤錯誤的樣本，或植入特定觸發特徵的後門 (Backdoor)，例如：內部惡意人員或外部駭客可能污染信用評分模型 (Credit Scoring) 的訓練資料集，使具備特定特徵的高違約風險借款人能自動被模型評定為「極低風險」，從而獲取巨額非正規無擔保信貸。

黃教授總結，AI 技術不僅延伸攻擊範圍，也將攻擊焦點從系統邊界轉移至模型本身的訓練資料、推論邏輯與自主決策機制，對金融機構的資安防禦帶來全新挑戰。

三、金融業的資安攻防：內外兼修的防禦架構

(一)外部攻擊武器化 (AI-Driven Threats)

外部威脅主要來自 Deepfake 身分偽冒、自動探勘與多態惡意程式以及超精準釣魚。黃教授建議應全面落實零信任架構 (Zero Trust Architecture，簡稱 ZTA)，推行 FIDO2 抗釣魚無密碼驗證與多模態活體檢測，並透過微隔離 (Micro-segmentation) 限縮單點突破後的爆炸半徑，避免攻擊者一旦入侵便能橫向擴散至整個系統。

(二)內部 AI 系統脆弱點 (AI-Native Vulnerabilities)

包括提示詞注入與 Agent 越權、資料投毒與模型後門以及敏感資料外洩。相對應防禦策略包含：部署雙向安全護欄 (Guardrails)，即時過濾注入指令並遮蔽敏感個資，同時限制 Agent 呼叫權限，並對高風險交易強制落實「人機協同 (Human-in-the-Loop, HITL)」覆核機制，確保關鍵決策仍有人為把關。

(三)供應鏈與體系風險 (Ecosystem & Governance)

涵蓋 AI 供應鏈漏洞，以及決策黑箱與幻覺究責問題。黃教授指出應建立第三方模型簽章審計與 AI 紅隊攻防演練機制，並遵循國際架構(如 NIST AI RMF) 建立模型可解釋性與幻覺監控體系，使 AI 決策過程具可追溯性與可課責性。

四、發展可信任 AI 技術：金融 AI 的六大基石原則

黃教授指出，金融機構若欲長期穩健地將 AI 技術導入核心業務，須遵循可信任 AI 的六大基石原則：

- (一)可解釋性：要求 AI 輸出下單依據並轉為白箱規則，以利交易員理解與監理審查。
- (二)穩健性：模擬黑天鵝與極端暴跌情境，確保策略在惡劣環境下不致崩潰。
- (三)隱私保護：防範模型洩漏訓練資料與分佈資訊，確保敏感交易紀錄不被還原外洩。
- (四)公平與去偏誤：消除群體不公平，確保不同客群享有平等的交易與定價權利，落實普惠金融。
- (五)治理與可審計：設置前置風控硬性護欄，阻斷超額違規下單，並完整存證決策軌跡，以滿足事後課責與金融檢查的需求。
- (六)人機協作：平時可由 AI 自動撮合以提升效率，但在突發重大事件時應由人工接管，並導入資深交易員的實務經驗，引導模型避開潛在的決策盲點。

五、金融產業導入 AI 的實務建議

黃教授強調未來人機協作的重要性，金融科技創新的本質並非打破法規或取代人類，而是在既有監理架構下賦能專業決策。建議初期先從小應用落地開始，鎖定高頻率、低風險的局部痛點，以小規模概念驗證 (PoC) 快速驗證價值，並累積快速成果以建立組織對全面轉型的信心，並逐步健全風控與治理。此外，黃教授提倡產學合作模式，企業可借助學界在前沿演算法上的研究，避免重複投入研發成本，並透過結合實務場景與學術創新，加速技術發展進程並培育金融科技人才。

意見交流摘要：

Q1：請問如何看待代理型 AI 的發展？金融業是否有落地應用的機會？

A1：代理型 AI 的核心運作機制，係建立在自動調派既有模組化工具的基礎上，若企業內部尚未建立成熟的自動化工具，不建議直接部署代理型 AI。考量金融業具備高度專業性與複雜度，企業可透過產學合作與專業人才的引進，建立 AI 技術正確認知，循序盤點既有工具與實際需求。

Q2：企業傾向於打造內部 AI 使用，請問您覺得未來 AI 會像微軟或 Apple 分為兩大系統嗎？如果是的話，最大的個資隱私問題應如何克服，有何具體解決方案？

A2：企業內部 AI 發展涉及核心競爭力與決策邏輯，故無法外包，僅基礎工具可評估外包或採用第三方服務以降低成本。針對金融業高度關切的個資隱私問題，實務上可透過聯邦學習 (Federated Learning) 等技術克服，使機構能在不洩漏或傳輸原始私密資料的前提下，達成跨機構數據訓練與協同合作。

Q3：請問 AI 產業會泡沫化嗎？

A3：以 2000 年網路泡沫為例，缺乏核心技術與商業模式的企業確實面臨淘汰，但技術本身不會消失，反而為未來技術奠定基礎。當前市場因業者害怕錯過(FOMO) 的心理，過度擴張資料中心等硬體資源並投入高昂運算成本，但模型尚無明確獲利模式。黃教授認為，AI 應與金融或醫療等實體產業深度結合、解決實質痛點並創造價值。

Q4：在金融業高度監理且須保留專業判斷與責任歸屬的前提下，您認為現階段 AI 最適合「提供建議」還是「實際參與決策」？如何與精準行銷或行為風險管理分析連動？

A4：基於金融與醫療領域均具備高度專業性、嚴格法規監理與建立人際信任之必要性，現階段 AI 最適合定位於提供建議與人機協作的輔助角色，而非完全取代人力或直接參與決策。在實務應用上，受限於法規禁止直接透過 APP 推薦商品，AI 可發展「解釋型推薦」賦能第一線理專，作為溝通輔助工具以強化客戶信任。

專題演講 2：

金融業 AI 治理、資安韌性與監理新趨勢

金管會資訊服務處 林裕泰處長

林處長指出，隨著前沿 AI 模型如 Claude Mythos 等代理型 AI 的問世，金融業所面對的資安威脅已從「單次內容生成」演變至「自主多步驟攻擊」：過去的生成式 AI 僅能就使用者輸入進行單次內容生成，然而新一代的代理型 AI 已具備目標規劃與跨系統操作能力，能自主執行檢索、呼叫工具、跨系統寫入與發送郵件等連續性行動。國際間對 AI 風險展現高度關注，例如：美國監理機關召集大型銀行執行長提醒相關風險，英國國家網路安全中心 (NCSC) 已展開對關鍵系統影響的討論，加拿大財政部與央行亦

將其列為重要資安議題，顯示 AI 風險已從單純的技術警示，正式升級為各國金融監理機關共同關注的系統性議題。

一、核心威脅：攻擊時間窗的極限壓縮

在傳統節奏下，漏洞自揭露、專家手工分析、概念驗證 (PoC) 到實際攻擊利用，往往需要數週時間，防守端有相對充裕的時間進行風險分級、測試與上線部署。然而，代理型 AI 能進行漏洞推演，將原流程壓縮至數小時之內即可進行規模化攻擊。此壓縮效應衍生四項風險：



- (一)武器化加速：自動化深度漏洞挖掘，大幅縮短攻擊武器化時間。
- (二)修補壅塞 (Patch Flood)：當關鍵供應商與開源元件在短時間內釋出大量修補程式時，將嚴重排擠金融機構的測試分流與上線資源。
- (三)攻擊鏈擴大：AI 能將零散的缺陷串接為完整的攻擊路徑，大幅降低攻擊門檻。
- (四)供應鏈曝險：開源套件與第三方元件的風險加速外溢，使攻防雙方的資源投入極度不對稱。

二、金融業的特有挑戰：高頻攻擊與穩定營運的正面衝突

林處長點出金融業因業務性質，在應對高頻攻擊時面臨獨有的結構性困境：

- (一)變更受限：核心交易系統難以頻繁停機，快速修補與回歸測試和既有排程衝突。
- (二)授權壓力：若缺乏預先授權機制，冗長的跨部門決策恐錯失黃金止血時機。
- (三)指標失真：傳統的弱點分級機制在高頻併發的情境下，難以反應真實的營運韌性衝擊。
- (四)資產盲區：軟體物料清單 (SBOM) 與相依關係透明度的不足，導致機構無法在短時間內精準界定受駭範圍。

林處長表示，大量自動化告警、跨單位協調溝通、監理通報義務與供應商追蹤等需求，往往匯集於少數掌握資安與維運知識的關鍵人員身上，若輪值與備援機制不足，極

易導致人員倦怠、決策延誤與品質下降等營運危機。若在事件發生後一小時內，技術應對、業務營運與法遵公關三條決策線若未能同步推進，將形成「決策癱瘓點」(Paralysis Point)，導致防禦黃金時間的流失。林處長強調，技術與資安管理並非由單一資安團隊獨立完成，而須高階管理階層共同承擔。

三、國際監理共識：從技術警示升級為董事會治理要求

面對 AI 威脅，國際間已逐漸形成共識，將資安韌性議題由單純的技術警示提升至董事會治理層級的要求。美國鼓勵銀行在受控條件下測試前沿模型的潛在影響，並將其視為系統性資安議題；英國由央行、財政部與國安部門聯合發布聲明，要求強化弱點管理與供應鏈控制整備；日本金融廳與日本銀行則聯名要求金融機構優先盤點關鍵系統曝險與營運韌性；而五眼聯盟 (Five Eyes) 資安機構更發布聯合聲明，明確定調「AI 已非未來議題，而是正在加速攻擊的現實風險」，並指出風險變革週期已由「年」縮短至「月」，領導者必須提升反應速度。五眼聯盟同時提出三項具體行動方向：回歸基礎並強化深度防禦、以 AI 強化防禦對抗威脅，以及落實「設計安全」與強效身分驗證。林處長強調，網路韌性已成為企業核心風險，直接影響業務持續性、市場信心與長期價值，理應由董事會層級負責並賦予資安主管相應的決策權。

四、金融業應對 AI 前沿模型之策略

(一)鞏固防禦基石，加速落實資安韌性藍圖

林處長指出，金融業可透過以下方式系統性強化金融營運韌性：

1. 零信任架構：透過抗釣魚多因子驗證與網路分段，限制 AI 攻擊的橫向移動。
2. 持續監控：藉由早期異常偵測與告警關聯分析，提升判斷與即時應對能力。
3. 資安左移：在軟體開發的早期階段便導入資安檢查，透過自動化檢測與回歸測試提升弱點治理效率。
4. 第三方治理：透過建立軟體物料清單快速定位受影響範圍。
5. 聯防協作：透過金融資安資訊分享與分析中心 (F-ISAC) 深化情資共享基礎，明確分級報送門檻與通報時限。

(二)短期整備，爭取黃金應變時窗

當修補時效無法追上攻擊速度時，金融機構須透過替代控管與韌性演練以降低實際曝險。短期整備可依循四步驟推進：

1. 盤點與可視化：識別優先對外服務並清理閒置帳號與已知未修補項目。
2. 資源排序與分流：評估修補量能並建立短期風險優先清單，避免僅依單一指標盲目修補。
3. 替代控管：當無法立即修補時，即時啟動網路分段、高權限帳號收斂與非必要服務停用等隔離措施。
4. 降級劇本與演練：事先取得授權並建立降級運作劇本，同時驗證離線與不可變備份機制的有效性。

(三) 前瞻規劃防禦性 AI，以 AI 抵禦 AI

林處長指出，運用防禦性 AI 建立可持續擴充的防禦能力，三大應用場景包含：一、增進軟體開發安全效率，運用 AI 輔助源碼檢視與弱點辨識；二、建立自動化弱點治理產能，整合漏洞發現、驗證至部署的完整流程；三、規模化監控與應對能力，透過 AI 進行告警關聯與脈絡整理，提升高頻攻擊情境下的分流與取舍效率。

五、代理 AI 安全治理的核心思維

隨著代理型 AI 廣泛應用於金融業，法遵、客服、洗錢防制及跨系統作業效率有所提升，卻也伴隨目標偏離、越權操作、資料外洩及錯誤擴散等風險。林處長建議，透過零信任架構建立權限邊界，明確規範代理型 AI 的使用目的、責任歸屬、資料來源、工具權限、輸出範圍及監控與停用機制，並透過工具白名單、參數檢查、資料外送限制、速率控制及緊急中斷機制等多層政策閘門，確保其自主行動均在可稽核、可監督及可隨時中止的安全範圍內。

六、風險分級與動態治理機制

為使代理型 AI 的風險控管與其自主程度及業務影響相稱，應從資料敏感度、動作影響、能力邊界與可擴散性三個面向進行評估，並依「就高不就低」原則將應用區分為 L1 至 L4 四個風險等級，分別配置相應的監控、政策閘門及人工覆核機制。風險分級亦不應視為一次性核可，而應隨著用例變更、模型更新及代理行為漂移持續重新評估，並

於正式上線前完成安全測試、韌性驗證及緊急人工接手機制建置，以確保高風險行動均具備適當授權、覆核及阻斷措施，避免代理型 AI 失控造成營運或客戶權益損害。

林處長總結，金融業應同步強化基礎資安韌性、短期應急能力及防禦性 AI 應用，並透過代理型 AI 分級治理及動態稽核機制，建立可理解、可監督、可課責的零信任安全邊界，在兼顧創新與風險控管的前提下，充分發揮代理型 AI 跨系統整合與提升營運韌性的潛力，並符合國際監理及董事會治理要求。

意見交流摘要：

Q1：針對小型金融機構面對日益高昂的資安防護成本與人才培訓困境，有何具體建議？

A1：建議小型金融機構可善用當前盈餘充實資本，專注於完善資安基礎設施與鞏固基本功，鞏固營運的根本之道。

Q2：從金管會的監理面向，對投信投顧業者使用 AI 的態度與監理原則為何？有哪些可做與不宜做的事項？

A2：投信投顧業務因直接涉及客戶權益與核心信託義務，在導入 AI（如機器人理財）的初期，人工審查與把關機制實屬必要且不可省略。隨著 AI 模型逐步穩定，監理上可允許將即時把關調整為批次審查，或是建立正常行為基準（Baseline），改為僅針對異常樣態進行抽查，藉此在控管風險的同時減輕基層人員的作業負擔。

Q3：隨著金融機構導入生成式 AI，已由知識查詢走向可自主呼叫 API 甚至執行交易的 AI Agent，未來監理上是否需將 AI Agent 視為一種新型「特權帳號或數位身分」進行管理？

A3：未來監理上確實有必要將 AI Agent 視為「需要受控管的數位員工」。若其涉入較複雜的業務流程，應比照特權帳號之管理邏輯，建立雙重驗證機制，例如由 AI 與人工各負責一道關卡。目前主管機關與金融業者雖均處於摸索階段，但絕不能對 AI 採放任態度，人為監督與控管是防範營運風險的關鍵措施。