中華民國第 65 屆中小學科學展覽會

作品說明書

高級中等學校組　電腦與資訊學科

052501

利用 ChatGPT 協助辨識詐騙簡訊及網頁

學校名稱： 新北市立丹鳳高級中學

作者：	指導老師：
高一 高浚誠 高一 徐翊庭 高二 陳宥維	柳心恬

關鍵詞： 人工智慧、詐騙、內容辨識

摘要

近年來隨著資訊科技發展，詐騙行為日益猖獗。儘管政府已建置防詐網站與專線提供人工審查，但面對大量詐騙資訊，其效率仍顯不足。為降低人工審查負擔，本研究探討提示工程技術（Prompt engineering），評估其是否可提升 GPT 模型對圖像與語音詐騙內容的辨識準確率。研究設計採用三種不同提示策略，分別應用於圖像與語音資料中，並比較其判斷正確數以分析準確率差異。結果顯示 Chain-of-Thought Prompting 在兩種媒介資料中表現皆優於其餘提示工程模型，顯示良好判斷效果。本研究基於 Chain-of-Thought Prompting 模型開發互動式網頁程式讓民眾可立馬使用網頁判別可疑的圖片、音訊，展現應用於基礎防詐之可行性，亦可為後續防詐系統提供設計參考。

壹、前言

一、研究動機

自資訊時代開啟，利用偽造文章、誘導訊息、虛假影音等進行牟利的詐騙行為愈加猖獗，根據內政部統計查詢網， 詐欺案件數由 2015 年的 21,172 件增加到 2023 年的 37,984 件，可見國家雖有成立內政部警政署 165 全民防騙網，因詐欺案件量過大，可見人工處理並不足以應對詐騙的遞增。現今資訊科技發達迅速，特別是現今 AIGC（Artificial Intelligence Generated Content）議題廣受討論，普通人無法自行製作程式來辨別詐騙圖片，所以本實驗想利用擁有龐大資料庫的 GPT 模型探討 AI 在詐騙上之應用，但在龐大且雜亂的資料量下，單純的命令可能無法使 GPT 模型無法精確判別而給出錯誤答案，若是能解決此問題並進行反詐應用程式的開發，或許能夠達成減少民眾受詐騙之目標。

二、文獻回顧

詐騙是對他人施以詐術，以虛構、扭曲的言語或是其他行為動作，而使其陷於錯誤，使他人對財產做出處分、移轉、交付等等動作，造成被害人財產上損害刑法第 339 條（普通詐欺罪）。

面對永無止盡的詐騙資訊，早期利用人工資料庫製作模型進行防範詐騙之工程，嚴浤元（2024）製作偽造語音辨識 APP，民眾在打電話時會自動偵測來電是否為合成語音再給予警告，許琇媛（2024）將來電者的語音即時轉換成文字，利用人工資料庫製作之模型自動偵測對話內容是否包含高風險的詐騙關鍵字，並給予警示，施瓊雯（2024）利用內政部警政署 165 全民防騙網，結合文字探勘技術，挖掘詐騙行為的關鍵字，旨在減少詐騙造成的損害。李承軒（2024）主要探討分析簡訊的號碼來源、網址連結及文本用語三個面向，透過隨機森林法建立模型識別台灣境內的詐騙簡訊。

但普通人大部分無法自行製作程式來辨識詐騙，所以我們利用提示工程技術（Prompt engineering）使用 GPT 模型，並使用機器學習中非監督式學習模式進行實驗。ChatGPT 是 OpenAI 公司基於 GPT 模型所開發的 AI 聊天軟體，而 ChatGPT 是使用 Transformer 架構所開發出的大型語言模型（Large Language Model，LLM）以下皆用 LLM 表示，也是屬 Pre-trained Model 的一種羅光志（2023）。Prompt Tuning 是可以讓 LLM 更容易泛化到下游任務的一種技術陳峰楷（2024）。蘇宥蓁等（2024）製作關於我與 ChatGPT 成為一家人的那件事

之實驗，針對提示工程對於GPT模型在家電上的運用進行討論，取得有效控制家電結果，證明 GPT 模型應用之可能性。

本實驗從 Boonstra, L.（2025）為生成式 AI 的提示工程給參考，共分為三種提示詞方向：直接提示（Direct prompting）、少量樣本提示（Few-shot prompting）、思維鏈提示（Chain-of-thought prompting）。根據 Vandierendonck, H., Rul, S., & De Bosschere, K.（2009），Accuracy、Precision、Recall、F1 score 是評估檢索成效的一項指標，評估資訊檢索系統的成效時，常將系統的判斷與人工的判斷做交叉分析，使用 Accuracy、Precision、Recall、F1 score 指標可以將我們判斷結果進行量化分析，雖然 Accuracy、Precision、Recall、F1 score 指標常常被當作實驗的一種判斷要素之一，但 He, H., & Garcia, E. A.【2009】提到，當樣本中有 95%的負面樣本時，若模型將樣本全部判別為負面時，則準確率仍為 95%。說明當樣本數極度不平衡，數據無法完全展現模型問題，所以 He, H., & Garcia, E. A.在實驗中增加了 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve）來驗證實驗結果，本實驗也將採用 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve）兩圖表判斷實驗的完整性及問題，避免後續詐騙樣本數不平衡等問題導致 Accuracy、Precision、Recall 指標無法當作判斷指標，利用 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve）也可以從更多方面來判斷模型問題，從 F1-score 以外數據判斷 GPT 模型判別結果。

三、研究目的

（一）探討不同 prompt 設計對 GPT 模型辨識詐騙圖像、語音的準確率影響，實驗出最佳範本。

（二）設計應用程式整合提示工程技術與GPT模型，希望能達成遏止惡意詐騙、減少民眾受詐騙等目標。

（三）提供未來擴展研究與研發程式樣本之參考，減少國人受騙之風險。

四、研究假說

本研究期待透過提示工程技術，能有效提升以GPT模型為基礎所開發之詐騙判別模型在詐騙判別上的準確性，並增進使用者對可疑資訊的理解與應對能力。本研究限制使用同樣的電腦硬體規格，軟體、模組版本、實驗樣本不變之情況下，將圖片或音檔匯入至本實驗製作之程式，若在給予更多和詳細 Prompt 的情況下，則預期圖像或語音資料之詐騙判斷準確率應得到相對的提升。

貳、 研究設備及器材

一、實驗硬體設備：

（一）CPU： AMD Ryzen 5 7500F 6-Core Processor

（二）GPU：NVIDIA GeForce RTX 4060（記憶體：32GB）

二、 實驗軟體設備：

（一）Python 程式語言

Python 為本研究主要環境，是個免費開源的程式語言，其優點為易於學習與開發，用戶只需在裝置上安裝好對應的環境即可使用，Python 也支持用戶自行開發模組，用戶也可以使用 Python Package Index 函式庫安裝第三方模組，例如 NumPy 模組可讓用戶執行較複雜的數學方程式、OpenCV 可讓用戶執行繪圖作業 。

（二）Chat GPT 4.1

Chat GPT 採用深度學習的技術來判斷圖片內的特徵以及內容，其中以積卷神經網路（CNN）為主要判斷方式，這種技術能夠自動地從圖像中提取不同層次的特徵，從簡單的像素模式到更高層次的形狀、顏色、物體等，這些技術背後的核心就是大量的數據訓練，讓模型能夠學會如何識別不同的模式和特徵。本實驗模型皆以該大型語言模型為基底設計，以下簡稱「模型」。

（三） Whisper AI v20240930

謝岳哲（2024）表示：語音辨識（Speech Recognition）是將語音訊號轉換成可辨識的文本或指令。Whisper AI 是由 Open AI 所訓練的語音辨識模型，是一個使用 Transformer 網路架構。可執行語音翻譯和語言識別等任務。

參、研究過程或方法

一、架構

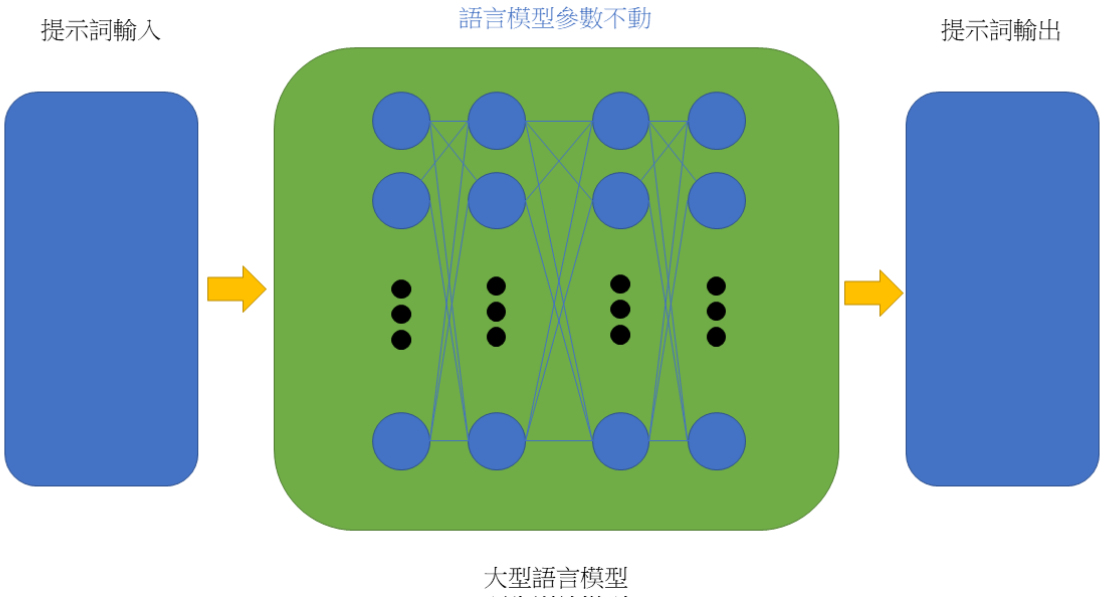
（一）提示工程 （Prompt Engineering）

當我們使用大型語言模型（Large Language Models， LLMs）進行任務時，模型在訓練階段所學得的參數於推論階段已固定，使用者無法透過傳統機器學習方式如微調來應對新的任務需求。因此我們必須改為設計不同的提示工程（ Prompt

Engineering），引導模型產生符合需求的回應。

提示工程（Prompt Engineering）是一種在不重新訓練模型的情況下，僅透過設計文字輸入以操控模型輸出的技術，具有彈性高、計算資源低、應用廣泛等優勢。此方法尤其適用於應對多樣性、變動性高的輸入情境，如本實驗所實驗的詐騙案例判斷。

圖 1：提示工程原理

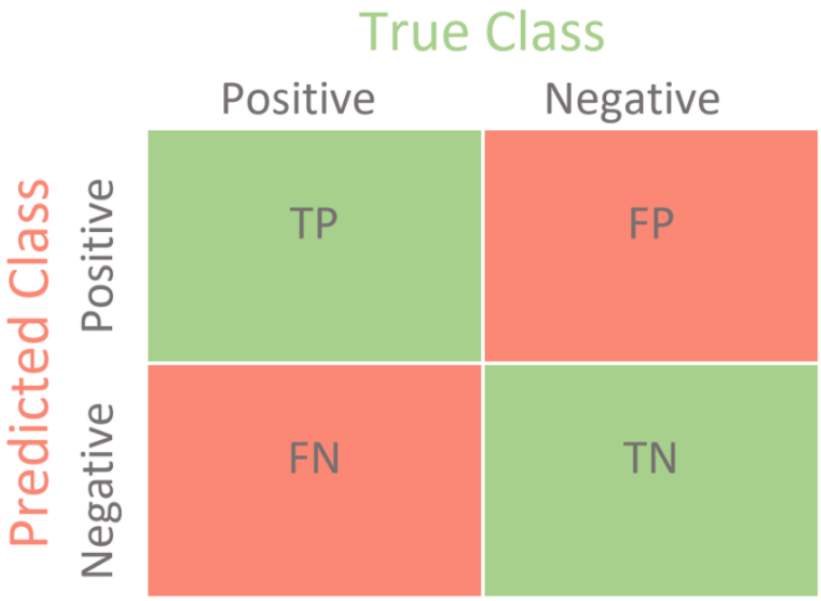


（圖片來源：本研究自製）

（二）混淆矩陣

本實驗針對音檔與圖片資料分別進行測試，並使用混淆矩陣（Confusion Matrix）進行模型表現評估。混淆矩陣是一種常見於二元分類問題中的評估指標，如圖 2 所示將能夠透過模型預測結果與實際標註結果的比較，產出四種分類情況：

圖 2：混淆矩陣圖表範例



（圖片來源：Confusion Matrix Zohreh Karimi（2021））

將混淆矩陣四項指標進行解釋：

1.TP（True Positive）：預測為正面，模型實際也判斷為正面，在本實驗中即人工預測結果為詐騙，且模型也判斷為詐騙。

2.TN（True Negative）：預測為負面，模型實際也判斷為反面，在本實驗中即人工預測結果為非詐騙，且模型也判斷為非詐騙。

3.FP（False Positive）：預測為正面，但模型實際判斷為負面，在本次實驗中即人工預測結果為詐騙，但模型判斷結果為非詐騙。

4.FN（False Negative）：預測為負面，但模型實際判斷正面，在本次實驗中及人工預測結果為詐騙，但模型判斷結果為詐騙。

依據上述四種情況，可進一步計算出混淆矩陣常見指標，如圖 3 所示：

圖 3：分類模型性能評估指標對照表及公式圖

指標(metrics)	用途	公式
準確率(accuracy)	衡量模型預測正確的比重	(TP+TN)/(TP+TN+FP+FN)
精確率(precision)	在預測為是的分類中，預測正確的比重	TP/(TP+FP)
召回率(Recall)	在預測正確的分類中，預測為是的比重	TP/(TP+FN)
F1-Score	同時考慮精確率與召回率(能更好的反應模型)	2*precision*recall/(precision+recall)

（圖片來源：本研究自製）

本次實驗中 Precision 的涵義為模型判定為詐騙的樣本數中，有多少是人工預測為詐騙的比例，當 Precision 越高，代表模型將愈少的非詐騙樣本判別為詐騙樣本。Recall 的涵義為人工預測為詐騙的樣本數中，有多少被模型判定詐騙的比例，當 Recall 越高，代表模型在判別詐騙樣本的準確度很高。F1 score 為 Precision 和 Recall 的調和平均，能綜合衡量模型在辨識詐騙時的準確性，避免 Precision 和 Recall 兩項數值偏差過大不易判斷實驗結果的情況，從兩折之中算出的數值可對模型有更全面判斷，因此本研究將以 F1-score 作為不同提示工程模型下的主要比較指標，能全面地反映模型在詐騙判斷任務上的實際效能。

ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve）經由混淆矩陣數值重新計算所做出的圖表，避免樣本數極度不平衡時導致對數值分析判斷錯誤。ROC 曲線（Receiver Operating Characteristic Curve）是以 False Positive Rate（FPR）為橫軸，True Positive Rate（TPR）為縱軸，其中 FPR＝FP/（FP＋TN），TPR＝TP/（TP＋ FN），根據 False Positive Rate（FPR）與 True Positive Rate（TPR）數值，模型數值在 ROC 曲線（Receiver Operating Characteristic Curve）左上方時代表模型判別詐騙結果越好。PR 曲線（Precision-Recall Curve）是以 Recall 為橫軸，Precision 為縱軸，根據 Recall 與 Precision 數值，模型數值在 PR 曲線（Precision-Recall Curve）右上方時代表模型判別詐騙效能越好。

二、實驗流程

本實驗將分成判別圖片與音訊，兩部分進行判別詐騙的實驗，其中音訊將以 Whisper AI 進行語音識別（Automatic Speech Recognition，ASR），將語音內容轉寫為文字，再輸入模型判斷。

（一）提示工程模型設計

本研究依據 Boonstra, L.（2025）所提出之提示設計邏輯，構建三種提示工程模型進行對比實驗：

1. 直接提示模型（Direct Prompting），以下簡稱為 Direct Prompting

本實驗將 Direct Prompting 模擬成一般使用者操作習慣，僅使用簡短的詢問語句，不提供範例或詐騙定義，模擬一般使用者的直觀操作情境。Direct Prompting 將作為對照組和 Few-shot prompting、Chain-of-Thought Prompting 進行對比。

2. 少量樣本提示模型（Few-shot prompting），以下簡稱為 Few-shot prompting

本實驗 Few-shot prompting 在 Direct Prompting 的基礎上加入詐騙與非詐騙的範例到提示詞中，期望透過示範學習提升判斷敏捷度與精確度。

3. 思維鏈提示模型（Chain-of-Thought Prompting），以下簡稱為 Chain-of-Thought Prompting

本實驗 Chain-of-Thought Prompting 將分成圖片、音檔兩種提詞，進一步嵌入多階段推理指引與結構化規則，從李承軒（2024）、許琇媛（2024）的論文中提取詐騙行為定義、常見詐騙網址、詐騙關鍵字等作為我們提示詞內容之，期望模型達成快速尋找詐騙特徵、判別等能力。

（1）圖片提示詞：本研究圖片中增加了李承軒（2024）論文中對於詐騙行為的定義、常見詐騙連結手法、不安全網址等，讓模型從多方面判斷圖片的情況下也可以尋找細微特徵減少判別時間過長的問題。

（2）音檔提示詞：本研究保留李承軒（2024）論文中詐

騙行為定義外，增加了許琇媛（2024）論文中詐騙關鍵字，在本次實驗音檔將使用 Whisper AI 轉換為文字再交予模型進行判斷，所以相較於圖片音檔模型只能從文句中進行判別，而許琇媛（2024）是將來電語音轉換成文字進行判別也取得判別良好的結果，故本實驗預期抓取關鍵字對 GPT 模型增加判別能力。

本研究設計思維鏈提示模型摘要（Chain-of-Thought Prompting），如表 1 所呈現之格式項目：

表 1：思維鏈提示格式內容摘要

項目	內容
格式開頭	簡短詢問語句
判斷結果選項	詐騙、非詐騙
建議	提供使用者行動建議
判斷依據	是否虛假消息、是否符合普通詐欺罪之定義等
詐騙定義	內容真實性、意圖性、法律定義
高風險	具有潛在詐騙可能，以保護使用者優先進行警告
判斷邏輯	延伸自詐騙定義，包含下列項目： 身分判斷邏輯 品牌資訊是否真實 官方文件特徵 網址、連結確認 個資填寫、金錢轉移 情境判斷 是否為廣告、教育宣導、 詐騙常用關鍵字

（資料來源：本研究自製）

（二）提示策略演化與設計思維：

本實驗所設計的 Direct Prompting 與 Few-shot prompting 為基礎型設計，前者模擬一般使用者詢問基本問題，不給予任何判別提示詞；後者則透過少量範例加以補充模型，兩者皆可提供初步判斷，本實驗預期此二者雖有一定的基礎辨識功能，但在尋找詐騙特徵與推理深度上可能仍有限，可能導致模型判斷正確性上有所欠缺。

Chain-of-Thought Prompting 是在前兩者基礎上發展而來，用提示疊代策略（Prompt iteration strategies），原有架構中加入明確的判斷定義、邏輯，面對複雜或

語境模糊的訊息時能維持穩定可解釋的推理過程。本實驗預期能進一步給予詐騙定義、細化出判別條件，使用者在面對複雜、模糊的詐騙訊息時，仍有清晰穩定的邏輯推理。

（三）詐騙與非詐騙樣本

資料來源與標註分為圖片與音訊

1.圖片樣本共 55 份，含詐騙影像 36 份與非詐騙影像 19 份。

2.音訊樣本共 25 份，含詐騙音檔 13 份與非詐騙音檔 12 份。

3.素材主要取自社群媒體貼文、新聞報導、事實查核中心資料庫，再經由本研究人員進行註記已確認是否為詐騙。

（四）系統整體架構說明

本研究之跨平臺詐騙判斷系統共有四層式架構，由上而下分為客戶端層、伺服器端層、服務層與執行環境層，本研究整合所需功能與自行研製的提示模型，開發出一套同時支援電腦本機端與行動手機端的應用軟體。

圖 3：系統整體圖

（圖片來源：本研究自製）

1.客戶端層

使用者可透過行動裝置與電腦兩種，可執行拍照、錄音及檔案上傳等操作，並即時接收系統回傳之詐騙判斷結果。

2.伺服器端層

本層整合四項後端服務。

（1）ChatGPT：負責圖文詐騙訊息之語意分析與推論。

（2）Whisper AI：將上傳音檔進行語音辨識後輸出文字。

（3）Discord Bot：提供多人協作或監控通報之通訊管道。

（4）Node.js：串接各 AI 模組並對外提供 RESTful API，處理請求排程與權限驗證。

3.服務層

系統核心功能模組，涵蓋拍照、錄音、檔案上傳與結果儲存與輸出，各模組皆以按鈕觸發機制串聯：

（1）使用者觸發拍照或錄音。

（2）客戶端將檔案上傳至伺服器。

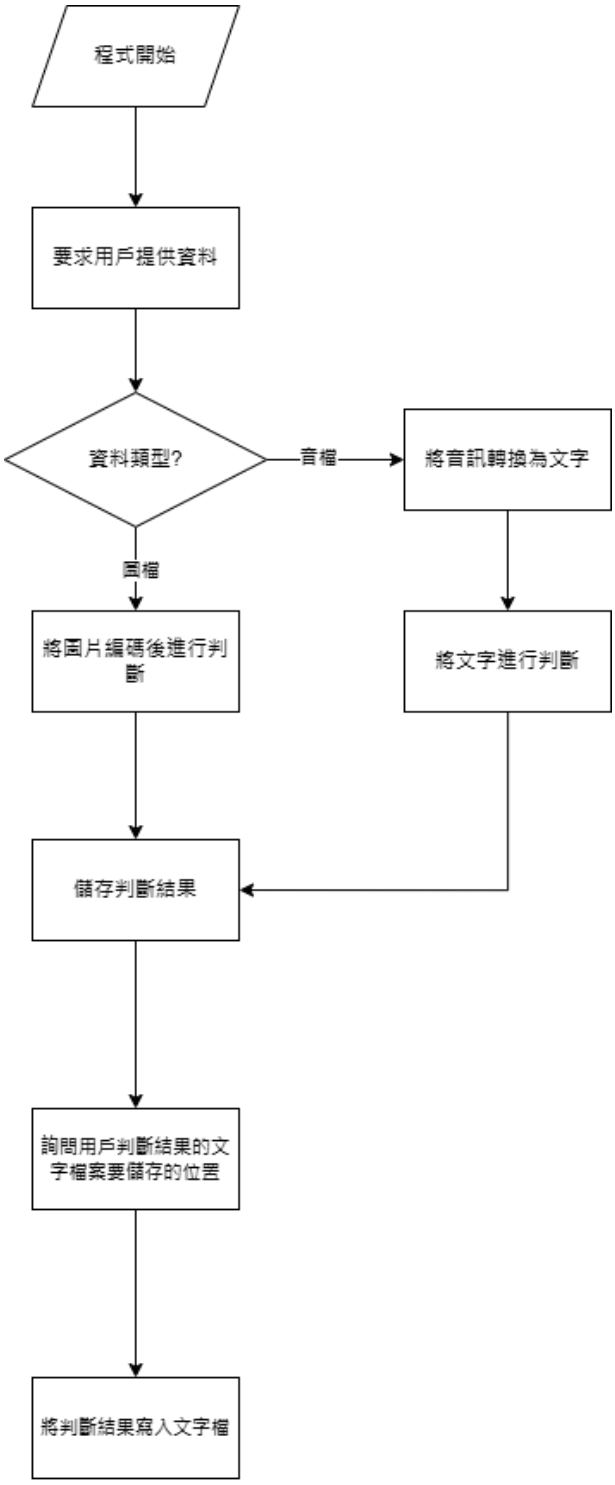
（3）伺服器完成推論後回寫判斷結果。

4.執行環境層

（1）Python： ChatGPT 與 Whisper AI 之呼叫、資料前處理與後處理。

（2）JavaScript：配合 Node.js 建立 API 服務與 Discord Bot，並支援 Web 端前端互動。

圖 4：實驗流程圖



（圖片來源自本研究自製）

本研究程式流程模型會先讀取檔案類型來，若為音檔則將音訊轉為文字並開始判別，若為圖片將圖檔編碼並進行判別，最後回傳判別結果。

圖 5：圖、音檔辨識函式圖

```python
import requests
import base64
import time

API_KEY = ""
GPT_URL = "https://api.openai.com/v1/responses"
WHISPER_URL = "https://api.openai.com/v1/audio/transcriptions"

Model = "gpt-4.1"

def encode_image(image_path):
    with open(image_path, "rb") as image_file:
        return base64.b64encode(image_file.read()).decode('utf-8')

def AudioDetect(AudioFile, PromptTextFile, IsMutiFile):
    start_time = time.time()
    files = {
        "file": open(AudioFile, "rb"),
        "model": (None, "whisper-1"),
        }
    AudioResponse = requests.post(WHISPER_URL, headers={"Authorization": f"Bearer {API_KEY}"}, files=files)
    with open(PromptTextFile, "r", encoding="utf-8") as PromptText:
        Response = requests.post(
            GPT_URL,
            headers = {
                "Authorization": f"Bearer {API_KEY}",
                "Content-Type": "application/json"},

            json = {
                    "model": Model,
                    "tools": [{ "type": "web_search_preview" }],
                    "input": f"{PromptText.read()}\n@{AudioResponse.json()["text"]}",
                    "temperature": 0.0
                }
        )
    if IsMutiFile == True:
        time.sleep(8)
    Response.raise_for_status()
    end_time = time.time()
    execution_time = end_time - start_time
    if Response.status_code == 200:
        return Response.json()["output"][0]["content"][0]["text"], execution_time
    else:
        print(f"Error: {Response.status_code}")
        return None

def PhotoDetect(PhotoFile, PromptTextFile, IsMutiFile):
    start_time = time.time()
    base64_image = encode_image(PhotoFile)

    with open(PromptTextFile, "r", encoding="utf-8") as PromptText:
        Response = requests.post(
            GPT_URL,
            headers = {
                "Authorization": f"Bearer {API_KEY}",
                "Content-Type": "application/json"},

            json = {
                    "model": Model,
                    "tools": [{ "type": "web_search_preview" }],
                    "input": [
                        {
                            "role": "user",
                            "content": [
                            {"type": "input_text", "text": PromptText.read()},
                            {
                                "type": "input_image",
                                "image_url": f"{base64_image}"
                            }
                            ]
                        }
                    ],
                    "temperature": 0.0
                }
        )
    if IsMutiFile == True:
        time.sleep(8)
    Response.raise_for_status()
    end_time = time.time()
    execution_time = end_time - start_time
    if Response.status_code == 200:
        return Response.json()["output"][0]["content"][0]["text"], execution_time
    else:
        print(f"Error: {Response.status_code}")
        return None
```

（圖片來源：本研究自製）

分別定義圖、音判別函式，並給予三個不同的參數，分別為判斷檔案路徑、Prompt 檔案路徑、是否為多個檔案布林值，透過 Request 將規定的標頭與 API key 發布至 ChatGPT 的網址，再將回傳值儲存至變數，判斷是否為多個檔案參數值，若為多個檔案參數值則等待八秒避免觸發 Token Per Minute 錯誤 ，若以上步驟皆成功回傳代碼為（200）則將函式回傳值設為回覆內容與運行時間，若以上步驟有任何一步失敗回傳代碼不為（200） 則將函式回傳值設為無（None）。

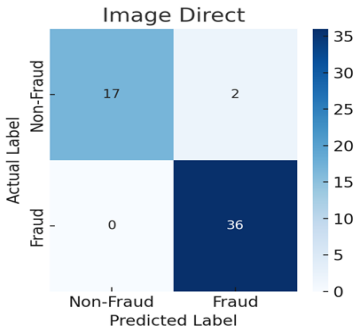
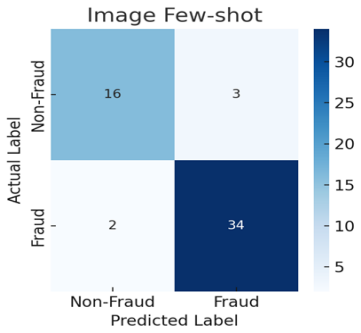
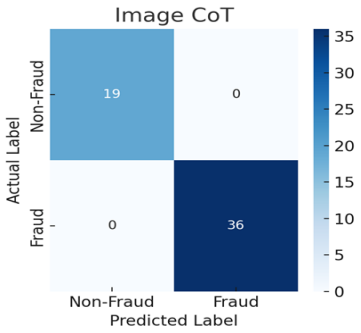
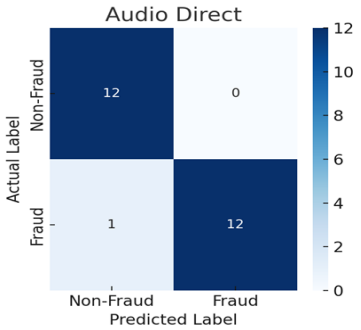
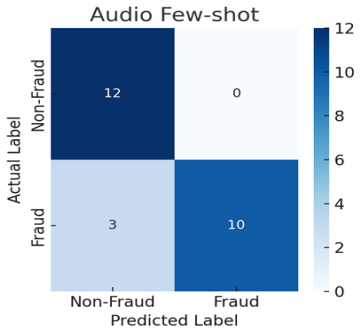
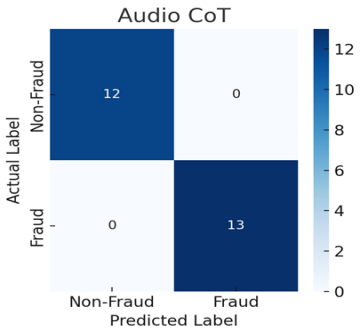
肆、研究結果

本研究實驗樣本圖片一共 55 張，其中詐騙樣本 36 張，非詐騙樣本 19 張。音訊共 25 份，其中詐騙樣本 13 份，非詐騙樣本 12 份。

製作出混淆矩陣後，本實驗利用混淆矩陣算出四項指標，並繪製出 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve），期望利用不同數據進階探討提示工程設計對模型判別詐騙準確率的影響。

一、根據圖實驗斷判結果將利用模型判斷（Predicted Label）對比人工判斷（Actual Label）分為四種情形並可根據公式計算出四項指標，並整理成以下混淆矩陣：

表 2：圖片、音訊混淆矩陣及其他推算數值

圖 6a：Direct Prompting（圖）	圖 6b：Few-shot prompting（圖）	圖 6c：Chain-of-Thought Prompting（圖）
		
Accuracy：0.963 Precision：0.947 Recall：1.0 F1-score：0.972	Accuracy：0.909 Precision：0.944 Recall：0.919 F1-score：0.931	Accuracy：1.0 Precision：1.0 Recall：1.0 F1-score：1.0
圖 6d：Direct Prompting（音）	圖 6e：Few-shot promting（音）	圖 6f：Chain-of-Thought Prompting（音）
		
Accuracy：0.960 Precision：1.0 Recall：0.923 F1-score：0.960	Accuracy：0.880 Precision：1.0 Recall：0.769 F1-score：0.870	Accuracy：1.0 Precision：1.0 Recell：1.0 F1 score：1.0

（表格與圖片來源：本實驗自行製作）

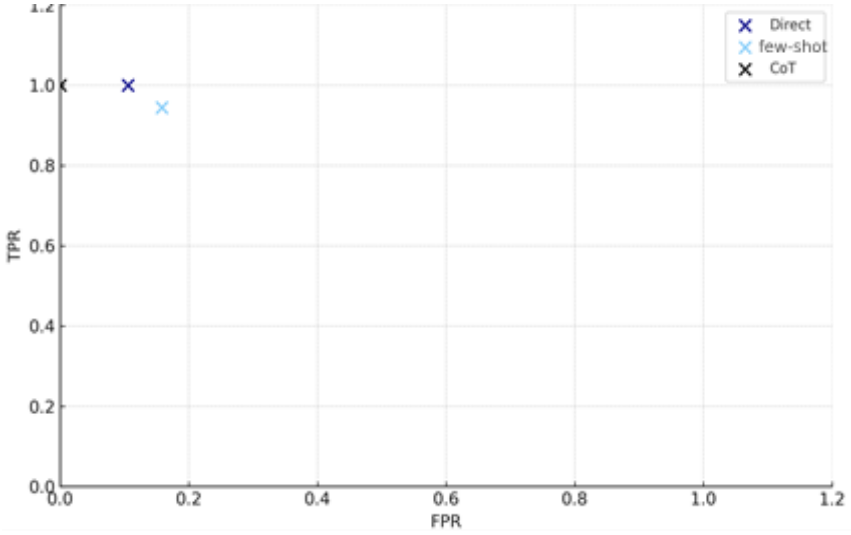
二、根據實驗結果，製作出的空間點圖和 PR 曲線（Precision-Recall Curve）的空間點圖表現每個提示工程模型在判別詐騙上的結果。

本實驗將使用 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve）但會以點圖的方式進行呈現，理由如下：

（一）基於本次實驗，因為我們提示工程模型種類較少，無法形成完整曲線，也難以觀察出模型判別表現的整體幅度，若強制畫出曲線，後面從前 3 點延伸出的曲線可能有失真或無意義的問題發生，且因為數值過少而無法計算 AUC 指標（Area Under Curve），所以未畫出曲線。

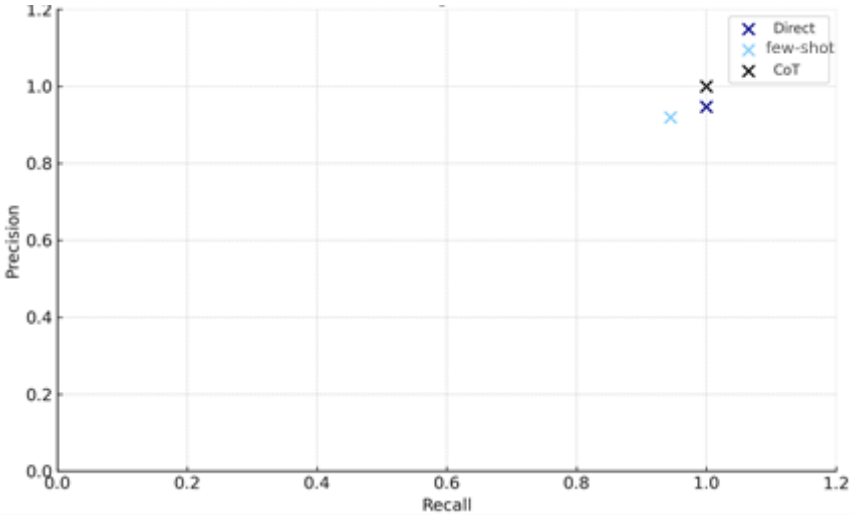
（二）本實驗雖用點圖方式呈現，但橫、縱軸還是對應 ROC 曲線（Receiver Operating Characteristic Curve）和 PR 曲線（Precision-Recall Curve），圖表仍有其意義不變，且以點圖直接呈現可清楚看出三個不同提示工程模型的判別表現，方便辨別其優劣。

圖 7：ROC（Receiver Operating Characteristic）空間點圖（圖片）



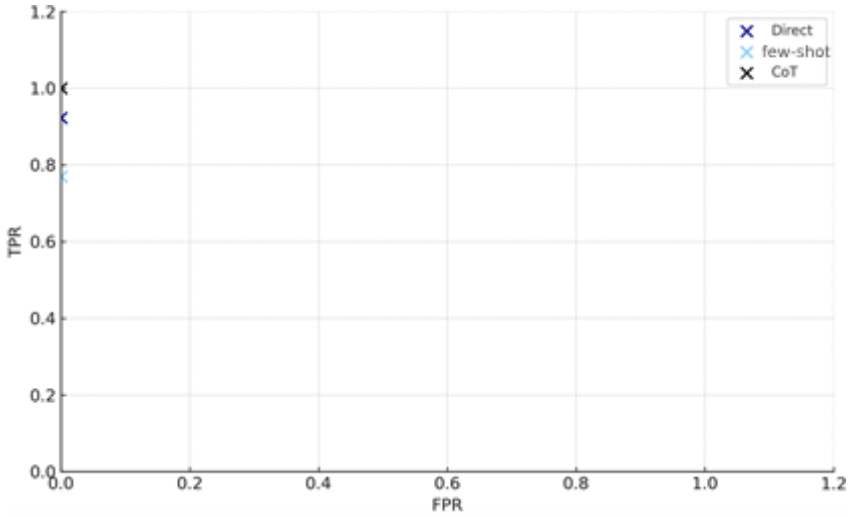
（來源：本實驗自行製作）

圖 8：PR（Precision-Recall）空間點圖（圖片）



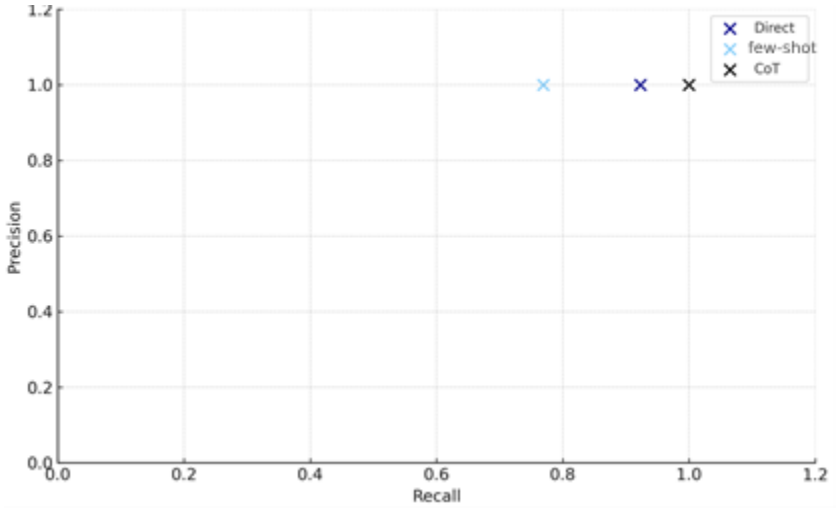
（研究來源：本實驗自行製作 ）

圖 9：ROC（Receiver Operating Characteristic）空間點圖（音訊）



（圖片來源：本實驗自行製作）

圖 10：PR（Precision-Recall）空間點圖（音訊）



（圖片來源：本實驗自行製作）

伍、討論

根據研究結果，下方將用研究結果探討提詞對 ChatGPT 判別詐騙的可能性。

一、圖片

（一）在 Direct prompting 中非詐騙已有不錯的準確率，但推測是因為沒給予提詞便讓 ChatGPT 進行判別，從判斷依據回復可以看出 GPT 模型只會根據圖片中一到兩個最大、明顯的特徵或文字進行判斷，可能導致將部分判斷成詐騙， 有誤導民眾可能，就算判斷正確回復依據也解釋不清楚，時常指出圖片內容便直接判別是否為詐騙。

（二）換成 Few-shot prompting 判斷正確數反而下降，與我們本來預期情況有些許不同，相較於 Direct prompting，Few-shot prompting 將兩個 Direct prompting 原本判斷正確的三張圖片判斷成錯誤。推測是因為案例的給予而限制 GPT 模型判別的影響，從回覆依據也可看出案例影響 GPT 模型對於引導、催促等字詞定義的理解。雖然本實驗和本來預期有所出入，但本實驗仍認為案例還是可以讓 GPT 模型增加判別、回復能力，從回覆依據發現 Few-shot prompting 的 GPT 模型相較 Few-shot prompting 能從圖片中給出更多合理的解釋、更多的細節，未來若要改善案例不佳的情況，我們可以從內政部警政署 165 打詐儀錶板等相關網站進行多去尋找案例，未來本實驗也將持續探討案例對 ChatGPT 判別能力的可能。

（三）Chain-of-thought prompting 是判斷正確率最高的，全部圖片中判斷皆正確，和本實驗本來預期結果相同，本實驗從李承軒（2024）尋找現今詐騙者如何詐騙的行為、找尋關鍵字定義，期望 GPT 模型能利用定義、前人已經確認可判別詐騙方式達到精確判別的目

的。從判斷依據中發現，判斷回復較 Direct prompting、Few-shot prompting 敘述多，回復較詳細且會給建議讓民眾知道遇到詐騙圖片時該如何防備且不傳給他人，ChatGPT 除辨別出圖片中的內容之外，還會簡潔說出內容細節，讓民眾可以略知圖片資訊。由 Chain-of-thought prompting 我們也可以更確認提詞對於 ChatGPT 的影響，回復依據常利用從李承軒（2024）中所提到定義（詐欺性、誘導性等）作為回復判斷依據，還有 GPT 模型會開始尋找圖片中是否有任何連結，並判斷連結是否安全。

（四）用 Accuracy、Precision 等數據可更明確知道提示工程對 GPT 模型判別詐騙影響。Accuracy 是可最直觀判斷提示工程效果，但我們從 Accuracy 只能知道提示工程判別正確數，無法得知如 GPT 模型將多少詐騙樣本判別成非詐騙樣本等詳細問題，所以 Precision、Recall 等數據可以將我們想知道的問題量化進行觀察，並可以有更細部判斷。從實驗數據可知，無論是 Precision、Recall、F1-score，Chain-of-thought prompting 的數值皆是最高，依序是 Direct prompting、Few-shot prompting，從 F1-score 發現提示工程給予越多的情況下 GPT 模型判別非詐騙樣本的能力提高，由圖 7、圖 8 可知不論是 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve），Chain-of-thought prompting 的判別能力最佳，依序是 Direct prompting、Few-shot prompting。

二、音訊

（一）從實驗數據發現，無論使用何種提示詞的模型在非詐騙音訊中皆能判別正確，推測是因為非詐騙音有完整敘述句，GPT 模型能理解音訊內容所述所以才能完全正確判斷，詐騙音訊相較於非詐騙音檔完整度較低，可能是導致詐騙判別正確率較低的主要原因。Few-shot prompting 的判斷正確數為三者中最低者，和圖片結果一樣，雖然 Few-shot prompting 相較於 Direct prompting 判別正確數較低，但 Few-shot prompting 在有給予案例的情況下，判斷依據較 Direct prompting 給出更多合理的解釋、更多的細節，本實驗仍認為案例可以讓 GPT 模型增加判別、回復能力。

（二）Chain-of-thought prompting 判別正確數為最高，從判斷依據發現，GPT 模型面對音訊會大量使用我們從許琇媛（2024）論文中提取的詐騙關鍵字進行判別的行為，推測是因為在我們這次判別音訊實驗中，因為利用 Whisper AI 將音檔轉換文字，在僅有文字情況下 GPT 模型並不會判別語氣、背景音等內容，不同於圖片 GPT 模型只能從文字這一類別進行判別，實驗數據也證明提示工程中從許琇媛（2024）論文中提取的詐騙關鍵字能有效幫助 GPT 模型判別詐騙音檔。

（三）從實驗數據可知，無論是 Precision、Recall、F1-score，Chain-of-thought prompting 的數值皆是最高，依序是 Direct prompting、Few-shot prompting，從 F1-score 發現提示詞給予越多的情況下 GPT 模型判別非詐騙樣本的能力提高。

ROC 空間點圖的 x 軸為 False Positive Rate （FPR ）、y 軸為 True Positive Rate （TPR ），所以在 ROC 空間點圖中，愈左上者為判別能力愈好者，由圖 9、圖 10 可知不論是 ROC 曲線（Receiver Operating Characteristic Curve）、PR 曲線（Precision-Recall Curve），Chain-of-thought prompting 的判別能力最佳，依序是 Direct prompting、Few-shot prompting。

表 3：*本研究與過往反詐欺模型之比較*

項目	本實驗	李承軒（2024）	許琇媛（2024）	施瓊雯（2024）	嚴浤元（2024）
所用模型	GPT-4.1	隨機森林法（Random Decision Forests）	MultinomialNB、Google Speech-to-Text 等	分層式模型（Hierarchical Explainable Network; HEN）、ARIMA 等	深度神經網路（DNN）、卷積神經網路（CNN）等
是否需要訓練	否	是	是	是	是
辨識內容	圖/音是否為詐欺	詐騙簡訊	向民眾發出警示以降低被詐騙風險	辨識詐騙行為	辨識深偽語音電話詐騙
實驗結果	合理判斷並給出依據	特徵組合與技術特徵於詐騙簡訊識別具顯著效果	能判斷出非詐騙文本	顯示詐騙手法多樣且社群媒體重要性	能辨識深偽語音與真實語音特徵差異
實驗數據	圖片、音訊 F1-score 最高為 1.00	100 則簡訊 F1-score 為 0.99	準確率為 0.84	無	合成型深偽語音準確率為 0.93

（資料來源：本研究自行製作）

四、反詐騙模型

本研究開發 Web 前端與 Android 應用程式 之原型開發，並建置可靈活調整之 AI Prompt 管理模組，語音、文件與系統層級提示詞均可因應時事或詐騙手法更新而即時修改。

1.反詐騙程式原型與功能流程

根據實驗結果，思維鏈提示模型（Chain-of-Thought Prompting）之 F1-score 表現最佳，故本系統以該模型作為核心推論引擎，並提供使用者三項偵測功能：

(1)檔案上傳判斷：用戶可直接上傳檔案並判斷是否為詐騙

(2)即時拍照判斷：呼叫裝置相機，並將影像以 JPEG 格式送至後端，後輸出判斷結果及關鍵依據。

(3)錄音判斷：透過錄音將錄下的音檔進行判斷

2.功能流程說明

(1)輸入階段

使用者完成上傳、拍攝照片 、錄音並按下按鈕後，前端將檔案以 JWT 認證之 RESTful API 傳送至伺服器。

(2)前處理階段

伺服器端針對圖片將進行清洗、尺寸標準化與 Base64 編碼，音訊以 Whisper AI 轉寫為 UTF-8 文字。

(3)推論階段

前處理結果與對應 Chain-of-Thought Prompt 組合後，交由 OpenAI API 執行推論。

(4)後處理與回饋

解析模型回傳的 JSON，提取判斷結論、關鍵佐證與建議行動，並把結果以卡片式 UI 呈現，並提供使用者警示文字。

3.系統特色

(1)Prompt 易更新：研究人員可於後台直接編輯並版本化提示詞，無須重新部署程式。

(2)多模態支援：單一 API 即可處理影像與音訊。

(3)可解釋回饋：輸出包含關鍵依據句與建議流程，提升使用者信任。

透過上述流程，本系統驗證了 Chain-of-Thought Prompting 在實務應用中的可行性，並為未來的跨模態詐騙偵測服務奠定技術基礎。

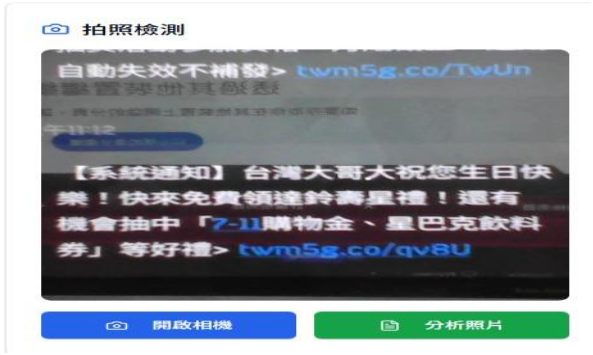
圖 4-1 主要程式介面架構

（圖片來源：本實驗自行製作）

圖 4-2 上傳檔案偵測功能介面

（圖片來源：本實驗自行製作）

圖 4-3 拍照功能介面



（圖片來源：本實驗自行製作）

圖 4-4 錄音功能介面

（圖片來源：本實驗自行製作）

圖 4-5 分析結果介面

（圖片來源：本實驗自行製作）

陸、結論

本研究所設計之最佳提示工程判別模型，能針對大多數圖片與語音資料進行有效辨識，並做出合理且邏輯性的推論。整體實驗模型具備快速反應、使用便利與分析詳盡等優勢，符合本研究降低民眾遭受詐騙風險的核心目標。

透過本次實證研究，我們驗證了具備明確定義與結構的提示詞能顯著提升 GPT 模型型對詐騙資訊的判別能力，進而有效預防詐騙事件的發生。

一、學習及感想

我們原本設想使用字典來製作自己的資料庫模型，後來基於成效不彰，判斷形式有限等問題，經由文獻探討，在多方查找資訊後，決議轉而使用擁有完整資料庫的 GPT 為判斷模型的基底，並探討提示工程對 GPT 模型判別詐騙的準確率，並參考蘇宥蓁等人（2024）提出的 GPT 模型提示工程確定 GPT 模型應用的可能性。

我們透過自學 Python 程式語言，熟悉各類模組的引用與整合，並以邏輯運算思維方式將判斷問題拆解為可測試的子項目，逐步完成 Prompt 設計、測試、驗證與修正。最終成功導出具備高度判別能力的模型架構。

而本研究採用生成式 AI 製作詐騙辨識軟體後，研究者驗證其可行性，經過校驗並修正後，再應用於實驗硬體設備並成功完成本次成果，成果豐碩。

二、改善及回顧

經過實驗發現，模型仍有其限制所在，例：對於判定邏輯可能不夠清楚，可能被誇大的真實廣告或新聞誤導，若僅提供部分資訊可能發生誤判等。基於上述各點進行改善，增加針對可能誤導之細節提示詞後，模型成功修正了部分出錯的結果，若未來再遇可能存在之隱患，也能依前述進行修正。

三、未來展望

本實驗期望於我們計畫導入 模型微調（fine-tuning） 技術，先於直接提示階段之前調整模型參數，藉此縮短提示詞長度、降低 token 耗用，進一步節省推論成本並提升執行效率。同時，我們亦將整合 即時網路搜尋（web search） 功能，針對影像與文字內容進行動態爬蟲與資料比對，並結合 檢索增強生成（Retrieval-Augmented Generation，RAG） 機制，使大型語言模型具備即時更新與領域知識擴充的能力，以因應日益多樣且快速演變的詐騙手法。

本研究雖然可以透過託管的網頁進行詐騙偵測，但為了各年齡層的需求，本實驗想再持續研究並設計出辨別詐騙的手機 APP，達成防範詐騙之目的。

柒、參考文獻

1.許琇媛 (2024)。語音辨識應用於詐騙關鍵字提取與防詐系統建立 (碩士論文)。國立臺北商業大學。

https://hdl.handle.net/11296/2vf93a

2.吳其龍 (2024)。 針對貨到付款的防詐騙系統 (碩士論文)。國立陽明交通大學。

https://hdl.handle.net/11296/sg7gq5

3.施瓊雯 (2024)。文字探勘應用於詐騙行為之研究 (2024)。淡江大學。

https://hdl.handle.net/11296/x3h2x7

4.李承軒 (2024)。應用機器學習於台灣詐騙簡訊之偵測 (碩士論文)。東吳大學。

https://hdl.handle.net/11296/xn3k55

5. 陳峰楷 (2024)。應用 LLM 與 Prompt Turning 填寫網頁表單以支援網頁應用程式測試之研究 (碩士論文)。國立臺北科技大學。

https://hdl.handle.net/11296/2nryxe

6. 楊子宜 (2024)。基於 ViT 與影像頻率域特徵之 AI 生成影像偵測 (碩士論文)。淡江大學。

https://hdl.handle.net/11296/8a595k

7. 廖曼伶 (2023)。應用人臉特徵與相似度於 AI 生成圖片之檢測 (碩士論文)。國立臺灣科技大學。

https://hdl.handle.net/11296/fr4725

8. 林雨蓉 (2017)。基於影像品質分析之生物辨識詐騙偵測 (碩士論文)。玄奘大學。

https：//hdl.handle.net/11296/s3by9z

9. 謝岳哲 (2023)。運用基於生成預訓練轉換器架構的 OpenAI Whisper 多語言語音辨識引擎之台語及華語語音辨識之實作 (2023) 。長庚大學。

https：//hdl.handle.net/11296/r2qg46

10. 蘇宥蓁、蔣亞棋、 (2024) 。關於我與 ChatGPT 成為一家人的那件事 。

https：//twsf.ntsec.gov.tw/activity/race-1/64/pdf/NPHSF2024-052502.pdf?0.2125046944907849

11. Python 學習筆記—常見的二元分類評估指標：混淆矩陣、ROC 曲線。

https://medium.com/@SCU.Datascientist/python%E5%AD%B8%E7%BF%92%E7%AD%86%E8%A8%98%E5%B8%B8%E8%A6%8B%E7%9A%84%E4%BA%8C%E5%85%83%E5%88%86%E9%A1%9E%E8%A9%95%E4%BC%B0%E6%8C%87%E6%A8%99-

%E6%B7%B7%E6%B7%86%E7%9F%A9%E9%99%A3-roc-%E6%9B%B2%E7%B7%9A-f214ecd84dab

12. 內政部統計查詢網 (2025)。詐欺案件統計。

https://statis.moi.gov.tw/micst/webMain.aspx?k=defjsp

13. Lee, B. (2025). *Prompt engineering [White paper]. Kaggle.*

https://www.kaggle.com/whitepaper-prompt-engineering?_bhlid=a2bfce2cac67662098bd85a241e7cb000576e5d4-

14.Karimi,Z.(2021). Confusionm Matrix. ResearchGate.

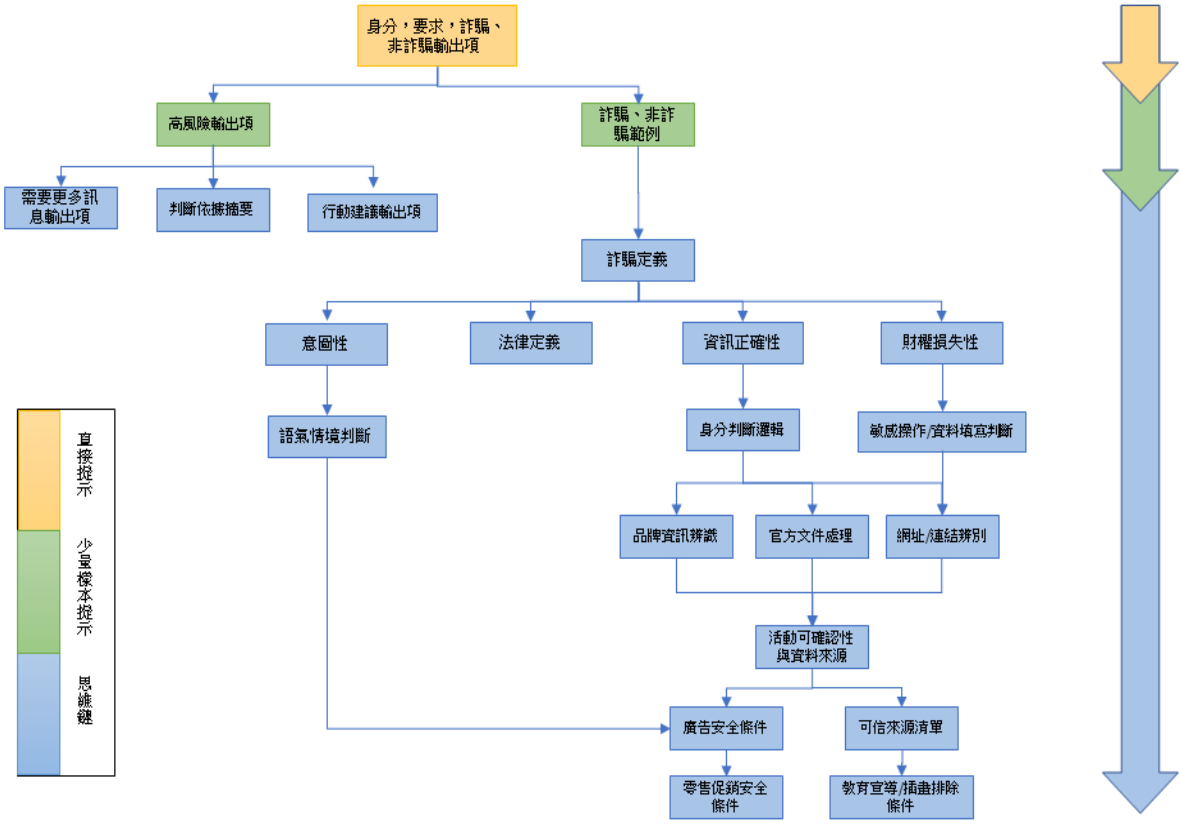
https：//www.researchgate.net/publication/355096788_Confusion_Matrix

15.He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, *21*(9), 1263-1284.

https://ieeexplore.ieee.org/abstract/document/5128907

附錄

圖 3：本研究提示工程模型設計的演進歷程。圖示中以箭頭方式呈現三種提示策略之邏輯流程與發展層次。



（圖片來源：本研究自製）

【評語】052501

本作品以 ChatGPT + Whisper AI 為主體，藉由不同的 prompt 設計判別圖像與語音資料中的詐騙內容，主題切合社會議題。

未來精進發展的建議：

1. 擴大資料集來源與語境，如加入跨語言詐騙內容，或即時社群截圖與留言判斷等情境。

2. 可考慮引入句法修復模型或前處理語音增強技術（Speech Enhancement），例如 RNNoise 等模組。

3. 思慮導入 LLM 微調或 RAG（Retrieval-Augmented Generation）技術以優化特定領域的精準回應的可能性。

作品海報

利用ChatGPT協助辨識詐騙簡訊及網頁

摘要

本研究探討提示工程技術（Prompt engineering），評估其是否可提升GPT模型對詐騙內容的辨識準確率。研究設計採用三種不同提示策略，分別應用於圖像與語音資料中，並依混淆矩陣比較其判斷正確數以分析準確率差異。結果顯示思維鏈模型表現皆優於其餘提示工程模型，顯示良好效果。本研究基於思維鏈提示開發互動式網頁程式讓民眾遭遇可疑資訊時可立即使用網頁判別可疑圖片、音訊，展現應用於基礎防詐之可行性，亦爲後續防詐系統提供設計參考。

1.研究動機

根據內政部統計查詢網， 詐欺案件數由 2015 年的 21,172 件增加到 2023 年的 37,984 件，可見人工處理並不足以應對詐騙的遞增。現今資訊科技發達迅速，普通人無法自行製作程式來辨別詐騙圖片，因此本實驗想探討利用擁有龐大資料庫的GPT模型探討AI在詐騙上之應用，但在龐大且雜亂的資料量下，單純的命令可能無法使GPT模型精確判別問題而給出錯誤答案，若是能解決此問題並進行反詐應用程式的開發，或許能夠達成減少民眾受詐騙之目標。

2.目的

（一）探討不同prompt設計對GPT模型辨識詐騙圖像、語音的準確率影響，實驗出最佳範本。
（二）設計應用程式整合提示工程技術與GPT模型，希望能達成遏止惡意詐騙、減少民眾受詐騙等目標。
（三）提供未來擴展研究與研發程式樣本之參考，減少國人受騙之風險。

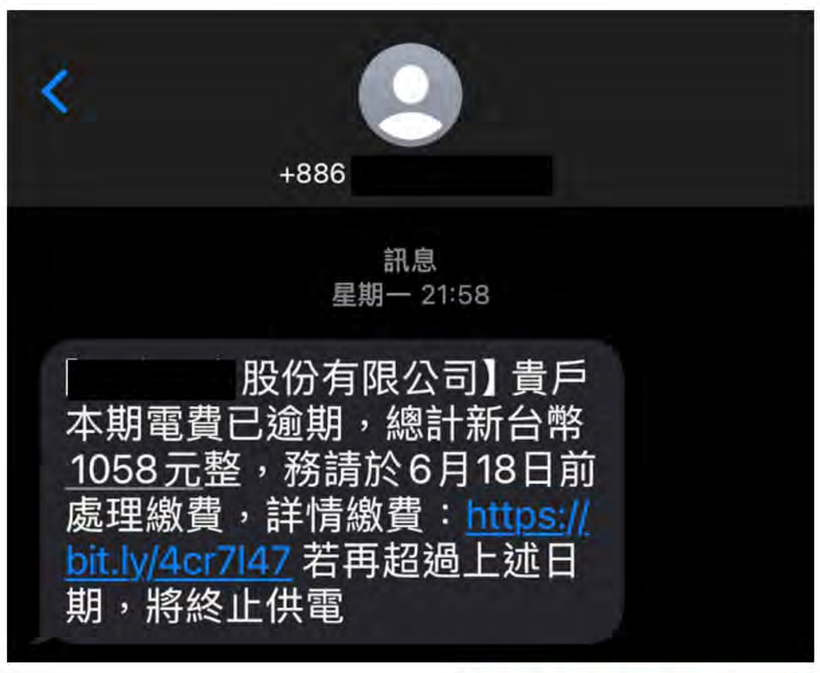
3.實驗步驟

本實驗依Boonstra（2025）將提示工程模型分爲三種：直接提示（Direct Prompting）、少量樣本提示（Few-shot prompting）、思維鏈提示（Chain-of-Thought Prompting），依序增加範例、詐騙定義、判斷方式等作爲實驗變因進行實驗。

本實驗將依圖像、音訊分開進行實驗。製作三種提示工程模型，將實驗樣本（圖檔、音檔）載入模型中，GPT模型會自行判斷檔案媒介，若爲音檔則將音訊轉爲文字並開始判別，若爲圖檔則直接進行判別，判斷完成後模型將回傳判別結果，本研究再依回傳結果計算出Accuracy、Precision、Recall、F1-score等所需數值對不同程度的模型進行對比與討論。

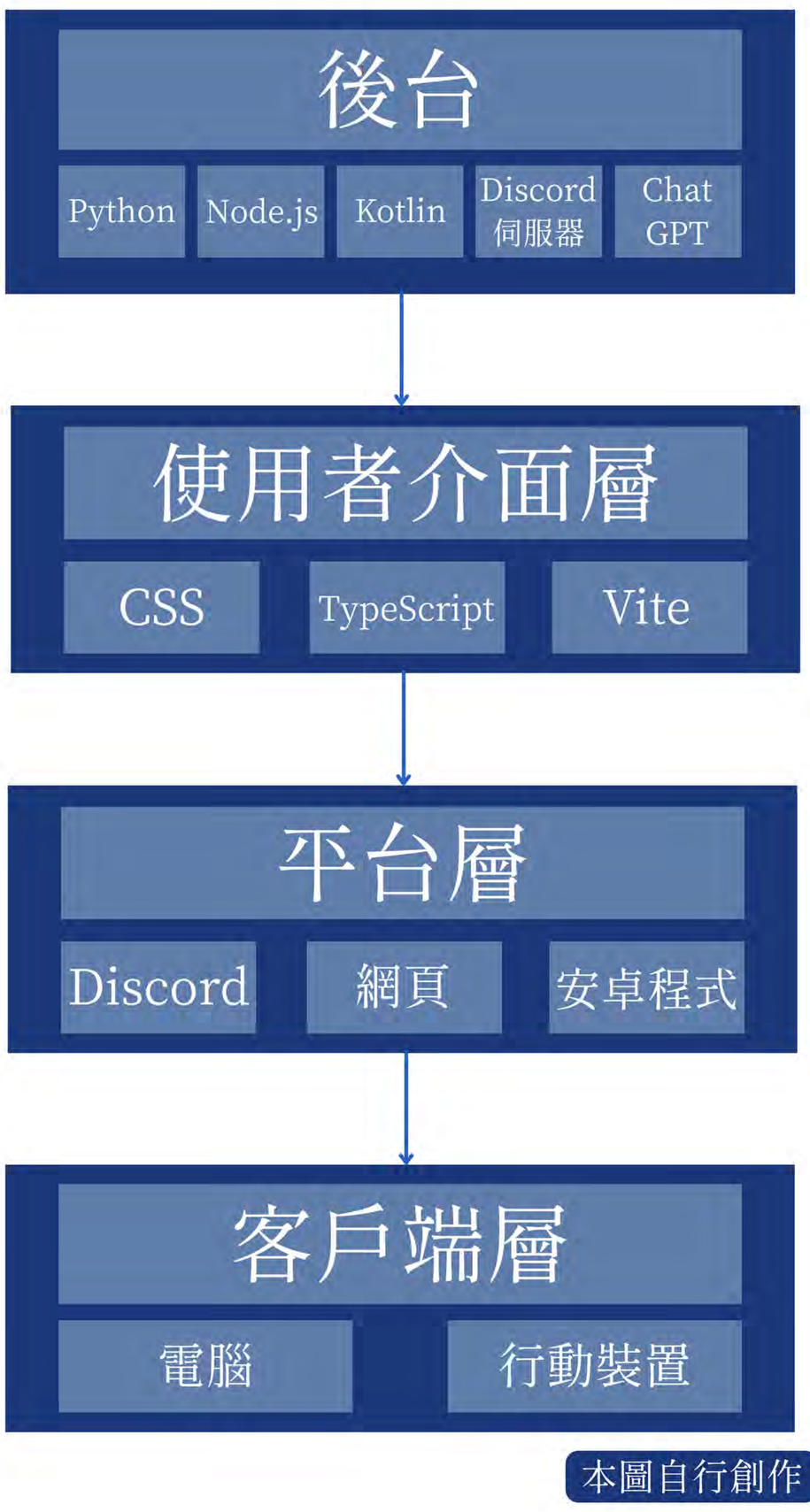
3-1實驗圖片範例

本研究自行截圖



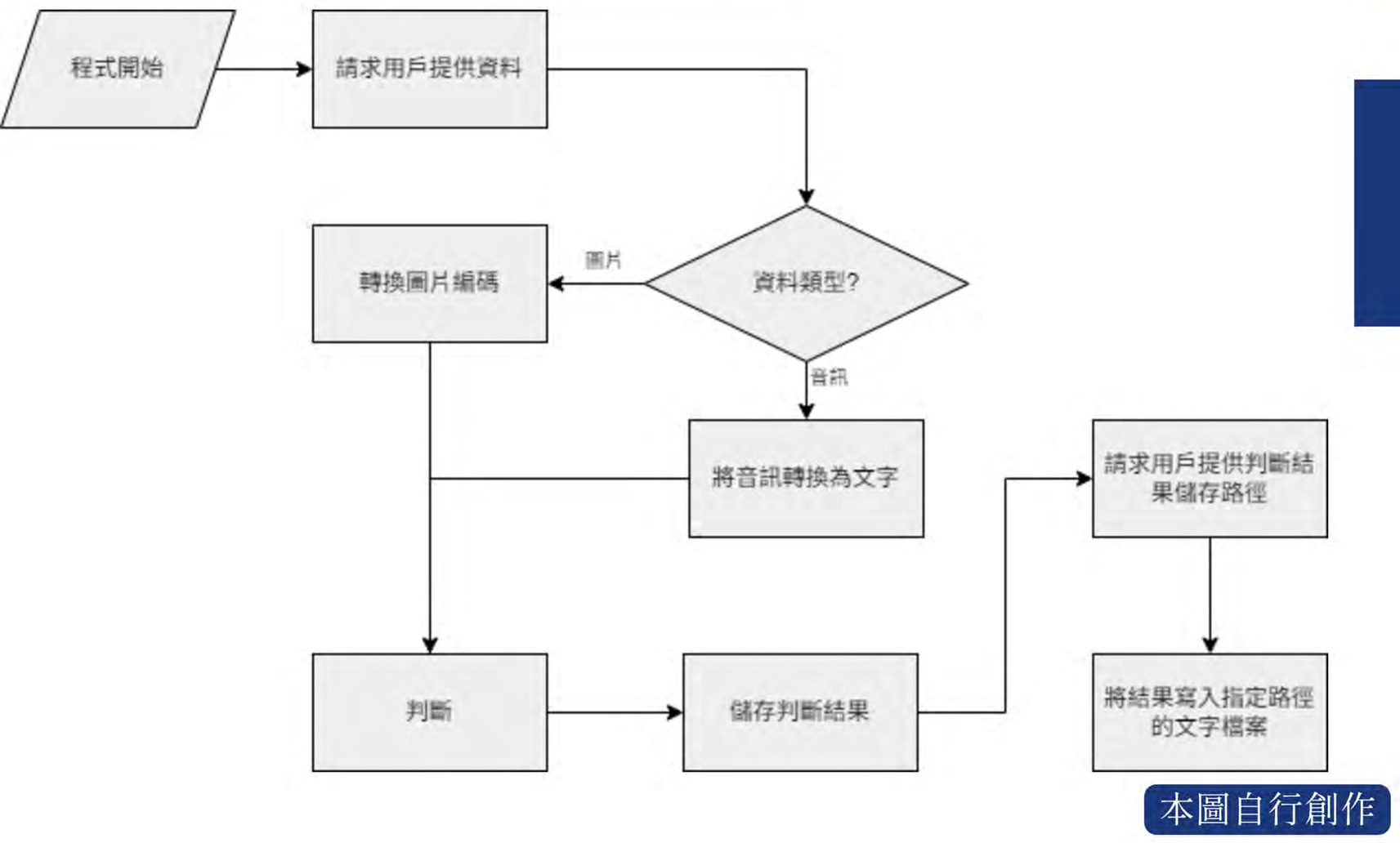
本研究自行截圖

3-3系統結構圖



本圖自行創作

4-2程式流程圖



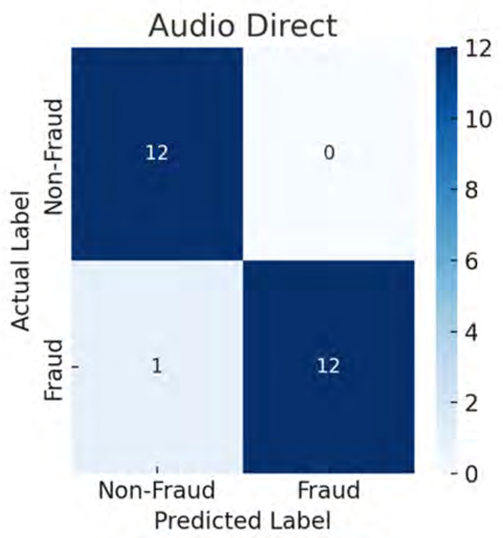
本圖自行創作

4.研究結果

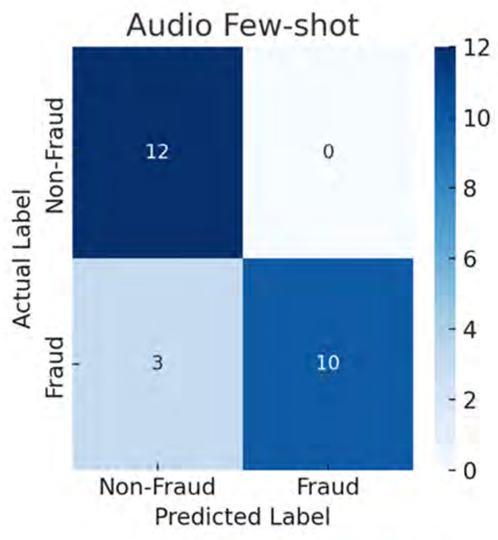
本研究實驗樣本圖片一共55張，其中詐騙樣本36張，非詐騙樣本19張。音訊共25份，其中詐騙樣本13份，非詐騙樣本12份。

製作出混淆矩陣後，本實驗利用混淆矩陣算出四項指標，並繪製出ROC曲線（Receiver Operating Characteristic Curve）、PR曲線（Precision-Recall Curve），期望利用不同數據進階探討提示工程設計對模型判別詐騙準確率的影響。

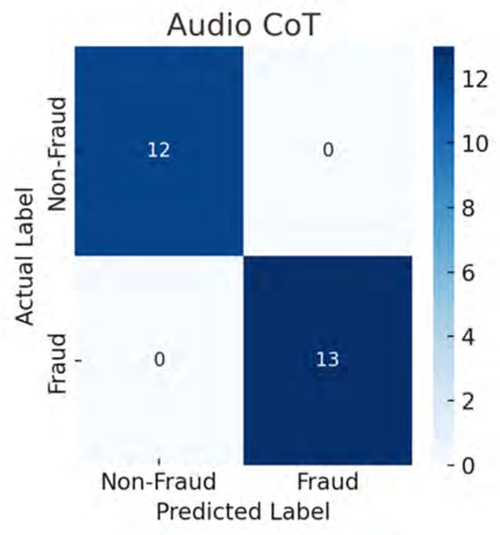
4-1混淆矩陣



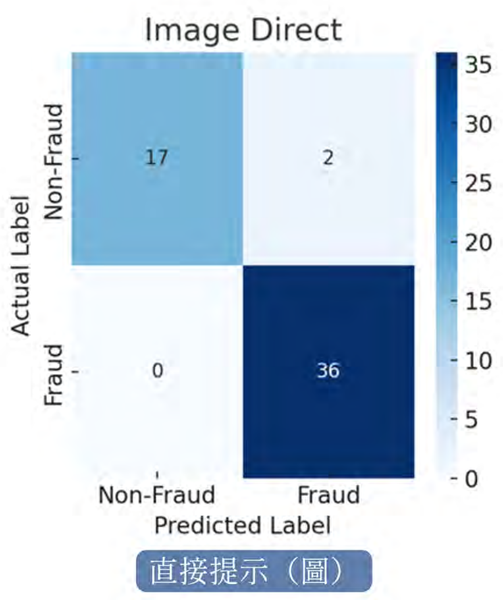
直接提示（音）



少量樣本提示（音）

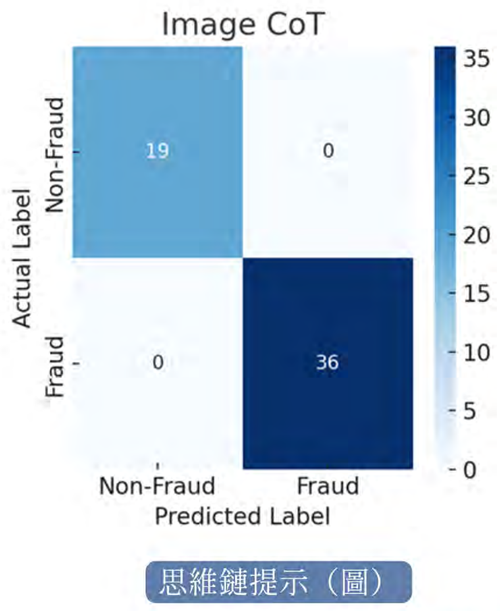


思維鏈提示（音）



直接提示（圖）

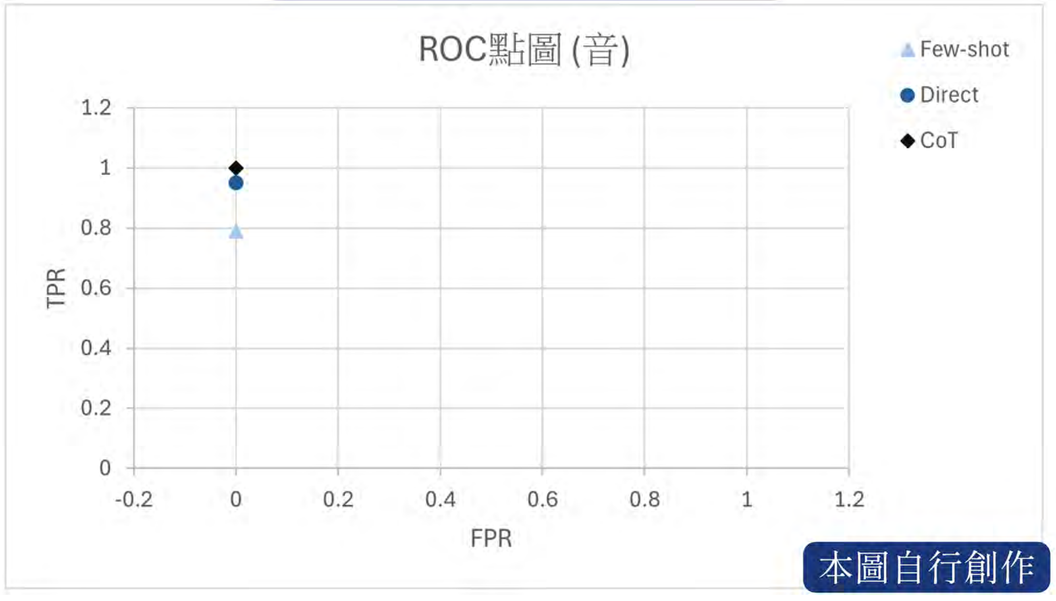
少量樣本提示（圖）



思維鏈提示（圖）

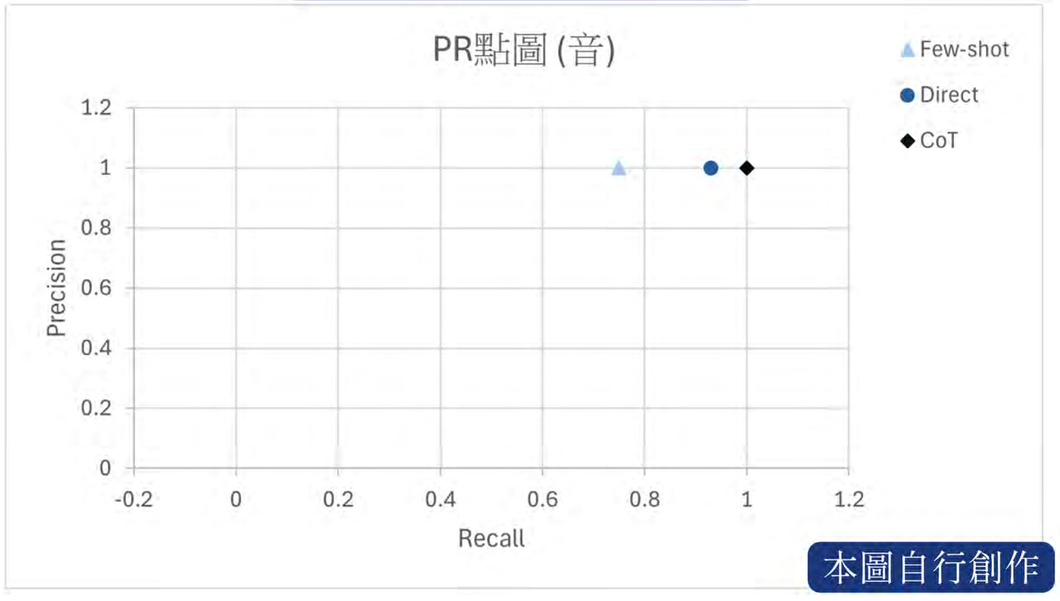
混淆矩陣皆自行創作

4-2ROC點圖（音）



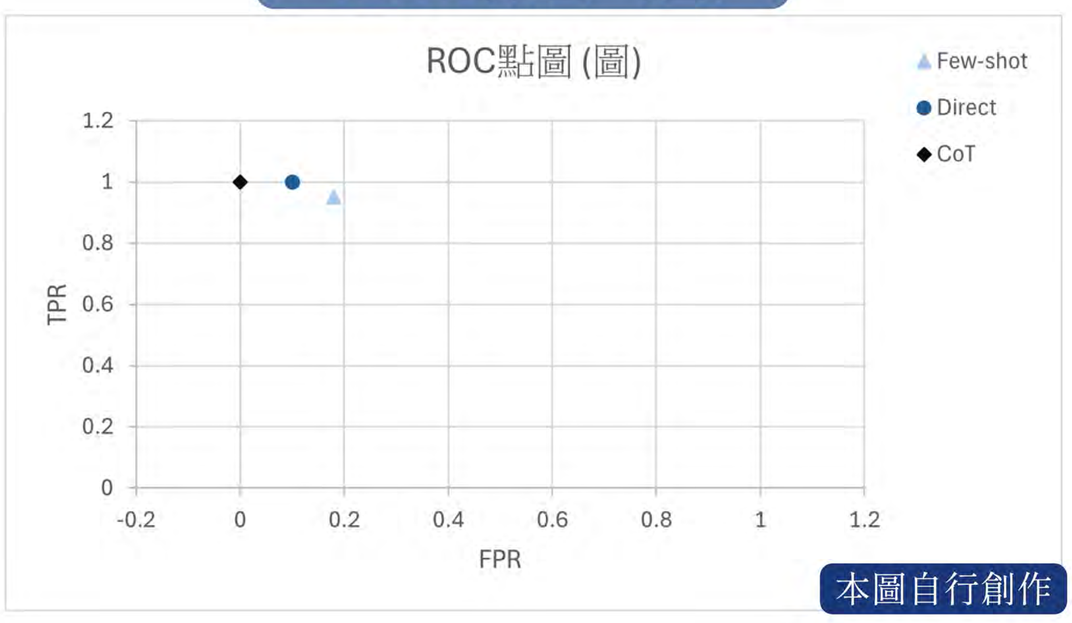
本圖自行創作

4-4PR點圖（音）



本圖自行創作

4-3ROC點圖（圖）



本圖自行創作

4-5PR點圖（圖）

本圖自行創作

實驗數據結果

本研究最終實驗數據不論圖檔、音檔，數據最高者皆爲思維鏈提示，其中兩者Accuracy、Precision、Recall、F1-score數值皆爲1.0。第二高者皆爲直接提示，音檔Accuracy、Precision、Recall、F1-score數值分別爲0.960、1.0、0.923、0.960，圖檔Accuracy、Precision、Recall、F1-score數值分別爲0.963、0.947、1.0、0.972。數據最低者皆爲少量樣本提示，音檔Accuracy、Precision、Recall、F1-score數值分別爲0.880、1.0、0.769、0.870，圖檔Accuracy、Precision、Recall、F1-score數值分別爲0.909、0.944、0.919、0.931。

5.討論

一、圖檔之判斷

根據實驗結果，圖檔判斷正確率由高至低依序爲思維鏈題示、直接提示、少量樣本提示。

（一）內容判別問題：

根據模型判斷回覆，直接提示及少量樣本提示多會直接根據圖片最爲明顯的特徵或文字進行判斷，在沒有詳細限制或指示的情況下，模型可能從而造成判斷錯誤、誤導使用者行動的結果。

（二）思維鏈提示回覆：

思維鏈提示在判斷回覆中清楚指出圖片中判斷的關鍵依據，本實驗利用詐騙定義、李承軒、許琇媛等人確認之判別詐騙方式所設計之prompt，能使模型能根據圖像中的細節進行權衡，得出正確的結果及完整的判斷依據。

二、音檔之判斷

根據實驗結果，音檔判斷正確率由高至低依序爲思維鏈提示、直接提示、少量樣本提示。

（一）音訊完整度問題：

根據模型判斷回覆，直接提示及少量樣本提示在處理音訊轉換文字時，由於詐騙音訊有節錄不全、音檔整體完整度不高等問題，導致模型常因資訊不足而判斷爲非詐騙，使詐騙音訊判斷錯誤，內容相對完整清晰的非詐騙音訊則能全數判斷正確。

（二）思維鏈提示回覆：

思維鏈提示在面對完整度不高的音檔時，由於針對模糊訊息的prompt設計，模型相較直接提示及少量樣本提示會傾向將資訊與情境進行連結，透過邏輯推導，正確識別潛在詐騙，提升音訊的判斷正確率。

5-1與前人反詐模型對照

表3：*本研究與過往反詐欺模型之比較*

項目	本實驗	李承軒（2024）	許琇媛（2024）	施瓊雯（2024）	嚴浤元（2024）
所用模型	GPT-4.1	隨機森林法（Random Decision Forests）	MultinomialNB、Google Speech-to-Text 等	分層式模型（Hierarchical Explainable Network; HEN）、ARIMA 等	深度神經網路（DNN）、卷積神經網路（CNN）等
是否需要訓練	否	是	是	是	是
辨識內容	圖/音是否為詐欺	詐騙簡訊	向民眾發出警示以降低被詐騙風險	辨識詐騙行為	辨識深偽語音電話詐騙
實驗結果	合理判斷並給出依據	特徵組合與技術特徵於詐騙簡訊識別具顯著效果	能判斷出非詐騙文本	顯示詐騙手法多樣且社群媒體重要性	能辨識深偽語音與真實語音特徵差異
實驗數據	圖片、音訊 F1-score 最高為 1.00	100 則簡訊 F1-score 為 0.99	準確率為 0.84	無	合成型深偽語音準確率為 0.93

本圖自行創作

三、互動式網頁開發

本研究最終以思維鏈提示爲基底開發網頁程式，其中網站民衆可因時事而進行自由管理、修改本研究設計之語音、圖片之prompt。根據實驗結果，本實驗決議將判斷正確率最高之提示工程模型（思維鏈提示） 投入反詐程式的實作，並預計給予使用者三項不同的判斷功能：

1.直接上傳檔案並判斷是否爲詐騙
2.透過開啟鏡頭拍攝照片進行判斷
3.透過錄音將錄下的音檔進行判斷

其中，錄音功能將有兩種模式供使用者選擇，分別爲使用麥克風錄音以及錄製使用者裝置的媒體音效。當使用者按下分析的選擇按鈕後，進行資訊是否爲詐騙的分析。

5-2網頁操作畫面

本研究自行截圖

5-3網頁操作畫面

本研究自行截圖

6.結論

一、結論

根據實驗結果，思維鏈提示無論圖檔、音檔皆全數判斷正確，和本實驗預測當prompt給予越多的情況GPT模型判斷正確數愈高的結果相同， 從判斷依據中發現，思維鏈提示判斷回覆較其餘prompt準確且清晰，且會給予建民衆如何防備詐騙訊息並阻止其進一步擴散。

透過該模型爲基底設計的反詐網頁程式，同樣能針對大多數圖片與語音資料進行有效辨識，並做出合理且邏輯性的推論。模型具備快速反應、使用便利與分析詳盡等優勢，符合本研究降低民衆遭受詐騙風險的目標。

透過本次研究，驗證了在給予愈多明確的prompt的情形下能顯著提升以GPT模型爲基底之模型對詐騙資訊的判別能力，進而有效預防詐騙事件的發生。

三、未來展望

本實驗已將模型打包並在Discord及網頁進行測試及實作，期望於後續發展中，將實驗模型包裝成可供下載使用之應用程式，並透過推廣達成降低民衆遭受詐騙機率的目標。此外，本實驗考慮引入微調技術（fine-tuning），在推論階段前對模型進行調整，以縮短提示語長度、降低 token 使用量，進一步節省成本並提升執行效率。亦考慮結合網路搜尋功能 (web search)，針對圖片與文字描述進行即時資料比對，搭配檢索增強生成技術 (Retrieval-Augmented Generation, RAG)，使大型語言模型具備即時更新與特定領域知識擴充的能力，以面對更多元的詐騙型態與變化。

7.參考文獻

1. 蘇宥蓁(2024) 。關於我與 ChatGPT 成爲一家人的那件事 。
2. 許琇媛 (2024)。語音辨識應用於詐騙關鍵字提取與防詐系統建立 (碩士論文)。國立臺北商業大學。
3. 陳峰楷 (2024)。應用LLM與Prompt Turning填寫網頁表單以支援網頁應用程式測試之研究 (碩士論文)。國立臺北科技大學。
4. 施瓊雯 (2024)。文字探勘應用於詐騙行爲之研究 (2024)。淡江大學。
5. 吳其龍 (2024)。 針對貨到付款的防詐騙系統 (碩士論文)。國立陽明交通大學。
6. 李承軒 (2024)。應用機器學習於台灣詐騙簡訊之偵測 (碩士論文)。東吳大學。
7. 內政部統計查詢網 (2025)。詐欺案件統計。
8. Lee, B. (2025). Prompt engineering [White paper]. Kaggle.
9. Karimi,Z.(2021). Confusion Matrix. ResearchGate.