IR & GenAI Team3 Final 心得

A1105505 林彧頎

這次的專題過程中，我負責整體架構設計以及第一、第二題的主要工作。過程充滿挑戰，但也讓我學到許多寶貴的經驗，以下我將依序分享過程中所得到的收穫與發現。

首先本專題的核心在於設計一個完善且高效的架構，以完成題目要求。我首先花了較多時間深入理解題目，因為題目的描述相對抽象，同組同學對問題的理解也不完全，因此整體方向需要我來明確規劃。我利用架構圖的方式進行初步規劃，清晰地將資料流程、模組之間的互動以及最終目標具象化，這讓團隊能迅速聚焦於正確的實作方向。

然而，設計架構並非一蹴而就，過程中不斷面臨例如 Token 使用量不足、生成內容與資料庫關聯性不夠等問題。我們透過動態優化提示工程（Dynamic Prompt Engineering），結合 RAG（Retrieval-Augmented Generation）與 Self-Reflection 的概念進行調整，這使得架構更為精煉，並有效解決問題。例如，在第二題的資料檢索與生成模組中，我們加入了 Retrieval Grader 與 Hallucination Grader，避免了與問題無關或幻覺式的回應生成。

再來是第一題，我其實一開始不是負責此題目的，我只有和負責題目的同學解釋過題目的要求，後續讓他們自由發揮，等我完成我負責的部分，我在幫忙增強，我以為他們一開始會做到至少某部分的自動化，結果只是某種使用者介面與文字分析，這很明顯與題目不符，因此我決定直接主導進行大幅度修改，將原本分散的功能整合為模組化流程，並使用多次迭代的提示優化方法，提升結果的準確性與穩定性。例如，我們針對 Together API 的 Prompt 不斷優化，從初版到最終版本經歷多輪測試，確保能自動提取所需的查詢條件並準確地抓取台灣期貨交易所的數據。

這部分的調整不僅讓我們的實作更符合要求，也讓我更深刻體會到規劃正確方向的重要性。透過大改，我們成功建立了一套穩定、可行的全自動爬取與數據處理流程，最後輸出 Excel 檔案，為後續分析提供基礎(這項研究最終有結合RAG)。

再者是第二題，它的核心挑戰在於架構的明確化。基於我的規劃，我們選用 RAG 方法結合多模組生成技術，並在架構設計的基礎上進行多層次調整。起初，同學在模組細節上的理解不夠清楚，導致流程的連貫性不足。我針對每個模組的角色進行明確定義，例如將 LLM（Large Language Model）模組的任務設定為生成與檢索相輔相成，並進一步優化動態提示生成邏輯。

整體流程從 Query Analysis 開始，經過多輪檢索與反饋優化（Self-Reflection），確保生成結果達到連貫性、創意性及相關性的要求。特別是在評估模組中，我們引入 Sentence-BERT 模型進行結構化評估，量化每次生成的結果，這極大提高了整體的精確性與效率。

這次的專題讓我對整體規劃的重要性有了更深刻的體悟。無論是第一題的大方向修正，還是第二題架構的細化，正確的規劃不僅能提高效率，更能幫助團隊在執行過程中保持一致性。此外，我發現，生成式 AI 雖然有一定的能力，但仍需要人類的介入，特別是在初始方向及迭代優化的過程中。同時，我也了解到團隊合作中溝通的重要性。當同學在技術細節或方向上有所迷茫時，我能迅速協助釐清問題並提出解決方案，這也促進了我們整體專題的順利推進。

最後，我認為這次專題的挑戰不僅在於技術本身，更在於如何平衡自動化流程與人工介入，並在有限的資源與時間內達成最佳結果。這些經驗對我未來的學術研究與實務應用都將有所幫助。