115 年 3 月份機關安全維護宣導

應用 AI 機制為核心 進行人頭帳戶識別的數位治理

（日期 115/3/2）

位帳戶，遭列入警示名單之比例明顯高於採用其他驗證機制的帳戶。部分金融機構的風險管理報告也指出，在列為高風險之帳戶樣本中，以自然人憑證開戶者之比例高達20%。此趨勢已促使多家銀行基於風險考量，陸續暫停接受以自然人憑證進行數位開戶之業務，主管機關亦隨之要求各金融機構全面檢視現行開戶流程，並強化風險控管機制之設計與執行。

此外，聯合國毒品與犯罪問題辦公室（United Nations Office on Drugs and Crime, UNODC）所發布報告中，雖未明確將臺灣列為詐騙活動主要發源地，但經過渲染或刻意扭曲解讀後，「臺灣為詐騙中心」之錯誤敘事迅速於社會輿論中擴散，不僅削弱社會大眾對自然人憑證制度及其信賴基礎之認同，亦反映出我國在數位金融治理面向仍有如缺乏制度性澄清與風險溝通機制、政策說明工具不完備等結構性的挑戰。

AI 防詐解決方案的提出與實踐：2025 內政部黑客松

隨著詐騙手法不斷推陳出新且快速演化，我國防詐策略亦逐步轉型，由傳統以「事後查緝」為主之被動式應對機制，朝向強調「源頭預警」與「風險預防」之前瞻性治理模式發展，而內政部警政署刑事警察局與高雄市政府警察局刑警大隊於「2025內政部黑客松AI應用競賽」共同提出之「AI驅動的詐騙集團人頭帳戶識別及警示解決方案」，即屬政府導入科技治理思維、強化人工智慧於金融詐騙防制中制度性應用之具體實踐例證。

該方案之核心理念是將人工智慧

（Artificial Intelligence, AI）技術嵌入金融機構「了解你的客戶」（Know Your Customer, KYC）流程中，期望在數位帳戶開立初期，即辨識潛在高風險帳戶，藉由預警與阻斷詐騙資金流向，實現事前防範之政策目標。其建構之AI模型是以實際破獲之人頭帳戶案例為訓練資料，結合戶政系統資料與警政情資，採用邏輯回歸演算法建構模組，以兼顧運算效率與模型可解釋性，並強調對帳戶開設環境與使用行為之監測。根據初步實證評估結果顯示，該AI模型召回率達85%，精確率亦達82%，能涵蓋大部分潛在之高風險帳戶，表現出良好的偵測效能，並在有效控管誤判風險的前提下，仍具高度辨識準確性。此外，預警系統可嵌入金融機構之數位開戶流程中，提供即時風險評分與預警功能，不僅有助於提升反洗錢作業效率，亦可支援跨銀行機構間對於可疑分散式開戶行為進行協同偵測與聯防治理。

AI 模型的挑戰與侷限：高擬態詐騙策略的興起

儘管人工智慧（AI）技術已廣泛應用於金融防詐領域，並在實務上展現初步成效，其辨識能力亦逐漸獲得肯定，然詐騙集團亦同步精進其手法，發展出更具隱蔽性之「高擬態式開戶策略」，對現行AI模

型形成實質性挑戰。此類策略主要特徵包括：一、由人頭帳戶本人親自於正規時間與地點完成開戶操作，以排除裝置特徵與IP位址等技術異常訊號；二、提供完整且真實之個人身分資料，有效規避身分偽冒偵測機制；三，寄件地址選擇使用如公司行號等具合理性的地址，難以被偵測為高風險指標；四、刻意避免使用VPN或境外IP等異常網路環境，使整體開戶行為模式近似一般用戶。此類手法已大幅降低AI模型於開戶即時辨識階段的判別敏感度。

造成AI模型辨識侷限之原因，主要可歸納為三點：首先，現行AI系統多基於「開戶當下」的靜態資料進行即時評估，缺乏對後續金流異常或交易行為的即時追蹤能力，導致風險偵測之時效性與完整性不足。其次，若該帳戶為人頭首次使用，其資料未曾列入任何警示清單或歷史黑名單，AI模型將無從辨識其潛在風險特徵。最後，單一帳戶的開戶行為往往難以獨立構成高風險事件，須透過跨帳戶、跨時間軸之圖譜關聯分析（graph-based behavioral analysis）方能識別其組織性犯罪特徵。上述因素顯示，若AI模型僅依賴孤立之單點

特徵進行辨識，將難以有效因應當前詐騙行為高度擬態化與分散化之趨勢。

此外，該AI防詐模型本身亦面臨實證基礎不足之問題，因其訓練資料涵蓋之真實詐騙案例僅約200筆，樣本數量明顯偏少，可能導致模型產生過擬合（overfitting）與樣本偏差（sample bias），進而限制其於多樣化詐騙情境下之泛化能力（generalizability），若未輔以持續性資料更新機制與獨立第三方之效能驗證，其實務應用成效將難以確保。更需關注者，在人工智慧日益受到政策部門與社會大眾青睞之背景下，若對其技術潛力寄予過度期待，而忽略其根本性限制與潛在誤判風險，將可能產生「治理幻覺」（governance illusion），使合法用戶因誤判而遭受不當處置，不僅侵害個人權益，更可能削弱社會大眾對政府數位治理能力之信任基礎，並危及相關政策措施之正當性與長期可持續性。

建構多層次防詐體系：技術深化與制度接軌的雙軌策略

面對日益複雜的高擬態詐騙手法及具延遲特性的異常交易行為，我國防詐體系未來應由傳統「單點式辨識」機制，轉型為結合「動態監控」與「跨域圖譜分析」的多層次風險治理架構。此一轉型不僅需仰賴人工智慧與時序資料分析等技術的進步，也必須同步調整監理制度，推動技術與政策的雙軌整合。唯有提升模型的彈性與風險感知能力，才能有效因應詐騙手法的快速演變與高隱蔽性。

在技術面，應優先強化現有模型架構，導入延時監控與圖譜分析能力。首先，可使用長短期記憶網路（Long Short-Term Memory, LSTM）等時序模型，針對開戶後24至72小時內的交易行為進行監測，辨識如大額跳轉、夜間高頻操作、短期棄用等風險行為。其次，透過圖神經網

路（Graph Neural Network, GNN）技術整合帳戶、IP、裝置、地址與電話等資料，分析跨帳戶關聯，辨識潛在詐騙網絡。此外，前端亦可引入更精細的使用者輪廓建模，包括裝置指紋、瀏覽環境、地理風險，以及滑鼠軌跡與鍵盤輸入等微行為特徵，以強化對可疑開戶行為的即時識別。

在制度層面，亦需同步調整以支撐技術應用的落地與合法性。首先宜由金管會主導建立跨金融機構的高風險情資共享平台，實現異常資料的即時交換，打破資料孤島以強化聯防。其次推動跨部門資料整合，將戶政、警政、電信與金融資料建構為多維風險圖譜，並運用聯邦學習[1]（Federated Learning, FL）與安全多方計算[2]（Secure Multi-Party Computation, MPC）等方法，確保數據可用但不可見，兼顧隱私與效能。最後，應修法賦予AI模型在KYC與AML流程中輔助決策之正式地位，並強調模型可解釋性，以利人工審查與風險管理，建立透明可信的「人機協作」治理架構。

結論

數位金融風險治理之核心關鍵，不僅在於先進技術之導入與應用，更需要奠基在社會大眾對制度體系之信任基礎。在詐騙手法日益進化與輿論環境趨於複雜的背景下，我國亟需以「科技治理化」與「法規數位化」作為雙核心驅動力，推動整體防詐體系之制度性轉型，不僅涉及金融監理機制的技術升級，更關乎法治架構與公民參與間的協同運作。

在提升社會信任方面，可從三大策略面向著手：第一，應推動防詐教育之制度化與常態化，透過多元管道全面提升國民之數位素養與識詐能力。第二，應健全法律救濟機制，強化受害者在遭遇詐騙事件後之法律協助、資金追討與身分修復等實質保障，以鞏固司法正義與社會公平。第三，應強化數位治理透明度，建立制度化的申訴處理與信任修復機制，使民眾在制度運作過程中感受到可預期性與回應性，進而提升對數位治理體系的整體信賴程度，唯有在法制明確、技術精進與社會信任三者共構之條件下，始能建構具前瞻性、韌性與自我演化能力之智慧型數位金融防詐治理架構。

1 是一種分散式機器學習方法，允許多個設備或機構在不共享原始資料的前提下，共同訓練一個模型的方法。

2 是一種密碼學技術，可讓多個參與方在不公開各自輸入數據的基礎上，偕同計算一個約定的函數，並只獲取各自的計算結果。

【資料來源：清流雙月刊 新北市政府警察局板橋分局偵查隊】