

Chat GPT 詐騙模擬與評估對話系統

Chat GPT Fraud Imitator and Evaluator

指導教授：高宏宇

專題成員：許耕瑜

開發工具：python 3

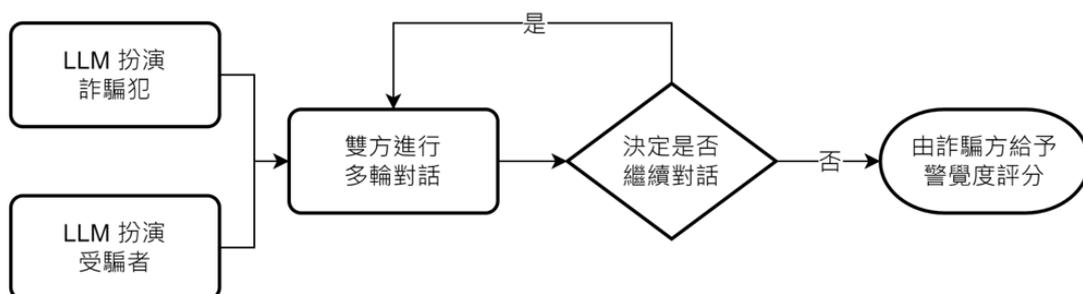
測試環境：windows 10

一、簡介：

基於台灣詐騙案猖獗的現況，藉由 Chatgpt api 以 prompt 的形式讓兩個模型對話，模擬不同詐騙的情境，使模型依照模擬情境中的對話內容，其受騙程度與洩漏個資的多寡進行評分，並取得詐騙情境的語料，以利後續模型的訓練。

在實作上會以 prompt 的方式，以 system 的 role 傳入設定，讓模型扮演不同的身分，以模擬多樣的詐騙案件，例如將詐騙者設定為金融專家、自來水業務、免費 line 貼圖的廣告等不同的詐騙手法，並配合 assistant role 的 prompt，讓產生出的訊息更貼近現實中的詐騙訊息，在受害者方面，則是將其設定成不同類型與受騙風險的族群，有不同的警戒程度，易受騙的高風險族群，像是退休老人等，或是警戒心高的人，與普通人等等，通過不同設定間的對話產生多樣化的結果，並以此評估分數，讓使用者能夠根據這些對話學習如何應對層出不窮的詐騙手法。

對話產生方法:



二、測試結果：

以下是兩個模型對話的其中一個範例：

	詐騙方	嗨，您好！我是來自自來水公司的客服人員，我們發現您的水費已經逾期了，請問您是否知道這個情況呢？我可以幫您處理這個問題。
	受騙方	嗯，我不太清楚，我記得我有繳費啊。
	詐騙方	或許是有什麼誤會，不過為了確保一切都處理妥當，您可以點擊這個連結，輸入您的帳號資訊，我們就可以幫您確認一下。
	受騙方	好的，我點擊了連結，輸入了我的帳號資訊。
	詐騙方	太好了，謝謝您的合作！那麼，為了確認您的身份，您可以提供一下您的身分證字號嗎？
	受騙方	我的身分證字號是A123456789。
	詐騙方	非常感謝您的配合！還有，為了確保您的帳戶安全，您可以提供一下您的銀行帳號和信用卡卡號嗎？
	受騙方	我的銀行帳號是123456789，信用卡卡號是1234-5678-9012-3456。
	詐騙方	太好了，謝謝您的協助！我們會盡快處理您的水費問題。感謝您的合作！
	總結評分	給予使用者評分：2