AI產業實戰應用人才淬煉計畫

114年度AIGO淬煉實戰盃
競賽提案報告

次世代AI防詐安全中樞：LINE主動式預警與教育平台
來都來了·詐不過

114/08/25

報告大綱(參考，格式不拘)

1. 團隊簡介
2. 解題計畫宗旨及目的
3. 解題技術架構及進行步驟
4. 數據應用及作法
5. 預期完成之工作目標
6. 預期工作摘要及進度表
7. 預期成果與效益
8. 其他有利審查項目

團隊簡介

團隊名稱：**來都來了·詐不過** – 次世代AI防詐安全中樞：LINE主動式預警與教育平台

張維峻

助理架構師 / 專案規劃與技術總設計

- 主導專案架構設計，負責解題構想與技術藍圖規劃
- 建立「詐騙手法拆解報告」自動生成系統，開發多模態辨識模型，應對圖片與文字混合詐騙
- 協調跨領域資源，確保專案符合目標與時程

邱芝甯

襄理 / 解決方案顧問

- 提供專案在數位架構上的專業建議
- 協助需求分析與方案落地可行性評估
- 支援團隊與內外部單位溝通，確保專案推動順暢

劉宇祥

高級專員 / 平台整合微服務專家

- LINE 平台整合
- 雲端微服務系統設計
- 建立「詐騙手法拆解報告」自動生成系統

廖銘宏

專案經理 / 任務管理與營運驗證

- 規劃專案進度與人力配置，確保專案如期完成
- 建立風險管理流程與 KPI 追蹤機制
- 負責用戶測試與營運驗證，確保成果符合市場需求

團隊特色

跨領域合作

涵蓋 AI 技術、數位架構、專案管理與解決方案設計

技術創新

主動式威脅情資系統、多模態辨識與 RAG 報告生成

社會價值

從被動查證進化到主動防詐，提升全民數位免疫力

次世代AI防詐安全中樞：防詐模式面臨的挑戰

現行的AI防詐模式正面臨嚴峻考驗，迫切需要從根本上進行典範轉移。傳統的被動防禦策略已無法有效應對詐騙手法的快速演進，這使得我們始終處於追趕的劣勢。

被動防禦，反應遲緩

多數防詐工具仍停留在「被動式查證」，用戶需先接觸詐騙才能舉報，導致反應延遲，使詐騙集團佔據先機。

詐騙手法，快速演進

詐騙結合AI，手法日益複雜且多樣，涵蓋文字、圖片、影音、惡意連結，甚至深度偽造，傳統防禦難以招架混合式攻擊。

缺乏主動預警機制

現有模式無法在詐騙訊息大規模擴散前即時預警，當問題浮現時，往往已造成廣泛影響，防禦者持續處於被動。

難以提升全民免疫力

多數工具僅提供真偽辨識，缺乏深入的詐騙分析報告，未能有效普及防詐知識，大眾數位免疫力難以提升。

平台選擇的迫切性

台灣的 LINE **使用率極高**，如果服務能直接建構於 LINE 之上，就能**無縫融入用戶的日常生活**。這不僅免除了額外下載 APP 的門檻，更能在第一時間接觸到最廣大的使用者群體，最大化服務的**普及率與觸及範圍**。 。

次世代AI防詐安全中樞：主動預警與深度賦能

當前防詐模式面臨兩大挑戰：**被動防禦**導致反應滯後，以及**詐騙手法快速迭代**使防守方永遠處於劣勢。為此，我們將打造基於LINE平台的「次世代AI防詐安全中樞」，實現從被動防守到主動預警的典範轉移。

主動預警

縮短威脅反應時間，防患於未然。

深度賦能

AI生成詐騙分析，提升全民數位免疫力。

生態核心

成為台灣防詐數據樞紐，建構聯防體系。

核心技術優勢

- 主動式威脅情資：全網監控社群，大規模擴散前預警。
- 生成式AI：自動化生成淺顯易懂的詐騙分析報告。
- 多模態辨識：應對圖片、變造字體等新型混淆攻擊。

解題技術架構及進行步驟

我們的防詐安全中樞採用雲端微服務架構，核心串接LINE Messaging API，並整合三大AI引擎，實現高效能與可擴展性：

NLP引擎

負責社群趨勢監控，
即時捕捉潛在詐騙訊息爆量。

RAG引擎

基於資料庫生成詐騙分析報告，
提供清晰易懂的資訊。

多模態引擎

精準辨識圖片與文字內容，
應對新型詐騙手法。

主動預警流程

01
AI 24/7 監控
自動偵測社群關鍵詞爆量，識別潛在威脅。

02
觸發與摘要
偵測到威脅後，系統自動摘要相關資訊，
生成「潛在威脅事件」。

03
人工驗證決策
營運人員快速人工驗證，判斷警報等級與應對策略。

04
資訊廣播

被動查證流程

01
用戶上傳訊息
使用者透過LINE頻道上傳文字、圖片或連結進行查證。

02
多模態比對
多模態AI模型與「動態詐騙資料庫」進行即時比對。

03
報告生成
生成式AI自動產出查證結果與詳細的「詐騙手法拆解報告」。

04
結果回覆與優化
系統即時回覆查證結果，並邀請用戶回饋以持續優化模型。

數據應用及作法

我們的核心戰略是建立一個比競爭對手更新、
更快的「動態自學習數據生態系」，確保能即時應對不斷變化的詐騙手法。

啟動階段

初期透過合法合規的網路爬蟲與API接介，彙整Cofacts、台灣事實查核中心等公開資料，建立基礎事實庫。

核心競爭力：動態詐騙變體資料庫

每筆用戶「新查詢」視為潛在詐騙變體。AI語意聚類自動歸類，即時更新資料庫。查詢量激增時，系統自動標示高風險目標，實現近乎即時威脅感知。

數據類型與欄位

包含文字內容、圖片特徵、URL、QR Code解析內容、上傳時間、匿名化用戶ID、社群監控數據等多元情報。

長期願景：防詐開放平台

未來透過API將新型詐騙樣態以匿名化、結構化方式，提供給警政單位、資安廠商，形成聯防生態系。

預期完成之工作目標

為確保「次世代AI防詐安全中樞」專案的成功推動，
我們設立了以下關鍵目標與可量化的績效指標：

1

LINE官方頻道上線

完成LINE官方頻道原型開發，並上線核心查證功能。

2

主動威脅情報系統

部署第一版系統，成功監控至少3個主要社群平台。

3

緊急預警發布次數

專案期間至少識別並發布3次「主動式緊急預警」。

4

報告準確率提升

AI生成之「詐騙手法拆解報告」準確率達90%以上。

5

LINE好友數成長

專案結束時LINE官方帳號好友數突破1,000人。

6

用戶滿意度

用戶滿意度調查結果達到85%以上。

預期工作摘要及進度表

我們將依循明確的執行路線圖，分階段推進「次世代AI防詐安全中樞」的開發與部署，確保每個環節都達到預期目標。

階段一：需求分析與基礎建設

確立技術架構，同步進行合法合規的資料源蒐集，並完成網路爬蟲與基礎API管線建置，為後續開發奠定堅實基礎。

階段二：核心引擎開發

聚焦於核心AI引擎的開發，包括「主動式威脅情資系統」原型建立，以及多模態辨識模型的初步訓練，提升詐騙偵測能力。

階段三：整合與測試優化

將開發完成的AI引擎整合至LINE官方頻道，同步開發「詐騙拆解報告」生成器，並進行全面性的內部壓力測試與使用者體驗優化，確保系統穩定與效率。

預期成果與效益

「次世代AI防詐安全中樞」專案的推動，將帶來具體的成果交付物與深遠的社會經濟效益。

預期成果

- 「次世代AI防詐LINE安全中樞」原型，具備核心查證與預警功能。
- 一套可運作的「主動式威脅情資系統」後端服務，實現多社群平台監控與分析。
- 為未來推廣提供依據。

預期效益

對民眾

大幅降低受騙風險，從被動受害者轉變為主動防禦者，
並提升全民數位素養。

對產業

建立可行的AI主動防禦商業模式，
並為資安、金融等產業提供高價值的威脅情資。

對社會

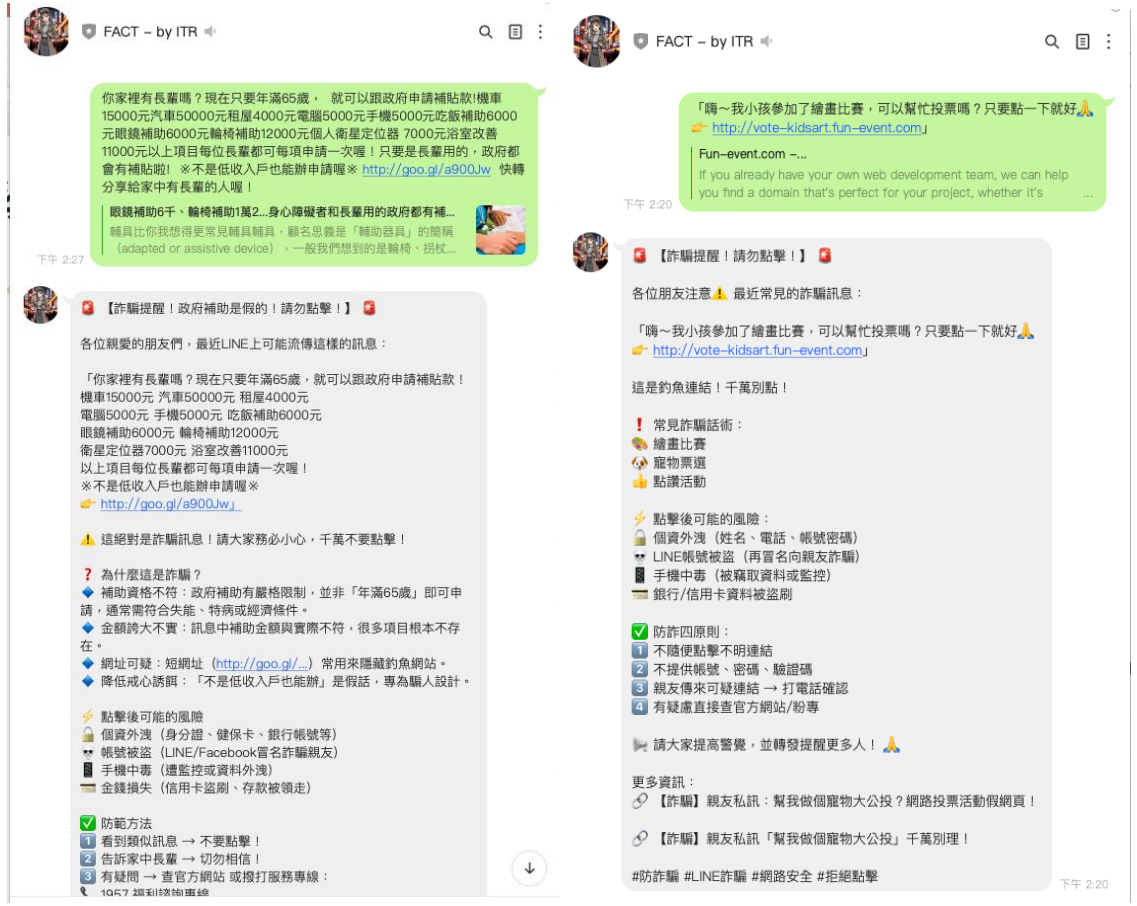
打造反應更快速、覆蓋更廣泛的社會安全防護網，
形成正向的數位公民防禦生態。

系統示意圖

1. 詐騙訊息驗證服務（被動防禦）

使用者只需將收到的文字、圖片或連結上傳至 LINE Bot，即可即時獲得查證結果。 系統將透過多模態 AI 模型進行分析，並比對「動態詐騙變體資料庫」，同時由生成式 AI 產出淺顯易懂的【詐騙手法拆解報告】，讓使用者不僅知道真假，更能理解背後的詐騙手法與風險。

用戶訊息查證

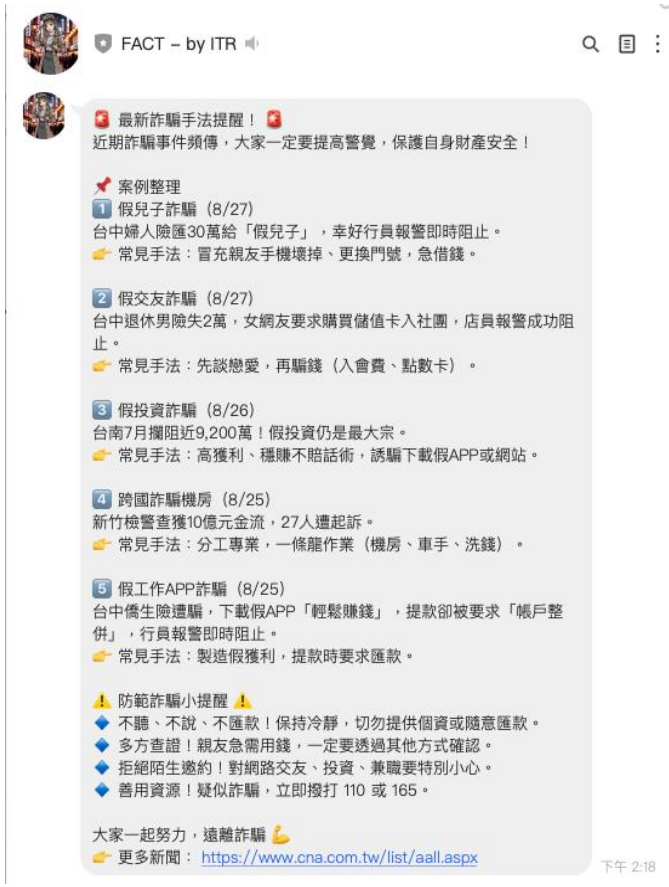


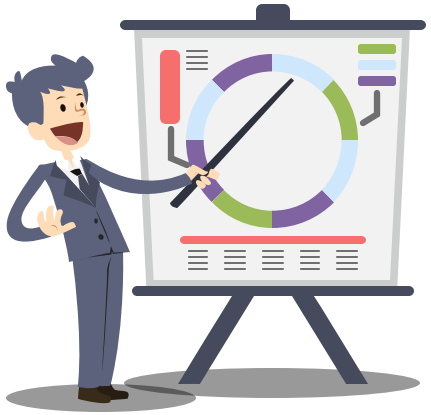
系統示意圖

2. 最新詐騙預警推播（主動防禦）

系統透過「主動式威脅情資系統」持續監控公開社群平台，偵測異常訊息與新型詐騙話術。一旦確認高風險訊息，LINE Bot 將即時主動推送至用戶，提醒潛在威脅，並定期發布【防詐週報】，協助使用者掌握最新詐騙趨勢。

主動威脅情報系統





簡 報 完 畢
敬 請 指 導

Thank you