陳定妤

系 別 / 年 級：資訊管理學系 / 四年級
學 校 輔 導 教 師：黃惒弘老師
實 習 機 構：大聯大控股股份有限公司
實 習 部 門：數據科學部 DSD
實習機構督導人員：Eric

大聯大控股股份有限公司

實習中印象最深刻的經驗

因為公司幾乎都是遠距上班，我們IT只有固定星期四是大家都要進入公司，剛開始蠻難以適應，因為剛接觸新環境，什麼都不知道，每天只會固定見到一位同事，而且因為我要去美國的關係，我2月3日就到公司報到，其他實習生2月17日才報到，所以我的教育訓練也是等了兩星期跟其他實習生一起開始。但待久了發現，大家要處理或交代事情都是直接線上meeting處理，還蠻方便的。

我的部門是 TSU-DSD 的 AI Team，剛開始幾天是在了解AI Team在做什麼，我們Team工作內容主要是幫公司內部的使用者開發一個wpgGPT，目前的功能有外部專家（外部專家資料源來自網際網路，類似 ChatGPT 或 Copilot 功能）、內部專家（內部專家來源於大聯大內部）、文案潤飾、文案摘要、文案媒推、會議摘要助手等功能，Mentor目前指派給我的工作主要是一些他們目前專案拆下來的小任務跟一些沒有人研究過的東西請我研究看看，這裡舉四個較大的成果來分享。

成果：

一、用azure openai api做一個聊天機器人，接著用pydantic 格式化輸出文章，將文章標題、日期、內文等資訊列出，當初做這個任務的時候因為對pydantic套件了解不深，查了很久的資料才開始，還有公司要的結構化數據大部分都要用json格式輸出，輸出文件則是用markdown格式。接著練習將寫好的程式打api，也就是測試這比請求是否能正常運作，這裡練習用postman的GET & POST功能。

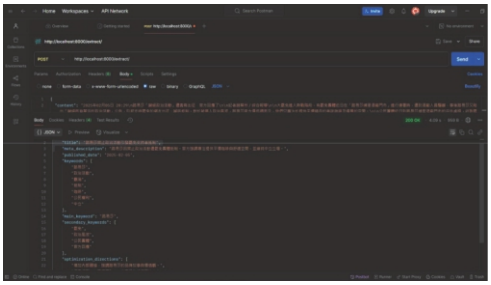

圖 1 postman打api成功畫面

二、需求是將外部專家中使用者詢問以及LLM回答的每筆資料呼叫azure openai api進行分類，先掃描Excel中的Question & Answer，再來下分類prompt真是一門學問，每調整prompt都會導致分類差別很大，這裡分成12類，重試機制是因為openai api有每分鐘的token限制，如果沒重試，這筆資料會被跳過，多線程可以一次處理多筆資料增加效率，傳回excel後，計算每一筆分類需要消耗多少tokens，這樣方便計算費用，下面為程式碼的流程。

圖 2 外部專家分類流程

跑完後將資料用Excel樞紐分析表交叉分析各欄位，得出分析結果，這部分主管有給我一些分析建議，可以往甚麼方面去延伸。

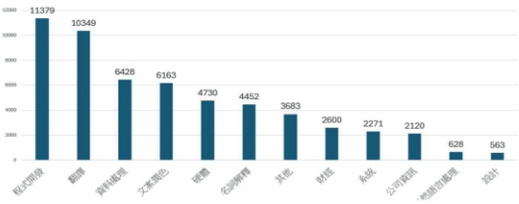

圖 3 各類別筆數

這個專案的後續，mentor有請我以這些類別為前提，寫一套system instruction，須具備全局性：System Instructions影響整個對話過程中的AI行為，而不僅僅是單個回應。規範性：指令明確規範AI應該如何回答問題或執行任務，確保一致性和專業性。角色設定：可以設定AI的角色和背景，使其回答更具針對性和專業性，我最後將十二個類別縮減到六個。

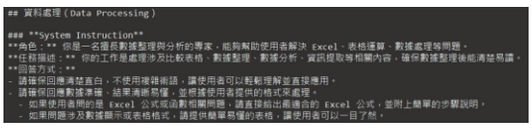

圖 4 System Instructions範例

三、研究apm效能監控，這部分使用Elastic cloud，需求是將mongoDB的data_chuck資料匯入，使用者輸入想查詢的資訊，用RAG搜尋相似文本後，找出最接近的向量所對應之文本，把文本 + user_msg 送給LLM回覆，最後透過apm監控每個步驟所花費的時間。

因為部門內沒有同事研究過這個，所以我所有功能都要自己研究，我不但找了ES Cloud官方，我還找了其他部門來幫忙。

四、需求：使用者輸入Data_package_id後，系統將對應的DataSource、DataChunk及DataLink資料刪除。

這個專案遇到最大的問題是邏輯很難處理，還有輸出資料要用json檔也是研究好久，因為很多層的情況下，有時候要用{ }，有時候要用[ ]，這都要慢慢研究，而且資料中的link_id_list欄位大部分是空值，這樣會導致data_link_id大部分資料都連結不到，結果會很醜。

講完專案，來分享一下在公司發生的有趣故事，我們部門在開發類似ChatGPT的東西，然後那時候功能開發到可以自己建立應用，主管們就迫不及待的想要建立一些有趣的應用，我們的大主管就想要建立一個他家狗狗反應的應用，因為他覺得他平常在上班的時候很想念他的狗狗，但因為那時候這個功能還在測試中，還不能用，所以主管就說：「我家狗狗不想理我」，我們就都在笑，主管突然就轉向離他最近的同事說：「你洗咧啥」，我們全部大爆笑，主管太可愛了。

還有一個是最近發生的事，我臨時進入一個別的部門的線上開會，會議快到尾聲的時候，主管就說是不是有人沒進來，會議主持人才想起來他忘記邀請那位同事，同事臨時被拉進來馬上要報告還一整個很矇，然後主持人邊說抱歉邊大笑，真的很有趣，我們主管管三個部門，竟然還能記得部門內都有誰，真的很厲害。

實習中最大的自我突破與成長

最大得突破還是跳脫舒適圈吧，之前就一直很想提早體驗看看職場到底都在做什麼，來了以後發現真的跟學校差好多，比如說學校是花錢學習，公司是邊賺錢邊學習，但也就是說要會自學很重要，畢竟同事課沒有這麼多時間教你，以前在學校報告時都是報告製作過程以及結果，但來公司以後發現主管只想知道結果，以及如何做更有效率，過程就顯得不是很重要了，而且因為我有紀錄每天做了什麼的習慣，主管聽到這件事時很驚訝，他說從來沒聽過有實習生會這樣做，還問我是不是很規律的人，不知道這樣是好是壞呢~然後因為公司使用的是scrum開發法，所以每天都有daily scurm要報告前一天做了什麼、今天打算做什麼，剛開始每天報告前都很緊張，到後來慢慢習慣這種報告方式，或許我趁機學習到很多溝通方式吧。還有公司每個月有一個IT Day，大概100多人左右，新來的要上去自我介紹，剛開始被部門同事恐嚇說可能會讓新人上去跳舞，害我很緊張，結果只是上去自我介紹就下來，但之後應該常常會面臨到要在很多人面前報告的情況，希望我能多學習一些應對技巧。在公司的期中報告結束後也學到很多，包括主管說得有時候學會問問題也很重要，問問題前先找答案，而不是一股腦地問，這樣反而學不到東西，還有主管強調很多次的作筆記，把之前的過錯與對策記錄下來，反覆釐清才能更熟練。

檢討與建議

做任何專案時需求一定要搞清楚，不要害怕詢問，我第一個禮拜來就試跑azure openai api，當時不知道api每次請求都要錢，傻傻地跑了21小時，我差點以為要賠在這裡，還好後來mentor說我沒有花到很多，而且我使用的模型很便宜;再來就是如果要動到公司資料庫的地方一定要備份，我當時在做第四個任務時，因為要刪除資料庫檔案，mentor是先幫我在資料庫中拷貝一份，又讓我拷貝一份到自己電腦，他自己的電腦也拷貝了一份，可謂是三重準備阿~看來真的很怕出事。

對學弟妹們的建議

這間公司是我當初104求職投遞的，也是我面試裡月薪給最高的一間，他雖然在臺北薪水不算最高，但真的是只要你學，你就學的到東西，這裡每年會收兩種工作的實習生分配到不同部門，分別是java助理程式設計師與數據助理程式設計師，我跟別的部門同事打聽，java那邊進去會考很多基礎考試，包括git、前端語法、sql等，有一個sop，通過考試才能開始接單；我這邊的部分入職大部分是做同事手邊任務拆下來的工作，幾乎用到的都是python語法，沒有要求你用甚麼方法完成任務，只要有成果就好，這裡相對較彈性，可以邊學邊做，就看想來的學弟妹喜歡哪種工作方式，這裡也不會讓你加班，下班馬上走都沒差，做六個月後可以遠距在家上班，每個星期四要來就好，真的是一個蠻自由的環境~但這裡訂便當超級貴的，除非你想每天吃全家，不然建議自己帶午餐。然後來這裡要自己帶筆電，建議筆電好一點的再來，我筆電在工作期間壞掉只能自掏腰包再買一臺，畢竟是要工作每天都要用到，但筆電帶來帶去其實真的蠻傷，下班帶回家記得要關機。