資訊人社會關懷獎學金關懷提案書

AI 全視聽防詐守護 App

提案人：徐筱雯

中央警察大學．資訊管理學系．四年級

共同提案人：田秉麒

中央警察大學．資訊管理學系．四年級

提案日期：113 年 10 月 17 日

AI 全視聽防詐守護 App

一、 社會關懷議題

(一) 關懷議題內容

在現代數位時代，詐騙手法日益精巧且難以察覺。詐騙者靈活運用社群平台、即時通訊軟體以及語音通話等工具來實施詐騙，通常透過語音和文字的互動逐步建立信任。他們的策略是通過長期且有計劃的溝通，使受害者難以在短時間內辨別真偽，最終引導受害者分享敏感信息或進行財務交易。

詐騙案件的增長顯示出詐騙行為愈加普遍，而受害群體的分布存在明顯的不均衡現象。各年齡層的受害者面臨著不同形式的詐騙，這使得針對不同群體設計專門的防護措施成為必要。受害群體按年齡分述如下:

1. 青少年和年輕人（18-29 歲）

   該年齡層的受害者往往活躍於網路平台，特別容易成為網路拍賣詐騙和解除分期付款詐騙的目標。詐騙者利用這一群體的消費習慣和對數位設備的依賴性，通過社交媒體或電子商務平台進行詐騙，使這些年輕受害者在不知不覺中陷入陷阱。

2. 青壯年群體（30-39 歲）

   這個年齡層的人群通常涉及更多的金融活動，並因此更易受到投資詐騙的影響。詐騙者經常透過偽造的高收益投資項目來吸引這一群體，利用其對財務增長的需求來設計誘餌。

3. 中老年群體（40 歲以上）

   中老年群體則是詐騙者在金融詐騙中特別青睞的目標，尤其是涉及健康、退休金和投資的詐騙。這些年齡段的人往往有較多的儲蓄，對於投資產品的防範意識較弱，因此容易成為大額詐騙的受害者。

112 年詐欺案件各年齡別之被害方式

年齡別	總計	投資詐欺	解除分期付款詐騙(ATM)	假網路拍賣(購物)	一般購物詐欺(偽稱買賣)	猜猜我是誰	假愛情交友	假冒機構(公務員)	遊戲點數(含虛擬寶物)詐欺	其他
總計	100.00	31.05	23.31	13.93	7.86	3.06	3.05	2.32	2.16	13.26
0-17歲	100.00	17.59	19.17	**21.34**	11.56	1.19	3.56	0.99	10.97	13.64
18-23歲	100.00	19.10	**33.66**	19.11	7.54	0.90	3.08	1.31	3.24	12.06
24-29歲	100.00	25.12	**29.26**	16.45	7.85	0.92	2.94	1.28	2.91	13.27
30-39歲	100.00	**29.29**	24.70	16.41	8.67	1.13	3.07	1.23	2.34	13.17
40-49歲	100.00	**32.23**	21.36	14.12	9.60	1.90	3.46	1.38	1.37	14.59
50-59歲	100.00	**43.47**	15.67	7.72	7.17	4.95	3.20	2.64	0.80	14.38
60-64歲	100.00	**47.46**	11.28	5.08	5.96	8.36	2.79	5.47	0.56	13.05
65歲以上	100.00	**46.50**	8.44	3.64	4.43	13.58	2.27	8.91	0.51	11.72

資料來源：本署刑事警察局。

這種對各年齡層的細緻分類和數據分析，能夠幫助我們更好地理解社會各群體面臨的詐騙威脅，並制定更加公平且全面的防護措施，以確保系統在廣泛的社會群體中都能有效運作，最終達到減少受害者數量並提高社會信任的目標。如果從詐騙類型去看，如下圖:

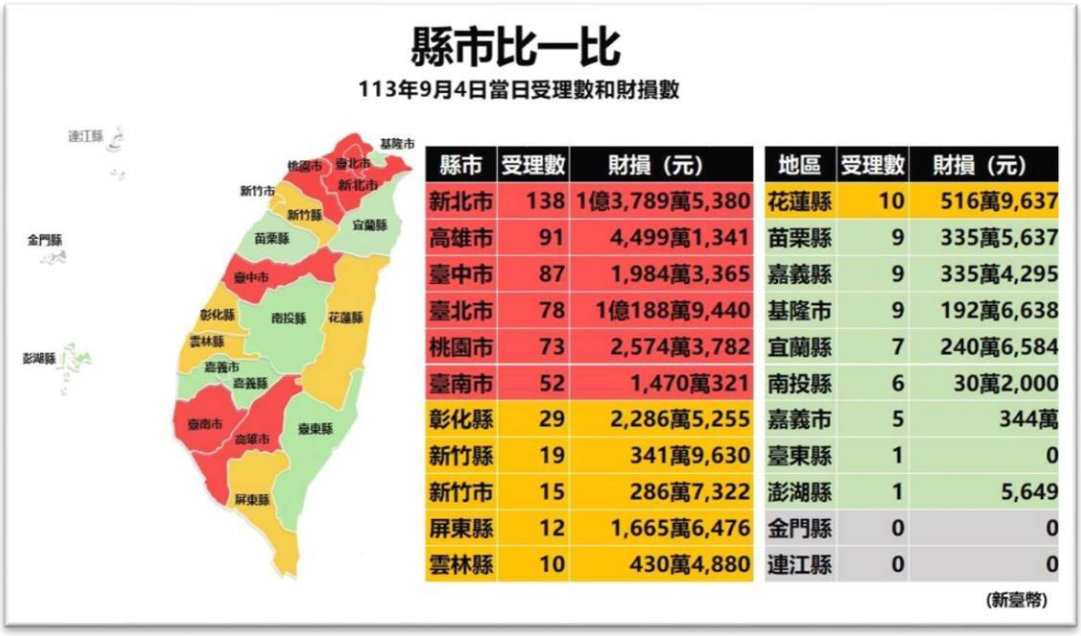
以新北市為例，依照詐騙手法的分布，主要集中在五種手法，財損金額相當驚人：

1. 假投資詐騙：此類詐騙以 206 件的案件數位居首位，造成的財損金額高達 3 億 8,87 萬元。詐騙者透過偽造的投資機會引誘受害者進行投資，並承諾高回報，最終使受害者損失大量金錢。
2. 網路購物詐騙：共有 80 件此類案件，導致 230 萬元的財產損失。詐騙者利用假冒的網路商店或偽造商品，誘騙消費者進行付款，但實際上從未交付商品。
3. 假買家詐騙：這類詐騙有 64 件案件，造成 733 萬元的損失。詐騙者假扮買家，利用交易平台或二手交易機會來騙取賣家的金錢或商品。
4. 假交友詐騙（投資詐財）：這類詐騙案件共有 41 件，損失金額達到 4,711 萬元。詐騙者通常會假扮為朋友或戀人，並利用情感依賴來說服受害者投資，從而騙取大額金錢。
5. 騙取金融帳戶詐騙：共有 33 件案件，導致 9 萬元的財產損失。詐騙者假冒銀行或金融機構，誘導受害者提供帳戶資訊或密碼，進行非法轉帳或提取。

綜合分析各年齡層與詐騙類型的數據，可以清楚看出詐騙行為隨著社會的數位化而變得更加精密、針對性強且多樣化。

(二) 關懷議題的社會影響度

在當今社會，詐騙問題已經滲透到人們生活的方方面面，不僅影響著個人財產安全，更對社會的信任基礎構成威脅。詐騙者以各種手段巧妙操控受害者的情感和判斷，使得各類人群無法在短時間內察覺到異常，最終落入陷阱。這種持續擴散的詐騙行為帶來了巨大的社會挑戰，而其中的具體影響可以從全台範圍的財損統計中一窺端倪。



縣市	受理數	財損（元）
新北市	138	1億3,789萬5,380
高雄市	91	4,499萬1,341
臺中市	87	1,984萬3,365
臺北市	78	1億188萬9,440
桃園市	73	2,574萬3,782
臺南市	52	1,470萬321
彰化縣	29	2,286萬5,255
新竹縣	19	341萬9,630
新竹市	15	286萬7,322
屏東縣	12	1,665萬6,476
雲林縣	10	430萬4,880

地區	受理數	財損（元）
花蓮縣	10	516萬9,637
苗栗縣	9	335萬5,637
嘉義縣	9	335萬4,295
基隆市	9	192萬6,638
宜蘭縣	7	240萬6,584
南投縣	6	30萬2,000
嘉義市	5	344萬
臺東縣	1	0
澎湖縣	1	5,649
金門縣	0	0
連江縣	0	0

(新臺幣)

從全台各地的詐騙案件財損統計圖表可以看出，詐騙案件在全國範圍內已經達到令人震驚的規模。例如，新北市的受理案件數量高達 138 件，財損金額超過 1 億 3,789 萬元，高雄市和台中市也同樣面臨巨大的財務損失，分別達到 4,499 萬元和 1,984 萬元。這些數據顯示，全台灣幾乎沒有一個縣市能夠免於詐騙的侵害，詐騙行為已經滲透到城市與鄉村、富裕與貧困地區。這些受害者不僅承受了財務損失，更可能面臨心理上的創傷，特別是因為信任被利用而帶來的失落感和無助感。

詐騙行為對社會的影響遠不止於金錢上的損失，許多受害者在發現自己被騙後，往往陷入深深的自責與焦慮之中，甚至發展成抑鬱症，導致悲劇性的自殺案例層出不窮。詐騙者巧妙地利用人們對他人的信任，逐步摧毀了受害者的心理防線，這不僅影響到個人，也對整個家庭和社會的信任結構帶來嚴重打擊。

雖然詐騙手法多樣且不斷演變，但其核心仍然無法脫離建立信任的過程，尤其是通過文字訊息和語音溝通來實施詐騙。詐騙者依賴這些溝通手段逐步取得受害者的信任，因此，如果能夠在這些關鍵的傳遞過程中進行有效的偵測與阻斷，將有助於減少詐騙成功的機會。透過技術手段分析並識別詐騙訊息或語音的可疑特徵，提前發出警示，便是一個可實施的防詐

騙方法，讓受害者能在信任被過度利用之前識別危險並做出應對。

二、 解決方案

(一) 現行 APP 介紹及其侷限性

目前市面上針對防詐騙的應用中，較為知名的 Whoscall 和網路詐騙通報查詢網，以下針對這兩個 App 做介紹。這些 App 的推出在一定程度上提高了大眾對詐騙行為的警覺性，並提供了相關的防範工具。

1. Whoscall

Whoscall 是一款來電識別 App，它能夠根據其龐大的數據庫，為使用者顯示來電者的可能身份。當使用者接到陌生電話時，Whoscall 能夠及時顯示出來電者是否可能是詐騙者，並提供過往的舉報記錄或用戶反饋。這樣的功能有助於使用者快速判斷來電的安全性，避免接聽可能存在風險的電話。

雖然 Whoscall 在來電顯示方面提供了有效的輔助，但其侷限性在於，這款 App 依賴的是已知的數據庫和過往舉報數據。如果詐騙者使用的是尚未被標記的新號碼或更換號碼，Whoscall 可能無法即時識別，導致部分詐騙電話仍有機會繞過系統。此外，來電的真實身份有時也會被詐騙者隱藏或偽造，僅憑顯示的資訊並不足以完全杜絕詐騙風險。App 介面如圖:

2. 網路詐騙通報查詢網

另一個現行的防詐騙平台是「網路詐騙通報查詢網」，它允許使用者將發現的疑似詐騙網址進行舉報，並等待官方驗證後進行相關處理。這一通報系統為防範網絡詐騙提供了一個公共參與的途徑，

讓更多人能分享疑似詐騙資訊並保護其他潛在受害者。

網路詐騙通報查詢網的主要侷限在於，它僅針對網址進行通報和查詢，無法即時防範其他形式的詐騙(如語音通話或訊息詐騙)。此外，使用者無法在第一時間獲得尚未被舉報的網址是否為詐騙的確認結果，且須等待調查確認是否可能為詐騙訊息，防護效果存在延遲，容易讓受害者錯過及時防範的機會。App 介面如圖:

總結來看，現行的防詐騙 App，雖然在特定領域能夠提供一定的保護，但其侷限性使得使用者仍然可能遭遇未被發現或標記的新型詐騙。因此，結合更多即時分析技術，如語音與文字的實時風險偵測，將是防詐騙應用發展的關鍵。

(二) AI 智慧防詐 APP 原理與開發

本 AI 智慧防詐 App 專注於即時識別詐騙行為，通過 AI 分析語音和文字訊息，來快速判斷是否存在詐騙風險。與目前市場上常見的防詐騙 App 相比，這款 App 不僅僅依賴於過去的數據庫或舉報記錄，更是結合了實時語音和文字內容的分析技術，能夠提前預測潛在的詐騙行為。這使得使用者能夠在詐騙者建立信任的初期，迅速識別風險，並採取行動。這種即時偵測與主動防護的能力，有效克服了傳統 App 依賴過往數據的局限，讓使用者在面對新型詐騙時，能獲得更高的安全保障，進而降低上當受騙的可能性。

詐騙者通常需要通過長期溝通來建立信任，而文字和語音則是其主要的溝通手段。由於許多社群軟體具有封閉系統的特性，無法直接分析平台內的內容，因此使用截圖來進行偵測變得至關重要。同時，語音訊息在詐騙過程中也扮演重要角色。結合 AI 進行語義分析，判斷語音中是否存在詐騙風險。讓使用者在詐騙者逐步建立信任之前便能及時採取行動，避免陷入更大的風險。以下分成語音和圖片的解決方案進行介紹:

1. 語音解決方案

(1).語音的詐騙判斷

語音經過 speech_recognition 技術將其轉換為可分析的文字內容。接下來，這段轉換後的文字會被送入 OpenAI 的 GPT-3.5-turbo 模型進行自然語言分析。通過這個模型，系統可以深入理解語音中的語境，並根據語句中的關鍵詞、語義結構以及常見的詐騙模式，判斷其是否存在潛在的詐騙風險。這樣的語音到文字再到詐騙風險的判斷流程，不僅實現了語音訊息的即時分析，還能夠根據語義判斷潛在的威脅，為使用者提供即時的風險提示，有效地提高了詐騙預防的準確性與效率。

(2).真偽語音判斷

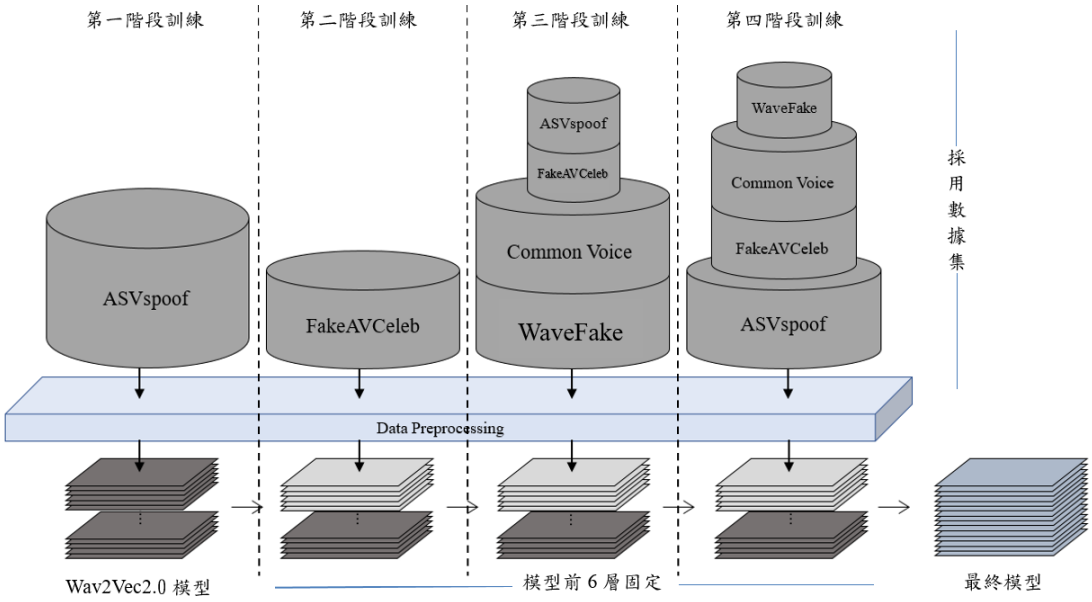
針對語音詐騙的複雜性，我們開發並訓練了一個基於 Wav2Vec2 的語音偽造檢測模型，專注於即時識別詐騙語音的真偽。該模型通過多階段訓練策略來增強其在不同語音偽造技術上的識別能力，並能夠即時通知使用者詐騙風險。模型的訓練過程涵蓋四個主要階段，使用多種數據集進行訓練，逐步提高模型的泛化能力和對偽造語音的精準檢測：

A. 第一階段：首先使用 ASVspoof 數據集進行初步訓練，幫助模型適應基本的二元區分任務，即真實語音與偽造語音的區分。這一階段的訓練為後續的深層偽造語音檢測奠定了基礎。

B. 第二階段：接著，使用 FakeAVCeleb 數據集進行更進階的訓練，專注於偽造技術的細微特徵識別。此階段中，模型的前六層卷積層被凍結，強化了高階特徵的學習能力，讓模型能夠識別更複雜的偽造模式。

C. 第三階段：模型進一步在多語言的 WaveFake 和 Common Voice 混合數據集上進行訓練，提升對多語言和高級偽造技術的識別能力。在這一階段，保持部分參數凍結，專注於學習多語言和更精細的偽造語音。

D. 第四階段：最終階段將所有數據集整合進行綜合訓練，通過持續學習策略提升模型的泛化能力，同時防止遺忘先前學到的關鍵特徵。各階段的訓練步驟圖如下:



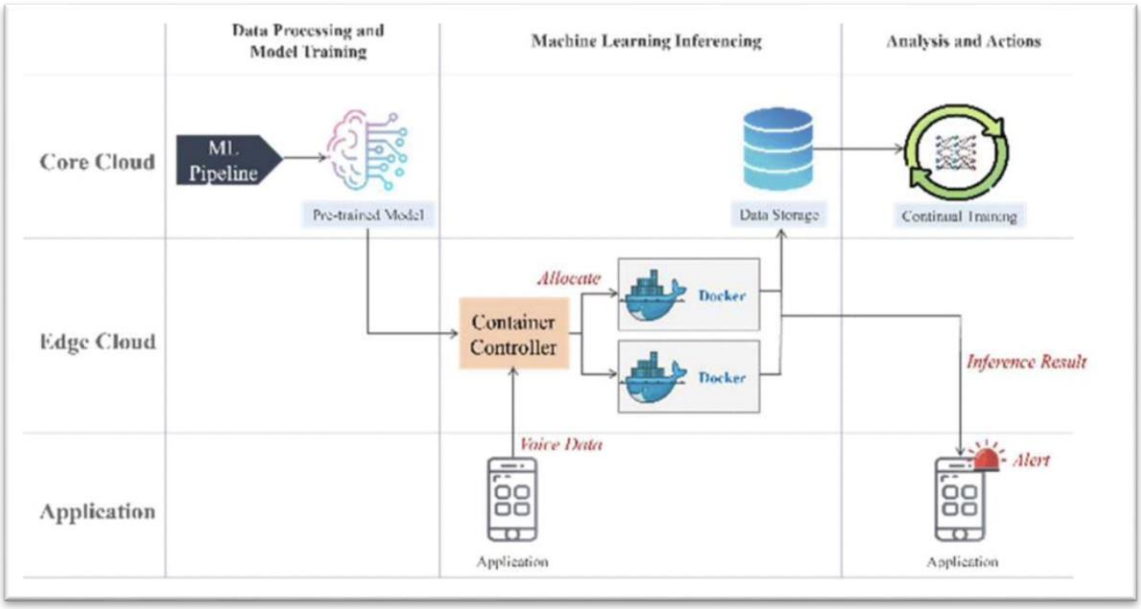
在這個模型中，我們結合多階段訓練和多數據集的使用，達到了較高的偽造語音識別準確率。對於測試數據集的預測結果如下:

數據集	ASVspoof	FakeAVCeleb	Common Voice+WaveFake
Accuracy	92.95%	95.45%。	98.35%
False Positive Rate (FPR)	0.0723	0.0704	0.0015
False Negative Rate (FNR)	0.0223	0.0328	0.0292

(3).即時語音判斷

在即時語音判斷方面，透過將 App 結合現有的電話語音錄製軟體 Cube ACR，實現了對通話錄音的自動化處理。當用戶進行通話時，Cube ACR 會自動錄製整段通話，並在通話結束後生成相應的語音檔。該應用程式透過腳本自動偵測 Cube ACR 的語音資料夾是否產生新的語音檔，一旦檢測到新的檔案，這些語音檔將立即被傳送至雲端，並進行以上兩個模型的分析。

如下圖所示，當新語音檔生成後，語音數據會通過容器控制器（Container Controller）分配至 Docker 容器進行處理，經過機器學習模型的推理後，若偵測到詐騙風險，透過簡訊功能，發出警示通知並將模型預測結果回傳給用戶，提醒該通話有可能涉及詐騙。



此系統可以擴展至更多的語音平台，並引入更精細的語音分析技術，進一步提升詐騙風險預測的精度和反應速度。同時，也可以考慮在 App 中開發自己的錄音功能，直接透過 App 錄製通話或語音片段，並即時分析其內容。這樣不僅可以擺脫對第三方錄音軟體的依賴，還能整合更多自定義的語音處理功能，為使用者提供更便捷且安全的詐騙防護體驗。

為了展示該技術的可行性，我們對通話語音內容「哈囉，你好嗎」進行了即時詐騙辨識。透過 App 的自動化操作流程，系統即時分析語音並判斷潛在的詐騙風險，隨後將結果以簡訊形式發送給用戶。傳回的簡訊內容如下所示：

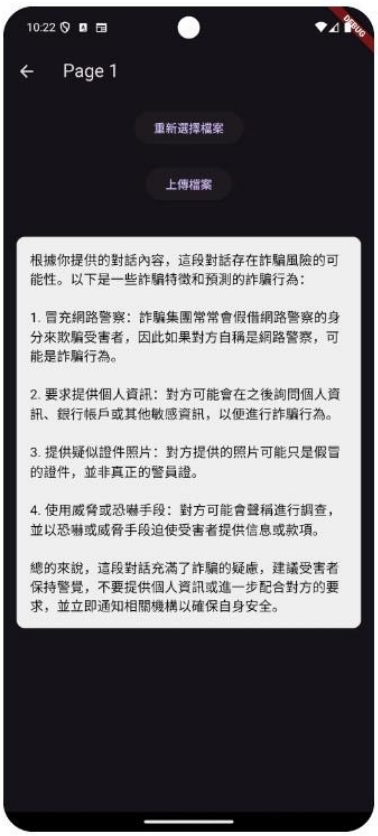
2. 圖片及文字的解決方法

詐騙手法不僅局限於語音或通訊內容，許多詐騙者還會透過圖像和文字進行欺騙，特別是在封閉式社交平台和網頁廣告中。為了提供更全面的防護，我們設計了一套針對圖片與文字詐騙的解決方案，讓使用者能夠即時判斷潛在的詐騙風險。例如，當使用者從 Line 等通訊軟體上傳對話截圖時，我們的系統結合了光學字符識別(OCR)技術，能自動提取圖片中的文字內容，並依據這些文字的上下文通過串接 OpenAI 的 GPT-3.5-turbo 模型進行深度分析，捕捉詐騙者對話的邏輯及其逐步引導受害者的過程。

此外，針對一頁式廣告的分析，系統也具備強大的檢測能力。根據台北市衛生局對詐騙一頁式廣告的七大特點，包括「強調貨到付款；售價明顯低於市場行情；標榜 7 天鑑賞期，不滿意可退費；限時限量促銷，活動長期倒數計時；網址拼音奇特；出現大陸用語或簡體字；未提供實體聯絡電話及地址，僅有電子信箱或通訊軟體帳號」等，系統能夠根據這些特點精確識別圖片與文字中的這些特徵，進行全面的風險評估，從而判斷是否存在誤導或詐騙行為，進一步提高防護的全面性。

系統的優勢在於不僅能夠檢測當下對話或廣告中的詐騙風險，還能夠通過對內容的深度分析，預測詐騙者接下來可能引導使用者進行的行為。系統會根據對話或廣告內容的特徵進行風險評估，並推測詐騙者的下一步策略，例如可能要求提供敏感的個人資訊、銀行帳戶號碼，或誘導用戶進行金錢轉帳等具體行為。這樣的預測能力不僅增強了系統的即時防護效果，還能提前警示使用者避免進一步落入詐騙陷阱。

以下是在安卓模擬器中模擬 App 對於 Line 對話截圖的結果：



透過該 App，不僅能有效應對多變的詐騙手法，還可以針對不同年齡層及常見的詐騙類型進行針對性保護。系統會在本地端儲存與用戶年齡和常偵測的詐騙類型相關的數據，並不定期發送警語提醒用戶，幫助他們保持警覺，特別是在易受攻擊的時期或面對相似的詐騙模式時。這種定期提醒機制不僅增強了 App 的防護效果，還能長期幫助用戶提升防詐騙意識。

三、 實踐策略

(一) 數據收集與管理策略

1. 數據存儲基礎設施
   系統需要配置高效且安全的數據存儲基礎設施，確保用戶數據能夠長期保存，並滿足高效存取的需求。
2. 數據備份系統
   需要投入經費購置高性能的伺服器和數據備份系統，確保在任何數據丟失或崩潰情況下能夠快速恢復。
3. 數據匿名化技術
   結合數據匿名化技術，確保數據處理過程符合隱私法規，進一步保護用戶隱私。
4. 專業技術支援
   需要專業的數據安全專家及 IT 團隊進行技術支援，確保數據存儲及處理過程中的安全性和合規性，並進行持續的系統維護。

(二) 技術開發與優化策略

1. 系統壓力測試
   系統需要投入資源進行多次系統壓力測試，以應對高流量下伺服器負載的風險，確保在高峰期仍能穩定運行。
2. 技術開發人員配置
   配置足夠的技術開發人員來進行模擬測試，持續優化系統性能並解決在測試過程中出現的問題。
3. 風險管理與應急計劃
   設立風險管理與應急計劃，確保系統在需求高峰期或出現意外情況時能夠迅速反應，保持穩定運行。
4. 定期系統維護
   技術團隊需定期對系統進行維護，並在系統運行過程中不斷進行性能優化，以應對未來的擴展需求。

(三) 使用者體驗與互動策略

1. 用戶界面與交互設計
   設計清晰的數據透明度與授權選項，讓使用者可以掌控個人數據的使用方式，增強對系統的信任。
2. 專業設計團隊支持
   依賴專業的設計團隊，根據使用者的回饋不斷改進操作體驗，確保界面易於使用且滿足用戶需求。
3. 行為模式分析展示
   向用戶展示系統的行為模式分析功能，讓他們了解系統如何預測和防範詐騙行為，進一步增強使用信心。

(四) 實時防護與通知策略

1. 即時風險偵測與通知系統
   系統需具備即時詐騙風險偵測能力，並結合多層次的風險提示，根據不同級別的風險進行快速反應。
2. 應急計劃設計
   設計針對不同風險級別的應對方案，確保系統能迅速處理從低風險到高風險的情況。
3. 與公安機構合作
   在發現高風險情況時，系統需與公安機構合作，提供即時聯繫和支援渠道，幫助用戶處理潛在的詐騙風險。
4. 運維團隊支持
   安排專業的運維團隊進行 24 小時在線系統監控，確保在任何緊急情況下都能立即響應並解決問題，保持系統穩定運行。

(五) 合作推廣與市場擴展策略

1. 多平台整合

利用模組化設計，確保系統能快速適應不同平台需求，並支持在多種環境中靈活運行。

2. API 整合

將系統通過 API 整合至社交媒體、通訊工具或電子商務平台，擴大其在市場中的應用範圍。

3. 建立合作伙伴關係

與相關企業建立合作伙伴關係，促進系統的推廣與應用，形成雙贏的合作模式。

4. 多語言與本地化適應

投資於多語言支持和本地化適應，確保系統能夠應對國際市場的需求，在全球範圍內推廣並獲得更大成功。

(六) 隱私與合規性策略

1. 法律合規性

本 App 的隱私保護及數據處理具備法律可行性，並符合相關法規要求。首先，依據《個人資料保護法》(Personal Data Protection Act, PDPA)，本 App 在收集、處理和使用個人數據時，必須強調隱私政策的透明度，清楚告知用戶其數據的使用方式，並保障用戶有權查詢、更正或刪除其個人數據。此外，本 App 處理通訊數據，根據《通訊保障及監察法》的規定，應確保用戶在語音和文字分析的過程中已獲得明確的知情同意，並保護其通訊隱私。最後，根據《電腦處理個人資料保護辦法》，本 App 將採取數據加密及其他技術手段，確保數據在傳輸和儲存過程中的安全性，防止資料外洩。透過遵循這些法律規範，本 App 在保障使用者隱私的同時，能夠合法進行詐騙偵測的數據分析，並滿足台灣法規的要求。

2. 合法性與安全性宣傳

透過公共媒體進行宣傳，強調系統的合法性與數據安全措施，讓公眾瞭解其在隱私保護方面的優勢。

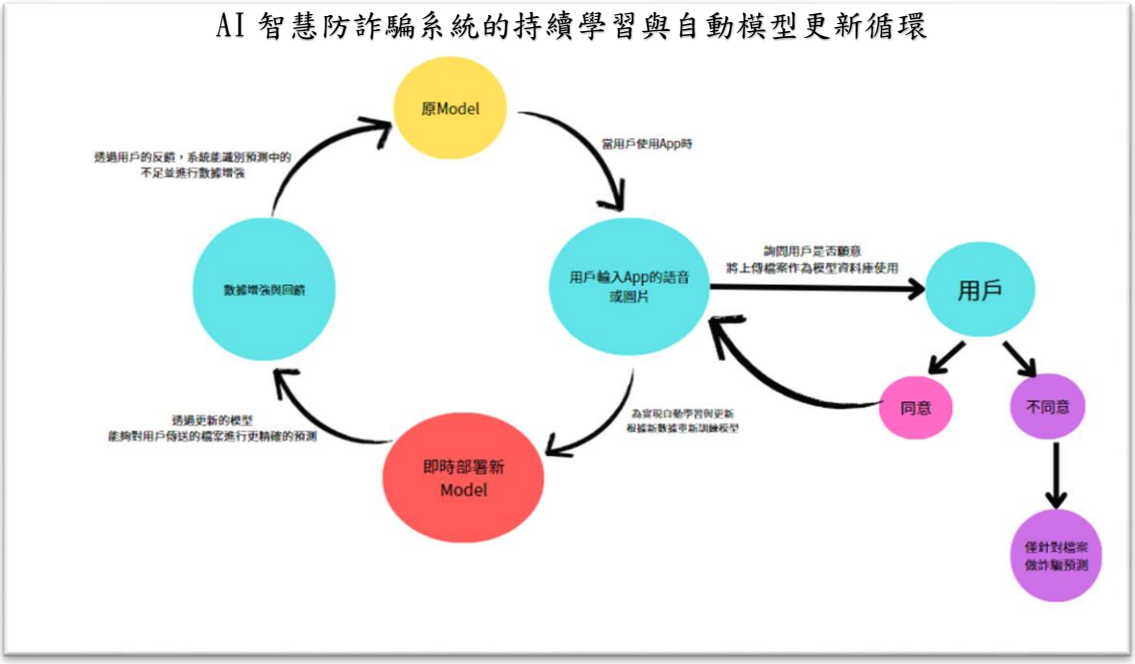
(七) 持續改進與未來發展策略

1. 自動模型重訓練機制

本 App 將設立自動模型重訓練機制，根據用戶偵測到的數據持續擴充資料庫，並在規定的時間內進行即時重訓練，確保系統能自動學習和適應不斷變化的詐騙手法，實現模型的迭代更新。此外，透過數據標註機制，系統能夠精確標記偵測過程中的詐騙案例，進一步提升模型的準確性，確保其在面對新型詐騙時具備強大的識別與處理能力。這些機制的結合，將使系統持續優化，保持對新興詐騙行為的敏銳反應。

2. 快速響應機制

建立快速響應機制，當新型詐騙技術出現時，系統能夠迅速更新並保護使用者免受新興詐騙手段的影響。如下圖展示了 AI 智慧防詐騙系統的自動化模型優化流程。當用戶上傳語音或圖片後，系統即時進行預測，並返回結果。如果用戶同意，數據將被用作模型訓練資源，整合進資料庫進行標註與增強，推動下一輪模型重訓練。新模型部署後，系統能夠更精確地預測後續詐騙風險，確保系統持續學習並提升防詐能力。



四、 結語

透過該 App 不僅能夠有效應對當前多變的詐騙手法，還具備高度的靈活性和可擴展性，為更廣泛的用戶群體提供了實時且精確的防詐騙保護。這款 App 的推出不僅有助於提升個人和社會對詐騙的防範意識，也能減少潛在的財產損失和心理負擔，從而對整體社會環境帶來積極的改善。隨著技術的持續發展與市場的推廣，這一解決方案將在防止詐騙行為、保障公眾安全和提升社會信任方面發揮重要作用，為打造更加安全的數位社會提供了可行的路徑。