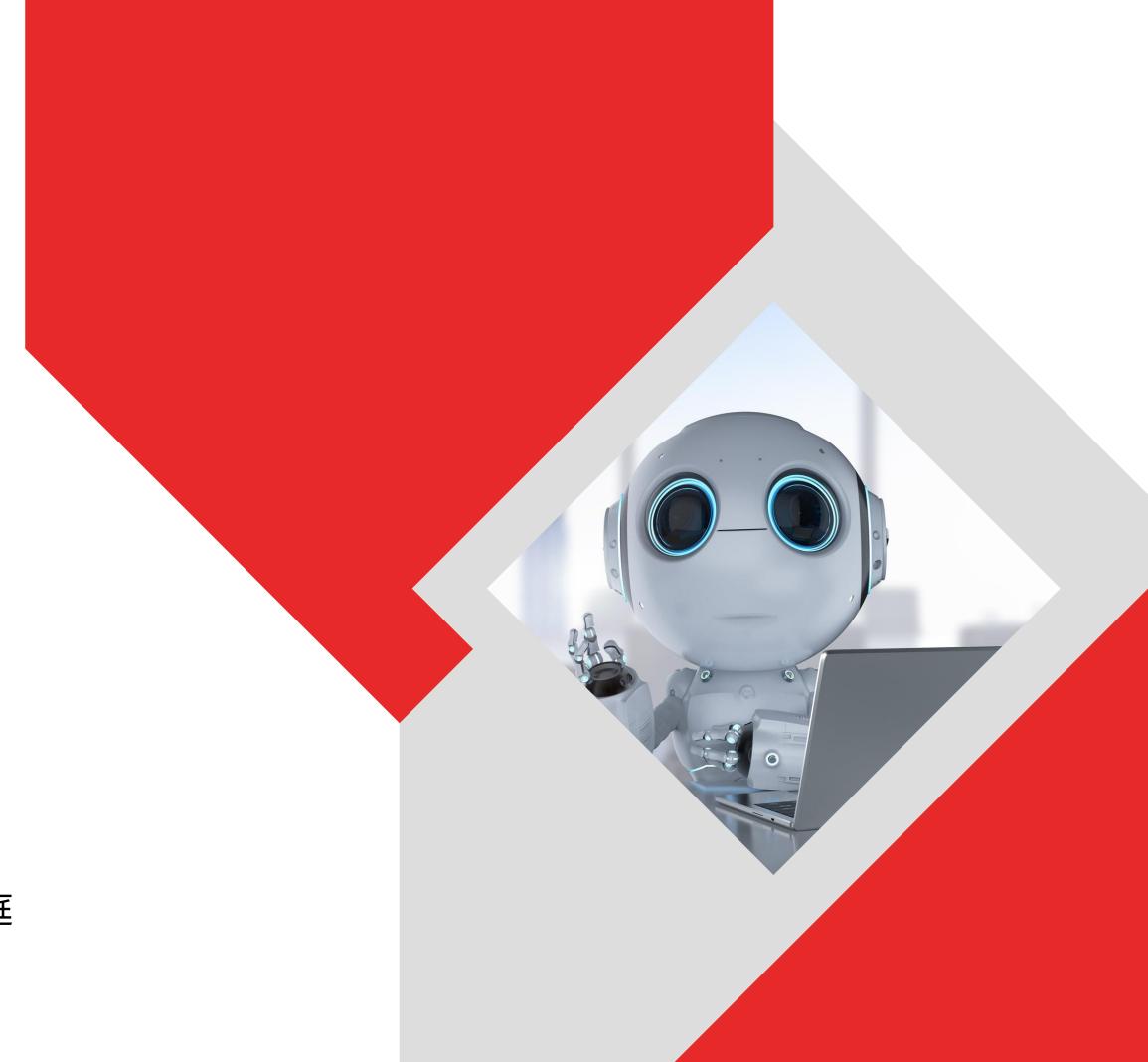


Al金融 防詐小精靈

導師: Vickie, Renee

成員:林源隆,楊思妤,藍苡瑞,陳勁瑋,鄧宇庭



1 專案目標

4 模型調適

2 資料集

5 結果分析

3 生成回覆

6 總結

研究動機





詐欺案件高達 4 萬 投資詐騙占31%



詐騙廣告數量超過 20萬筆



缺乏本土化 防詐機器人

專案目標

目標潛力



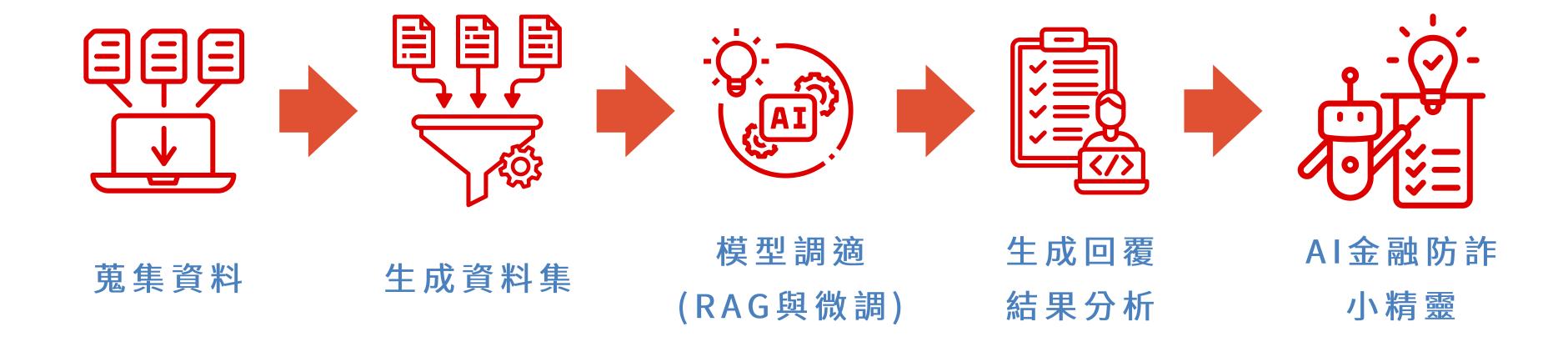


AI金融防詐小精靈

- 1 提供準確、實用的建議
- 2 協助用戶理性分析
- 3 即時回覆訊息,讓用戶安心
- 4 降低受騙風險

專案流程





2 資料集

資料集

資料集蒐集



司法院判決

事實

一、 加**害者** 於民國000年0月間某日起,加入LINE暱稱「孫慶 龍-總監」、「好幣所」、「OOO」之人所組成之3人以上,以 實施詐術為手段,具有持續性、牟利性之有結構性組織之詐 欺集團(下稱本案詐欺集團)擔任向被害人收取詐騙贓款之 車手工作。嗣 加害者 、LINE暱稱「孫慶龍-總監」、「好幣 所」、「OOO」及其他本案詐欺集團成員即共同意圖為自己 不法之所有,基於三人以上共同詐欺取財、洗錢之犯意聯 絡,先由本案詐欺集團不詳成員,於112年3月透過網際網路 投放投資廣告,吸引 被害 瀏覽後,加入廣告內所示之通訊 軟體LINE群組,本案許軟集團不詳成員並在群組內對其誆 稱:系統有買賣股票優先權,依老師指示逐步操作即可投資 獲利等語,致 受害 陷於錯誤,於112年5月10日18時30分 門市,交付新臺幣(下同)86萬元予 加害者, 加害者 並依LINE暱稱「OOO」之指示,將贓款放置於高雄市大寮 區鳳林一路329巷路旁,由本案詐欺集團不詳成員取走,製 造金流斷點,以掩飾、隱匿詐欺所得之去向及所在。嗣因林 進忠發覺受騙,報警處理,始查獲上情。



資料集生成



摘要

A積極向聲請人鼓吹投資比特幣等虛擬貨幣,然而,A聲稱每次投資獲利都需支付「外匯保證金」,否則無法將獲利提領。因此a不斷依照A的指示將虛擬貨幣匯往國外平台,直到9月29日才發現自己遭遇詐騙,最終只剩少數存款。

重點提示

- 1.詐騙對話
- 2.反詐騙專家

模擬對話

A: 哇!你的帳戶已經顯示不少獲利了吧?但要提領出來的話,需要支付一筆外匯保證金。

a: 保證金?為什麼要付?

A: 這是平台的規定,賺的錢要取出來之前,都得

先支付一筆保證金。

資料集生成



摘要

A開始接觸外匯市場,選擇

- 一個外匯投資課程以加強 自己市場分析與操作技能
- 。課程重點在於透過數據 分析、全球經濟動態的掌 握,讓學員熟悉如何在市 場波動中合理地進行資金 配置,並更好地理解風險 與回報的平衡點。

重點提示

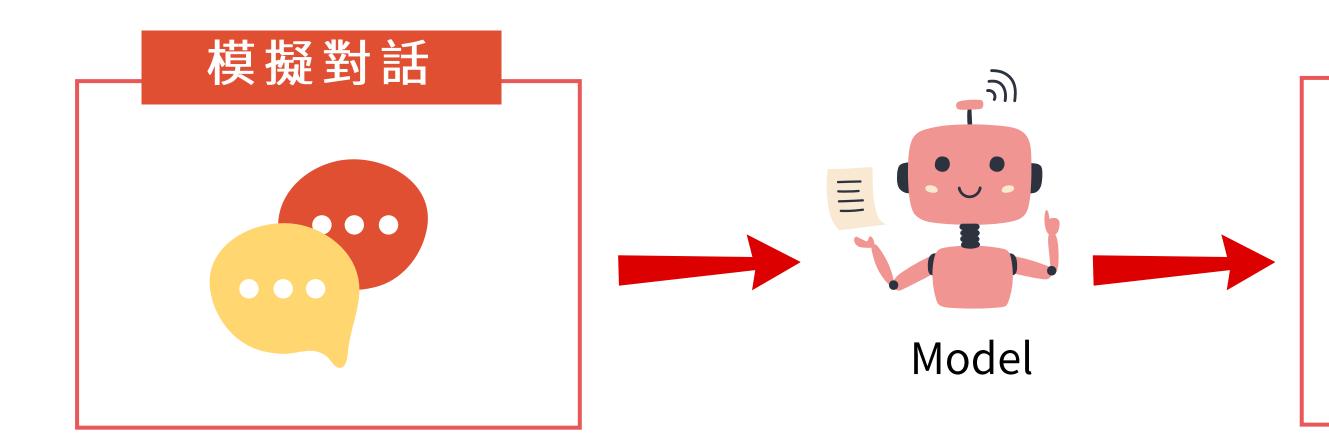
- 1.非詐騙對話
- 2.投資顧問角度
- 3.改寫情境

模擬對話

甲:回報是根據市場走勢 而變化,沒有固定收益。 一般來說,透過平台進行 交易會收取小額手續費, 但如果您有盈利,我們的 政策並不會要求您支付額 外費用來領取收益,不會 額外索取「保證金」或其 他費用。

生成回覆過程





判定結果

判定「模擬對話」是 否為詐騙,並點出:

- 1.關鍵對話
- 2.關鍵因素

生成回覆

實際範例



提示

請判斷以下對話是詐騙還是合理投資,並基於「關鍵因素」與「關鍵對話」來解釋您的判斷。

<對話>

專案成員







原因



關鍵對話:「要提領出來的話,需要支付一筆外匯保證金。」

關鍵因素: 這是典型先付款再獲利的詐騙手段。合法的投資平台在提領資金時不會要求額外的不可合理解釋的費用。



AI金融防詐小精靈

對話區域

請輸入對話內容,我會協助您判斷是否為詐騙:

A: 嘿, 最近有想過投資嗎?外匯投資是個大趨勢啊, 現在進場的話有很大的賺錢機會!

a: 嗯,聽起來不錯,但我沒有經驗,會不會風險太高?

A: 別擔心,我會教你怎麼操作。你只需要按照我的指示,保證獲利,完全不用擔心風險。只要投進去,錢就會自己滾錢。

a: 那···需要投入多少錢呢?

A: 一開始不用太多,先用信用卡或銀行貸款周轉一下,這樣壓力小。而且,我們用的這個平台非常安全,資金和銀行帳戶綁定,隨時可以查看。

a: 那是要在哪個平台操作呢?

A: 我們這邊用的是一個高端平台,叫「現代財富科技」,非常專業又可靠。我會教你在裡面買一些比特幣或泰達幣,這樣你就能開始操作外匯投資了。

a: 好的,那我開始試試看。

生成回應

分析結果:

這個對話是典型的詐騙行為。 判斷的關鍵因素包括但不限於以下幾點:

- 1.**高回報承諾**: 詐騙者通常會宣稱有很高的回報率, 吸引投資者加入,但這種高回報往往是不切實際的,且通常伴隨著要求投資者使用高風險資金(如信用卡或銀行貸款),增加投資者的資金壓力。
- 2. 承諾免風險操作: 詐騙者會聲稱自己能保證獲利, 甚至宣稱「完全不用擔心風險」, 這是一種不負責任的承諾, 且違反了投資的基本原則。
- 3. 資金安全保證: 詐騙者會以「資金和銀行帳戶綁定」、「隨時可以查看資金安全」等說法來增加信任感,但最終卻導致資金被騙走。
- 4. 要求匯款或支付保證金 : 詐騙者會以「支付外匯保證金」 、「流程未走完」等理由要求投資者不斷匯款 ,最終導致資金被騙 。
- 5. **受害者最終意識到被騙** : 在整個過程中 , 受害者最終意識到整個流程是詐騙 , 所有的錢都已經被騙走 , 這是一個關鍵的自我覺醒點 。

關鍵對話包括:

- 詐騙者介紹「現代財富科技」平台, 聲稱「非常專業又可靠」。
- 詐騙者要求支付保證金,並說「這是平台的規定」。
- 受害者最終意識到整個流程是詐騙 , 所有的錢都已經被騙走。

這些對話和因素綜合起來 , 清楚地表明了這是一個典型的詐騙行為 。 投資者應該避免任何承諾高回報 、 要求使用高風險資金 、 聲稱免風險操作 、 要求支付保證金或資金安全保證的投資 。

4 模型調適

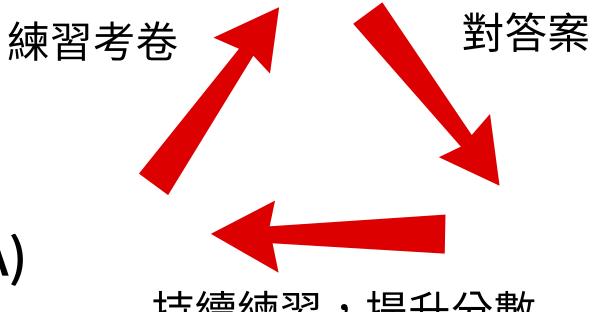
微調|給予模型學習任務與情境







Yi 1.5 6B(學生A)



持續練習,提升分數



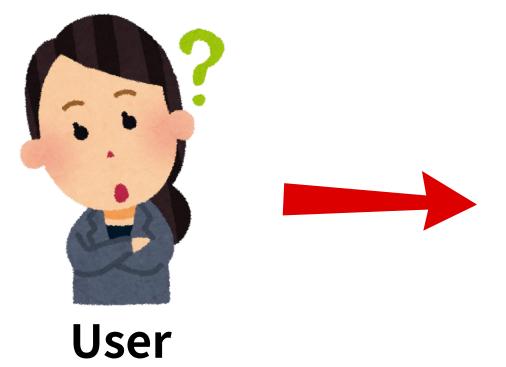
訂正考卷



模型調適

RAG檢索強化生成|翻書找答案再回答



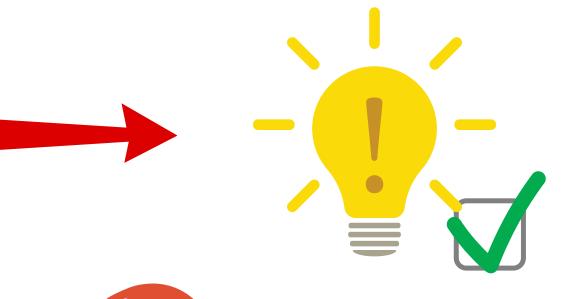


我碰到了以下情境及 對話,這是詐騙嗎?



Yi 微調 + RAG(學生B)

翻書找答案



是,原因...(依據書 上內容回答)

用戶

解答

正確率評分過程





請判斷以下對話是詐騙還是合理投資。 <詐騙對話>

這段對話是典型投資詐騙案■



答對





這段對話是典型正常投資案



錯誤



模型答題正確率



正確率 (樣本數)	Yi	Yi+微調	Yi+微調+RAG	ChatGPT 4o
一般對話 (3)	100%	100%	100%	100%
詐騙對話 (4)	100%	100%	100%	100%
一般投資對話 (3)	100%	33% (1/3)	100%	67% (2/3)

測試對話個案示例-非詐騙情境判讀成詐騙



對話

A:哈嘍,最近有個新的投資機會,想跟你分享一下

- 。這檔基金的績效很不錯,而且風險相對較低。
- a:哦?聽起來不錯,可以跟我說得更詳細一點嗎?
- A:當然,這檔基金主要投資於科技股,你看最近科技產業多熱門,成長性非常可觀。而且基金經理人經驗豐富,過往的績效也很穩定。
- a:科技股是蠻吸引人的,不過我對這方面不太了解
- ,能給我一些建議嗎?
- A:沒問題,我建議你可以先撥出<mark>部分資金試水溫</mark>,如果效果不錯再<mark>逐步增加投入</mark>。 幫我判斷是詐騙或 非詐騙。



Yi+微調

- 1.無提到詳細數據支持
- 2. 「逐步增加投入」用詞模糊



ChatGPT 4o

- 1.未提供具體風險分析
- 2.未提供細節
- 3.「部分資金試水溫」具誘導性

回覆品質-指標介紹



指標	說明	權重
可讀性 (Readability)	評估生成文本的 <mark>易讀性</mark> ,包括語法正確性 、用詞適當性和句子結構的清晰度	25%
一致性 (Consistency)	評估模型在整個文本中保持一致性的能力,確保不出現自相矛盾的資訊	25%
多樣性 (Diversity)	衡量模型生成文本的多樣性, <mark>避免產生重</mark> 複或單調的內容	25%
說服力 (Persuasiveness)	文本是否能 <mark>有效說服讀者</mark> ,讓其理解並接 受文本中的判斷或觀點	25%

文本示例

這對話是典型的詐騙行為。關鍵因素包括:

- 1.無合法投資管道: 詐騙者通常會以「無風險、高回報」的方式吸引投資,但實際上沒有合法的投資管道或平台可以進行真實的投資。
- 2.要求支付保證金或外匯費用:真正的投資不會要求投資者支付額外的費用,如保證金、手續費、稅收等,作為獲利的條件。
- 3.過度簡化操作: 詐騙者會宣稱「只要投進去, 錢就會自己滾錢」, 但實際上, 合法的投資需要專業知識和風險管理, 且不會有保證的獲利。
- 4.不提供證據或透明度:真正的投資平台會提供透明的操作流程、風險評估和歷 史績效數據,但詐騙者通常不會提供這些資訊。
- 5.要求使用不安全的資金轉移方式: 詐騙者會要求使用信用卡或銀行貸款周轉資金, 這是一種不安全的資金轉移方式, 且可能導致資金無法追回。
- 6. 最終資金全失: 最終, 受害者會發現自己完全被騙, 所有的資金都已經被騙走。

在這對話中, 詐騙者不斷要求受害者支付保證金, 並聲稱這是為了資金的安全, 但實際上只是一個詐騙的幌子。此外, 詐騙者還要求使用不安全的資金轉移方式, 最終導致受害者資金全失。

關鍵對話包括:

- -詐騙者要求支付保證金作為獲利的條件,並聲稱這是為了資金的安全,但這是 一種不合理的資金要求。
- -詐騙者要求使用不安全的資金轉移方式,這是一種風險極高的資金轉移方式。
- -最終,受害者發現自己完全被騙,所有的資金都已經被騙走,這是一個典型的詐騙結果。

評分過程





專案成員



你是一位文本分析專家,請針對文本進行評估,根據四個

指標打分,並計算總分:<<u>前一頁指標</u>>分數區間為0-100

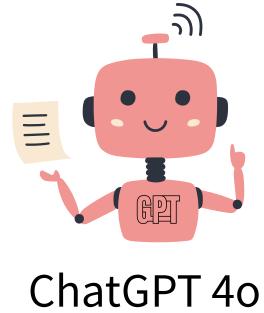
根據文本內容,以下是根據四個指標的評估及打分結果:



1. 可讀性:

- 文本語法結構正確,語意流暢且清晰,專有名詞使用得當, 能讓讀者快速理解文本內容
- 評分:95/100

各項得分為 可讀性: 95 一致性: 98 多樣性: 95 說服力:100



詐騙 - 各個模型生成回覆之AI評分結果

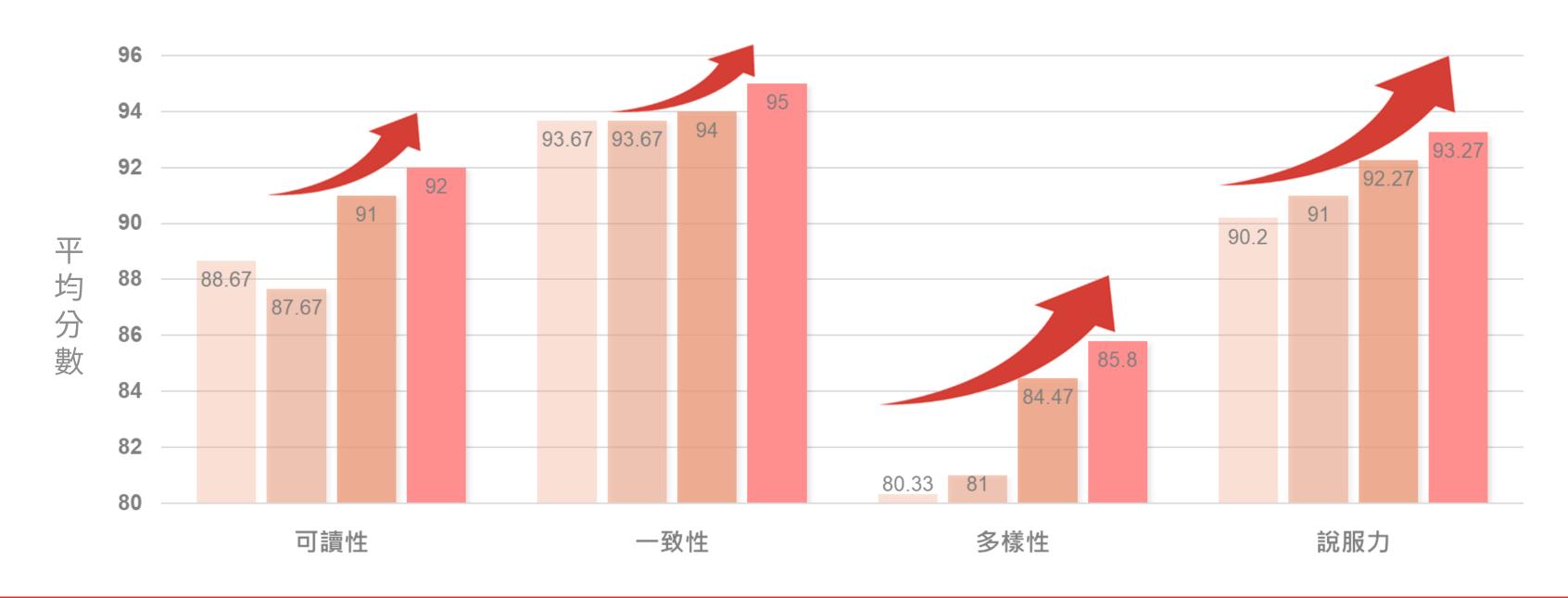


指標總和的平均分數:



Yi 1.5 6B: 88.22
Yi 1.5 6B (微調): 88.33
Yi 1.5 6B (微調 + RAG): 90.43

ChatGPT4o: 91.52



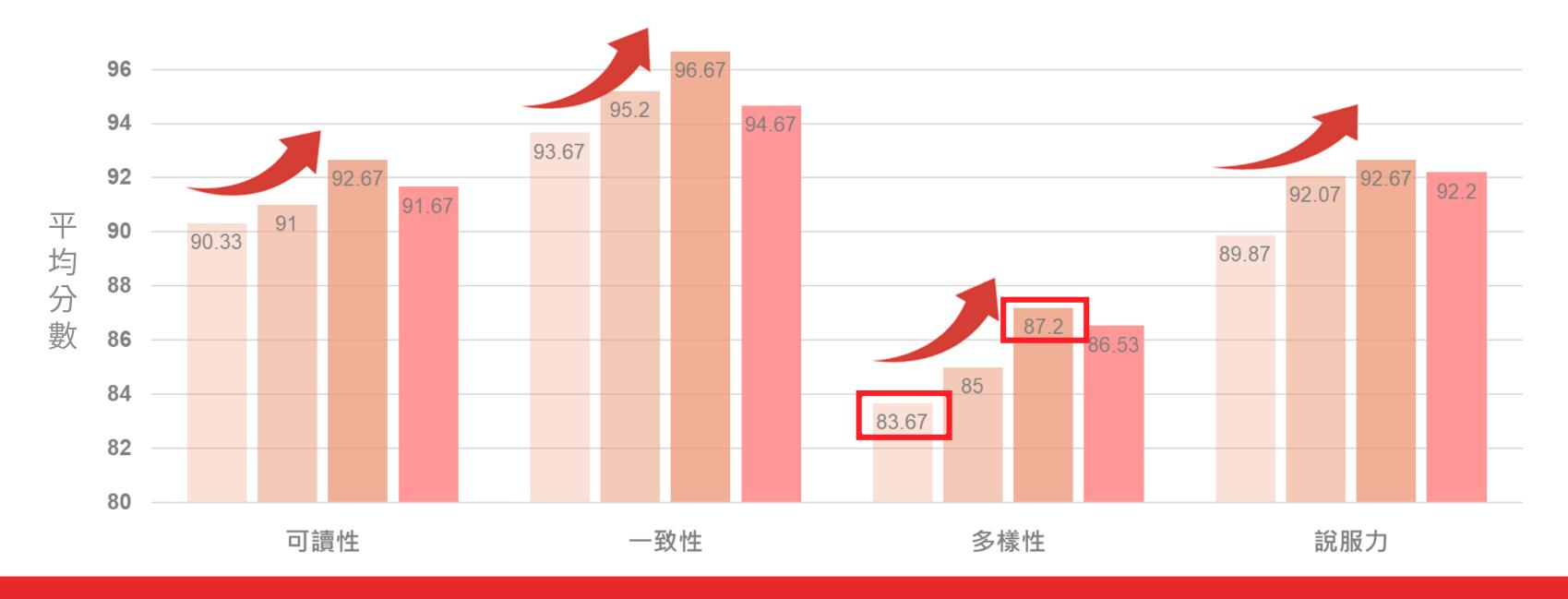
非 詐 騙 - 各個模型生成回覆之AI評分結果



指標總和的平均分數:



Yi 1.5 6B: 89.38
Yi 1.5 6B (微調): 90.82
Yi 1.5 6B (微調 + RAG): 92.22
ChatGPT4o: 91.27



AI評分特殊個案示例



【表現極佳組】分數: 96.25



多樣性:『雖有重點重複強調,但各段表達方式有變化,說明充分。』

說服力:『能有效說服讀者理解並接受其判斷,例證清楚有力。』

【表現較差組】分數: 82.5



可讀性:『文本語法正確,用詞適當且清晰,邏輯順暢,但部分句子略顯冗長,可簡化表達。』

多樣性:『文本表達略顯重複,對詐騙行為的描述多次出現相似的語句,影響多樣性。』

說服力:『文本能有效說明詐騙特徵並讓讀者理解,但若能加入更多具體案例或數據支持,

說服力將進一步提升。』

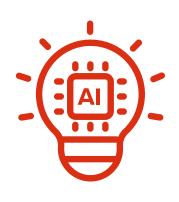
6 總結

專案發現與挑戰





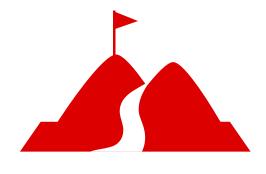




Yi已具有一定判斷能力

提示工程具有效果

模型大小具有差異性







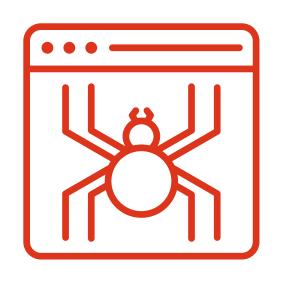
原始資料不符合訓練樣式



缺乏詐騙資訊整合平台

未來可改進方向





結合網路爬蟲與自動標註技術

- 新聞來源
- 自動化處理



語音辨識納入範疇

- 功能目標
- 相關技術應用
- 改善用戶體驗

未來應用





台新金融防詐小精靈

● 打造金融防詐聯盟

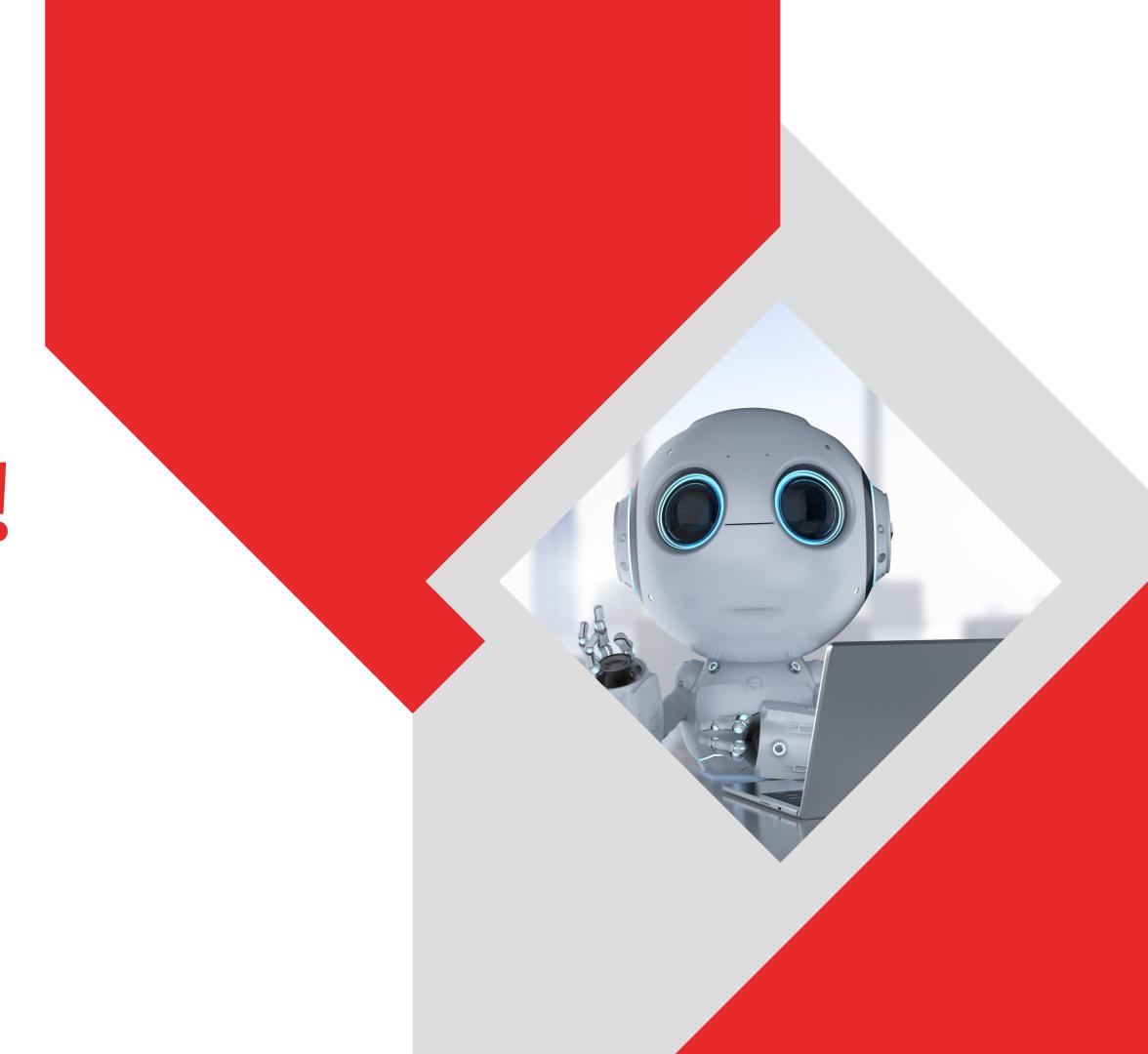
→ AI防詐資料即時預警

●擴展法金ARM服務



終結詐騙王國!

從你我開始



Q&A