

靜宜大學資訊工程學系畢業專題計畫書

一、封面內容包括：

專題名稱：以支援 AI 和聊天機器人為主的一個校園互動平台研究

指導教師：謝孟諺 老師

專題學生：

資工三 A 411000554 陳秉杰 s1100055@gm.pu.edu.tw

資工三 A 411004134 黃哲偉 s1100413@gm.pu.edu.tw

資工三 A 411017666 吳昀霖 s1101766@gm.pu.edu.tw

資工三 A 411017925 張致鋹 s1101792@gm.pu.edu.tw

繳交日期：2024/2/27

二、內容包括：

● 摘要

隨著數位科技的快速發展，資訊傳遞交流方式日益多元化，在學校中，學生面臨著資訊不對稱、資訊取得不易等挑戰，而現有的資訊平台仍存在著資訊分散、互動性不足等狀況，為解決上述問題，本研究提出一個全新的校園互動平台計劃。

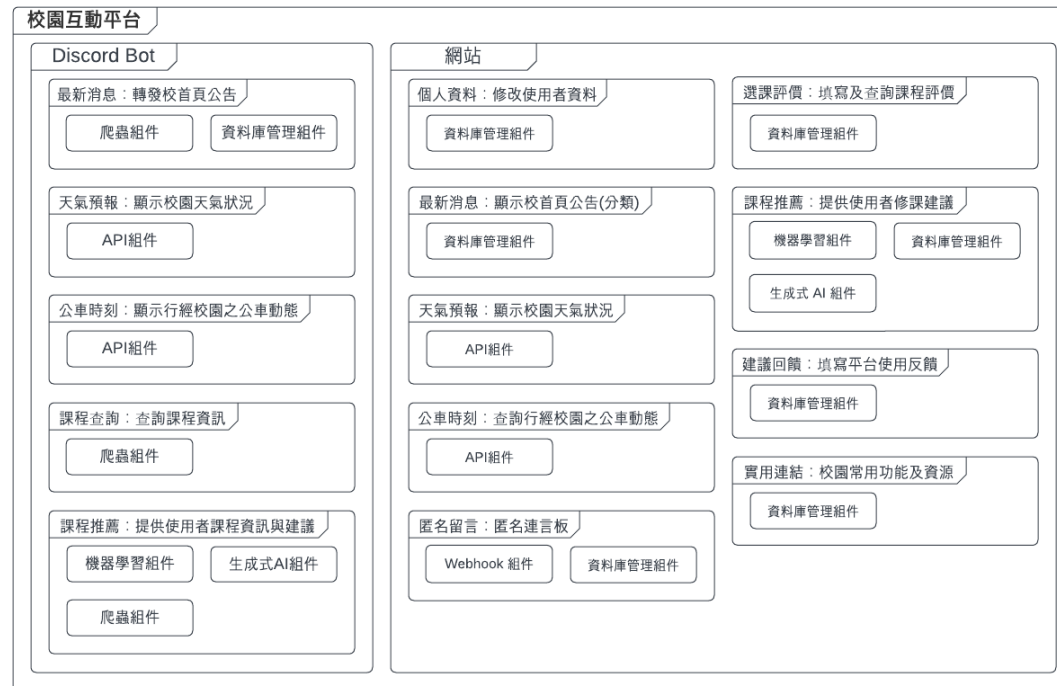
本研究計畫旨在於建立 Discord 伺服器 and 應用程式(APP)。其中 APP 支援聊天機器人，以及研究如何以較好的方式來引用生成式 AI 技術，建立一個有效且快速地提供校內師生們取得課程資訊以及相互交流的校園平台。

在此校園互動平台 (Discord 伺服器)，建立公共頻道和各科系頻道，促進資訊流通。透過機器人互動方式，將提供 (1) 最新消息、(2) 天氣狀況、(3) 公車動態 (4) 課程查詢、(5) 選課評價等功能。同時，深度地整合生成式 AI 技術，讓機器人能夠更有智慧地，並準確的回應使用者的需求，提供更個人化的互動體驗。除了 Discord 伺服器和 APP 外，也將功能發展出一個專屬的網頁平台，延伸機器人的各項功能，提供更多元的資訊和功能，滿足桌面上使用者的需求。

● 進行方法及步驟

1. 系統架構圖

圖一，是本計畫提出的系統架構圖，展示對校園互動平台的規劃。該平台主要分為兩大子系統：Discord 聊天機器人（Discord Bot）和網站。



圖一、校園互動平台之系統架構圖

在 Discord Bot 子系統中，共有 5 個模組，各個模組負責不同的功能。最新消息（模組 1）透過爬蟲組件從網頁上抓取最新消息並存入資料庫，使得用戶能夠及時獲取校園內的重要資訊。天氣預報（模組 2）則通過 API 組件向外部天氣 API 發送請求，提供即時的天氣預報。公車時刻（模組 3）和課程查詢（模組 4），同樣透過 API 組件或爬蟲組件提供公車時刻表和課程查詢功能，方便用戶查詢相關資訊。課程推薦（模組 5），則結合了爬蟲組件、機器學習組件和生成式 AI 組件，通過分析用戶的興趣和需求，提供個人化的課程查詢與推薦建議。

在網站子系統的部分同樣擁有多個模組，這些模組不僅提供了與 Discord Bot 子系統相似的功能，還擴展了一些額外的功能，如個人資料模組、選課評價模組、匿名留言模組等。這些模組通過資料庫管理組件來管理相應的數據，保證用戶的資料安全性和一致性。

2. 平台融入的前瞻技術

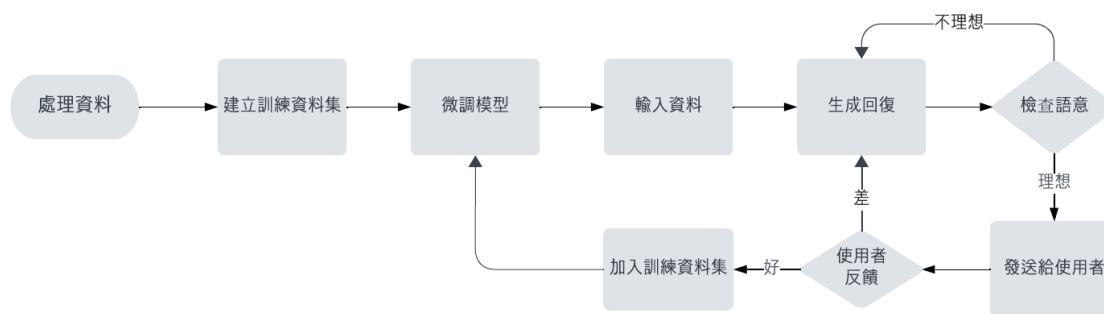
(1) 生成式 AI 組件

在使用生成式 AI 技術，特別是機器人問答和訊息生成方面，面臨著選擇適合的大型語言模型的挑戰。目前，主流模型包括 GPT-4 Turbo、Llama 2 和 Gemini Pro。這些模型各有優缺點，需要比較最符合需求的模型。

GPT-4 Turbo 擁有巨大的模型規模和廣泛的語言支援，但是它不開源且僅能透過 API 呼叫，使用成本相對較高。相比之下，Llama 2 雖然模型規模中等，但它的簡潔高效、易於訓練和部署，且不收費。而 Gemini Pro 具有強大的性能和易用性，且不收費，但模型規模相對較小且僅能透過 API 呼叫。綜合考量後，將選擇 GPT-4 Turbo 或 Gemini Pro 模型作為機器人訊息生成的基礎。

接下來，將著手處理資料和建立訓練資料集，例如、確定問題的主題範疇以及收集常見的使用者與機器人問答。透過這樣的資料集，能夠讓模型更深入地理解問題並生成合適的答案。

雖然模型經過微調後，仍無法保證其生成的內容都符合期望。因此，將透過自然語言檢查模型生成的答案，確保其品質。同時，積極收集使用者的反饋，並根據反饋不斷調整和優化模型，例如、若反饋為好的，就將其對話納入訓練資料集用於模型的微調；若反饋為差的，則透過模型重新生成新的內容。圖二，為強化生成式 AI 組件之流程圖，展示基於自然語言及使用者反饋調整生成式 AI 模型。

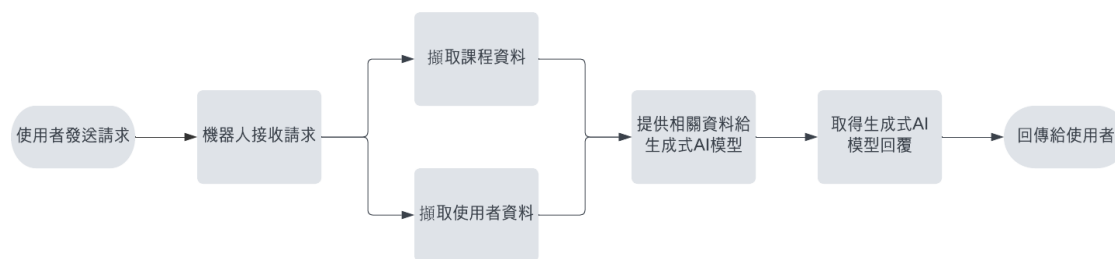


圖二、強化生成式 AI 組件之流程圖

以課程推薦功能（模組 5）為例，圖三，為應用生成式 AI 組件進行課程推薦之流程圖，展示使用者透過機器人取得課程資料的流程。機器人接收輸入或關鍵字訊息，透過資料庫中既有的資料提供回應資料，再將撈取到的回應資料，當作成輸入資料至生成式 AI 組件，針對使用者學習需求的個人化課程提供資料和建議。這使得機器人能以更自然的回應使用者，提供更精準的答覆。例如、當學生輸入“我想修『動態網頁設計 Django』程式，可以給我一點建議嗎？”，此時平台會（1）擷取關鍵字『動態網頁設計 Django』，透過查詢資料庫或爬蟲將既有課程資料當作參考資料，（2）平台會根據學生是否有修習過相關課程（『網頁前端程式設計』或『python 程式設

計』)，來產生適當訊息。最後將 (1)、(2) 所產生的訊息，結合一段將詢問意見的輸入問題，輸入至生成式 AI 組件。例如、輸入至生成式 AI 組件的訊息可能如下：

我已經修過 ["網頁前端程式設計","Python"] 課程，我是否具備修習 ["動態網頁設計 Django"] 課程的基礎？["動態網頁設計 Django"] 課程是否適合我？其中 ["動態網頁設計 Django"] 課程摘要如下，... (由資料庫撈出來或是爬蟲取得的參考資料)。



圖三、應用生成式 AI 組件進行課程推薦之流程圖

(2) 機器學習組件

首先，需要收集個人基本資料（如就讀科系、在學年級）、修課紀錄和職業興趣偏好，以及職能診斷報告。這些資料將以結構化的格式儲存在資料庫中。接著，進行資料預處理，包括數據清洗、去除不完整或不準確的資料，以及特徵選擇或提取，以將原始資料轉換為可供機器學習算法處理的格式。

然後，使用監督式機器學習方法來建立模型，使用決策樹或支持向量機模型，用於預測使用者對課程的興趣程度或推薦特定課程。

下面將分析上述兩種模型之比較：

支援向量機(Support Vector Machine)：

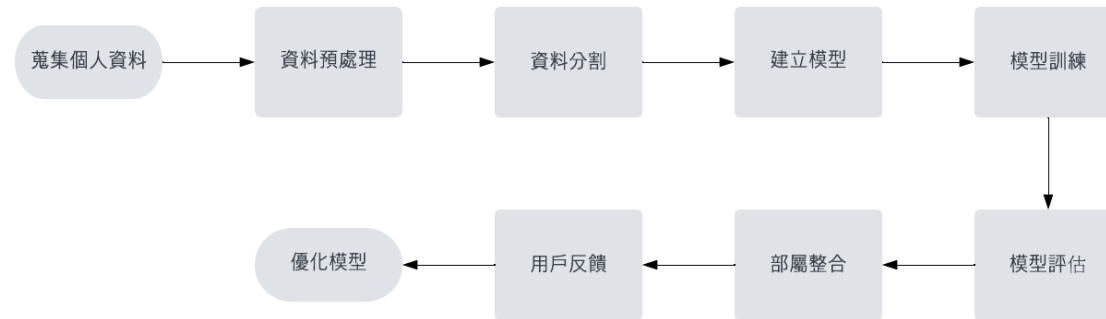
支援向量機(以下簡稱 SVM)是分析資料的機器學習算法，通過建構複雜的決策邊界來捕捉使用者的興趣模式。SVM 通常通過找到最佳的分隔超平面來將數據分類為不同的類別。在課程推薦的任務中，SVM 可以根據使用者的修課紀錄和興趣特徵來構建一個模型，從而將使用者分類為對某些課程感興趣和不感興趣的群體。由於 SVM 能夠處理高維度的資料和非線性問題，因此它可以有效地捕捉使用者的複雜興趣模式。然而，SVM 的計算成本較高，對於大型資料集可能需要更多的時間和計算資源。

決策樹(Decision Tree)：

決策樹是輔助決策的機器學習算法，通過將資料分割成不同的子集來建立一系列的分類規則，從而預測用戶對於某些課程的興趣程度。在課程推薦的任務中，決策樹可以幫助理解哪些特徵對於預測使用者興趣最為重要，並且提供了一個易於理解的模型結構。另外，決策樹對於混合型數據和缺失值的處理能力較強，並且不需要太多的數據預處理。然而，決策樹容易出現過度擬合的問題，特別是在處理高度非線性的問題時，因此可能需要進行樹的修剪或者使用集成學習方法來改進效果。

在訓練模型之前，需要進一步進行資料分割，以便將資料分成訓練集和測試集以進行驗證。接著進行模型訓練，將資料提供給模型，並通過調整模型的參數，以最小化預測與實際結果之間的誤差。在模型訓練完成後，使用測試集對模型進行評估。這可以通過計算不同評估指標（如準確度、F1 分數等）來完成，以評估模型的性能。

最後，將訓練好的模型部署並整合到課程推薦功能中。當使用者發出查詢時，系統將使用者的個人資料作為輸入，並通過模型進行預測，從而生成個人化的課程推薦結果。同時，為了不斷提升系統的準確性和用戶滿意度，需要持續收集使用者的反饋資料，並根據這些資料來優化模型。圖四，為機器學習組件運作之流程圖，展示從蒐集資料開始到訓練、優化模型的流程。



圖四、機器學習組件流程圖

(3) 爬蟲組件

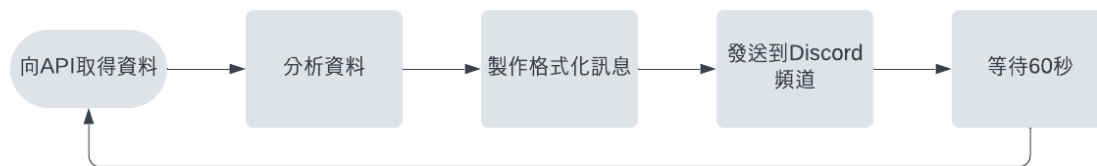
使用爬蟲技術週期性地收集和更新相關的資料，並將這些資訊整合到機器人的回應中，提供更即時和準確的資訊。以課程查詢（模組 4）為例，使用者發送課程搜尋請求，機器人會至學校的網站爬取課程資料（例如：課程簡介、評分方式... 等），再製作嵌入式訊息回傳給使用者。圖五，為爬蟲組件運作之流程圖，展示使用者利用機器人來取得如課程、校園相關資訊的流程。



圖五、爬蟲組件流程圖

(4) API 組件

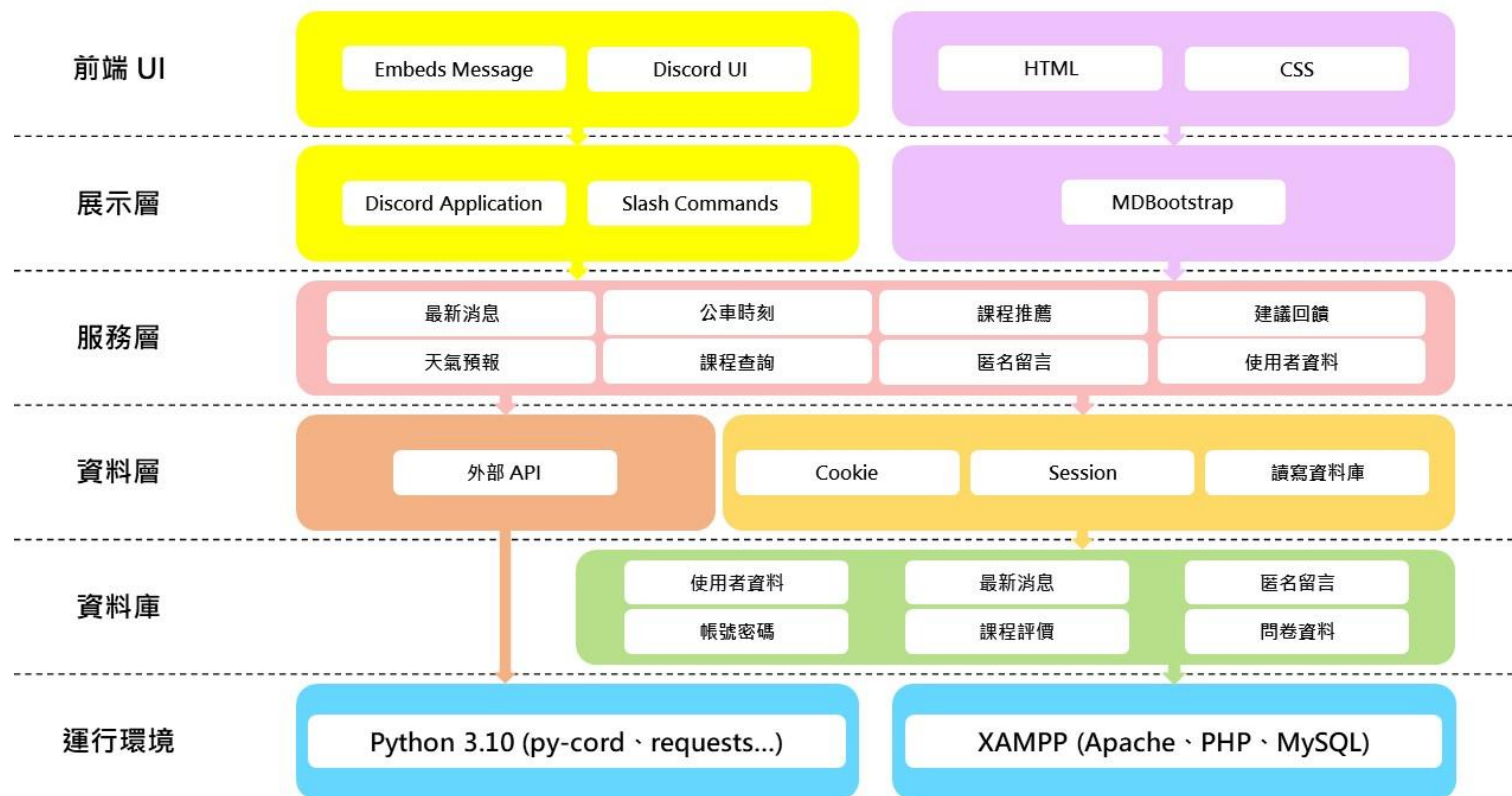
利用開放資料（OPEN DATA）來取得天氣資料、公車狀態等資料，以增強機器人的功能。透過 OPEN DATA 提供的公共資料，獲取各種實時或者預測性資訊。這些資料可以直接用於機器人的功能模組，獲取最新的資訊並提供給使用者。透過整合這些開放資料，我們可以使機器人提供更豐富、更實用的服務，提升使用者體驗。圖六，為 API 組件運作之流程圖，展示機器人從 API 取得資料並發送訊息的流程。



圖六、API 組件流程圖

3. 平台技術與部屬

圖七、為平台分層架構圖，展示網頁平台和聊天機器人的組成部分、使用的技術以及產生的資料與服務。



圖七、分層架構圖

(1) 網頁平台

運用 XAMPP 或類似工具在實驗室的電腦上搭建伺服器環境，包括 Apache 伺服器、MySQL 資料庫和 PHP 解析器。XAMPP 提供了一個完整的開發環境，這樣的方式能夠在開發和測試階段更靈活地進行操作，確保與 MySQL 資料庫的連接和操作順利進行，以儲存和檢索用戶資料。同時，利用 mdbootstrap 前端框架，快速建置美觀的網頁介面。這個框架提供了許多現成的 UI 元件和樣式，有助於提高開發效率和網頁的視覺吸引力，也能實現在不同尺寸和設備上提供一致的使用體驗。

在部署網頁應用之前，需要進行最後的工作。首先，需要購買網域名稱，以便使用者能夠通過網址訪問應用。網域名稱就像是應用在網際網路上的門牌號碼，方便使用者找到應用。再來，將申請一個 SSL 證書。SSL 證書是一種數字證書，用於在網際網路上建立安全的連接。通過將 SSL 證書用在網站上，可以實現 HTTPS 協議，這是一種安全的通訊協議，用於在使用者瀏覽器和網站之間進行加密通訊。這意味著使用者在與網站互動時，他們的敏感信息將被加密，從而保護他們的隱私和安全。

(2) 聊天機器人

使用 Python 作為主要開發語言。並整合 requests 和 beautifulsoup4 爬蟲套件這兩個套件的功能主要為，前者用於發送外部 API 以及其他請求，像是獲取天氣資訊、公車時刻、網頁原始碼等數據；後者從網頁中爬取所需的資料，例如最新消息、課程資料等，以供機器人使用。同時，使用 py-cord Discord 套件進行 Discord 聊天機器人的開發，此套件提供了與 Discord 伺服器進行連接和互動，提供多元化的功能，實現與用戶的互動和功能操作。此外，也整合 OpenAI 套件，如 GPT-4 Turbo 或 Gemini Pro，提高聊天機器人的對話理解能力和回應自然度。

- **設備需求** (硬體及軟體需求)

軟體需求: Xampp (Apache、MySQL、PHP)、Python

硬體需求: 伺服器 (i7-12700K/RTX3080/32G/1TB)

● 經費預算需求表 (執行中所需之經費項目單價明細)

項 目 名 稱	說 明	單 位	數 量	單 價	小 計	備 註
				臺幣(元)	臺幣(元)	
雜支	掃描、影印費、文具用品、國內研討會差旅費、郵電費	批	1	5000	5000	
消耗性器材	碳粉匣、墨水、影印紙、電腦週邊、隨身碟、硬碟、記憶體、3C 耗材	批	1	6000	6000	
書籍	相關專業書籍	本	4	800	3200	
網站域名	網域名稱註冊管理費	年	1	800	800	
資料使用費	付費 API(大型語言模型、TDX 等相關費用)	月	12	800	9600	自行負擔
論文發表費	國內論文註冊、出版、和出席論文發表等相關支出費用	批	1	5000	5000	
共 計					86100	

● 工作分配 (詳述參與人員分工)

陳秉杰：主機及伺服器環境設定、整合網頁及聊天機器人各項功能、測試

黃哲偉：網頁功能開發

吳昀霖：聊天機器人開發

張致鏞：聊天機器人開發

● 預期完成之工作項目及具體成果

當校園互動平台順利建置後，我們預期會達到多個顯著的成果。圖八，為平台介面設計之雛形，展示平台介面之整體架構與功能配置。首先，校園互動平台將成為校園內學生溝通的重要的管道，提供了即時、便捷的資訊服務。這有助於改善校園內的溝通效率，減少資訊不對稱和誤解，促進校園社群的凝聚力和合作氛圍。

透過機器學習技術提供的課程推薦功能，以及平台所提供的多元功能模組，例如最新消息、天氣預報、公車時刻表、校園活動資訊、社團交流等，期待能夠幫助使用者更有效率地選擇符合他們興趣和學習需求的課程，並提供除課程方面外更豐富的服務。使校園互動平台成為學生生活中不可或缺的一部分，並顯著提升滿意度和成效。

此外，校園互動平台的開放性和可擴展性也為未來的創新和發展提供更多可能性。校園社群的成員也能積極參與到平台的建設和發展，共同創造一個更加豐富、有價值和永續的校園互動生態系統。