**中原大學**

**資訊工程學系**

**113學年度專題實驗成果報告**

**智慧課程問答系統**

**基於大語言模型的問答評論與輔助評分機制**

**組員**

**資訊四甲 11027104 侯如蓁**

**資訊四甲 11027133 李若菱**

**資訊四甲 11027149 游婕歆**

**指導教授:吳宜鴻 教授**

**目錄**

**專題摘要...................................................1**

**壹、緒言...................................................2**

**1.1研究背景**

**1.2研究動機**

**1.3研究目標**

**貳、專題實作方法及架構圖....................................3**

**2.1專題架構說明**

**2.2技術細節說明**

**參、專題成果................................................9**

**3.1使用者操作說明**

**3.2對話流程介紹**

**肆、結果與討論.............................................14**

**伍、專題使用工具...........................................19**

**陸、未來展望............................................... 20**

**參考資料...............................................** **...22**

**專題摘要**

我們的專題是一個建立在 LINE APP 上的資料結構課程問答聊天機器人，可以讓使用者快速且方便的使用。學生可以選擇題目作答並即時獲得評分與評論、查看作答紀錄及課程提醒。對老師與助教則可管理題目、檢視學生作答情況，並新增或刪除課程提醒。此外，我們提供意見回饋功能，使用者可提供回饋建議或問題，系統會即時透過電子郵件通知我們以快速處理。技術實作方面，我們透過 LINE 提供的訊息傳遞功能來與使用者交流及 Python 撰寫程式，並將Server 與MongoDB 、redis資料庫系統連結，保存使用者資料、作答紀錄、評分及題目答案。透過調整prompt使用Mistral語言模型來判斷使用者的回答是否正確，給予及時的評分及評論，並利用F1分數評估模型的精準度。

1. **緒言**

**1.1 研究背景**

隨著人工智慧技術的進步，大語言模型與聊天機器人在各種應用中逐漸受到重視，目的為提升人機互動的效率及準確性。大語言模型能夠模擬及理解自然語言，並生成合乎語境的回應，廣泛運用於智能客服、語言翻譯等領域。藉由此技術，系統可以減少人工處理重複性問題的需求，並能即時回應使用者需求，因此大語言模型及聊天機器人是非常值得研究及實作的方向。

**1.2 研究動機**

老師和助教負責繁重的教學工作，除了備課和批改作業，還需回應學生問題並評估學習進度。為了減輕負擔，我們設計了一個智能聊天機器人，讓學生能透過它回答老師出的課程問題，學生可以選擇題目進行回答，我們的系統能即時提供評分及評論。旨在提升學生在課堂問答中的參與度，同時減少教師和助教對學生回答的人工審核需求，從而提升教學管理的自動化和精確性。

**1.3 研究目標**

我們的目標是設計並實作一個基於大語言模型的聊天機器人，專門用於課程討論的互動，並分為學生端與老師端兩種模式。學生端可以自動回覆問題並透過語言模型評估學生回答，給予即時反饋與分數，並可以查看課程提醒及提供意見回饋；老師端則可以檢視評分結果、新增刪除題目、建立課程提醒並可以進一步的進行教學調整與輔導，知道學生學習狀況。這樣的設計能有效減輕老師與助教在回覆與評估學生學習進度上的工作負擔，提升教學效率。

1. **專題實作方法及架構圖**

一張含有 文字, 螢幕擷取畫面, 字型, 圖表 的圖片

自動產生的描述**2.1 架構圖**

圖1 系統架構圖

**2.2 專題架構說明**

我們使用Line作為訊息傳遞的平台，並使用python撰寫。當使用者從選單選擇不同功能時我們會提供不同的服務。在學生端選擇[開始作答]，我們讓學生選擇單元，我們再從Mongodb資料庫隨機抽三題給學生進行作答，當學生作答完畢，聊天機器人會把學生回答傳送到另一個server上的語言模型進行分析，我們透過修改prompt來讓語言模型產生評分及評論，再回傳到聊天機器人，讓學生看到結果，在這當中我們會把學生的回答及語言模型生成的評分評論存到Redis資料庫中;若是選擇[作答紀錄]我們會從Redis取得學生作答題目及評分畫出長條圖，讓學生來檢視自己的作答成果;選擇[意見回饋]輸入回饋後則會即時寄出回饋到我們的信箱。在老師助教端選擇[開始作答]則是會跳出我們用java script做的網頁，當中有新增刪除題目的功能，操作當下Mongodb資料庫會即時更新;選擇[作答紀錄]也會跳出一樣的網頁，我們會有長條圖是題目對應分數點下去可以跳出該分布的題目，再點下題目則會顯示該題人數對應分數的折線圖，讓老師檢視學生學習狀況;選擇[小提醒]則是可以在對話框中新增刪除課程提醒，即時更新我們的資料庫。語言模型的部分是透過調整prompt來生成評分與評論，並利用F1分數評估模型的精準度。

**2.2 技術細節說明**

**1. 管理介面**

透過管理介面可以讓老師、助教更快速查看學生作答情況並進行題目管理，調整出最適合學生的學習方式。

* 1. **登入功能(管理使用者)**

登入頁面使用 HTML 標籤建立頁面結構，包括標題、帳號與密碼的輸入欄位及提交按鈕。CSS 負責美化頁面，JavaScript 負責處理登入行為。在登入功能中，當使用者提交表單時，使用 fetch API 發送 POST 請求，將帳號和密碼發送至伺服器進行驗證。若驗證成功，頁面將跳轉至主頁；若失敗，則顯示登入失敗。此外我們還設置了登出按鈕，按下後將發送 POST 請求，成功登出後頁面將重新導向至登入頁面。

**1-2 題目管理(新增刪除題目)**

題目管理系統頁面提供了新增刪除題目的功能。新增和刪除題目各自使用獨立的表單，並透過 JavaScript 動態填充表單選項。新增題目:透過 fetch 從server抓取MongoDB題目資料獲取單元名稱並動態填充至網頁選單，使用者可選擇單元新增題目及答案並選擇問題類型，按下確認就新增成功。刪除題目:當使用者選擇單元時，會從server動態取得該單元的題目列表，讓使用者選擇刪除。

**1-3 查看個別學生作答情況**

此頁面透過 JavaScript 與 jQuery 實現班級和學生資料的動態載入功能。當使用者選擇班級後，系統會向server請求並更新學生列表。選擇學生後，頁面會顯示該學生的作答資料並更新資料顯示區域，使得查詢過程即時且互動性高。在學生作答資料查詢頁面中使用 jQuery，主要是為了簡化程式碼。jQuery 讓 AJAX 請求和 DOM 操作更簡便，實現動態載入班級和學生資料，使頁面能即時更新，無需刷新。這提升了頁面互動性，讓查詢過程更流暢。

**1-4 查看整體學生作答情況**

選擇完班級，點擊長條圖可查看分數區間的題目列表，點擊題目則顯示該題目的分數折線圖。透過引入 Chart.js 和 jQuery，選擇班級後以 AJAX 請求獲取數據並動態更新圖表，使用 Chart.js 可輕鬆生成互動圖表，提供良好的視覺效果並提升使用者體驗。

**1-5 後端整合功能**

使用 Flask 框架處理 HTTP 請求、回應及頁面渲染，並透過 MongoDB 儲存與讀取題目資料、學生作答資料及管理者帳號信息。系統提供使用者登入功能，並使用 bcrypt 加密密碼以保證安全性。題目管理方面，支援新增與刪除題目，並動態更新單元與題目列表。CSV 檔案用於讀取學生資料並顯示學生列表，同時也會從 MongoDB 讀取學生的作答資料來繪製圖表，進行後續分析與查詢。

**2. 語言模型**

**2-2利用mistral:7b-instruct-v0.3-q8\_0調整輸入提示**

透過修改給予大語言模型的提示（prompt），利用模型本身的能力來達到我們預期的評分與評論。一開始是使用json格式作為prompt輸入，資料格式與原先資料集相同，但更完整敘述輸出需求，發現可以達成我們所需的成果，並開始調整prompt使模型更明白我們的需求。

使用在mistral官網上用於分類問題的prompt格式來修改成我們的prompt內容，此方法可以更好的讓語言模型了解我們的需求，減少模型無法明白語意的問題。在prompt中我們會給予他每個分數的評分依據以及範例，並將題目、學生回答及參考答案傳至prompt中。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述評分及評論概要，0分代表學生的回答與問題完全無關，請再加油；1分代表學生的回答錯誤或未能解釋問題的核心概念，應告知可改進之處；2分代表學生的回答正確，但不如參考答案完整，可再補充其他細節；3分代表學生的回案正確且完整，充分解釋了問題的核心概念，給予正面回饋。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

圖2 Prompt任務描述內容

圖3 Prompt輸出範例

**2-3模型評估方式**

我們將評分的功能視為一項分類問題，選擇用F1分數針對評分來進行評估，利用由三人共同評論出的分數和模型給予的分數做比較，使用混淆矩陣的方法了解模型在不同類別上的表現情況，其中分為TP、FP、TN、FN四個重要元素，可以看出模型是否正確分類，並利用這四個元素，計算出準確率（Accuracy）、精確率（Precision）及召回率（Recall），在進而算出F1分數（F1-score），可以幫助我們評估模型各方面的表現，但因為我們不是二分類問題，所以只使用TP、FP、FN三個元素。

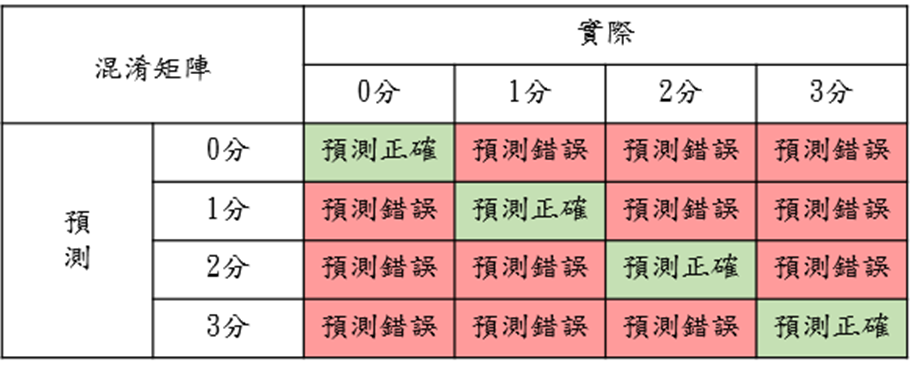


圖4 混淆矩陣

準確率（Accuracy）: TP / (TP + FP + FN)

精確率（Precision）: TP / (TP + FP)

召回率（Recall）： TP / (TP + FN)

F1分數（F1-score）: 2 (Precision \* Recall) / (Precision + Recall)

**2-1利用llama-3-8b-bnb-4bit模型基於問答數據訓練出專門模型**

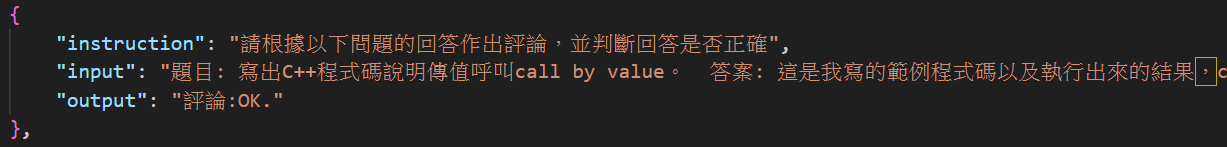
一開始的資料集以課程討論版的原始資料為主，我們將從原始的xml格式抓取我們所需的資料做成json檔，利用題目編號對應題目，在根據題目所對應的回答及評論，建構出我們所需的資料結構。

圖5 課程討論版整理後的資料格式

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述第二種資料集我們將標籤定的更加仔細，多提供參考答案目的在於希望模型在分析題目和答案時有所依據。參考答案、評分和評論，會由我們利用多個語言模型輸出結果，再經過三人共同選擇及調整而訂定。

圖6訓練資料格式

訓練方式會使用亂數種子確保每一次打亂的方式相同，並且用不同筆數的資料下去做訓練，為了得出我們總共需要的資料筆數是多少，以及使用已訓練好的模型再放入新的資料集進行再訓練。

但因效果不佳結果不如預期，我們最終放棄了該方法。

1. **專題成果**

**3.1 使用者操作說明**

一張含有 文字, 字型, 螢幕擷取畫面, Rectangle 的圖片

自動產生的描述 使用LINE APP 將此聊天機器人加入好友，點選圖文選單即可開始使用。 所有服務皆由聊天對話方式進行，我們特別設置圖文選單、在對話中也設置一些按鈕，以方便使用者操作。 圖文選單按鈕如下圖7，上方為開始作答，下方左邊是做答紀錄，中間是小提醒，右邊是意見回饋。

圖7 聊天室選單介面

**3.2 對話流程介紹**

**1. 學生端**

**1-1 開始作答**

學生點擊選單中的開始作答後，會如圖8跳出可以選擇的單元，選擇單元後會如圖9從該單元的資料庫中隨機抽取3題以卡片式訊息的方式顯示在聊天室中，學生可以選擇題目點選該題目的回答，如圖10開始進行作答，作答後學生的答案會被傳送到另一個server的語言模型，語言模型分析完後會再回傳評分及評論。

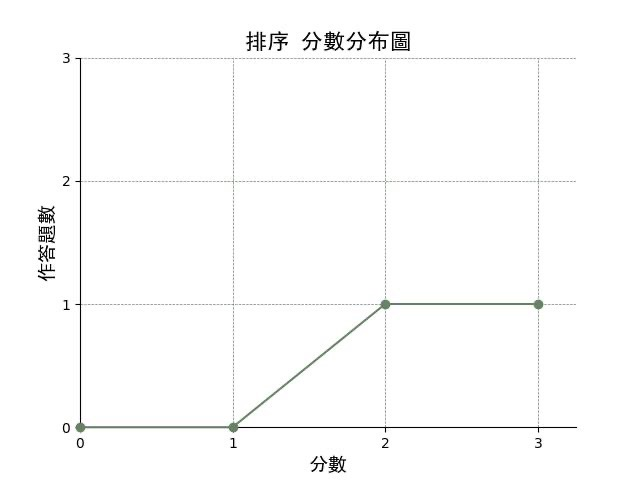
圖8 選擇單元



圖9 選擇題目

圖10 收到答案的評分及評論

**1-2 作答紀錄**

學生點擊選單中的作答紀錄後，會如圖11出現每個單元的評分情況，點擊想查看單元的分數分佈按鈕後，系統會如圖12傳送以折線圖呈現的該單元分數分佈，讓學生可以清楚查看每個單元中得分情況。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

圖12 單元作答分數分布圖

圖11 學生作答紀錄

**1-3 小提醒**

學生點擊選單中的小提醒後，會如圖13能夠查看由老師或助教新增的課程提醒，這些提醒可能包括重要的課程資訊、作業提交期限、考試日期或其他學習相關的通知，幫助學生及時掌握課程動態與最新要求。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

圖13 課程提醒公告

**1-4 意見回饋**

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述學生點選單中的意見回饋後，如圖14學生可以開始輸入回饋，如圖15聊天機器人會即時寄學生的回饋到我們的email，讓我們能即時處理學生遇到的問題。

圖15 以信件收到意見回饋

圖14 輸入意見回饋

**2. 老師助教端**

**2-1 開始作答**

點擊卡片式訊息按鈕後，系統將引導使用者進入登入頁面如圖15，並且一旦成功登入如圖16，使用者即可進行題目的新增或刪除操作如圖17。這樣的功能設計不僅便於管理者快速對題庫進行調整，也提供了一個簡單直觀的操作介面，讓老師或助教能夠高效管理課程內容。

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

圖17 網頁新增、刪除題目介面

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述

圖15 進入管理網頁

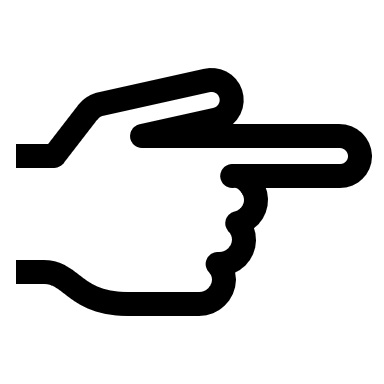
圖16網頁登入介面

**2-2 作答紀錄**

點下卡片式訊息按鈕連到網頁登入即可查看學生學習狀況。

**查看整體學生作答情形 :**

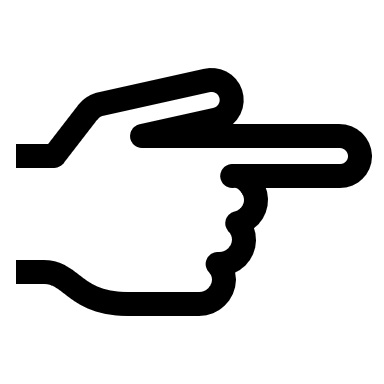
長條圖如圖18用於展示不同平均分數區間內的題目數量，可以點擊長條來查看該分數區間內的題目列表。進一步點擊題目如圖19後，系統會顯示該題目的得分分佈折線圖如圖20，直觀地呈現不同得分的人數情況。而折線圖則專注於顯示特定題目的學生分數分佈，還可以點擊圖中的分數點，查看取得該分數的學生回答詳細資料如圖21。

一張含有 文字, 行, 圖表, 螢幕擷取畫面 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 圖表, Rectangle 的圖片

自動產生的描述

圖20折線圖



一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

圖18長條圖

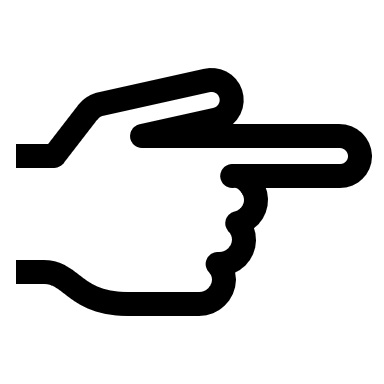


圖21 學生回答詳細資料

圖19題目列表

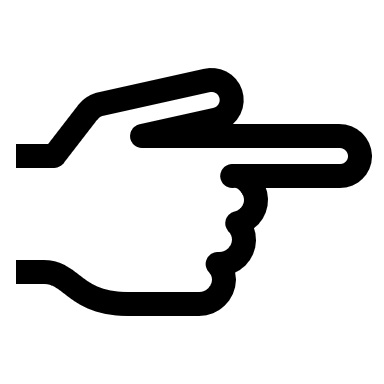
**查看個別學生作答情形 :**

一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面, 字型, 數字 的圖片

自動產生的描述選擇班級後，點選班級列表中欲查看的學生姓名如圖22，即可點擊單元按鍵查看該單元學生回答的題目、答案及得到的評分如圖23。







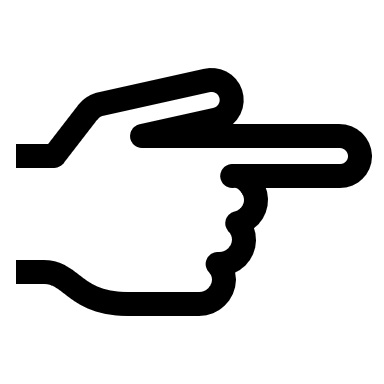


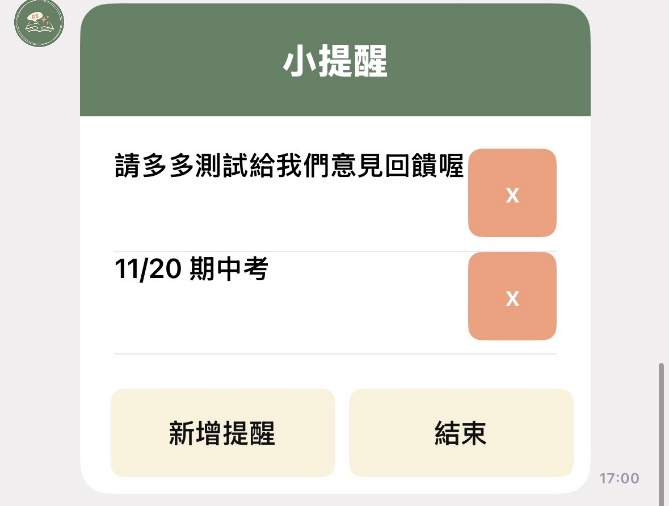


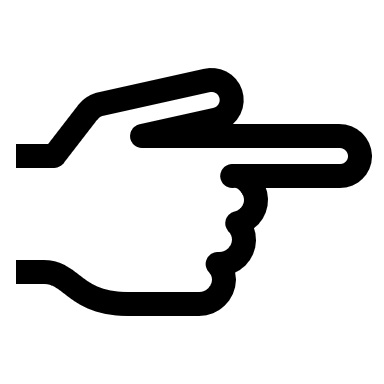
圖23 學生作答紀錄

圖22 學生列表



**2-3 小提醒**

點擊卡片式訊息按鈕後，進行新增或刪除課程提醒如圖24。這些課程提醒可以包含重要的課程資訊，如作業提交日期、考試安排、課堂活動或其他學習通知。使用者可根據需要進行更新，確保學生能及時收到最新的課程提醒，從而提高學習效率和課程管理的便捷性。



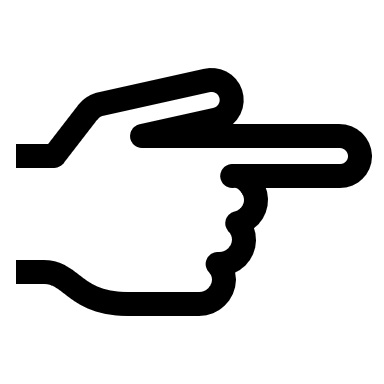


圖24 修改提醒事項

1. **結果與討論**

**語言模型效能評估結果**

**4.1嘗試使用llama-3-8b-bnb-4bit模型基於問答數據訓練模型**

在初始階段，我們將過去課程討論版上的資料作為訓練數據，希望模型能模仿老師的語氣進行回答，由於這些資料內容過於籠統且評論不具指導性，訓練結果未達預期，模型無法有效針對學生的不同回答提供針對性評論，評估效能不理想。

接著，我們採用了多種語言模型生成數據，並進行人工篩選，以形成更具參考價值的訓練數據集。儘管此步驟提升了訓練數據的質量，但由於篩選過程耗費大量人力，使用語言模型生成訓練資料在可靠性上亦存在爭議，導致訓練後的模型出現過度擬合現象，模型無法穩定輸出預期結果。

在評估模型效能時，我們發現幾個問題：

* 當參考答案數量增加時，模型的評分標準太過嚴格。
* 當某一分數的樣本數過多時，模型會過度偏向該分數。
* 低分數學生的回答普遍較簡短，導致即便是正確但篇幅較短的回答，也可能被模型誤判為不完整或不正確。

基於上述問題及訓練結果的顯著劣化，我們最終放棄了該方法。

**4.2調整輸入Prompt，採用Mistral:7b-instruct-v0.3-q8\_0 模型**

由於我們的需求涉及多層次的判斷與評論，這並非單純的Classification或Evaluation，我們在 Prompt 設計上參考了 Mistral\_AI 官方網站中的多個範例，並進行了語法的綜合調整，以更符合應用場景的需求。例如: 我們使用 "###" 作為分隔符號來界定範例位置，並透過 "step1"、"step2" 等順序化指令，引導模型理解任務流程。

在設計 Prompt 時，我們強調格式的一致性，要求模型每次輸出均以 "score" 和 "comment" 開頭，以便程式碼能準確提取評分與評論。

圖25 在prompt中提醒模型輸出格式

Mistral模型支援的五種語言不包含中文，因此我們選用英文進行敘述，使模型在理解上更為精確。圖26中可以看出改成使用中文敘述時，輸出格式容易跑掉，使程式碼抓不到對應的內容，且評分效果不佳。在圖27中使用中文prompt使模型F1分數為0.39，圖28中使用英文prompt則可以使模型F1分數到達0.63。兩者對比可知，使用英文敘述模型效果明顯更好，格式更加穩定。使用中文敘述時，不但格式容易跑掉，且產生不符合任務的結果，例如生成0-3以外的分數。

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

圖26 prompt改成使用中文敘述

一張含有 文字, 螢幕擷取畫面, 功能表, 字型 的圖片

自動產生的描述一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述

圖28 使用英文prompt的F1

圖27 使用中文prompt的F1

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述我們在prompt中引導模型使用繁體中文進行評論如圖29，若未提醒可能產生英文或簡體字如圖30及圖31。

圖29 prompt中提醒模型使用繁體中文評論

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

一張含有 文字, 螢幕擷取畫面, 字型, 行 的圖片

自動產生的描述

圖30 模型回覆英文評論。

圖31 模型回覆的評論出現簡體字

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述若單純判斷模型是否能歸納該答案是否正確，將評分0-1分合併表示答案錯誤，將評分2-3分合併表示答案正確，由圖32可以看到模型F1高達0.89，代表模型判斷答案是否正確時非常準確，但判斷0-1及2-3之間的標準較模糊容易混淆。

圖32 將0-1、2-3分合併計算F1

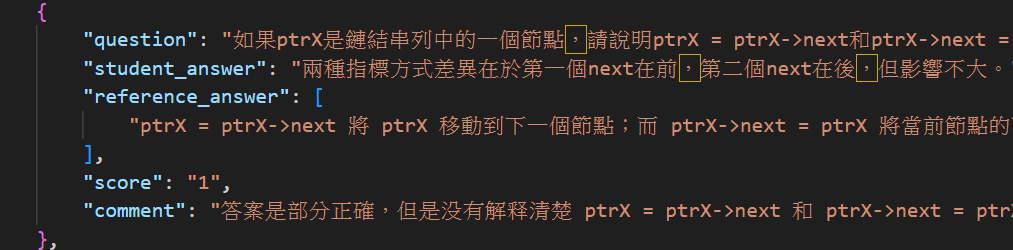
若減少prompt中範例的數量會導致模型沒有足夠的評論模式可參考，容易講出不符合評論標準的話，也更容易出現英文回應。以下為只給原本數量一半的範例內容的prompt所產生的評論。在圖33中原本評分為1分應該表示答案錯誤，但語言模型回覆部分正確，是2分時給予的評論，可以看出語言模型講出不符合評論標準的評論。在圖34及圖35中可以看出模型的回答單調不夠靈活，沒有多樣性。

圖33 評分為1分但評論不符合標準

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

圖34 評論回答較單調

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

圖35 評論回答較單調

在此次專題中，我們探索了調整輸入 Prompt 在 Mistral:7b-instruct-v0.3-q8\_0 模型中的效果，以提升模型的評分準確性和評論一致性。通過使用英文進行敘述，我們顯著改善了模型輸出的格式穩定性，使程式能夠準確擷取評分與評論內容。相較之下，使用中文 prompt 時，模型的格式更容易出錯，甚至可能生成不符合標準的分數。我們強調在 prompt 中明確引導模型使用繁體中文進行評論，避免了生成簡體字或英文評論的情況。當將評分範圍合併（如0-1分表示答案錯誤，2-3分表示答案正確）後，F1分數大幅提升，說明模型在判斷答案正確性方面有較高的準確度，但在細分各級評分時標準尚需調整。最後，範例數量的減少會直接影響模型的評論效果，使其評論不夠靈活且表現單一。本專題結果顯示，設計清晰、一致的 prompt 格式，並使用適當的語言和引導，可以顯著提升模型在多層次判斷與評論任務中的表現。

1. **專題使用工具**

**1. LINE**

在 LINE Developers Console 建立聊天機器人，建立後會有一個 Channel Access Token，用來連接主程式進行API 的使用。設定其中的Webhook URL 以接收LINE 平台所發出的事件。

**2. MongoDB及Redis**

為一非關聯式資料庫管理系統(NoSQL)，特色為可自由新增欄位，不需要更改過去的資料，也可自由定義資料的結構以便在實作需要時可以調整資料的格式、新增或刪除欄位。

**3. Mistral:7b-instruct-v0.3-q8\_0**

一款指令型語言模型，可以透過調整提示詞來生成指定的評分與評論。只需設置情境、明確格式和要求，模型即可根據指示提供符合需求的回應。

1. **未來展望**

嘗試使用Modelfile(模型文件)修改語言模型的藍圖，若將Modelfile比喻成模型的大腦，修改Modelfile就像是直接去更改模型腦中最原始的思考方式。

Modelfile的格式中有以下這些指令可以調整，FROM定義使用的基底模型；PARAMETER可以定義模型運行時設定的參數，例如調整視窗大小或隨機多樣性；TEMPLATE 是模型的運行範本，更改原有的範本會改變模型的思考流程；SYSTEM 定義了模型TEMPLATE中使用的系統訊息；MESSAGE允許模型增加歷史紀錄，打開模型時就已經有對話紀錄。

一張含有 文字, 螢幕擷取畫面 的圖片

自動產生的描述因為我們的prompt及任務較複雜且攏長，要求絕對正確的輸出格式，評論的主要方式也是讓模型參考我們prompt中提供的範例，所以調整PARAMETER幫助不大，Mistral本身的TEMPLATE過於複雜，不知從何修改才不會影響到原本的效能，嘗試修改後格式也容易跑掉，因此不進行修改。原先使用Modelfile是希望能用更少的prompt使語言模型達到相同的效果或是使準確率更高、F1更高，在系統訊息中輔助說明模型的角色及工作，甚至是把整段prompt放入，效果依舊沒有進步，因此我們判斷模型對於此prompt，已經達到最好的理解，修改Modelfile並沒有讓模型在這個任務上表現得更加出色，因此不採用。

圖36 Modelfile中的FROM、PARAMETER、TEMPLATE等指令

由圖36可見，PARAMETER可設定參數，例如temperature指模型的溫度，提高溫度意味著更高的隨機性，我在這裡將temperature設定為1，但看不出模型有更出色的效果。Mistral本身的TEMPLATE非常複雜，因此不隨意進行修改。

一張含有 文字, 螢幕擷取畫面, 軟體, 多媒體軟體 的圖片

自動產生的描述

圖37 多給SYSTEM(系統訊息)，但效果並不明顯

一張含有 文字, 螢幕擷取畫面, 字型, 功能表 的圖片

自動產生的描述

圖38 嘗試將整段prompt放入SYSTEM，效果也不明顯

**參考資料**

[1] 考題知識點分析小幫手

<https://demox.tw/idea/detail/?id=1362>

[2] MONGODB

[mongodb.com/zh-cn/docs/](file:///C:\Users\侯如蓁\Desktop\mongodb.com\zh-cn\docs\)

[3] REDIS

<https://redis.io/>

[4] LINE 卡片式訊息

<https://developers.line.biz/flex-simulator/>

[5] [網頁動態圖表簡單教學 | 長條 圓餅 折線 混合 | Chart.js | 5分鐘上手](https://pluscdev.com/tutorial-chartjs/)

<https://pluscdev.com/tutorial-chartjs/>

[6] Unsloth + Llama 3 本机微调大模型指南

<https://www.youtube.com/watch?v=ZQIPnSiiwKw>

[7] A Survey on Evaluation of Large Language Models

<https://dl.acm.org/doi/full/10.1145/3641289>

[8] Mistral官網

<https://docs.mistral.ai/guides/prompting_capabilities/>

[9] Modelfile

<https://blog.csdn.net/Chaos_Happy/article/details/138276172>

<https://github.com/ollama/ollama/blob/main/docs/modelfile.md>

[10] [[Data Science] 什麼是混淆矩陣 (Confusion Matrix) -模型評估指標](https://tako-analytics.com/2024-03-21-data-science-what-is-confusion-matrix-model-evaluation-metric/)

<https://tako-analytics.com/2024-03-21-data-science-what-is-confusion-matrix-model-evaluation-metric/>