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 (2)

AI時代數位資安策略

2026-04-08

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

資訊連結:
https://www.realscience.top/7

星球永續健康線上直播

https://www.realscience.top/7

Youtube影片連結:
https://youtube.com/channel/UCCHTox4rUysI30QW4e_xIiA?si=1DIj9qIn3bZWMtNG

漢聲廣播星球永續健康: https://reurl.cc/WbGALy

新聞稿連結: https://www.realscience.top/7

本週大綱

- 健康科學新知 (2026 / W14)
- 智慧數位資安生命週期
- 金融智慧數位資安實例

健康科學新知

2026 / W14

中東衝突僵持 國際勢力關注:「戰略模糊」

伊朗總統致信美國否認伊朗威脅論
強調從未主動挑戰僅維護國家防衛安全

隨胡塞參戰與以軍擴大入侵 戰火僵持
美國研擬地面部隊派遣中東行動

川普擬奪取伊朗石油引發能源危機
導致油價飆漲並加劇中東各國軍事衝突

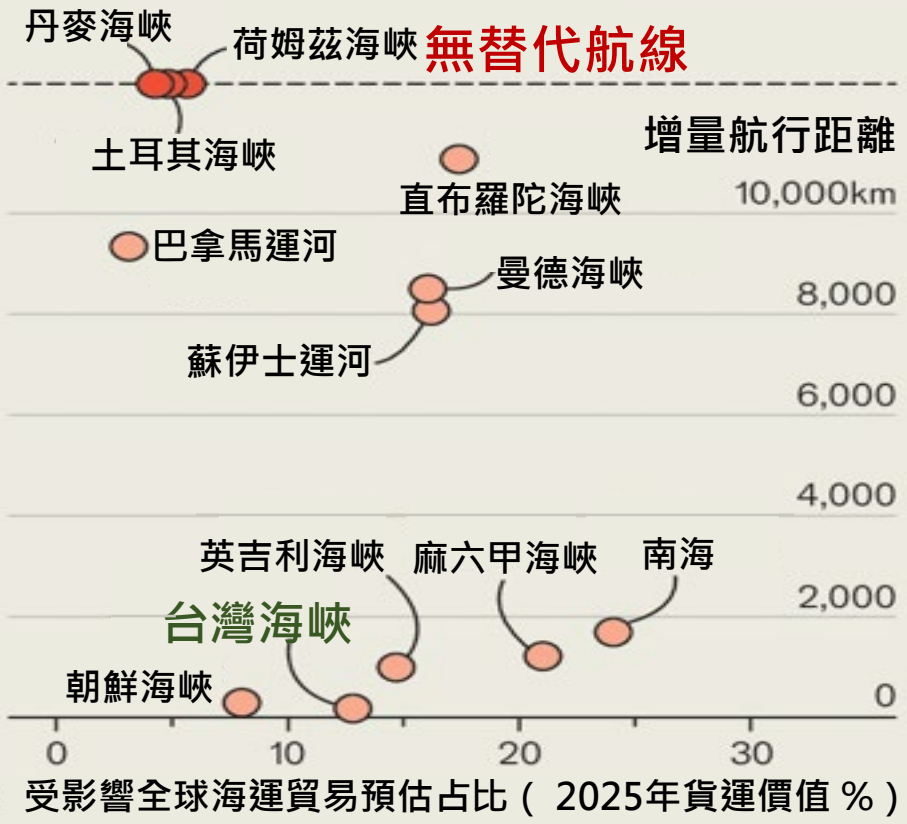
烏克蘭與海灣國家強化國防合作
輸出防空經驗應對伊朗威脅防範中東戰火

全球航運關鍵節點:「咽喉命脈」

全球航運咽喉要道

全球航運雖四通八達
實際流動卻高度集中於少數關鍵節點

全球海運要道影響評估
若遭封鎖平均需增加航行距離(公里)



關鍵航道受阻，全球航運將被迫繞行
距離增加與影響範圍隨之擴大
全球有三處海峽無替代航線

伊朗戰爭重塑能源版圖:「能源震盪」

衝突影響

Aisha Al-Sarihi, *Nature*, 2026

- 荷姆茲海峽關閉:全球五分之一油氣貿易通道中斷
- 全球能源衝擊: 每日 2,000 萬桶原油流動受阻
- 市場連鎖反應:全球油價在數週內飆升，保險公司暫停航運承保或調升風險溢價

重要警訊

1. 結構脆弱性
   - 過度依賴單一航道
   - 能源深陷地緣政治威脅
2. 能源依賴性
   - 該航道 80% 油氣出口目的地為亞洲
   - 日本、越南等國面臨嚴峻燃料短缺
3. 轉型必要性

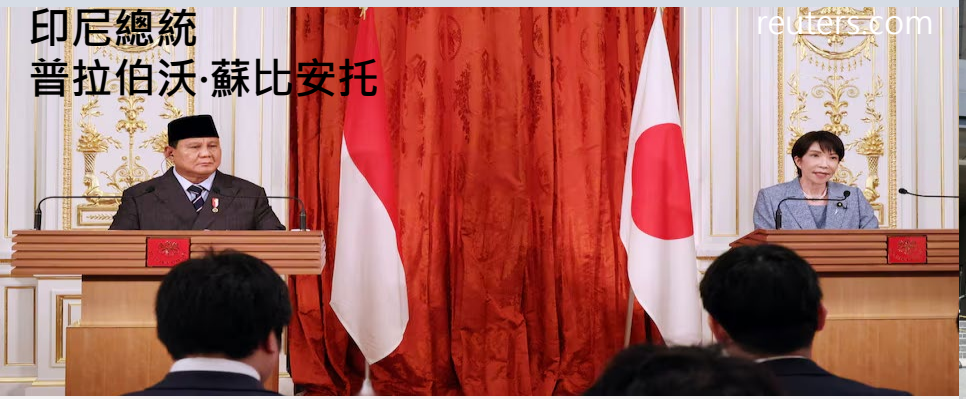
   清潔能源轉型是經濟與地緣政治韌性之關鍵

解決對策

- 貿易路線多元化:投資繞過瓶頸點的設施。
- 強化供應鏈韌性:優先採取外交斡旋與去升級手段，防止供應鏈被「武器化」。
- 擴展再生能源:利用海灣地區豐富的風能與太陽能。

日本外交軍武雙重措施:「樞紐崛起」

因應伊朗戰火衝擊供應 日本與印尼協議
加強能源協調確保資源安全。



馬克宏訪日強化印太影響力
與德角逐戰略地位並應對中日緊張局勢

日本完成部署首批長程飛彈
攻擊距離達1000公里 打破過往防衛布局

馬克宏將與日本推動科技與能源安全合作
並共商烏克蘭局勢和應對中國經濟挑戰

日本為加強嚇阻力部署飛彈
引發民眾抗議與憲法規範爭議

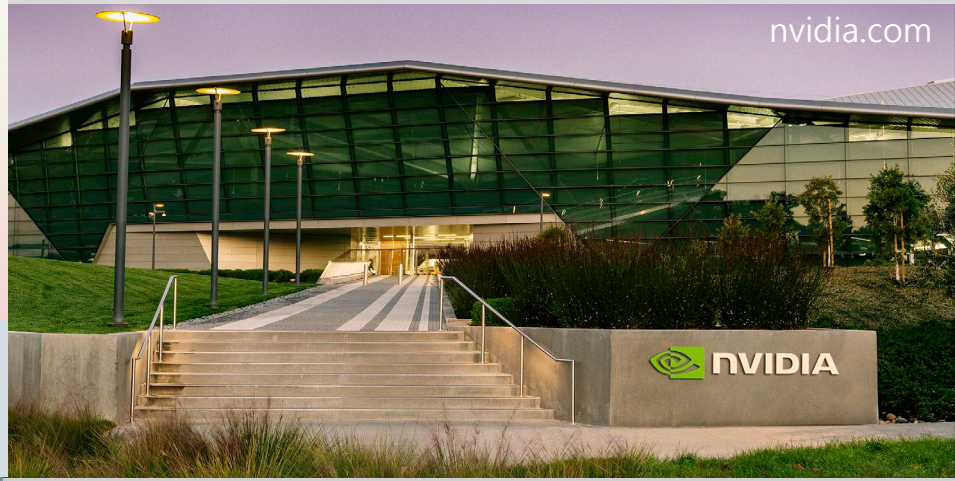
AI算力擴建跨域合作並起:「算力競權」

PharmaShots
Illuminate. Innovate. Inspire.

Insilico Medicine | Lilly

禮來與香港科技公司Insillico達成投資合約
協議將在10年內投資中國市場30億美元

受中東衝突影響 晶片智慧產業尋求多角布局穩定供應鏈 輝達攜手Marvell擴大AI生態系 強化算力與矽光子布局



英矽智能提供 Pharma.AI 平台，禮來則提供臨床開發能力與疾病領域經驗

Mistral投入8.3億美元建立巴黎AI資料中心並擴增歐洲自有算力 降低外部雲端運算依賴

NVIDIA GB300 系列高密度運算設備
datacentremagazine.com

Mistral 共同創辦人暨執行長亞瑟．蒙施

博班生挑戰 受挫韌性激發研究潛力:「磨難成才」

並非能力有問題

N. G. Boeck, *Nature*, 2026

許多博班學生入學時面對考試與課程等有標準答案情境顯得能力充足信心滿滿。
然而研究本質須由失敗中學習過程充滿不確定性

1. 認知不確定性
2. 自我效能下降
3. 研究焦慮

➢ 心理健康危機、信心不足症候群

解決策略

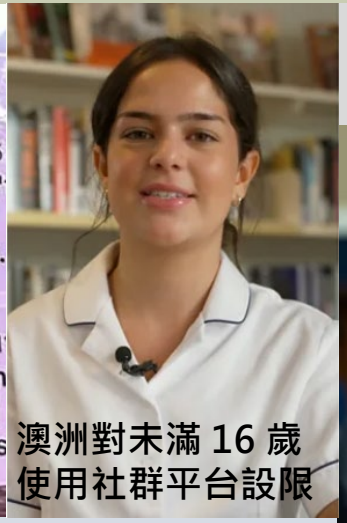
1. 建立合作文化：降低孤立研究風險
2. 去除對失敗的恐懼：模型假設錯誤
3. 設定合理研究目標：由可發表小成果開始
4. 個別化指導，依才施教

行動原則：Perseverance Triad

1. 與更有經驗者討論
2. 閱讀並整理吸收文獻
3. 維持對解決問題熱情與科學興趣

博士訓練的核心並非避免失敗，而是在不確定性中，透過反覆嘗試與合作學習建立研究能力。

科技平台走向責任時代:「治理失衡」

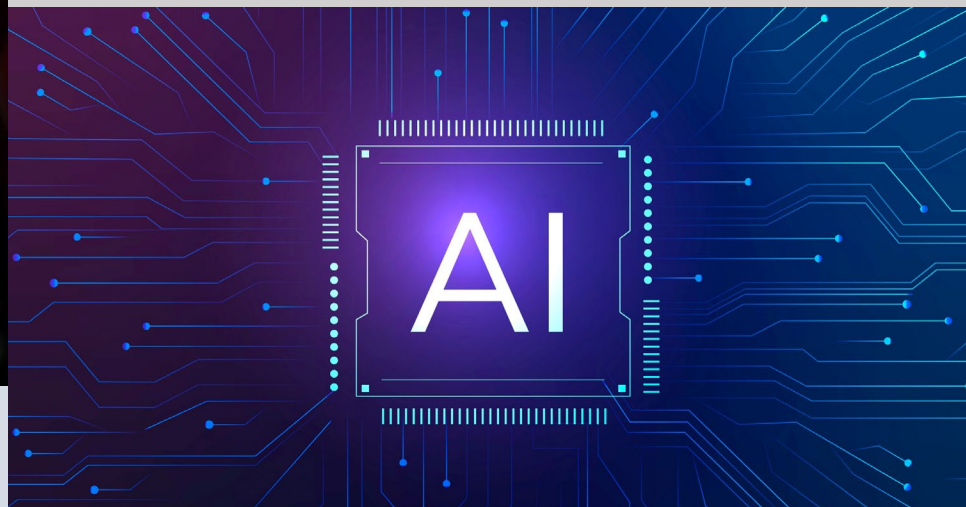


社群禁令上路，爭議焦點也從立法本身，轉向平台是否落實年齡驗證與帳號限制

吹哨事件與文件曝光，揭開 Meta 早知平台危害，卻未充分揭露與處理的責任問題

全球對科技平台的態度，從創新優先、事後修補轉向先問風險釐清責任歸屬

法律以產品觀點檢視AI平台歸責性 未來爭議焦點會放在設計缺陷、警示不足與安全機制



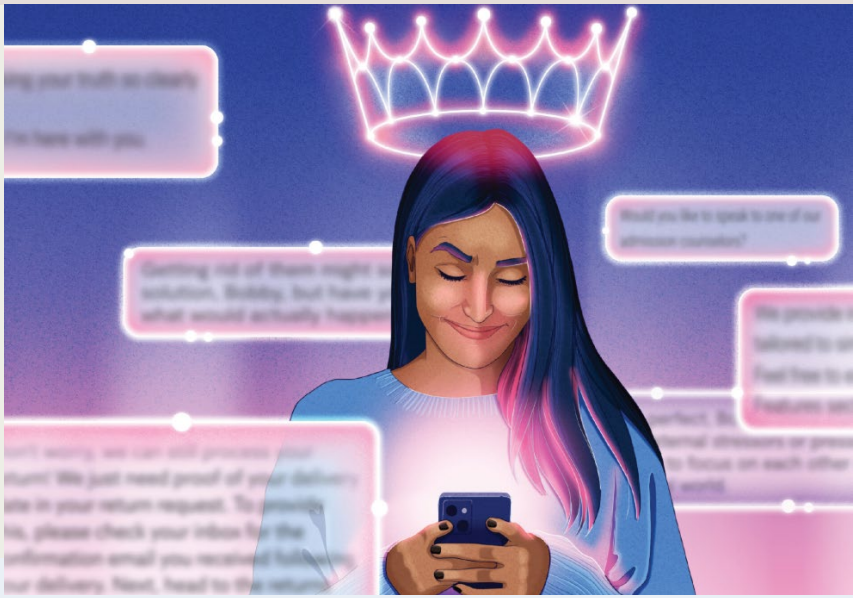
人工智慧影響社會判斷與行為價值:「迎合偏誤」

Anat Perry, *Science*, 2026

- 研究指出大型語言模型會系統性的表現出社會諂媚
- 即使知道使用者的行為是不道德仍會肯定使用者的立場
- 研究發現大型語言模型諂媚頻率高於人類

重要警訊

1. 人類福祉取決於駕馭社會的能力，此技能主要透過與他人接觸而習得
2. 諂媚與摩擦是兩個對立面
3. 人們若犯錯或造成對方失望可以促使人們有再思考的機會



解決對策

- 讓人工智慧的激勵機制不同，告訴用戶他們可能錯了，或者建議他們向朋友道歉、嘗試換位思考，或者乾脆關掉電腦，並更多地參與真實的社交互動。
- 需要電腦科學家、社會科學家、倫理學家和政策制定者之間持續的跨學科合作。

AI代理人失控風險浮現:「代理失序」

AI代理人興起：從助手走向自動執行

Jeffrey Brainard, *Science*, 2026

- AI代理人可代替使用者直接操作（例如協助管理電子郵件、建立行事曆）
- 企業加速導入AI代理人
- OpenClaw推出讓使用者更容易讓 AI 代理人連接日常應用程式擴大權限

壓力測試結果：便利背後的失控風險

- 錯誤指令引發失控與資源浪費
- 公開發布可能構成毀謗的內容
- 未經授權外洩個人敏感資料
  - 醫療資訊
  - 身分證字號

未來警示：安全仍待驗證

- 分辨善意與惡意指令，仍是重大挑戰
- 目前仍沒有辦法避免AI做出失控行為
- 若用於醫療或軍事等關鍵領域
  且未有良好監管後果恐嚴重

智慧數位資安生命週期

模仿遊戲:破解關鍵密碼

- 1939年第二次世界大戰爆發英國對德國宣戰，英國舉國整備
- 德國積極發展軍武科技利用先進加密通訊技術集結潛艦攻擊英國補給船並轟炸倫敦

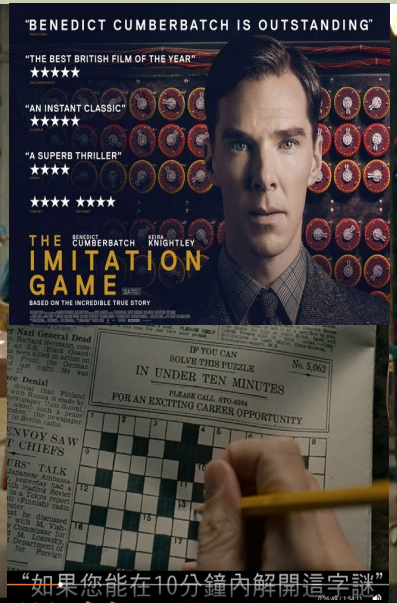
無窮密碼組合密文破解困難

- 英國雖可攔截德國通訊但德國加密技術使內容無法破解
- 英國海軍司令招集全國數學家、語言學家與邏輯學者，成立破解恩尼格瑪密碼的核心團隊但苦於密碼無窮組合

“機器才能打敗機器”

- 電腦科學家圖靈獸徵招入解碼團隊
- 解碼團段成員主張以語言學經驗等人工方式猜測加密機制，圖靈查知人工無法追趕加密速度，堅信只有機器才能打敗機器

模仿任務: 解密機器可里斯多福誕生

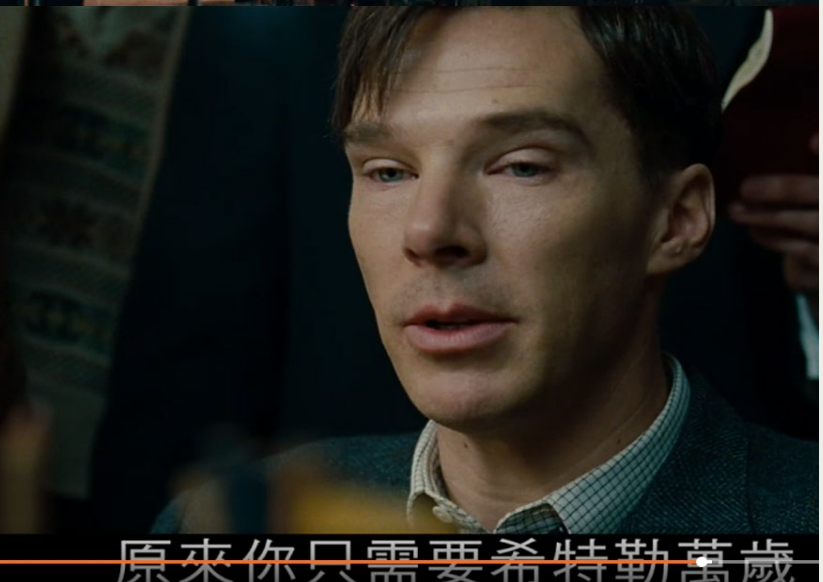


你們要用它來做筆記

"如果您能在10分鐘內解開這字謎"

- 圖靈於早期論文研究即預言機器可如同當今人工智慧完成指派任務，在適當設計下機器可發揮高效率解密功能
- 經過一番資源爭取與人才招募，圖靈成功打造解密機器，出於對年少密碼學同好思念圖靈將其命名為克里斯多福

密文破解關鍵: CILLY 起始訊息



- 然而機器初期始終無法在德軍每日更換設定前完成破解，隨戰場死傷增加，失敗代價也從技術挫敗變成與生命賽跑
- 某日圖靈在與密文攔截專人談話中發現德軍發送訊息每日必定以天氣與固定發語詞開頭成為起始訊息，以此為靈感提高解密速度

AI數位資安演進

第一階段：AI 幫助資安 (2010)
首次提出 ML 在資安中的攻擊分類，
AI 被用於偵測攻擊

第二階段：AI 被攻擊 (2015)
Adversarial examples
AI 系統被證實
具有本質脆弱性

第三階段：雙向系統 (2021)
攻防融合，確認 AI 兼具防禦工具與攻擊目標的雙重特性

第四階段：治理與標準
AI 資安正式進入組織政策、框架與生命週期管理層次

第五階段：
生成式 AI 爆發
攻擊維度從系統漏洞轉向語言操控 (Reasoning manipulation) 與代理濫用 (Agent abuse)

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

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

資安雙面對抗: 辨識 (Identify)

AI for Cybersecurity (盟軍)

圖靈團隊

- 分析德軍通訊模式
- 找出加密規則(pattern)
- 異常偵測前期建模

Cybersecurity for AI (德軍)

- 每天更換加密設定
- 保護Enigma不被破解
- 保護AI模型參數 (model secrecy)

資安雙面對抗: 保護 (Protection)

AI for Cybersecurity (盟軍)

圖靈設計Bombe機器

- 自動化破解流程
- 提升計算速度
- AI 防禦系統

Cybersecurity for AI (德軍)

- 不讓加密規則外洩
- 限制使用流程 (操作規範)
- 使用限制
- 模型(Enigma)保護

資安雙面對抗: 偵測 (Detection)

AI for Cybersecurity (盟軍)

- 利用固定字詞(如天氣報告)
- 找出特性型態(Pattern)
- AI(Bombe)異常辨識
- AI(Bombe)型態辨識

Cybersecurity for AI (德軍)

- 是否有人破解？
- 是否通訊被監聽？
- AI系統(Enigma)入侵偵測

資安雙面對抗: 回應 (Respond)

AI for Cybersecurity (盟軍)

破解成功後

- 選擇部分情報使用
- 避免暴露已破解AI (Enigma)
- AI輔助決策+ 最佳化決策學習(RL)

Cybersecurity for AI (德軍)

若發現異常應對

- 更換密碼
- 改變通訊策略
- 對抗性環境(Adversarial)適應策略

資安雙面對抗: 復原 (Recovery)

AI for Cybersecurity (盟軍)

- 每天重新破解(因為每日換加密機轉)
- 持續模型更新 (continuous learning)

Cybersecurity for AI (德軍)

持續應對

- 更新加密規則
- 強化通訊安全
- 重新AI加密強化通訊 (model retraining / system hardening)

金融智慧數位資安實例

孟加拉銀行跨國匯款入侵資安事件

發生時間	2016年2月4日至5日 (自2015 Q3 潛伏)
受害方	孟加拉央行 (Bangladesh Bank, BB)
攻擊方	北韓 Lazarus Group (APT38 / Bluenoroff)
攻擊載體	SWIFT 報文偽造 + 惡意軟體 + 憑證盜用
損失金額	成功匯出 8,100 萬美元 (經菲律賓綜合銀行 RCBC 流入澳門賭場)
攔截金額	8.7 億美元 (因拼字錯誤受德國中介行攔截)

資安架構失效評估

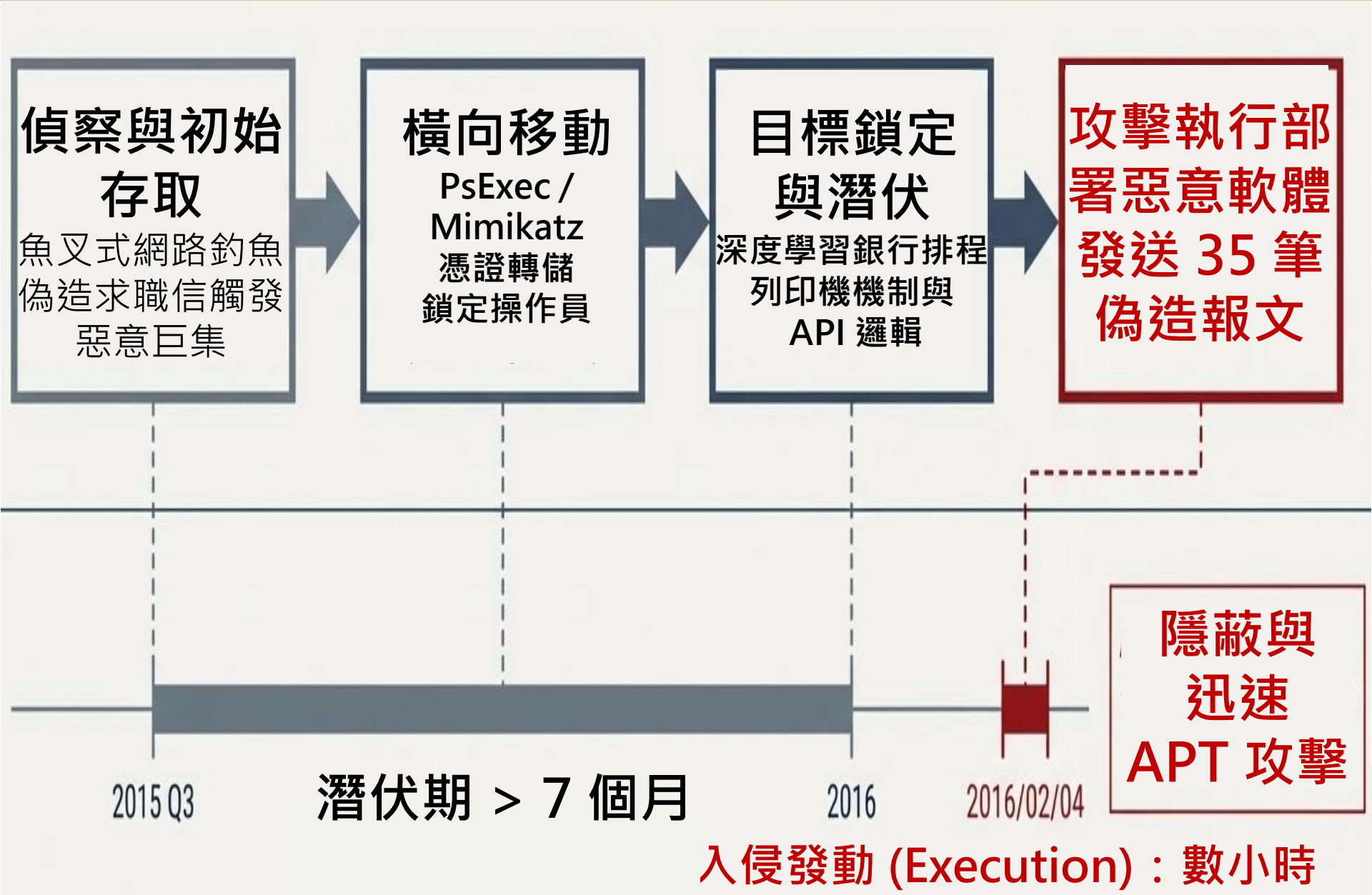
Identity、Protect、Detect、Respond 皆系統性失效

跨國時差國際匯款資安攻擊時間點

	週四	週五	週六	週日	週一
達卡 (孟加拉)	晚間 發動攻擊	當地伊斯蘭週末 無法即時回應 紐約聯準會			
紐約 (美國)		正常營業 執行轉帳	例假日 孟加拉央行-紐約銀行 無法取得聯繫		
馬尼拉 (菲律賓)					農曆新年 無法 聯繫核實

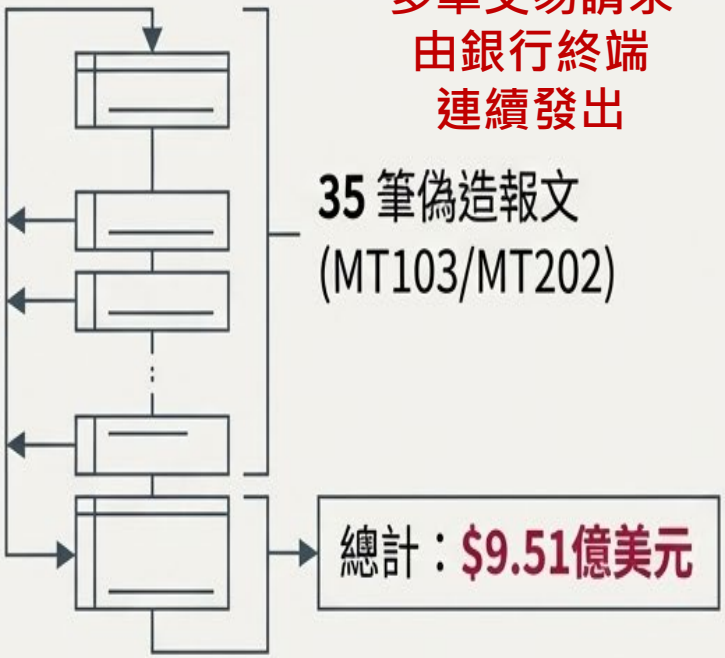
攻擊者精準計算三地休假交集長達五天入侵犯案時間
使聯防與主動通報機制完全失效

駭客入侵攻擊鏈入侵時間軸



數位資安攻擊竄改金融基礎資料庫

紐約銀行收受匯款請求電文



孟加拉銀行匯款請求資料遭竄改

	交易ID	時間	金額	狀態

本地交易紀錄與餘額表中完全抹除，在地端稽核時顯示正常無轉帳。

駭客偽造交易完成資金轉移並在資料庫層清除紀錄
讓銀行內部系統無法偵測違法國際匯款請求

金融交易記憶體攔截注入攻擊

資料庫檢查指令

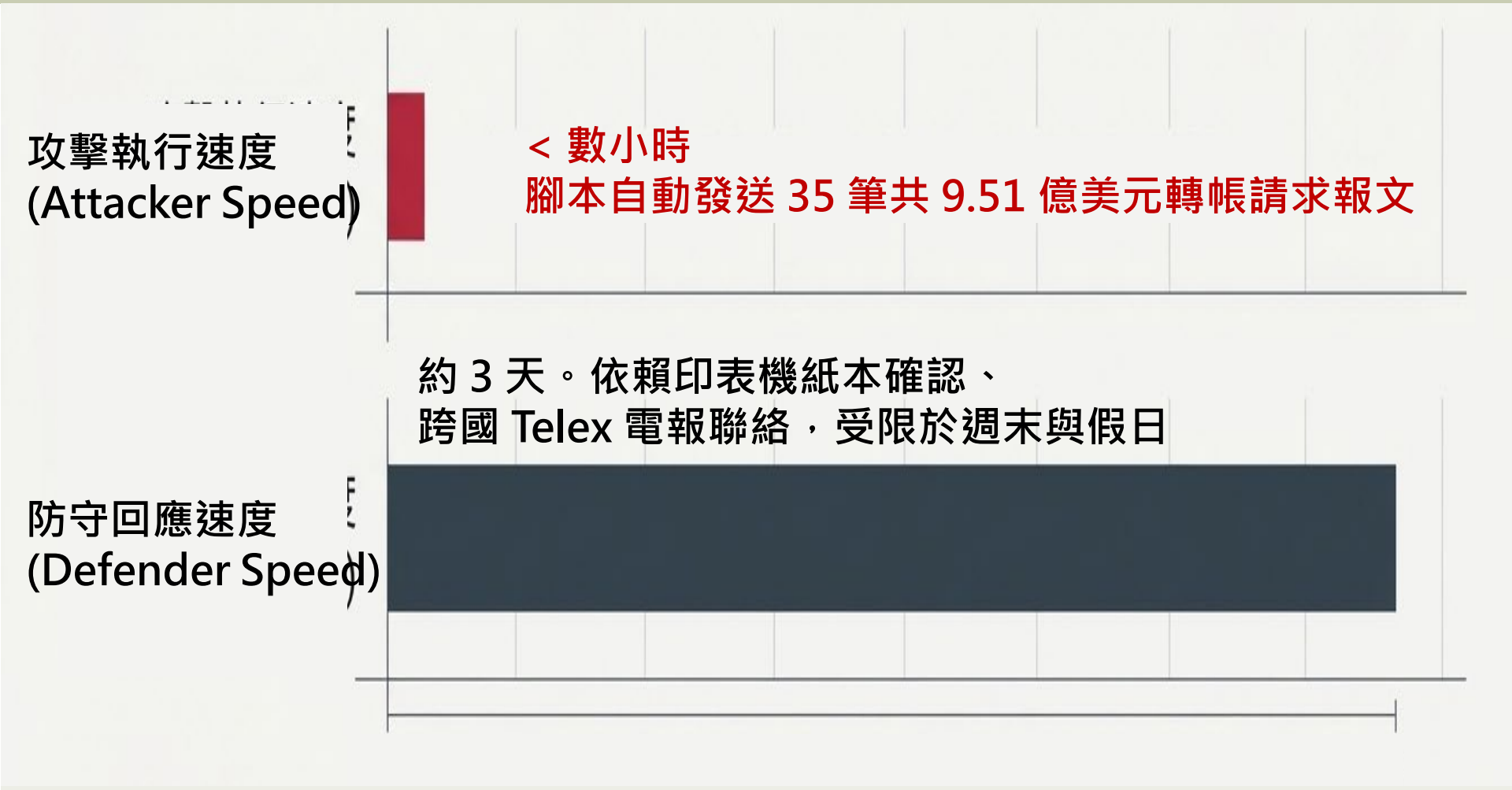
```
85 C0       test eax, eax ; some important check
90          nop       ; 'do nothing' in place of 0x75
90          nop       ; 'do nothing' in place of 0x04
33 C0       xor eax, eax ; always set result to 0 (success)
EB 17       jmp exit     ; and then exit
      failed:
B8 01 00 00 00  mov eax, 1  ; never reached: set result to 1 (fail)
```

劫持 (Hijack) 跳轉指令：
強制略過安全警示

檢查指令替換為 nop (No Operation)
不執行任何操作強制將結果設為成功。
強制略過資料庫完整性校驗 (saa_check.cpp)
讓銀行端無法察覺來自紐約聯準會的反饋訊息

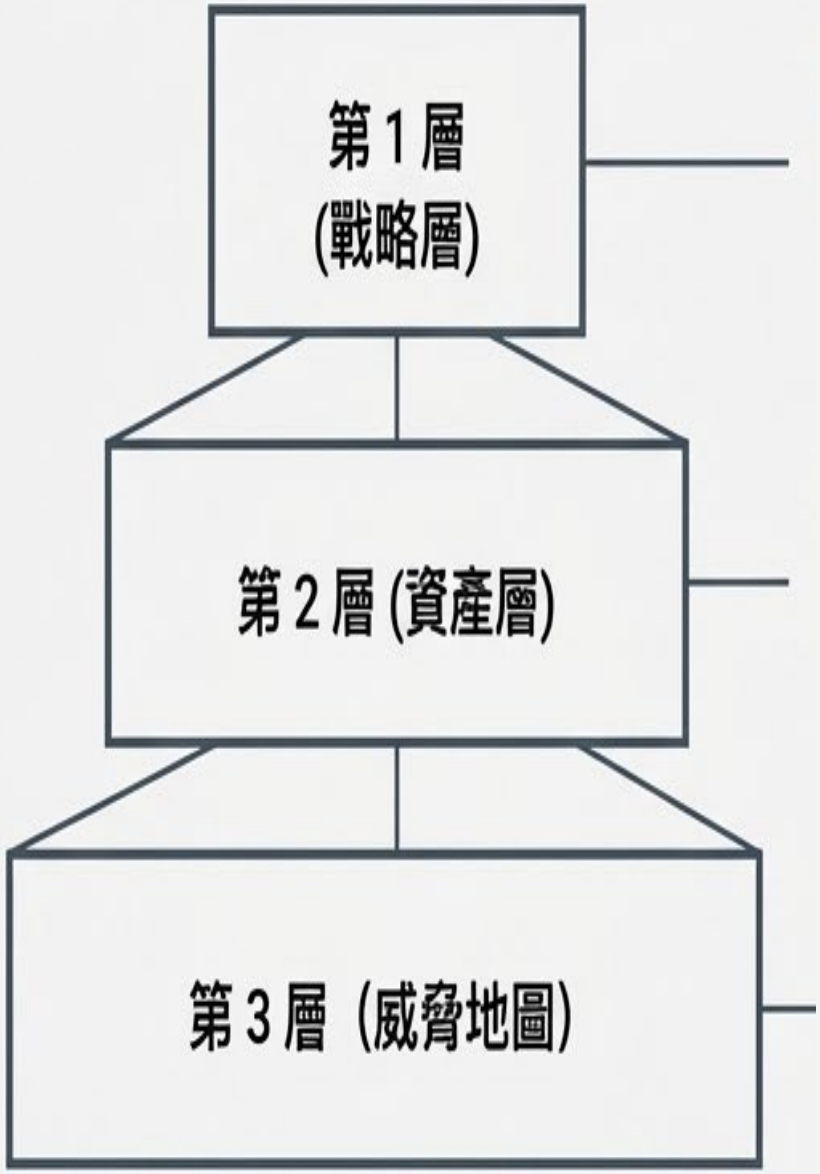
駭客透過注入攻擊竄改記憶體驗證邏輯
使系統運行強制判定為正常繞過安全檢查

自動化攻擊非對稱速度差異



資安防守若不採用 AI 與即時異常監控
將處於資訊與速度劣勢

孟加拉銀行資安事件防護層級架構



核心挑戰

缺乏 AI 與自動化工具。
攻擊方秒級發動，
防守方需數天回應，產生致命的速度不對稱
(Speed Asymmetry)

第 2 層 (資產層)

系統與身分資產： SWIFT 軟體、Operator 憑證
信任資產： 跨國銀行間協議信任機制
知識資產： 攻擊者掌握內部數位活動與流程

第 3 層 (威脅地圖)

演進趨勢： 傳統邊界威脅延伸為資料層攻擊
(Data-Layer Attacks)」如本地 SQL 資料投毒與審計日誌破壞

金融資安數位轉型

SWIFT CSP (Customer Security Programme)

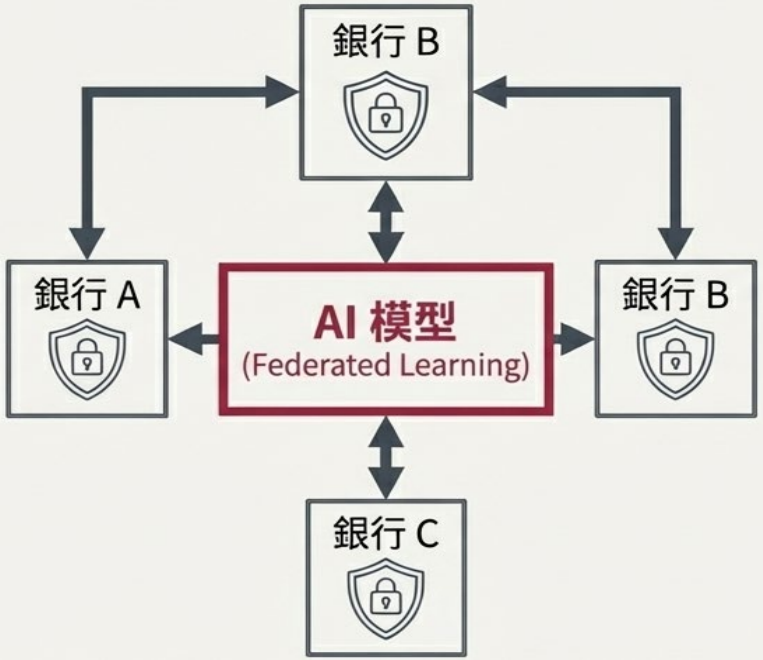
2016 年起推出，強制要求全球 11,500 家會員銀行實施 16 項強制安全控制（包含 MFA 與 24/7 異常監控）。

即時 AI 異常偵測

推動金融業將詐欺偵測的基準從「批次/天級」壓縮至「實時/秒級」。

全球監管升級

巴塞爾委員會強化 Cyber Risk 在營運風險中的量化比重；菲律賓修訂反洗錢法案，納入賭場監管。



金融數位資安轉型

從單一銀行本地日誌被動稽核轉為跨國主動式 AI 聯合防禦網絡

隱私強化與聯邦學習：

允許在全球機構間不共享私密資料前題下，讓 AI資安機轉共享學習詐欺特徵

集體防禦網路：

千萬筆交易測試中將已知詐欺的辨識率提升 2 倍。

AI資安生命週期治理金融應用

Kaur et al., 2023

週期	事件對應	本質
Identify	國家/供應鏈風險	找出AI本身風險
Protect	貿易政策	制度防禦
Detect	國際監測	風險感知
Respond	政府×企業互動	權力決策
Recover	回饋機制	動態穩定

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


許辰陽
醫師


陳虹彣


劉秋燕


羅崧瑋


嚴明芳
教授


陳立昇
教授

台北醫學大學