

元智大學資管系
學術類畢業專題頂石課程(二)
期末報告

生成式 AI 在升學備審文件製作之輔助初探

Exploratory Investigation into the Application of Generative AI for Enhancing the Preparation
of University Admission Documents

姓名：林芝仔

分組代號：ZE2

指導教授：邱昭彰 博士

中華民國 114 年 11 月

November, 2025

目錄

壹、摘要.....	3
貳、研究動機與研究問題.....	3
一、研究背景與動機.....	3
二、研究目的.....	4
1. 智能檢索大學招生簡章與審查標準指引.....	4
2. 提供自傳與學習歷程自述之撰寫建議.....	5
三、研究挑戰.....	5
1. 資料擷取與處理的挑戰.....	5
2. 語意檢索與文本生成的挑戰.....	5
參、文獻探討與回顧.....	5
肆、研究方法與步驟.....	7
一、研究架構.....	7
1. 需求分析與功能規劃.....	7
2. 採用之技術.....	7
二、系統實現.....	8
1. 系統架構.....	8
2. 技術實現.....	9
伍、實驗結果與系統展示.....	10
一、系統設計與功能實現.....	10
1. 備審資料指引查詢模組.....	10

2. 備審段落生成模組.....	12
二、技術應用與成效分析.....	15
1. Skeleton of Thought (SoT)與多階段生成流程設計.....	15
2. RAG 技術應用與資料檢索優化.....	18
3. 個性化生成與評估指標分析.....	19
4. 整體成效.....	19
陸、 結論.....	20
柒、 參考文獻.....	21
附錄 A. 專題工作內容.....	23
附錄 B. 專題心得與建議.....	24

壹、摘要

臺灣現有的大學多元入學制度強調備審資料，然而學生面臨各校系招生簡章與審查標準資訊分散、整合困難的挑戰，耗時且影響申請品質。輔導老師亦難即時掌握最新資訊。生成式 AI 雖具潛力，但其「幻覺」問題限制了應用的可靠性。

因此本研究旨在開發一套自動化的個人申請 AI 輔導系統，希望達成以下目的：(一) 利用智能檢索技術，整合並提供即時的大學招生簡章與審查標準；(二) 運用生成式 AI，依據真實的校系要求，提供學生個人化的自傳與學習歷程撰寫建議。

本研究建構三層式架構，其中資料層透過網頁爬蟲與 OCR 技術擷取並結構化招生文件，應用層導入檢索輔助生成(RAG)技術，藉由 LangChain 與向量資料庫，將建構真實簡章知識基礎在大型語言模型(LLM)上，以緩解幻覺，用戶層則採用「思維骨架」(SoT)生成流程，先建構大綱，再透過多模型並行生成，確保內容邏輯。

經樣本測試，導入 SoT 與 RAG 的架構顯著提升文本品質。系統生成內容的「邏輯性與結構」得分高，模型「幻覺率」顯著降至 4.5%，內容一致性達 93%。並行生成策略亦證實兼具效率。本研究成功開發一站式備審資料輔助系統，證實結合 RAG 與 SoT 技術的生成式 AI，能有效解決升學輔導中資訊分散與內容品質的痛點，提供兼具效率與事實性的實用工具。

貳、研究動機與研究問題

一、研究背景與動機

台灣大學多元入學制度強調個人申請時需提交學習歷程檔案與備審資料，然而學生常因資訊分散而必須從不同網站查詢各校校系分則資訊(圖一)與審查資料準備指引(圖二)，導致資料整合困難與準備時間過長(葉玉玲&丁一顧, 2021)。由於各校對備審文件要求不盡相同，學生在撰寫自傳、學習歷程與讀書計畫時也常感到困惑，此外，輔導老師在協助學生準備備審資料時，往往面臨各校系招生資訊更新不及時與資料分散的問題，使得老師亦難以根據最新審查標準及趨勢提供有效輔導(江宜芷&林子斌, 2020)。

元智大學 資訊管理學系		學測、英聽篩選方式		甄選總成績採計方式及佔總成績比例				甄選總成績同分參酌之順序	
		第一階段		第二階段					
		篩選	學測成績採計方式	佔甄選總成績比例	指定項目		檢定		佔甄選總成績比例
項次	審查資料	項目內容代碼對應							
1	修課紀錄 (第 11)	A. 修課紀錄		*1.00 *1.00 *1.00	40%	審查資料 面試	-- --	30% 30%	一、面試 二、審查資料 三、學測國英數B級分總和
2	課程學習 成果 (第 12)	D. 自然科學領域探究實作成果，或特殊類型領域之相關課程學習成果 E. 社會領域探究活動成果，或特殊類型領域之相關課程學習成果							離島外加名額縣市別限制 (無)
3	多元表現 (第 13)	F. 高中自主學習計畫與成果 G. 社團活動經驗 H. 積極幹部經驗 I. 服務學習經驗 J. 競賽表現 K. 非修課紀錄之成果作品 L. 研究發現 M. 特殊優良表現證明 N. 多元表現暨心得							備紀錄(A)、課程學習成果(B・C)、多元表現(G・H・M・N)、學習歷程自述(O・Q) 項目內容請參照本簡章「甄、分則」乙、審查資料項目內容對照表(第20頁)。
4	學習歷程 自述 (第 14)	O. 高中學習歷程反思 P. 自我評鑑 Q. 未來學習計畫與生涯規劃							「面試準備指引」請至元智大學首頁→招生資訊網頁查詢。
5	其他 (第 15)	R. (統系自行輸入限 10 字) S. (統系自行輸入限 10 字) T. (統系自行輸入限 10 字)							5類1名，優先錄取達最低錄取標準之中低收入戶或低收入戶或特殊境遇家庭或新住民子女學生(相關證明繳交方式詳見「事項說明」)。
同級分(分數)超額篩選方式		一、學測國文、英文、數學B之級分總和 二、學測英文級分							
備註		龐大數據、人工智慧與資訊安全等對資訊管理專才之需求，本系致力於： 1. 人工智慧及大數據分析能力； 2. 資安與網路攻防能力； 3. 業界專業實習，培養學生就業軟實力； 4. 海外移地教學，提升學生國際移動力； 5. 參加學術競賽，提升學生專業競爭力； 6. 自主學習社群，激發學生創新能力。 網址：www.mis.yzu.edu.tw 聯絡方式：(03)4638800分機2604。							

圖 2-1-1 校系分則資訊 資料來源：大學甄選入學委員會

(040112)資訊管理學系 審查資料準備指引 下載檔案 面試準備指引 下載檔案 校系網頁 請點選 校系分則 請點選	
元智大學 資訊管理學系 113 學年度 申請入學 審查資料 準備指引	
學系定位與特色： 1.本系隸屬於資訊學院，教師專長涵蓋資訊技術、計量決策、資訊管理、企業管理及社會創新等領域。本系以大數據分析及人工智慧應用為研究發展主軸。並以商業智慧、生醫資訊、社群媒體分析與智慧製造作為特色應用領域。 2.因應產業對大數據、人工智慧與資訊安全的資管專才需求，本系致力於： (1)開設人工智慧及資安跨領域課程，建構學生大數據分析及資安防護能力； (2)強調學用合一，指導學生參加國內外競賽，強化學生實作能力，提升專業競爭力； (3)歷史悠久的專業實習制度，平均每位學生2個以上業界實習機會，高達60%學生至全國前一千大知名企業實習； (4)多元的國際交流資源，如：美國羅德島大學及英國艾塞克斯大學雙聯計畫、海外移地教學及海外參訪活動，培養學生宏觀國際視野，強化國際移動力； (5)豐富的學生活動，如：學生自主學習社群、實戰工作坊、資訊週、資管之夜等。 (6)眾多傑出校友任職於國內外知名企業，如：Microsoft、台積電、日月光、緯晶科技、Salesforce、SAP、遠東集團相關企業等。	
審查資料評核能力	學習與資訊科技能力、溝通表達與合作能力、自我精進與問題解決能力

圖 2-1-2 審查資料準備指引 資料來源：元智大學招生資訊網

現行備審資料的準備方式高度依賴學生的個人資訊檢索與經驗累積，這不僅增加了準備的困難度，也可能因資訊錯誤或不完整而影響申請結果(陶宏麟 & 吳澤玫, 2019)。由於各校審查標準與學系特色不同，學生往往需要大量時間查閱公告，在面對龐雜資訊時可能產生資訊過載，進而錯失適合的申請機會(Bengio, Courville, & Vincent, 2013; Lan et al., 2024)。

近年來，生成式 AI 技術的發展使得資料整合、文本生成與個人化輔導變得可行(Devlin et al., 2018; Lewis et al., 2020)。透過網頁爬蟲技術(Web Scraping)與智能檢索的結合，可以有效解決資訊分散的問題，提供一站式的個人申請輔導方案，幫助學生快速掌握各校招生資訊，提升備審資料的完整度與競爭力(Gao et al., 2024)。然而，傳統的文本生成模型(如 GPT-4、LLaMA 3)仍然面臨「幻覺」(Hallucination)問題，即模型可能產生不準確或不符合事實的內容，影響其作為備審資料建議工具的可靠性(Pintrich, 2004; Proceedings of the 2019 ACL Workshop, 2019)。為了解決此問題，本計畫引入 RAG 增強檢索技術，透過「檢索+生成」的方式，確保 AI 生成的內容基於真實的招生簡章與學系標準，從而提升準確性與可信度(Bengio, Courville, & Vincent, 2013; Hou, 2024)。

二、研究目的

本計畫旨在開發一個自動化的個人申請 AI 輔導系統，運用網頁爬蟲技術擷取分散於不同網站的招生簡章與審查標準指引，並透過檢視申請者個人相關資料，並藉生成式 AI 機制產生備審資料及自傳與學習歷程自述撰寫建議。此外本計畫也將開發智能檢索機制，讓使用者有效查詢招生簡章與審查標準指引內容，以提升學生的校系申請文件準備效率。具體目標如下：

1. 智能檢索大學招生簡章與審查標準指引

利用爬蟲技術擷取各大學校系的招生簡章與審查要求之內容，並藉 RAG 增強檢索技術，開發一個智能輔助查詢系統以提高使用者查詢效率。

2. 提供自傳與學習歷程自述之撰寫建議

透過生成式 AI 基礎，並根據科系提供之審查標準指引，自動生成適合該校系申請需求的撰寫建議，幫助學生精準撰寫申請文件，以符合校系要求，提高科系錄取機率。

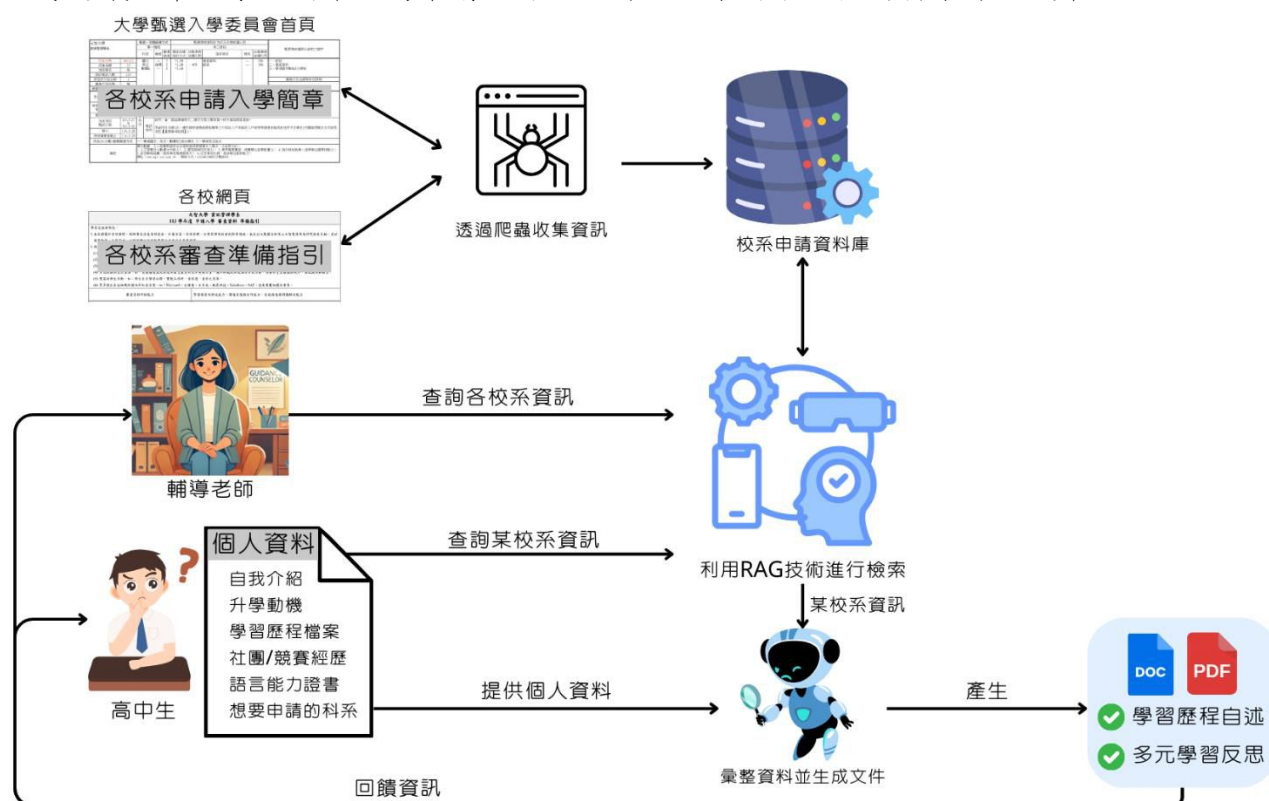


圖 2-2-1 提供自傳與學習歷程自述之撰寫建議示意圖

三、研究挑戰

1. 資料擷取與處理的挑戰

目前各大學的招生簡章與審查標準分布於不同網站，且格式各異(如 HTML、PDF、Excel)，部分學校甚至可能採用掃描文件的 PDF，使得解析難度增加。本研究需透過網頁爬蟲技術(Scrapy / Selenium)與 OCR(光學字符辨識, Tesseract OCR)處理非結構化數據，確保資料可被正確解析與索引。

2. 語意檢索與文本生成的挑戰

不同學校、學系對學生的學習歷程與備審資料有不同的資料要求(江宜芷&林子斌, 2020; 陶宏麟&吳澤玟, 2019)，格式也略有不同。本研究需透過 RAG 檢索技術，將招生簡章與審查資料要求結合，並透過生成式 AI 機制，生成適合該校系申請需求的撰寫建議。

參、文獻探討與回顧

台灣自實施多元入學制度以來，個人申請於高等教育選才機制中扮演日益重要的角色。其中，個人申請尤其強調學習歷程的展現，包括修課紀錄、自主學習計畫與競賽參與等多樣化資料。然而，這些資料常散落於不同網頁與檔案，學生在彙整與整理時往往面臨資訊分散、不易管理以及缺乏有效指導等困境(葉玉玲&丁一顧, 2021)。同時，研究顯示學生的學習歷程表現會

影響大學審查結果，但若缺乏輔助工具，學生必須投入大量時間與心力準備申請資料(江宜芷&林子斌, 2020)。從自我調節學習(Self-Regulated Learning, SRL)觀點來看，若能提供適當且高效率的備審建議，不僅有助於學生規劃學習方向，也能改善其備審資料的整體質量(李浩仲等, 2016; Pintrich, 2004)。

近年來，生成式人工智慧(AI)技術的迅速發展為解決上述問題提供了新的契機。透過結合大數據、機器學習與深度學習方法，AI 能夠自動擷取並整合來自不同來源的資訊，進而支持文本生成與個人化建議的應用(Bengio & Courville, 2013；Hou, 2024)。例如，Transformer 模型的提出(Vaswani & Shazeer, 2017；Allard, 2019)使得自然語言處理取得突破性進展，而 BERT 等預訓練模型(Devlin & Chang, 2018)則進一步提升了文本理解與生成的效果。儘管如此，傳統生成模型在知識密集型任務中仍存在「幻覺」現象，可能產生不準確或與事實不符的內容(OpenAI, 2023；Pintrich, 2004)，因此需要進一步強化機制來提升文本生成的準確性與可信度。

在此背景下，檢索式輔助生成(Retrieval-Augmented Generation, RAG)技術受到廣泛關注。該技術通過結合外部知識檢索與內部文本生成，能夠根據用戶查詢從大規模知識庫中提取相關文檔，並將這些信息作為輔助資料輸入生成模型(Lewis & Perez, 2020；Gao & Xiong, 2024)。此方法不僅顯著降低了生成文本中的偏誤與虛構現象，還提升了生成內容的事實依據和靈活性。相關平台如 LangChain(LangChain Developers, 2023)進一步提供了高效串接 RAG 技術的框架，而對向量化資料庫技術的研究則為高精度檢索提供了理論與實踐依據(Han & Liu, 2023)。此外，近期一些研究也開始探索多模態檢索與生成應用，如利用圖像、文本等多來源資料進行融合，進一步提升系統在複雜查詢場景下的效能(Yasunaga & Aghajanyan, 2023；Tang & Zhang, 2025)。圖四中可見檔案讀取與切分、文向量儲存與檢索，以及最後在本地 LLM 進行生成的完整過程，能有效強化模型回答時的事實依據，以降低幻覺問題。

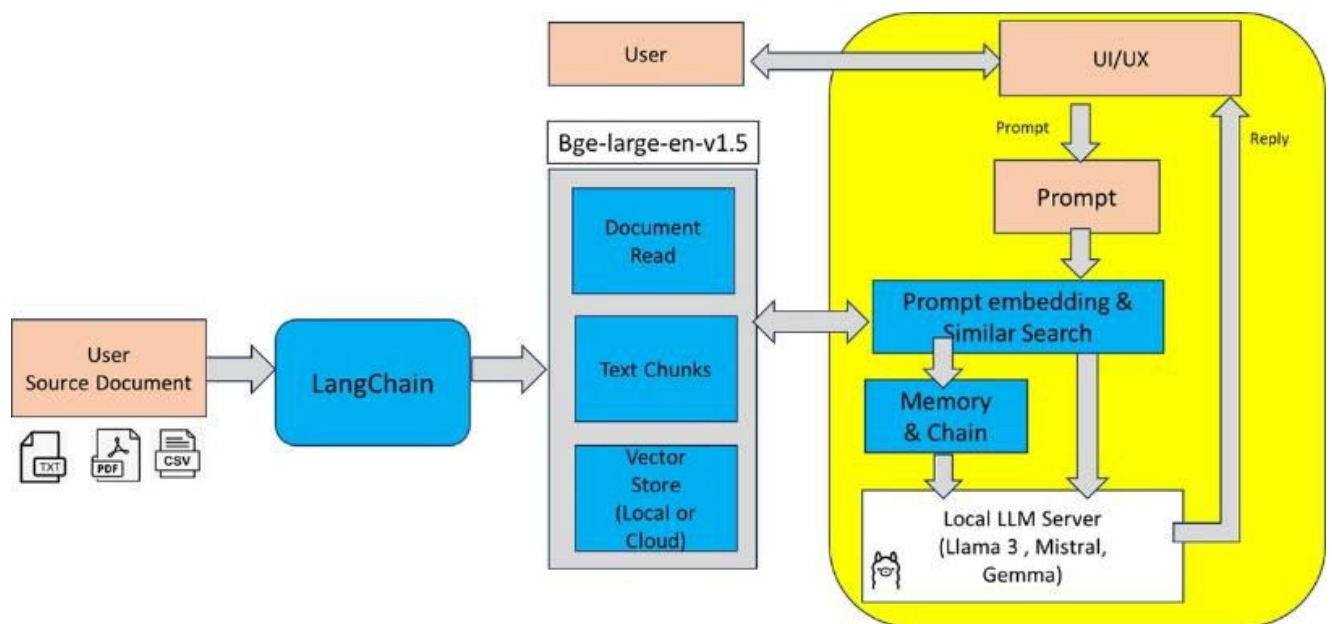


圖 3-1-1 RAG 架構示意圖 資料來源：CIO Taiwan(2024)

最後，AI 文本生成與 RAG 的發展為資訊檢索與文本生成技術帶來了更高的準確性與靈活性。因此，本計畫將聚焦於如何整合爬蟲技術與 RAG 技術，運用網頁爬蟲技術擷取分散於不同網站的招生簡章與審查標準指引，並透過生成式 AI 機制產生個人化的備審資料、自傳與學習歷程自述撰寫建議，以協助考生提升申請文件準備的效率與品質；同時為輔導老師提供精準的招生資訊查詢工具，達到雙向優化輔導流程的目的。

肆、研究方法與步驟

一、研究架構

1. 需求分析與功能規劃

- (1) 蒐集招生資訊：利用自動化資料處理技術，將散落於各大學招生網站、公告系統與官方平台中的招生簡章、審查標準與備審資料解析成結構化數據，便於後續資料檢索與運用。
- (2) 生成個人化撰寫建議：基於生成式 AI 與檢索輔助生成(RAG)技術，依據各校系最新審查標準，提供學生個人化的備審資料撰寫建議，幫助學生更有效地展現自身學習歷程。
- (3) 輔導老師端介面：輔導老師端介面除了提供基本的招生資訊查詢外，還將整合數據分析模塊，呈現各校系的歷史變化趨勢、錄取數據等。

2. 採用之技術

(1) 資料處理模組

資料處理模組負責統整與解析各大學招生簡章、審查標準及備審資料要求等格式各異的文件。透過標準化的解析流程，系統可自動從 PDF 等格式中提取關鍵文本內容，並轉換為結構化資料格式，如 JSON 或資料庫表格，內容涵蓋學校名稱、科系要求、備審文件類型、提交期限與評量標準等欄位。

這些結構化資料作為後續「檢索增強生成」(Retrieval-Augmented Generation, RAG) 技術的基礎，支援語言模型進行向量化處理與相似度檢索，以生成貼合目標學校需求的個人化備審資料建議。此模組不僅提升資料處理效率，亦確保生成內容具有針對性與一致性。

(2) 文本生成與檢索輔助生成(RAG)模組

模組的理論基礎主要源自 Retrieval-Augmented Generation (RAG) 技術，其核心思想在於將信息檢索與文本生成有效結合。根據 Gao et al. (2024) 的研究，RAG 模型在整合外部知識檢索與內部文本生成方面展現了顯著優勢。為了實現這一技術，本研究引入了 Langchain 平台作為核心串接框架，該平台利用向量化技術，從大規模知識庫中精確提取與用戶查詢高度相關的文檔片段，並將這些檢索結果與查詢上下文一併輸入生成器中，從而提高生成文本的準確性。(圖六)中，Langchain 進行 RAG 串接的架構圖顯示，檢索器依據相似度計算與向量匹配機制，從知識庫中提取出高相關性的信息，作為生成

器的重要輔助資料；而生成器則依托多模態自回歸生成架構，結合檢索結果與查詢上下文進行自回歸文本生成。Lewis et al. (2020)的實證研究亦證明，在知識密集型任務中，採用 RAG 模型能顯著降低生成文本中的偏誤與虛構現象，從而提高整體生成品質。

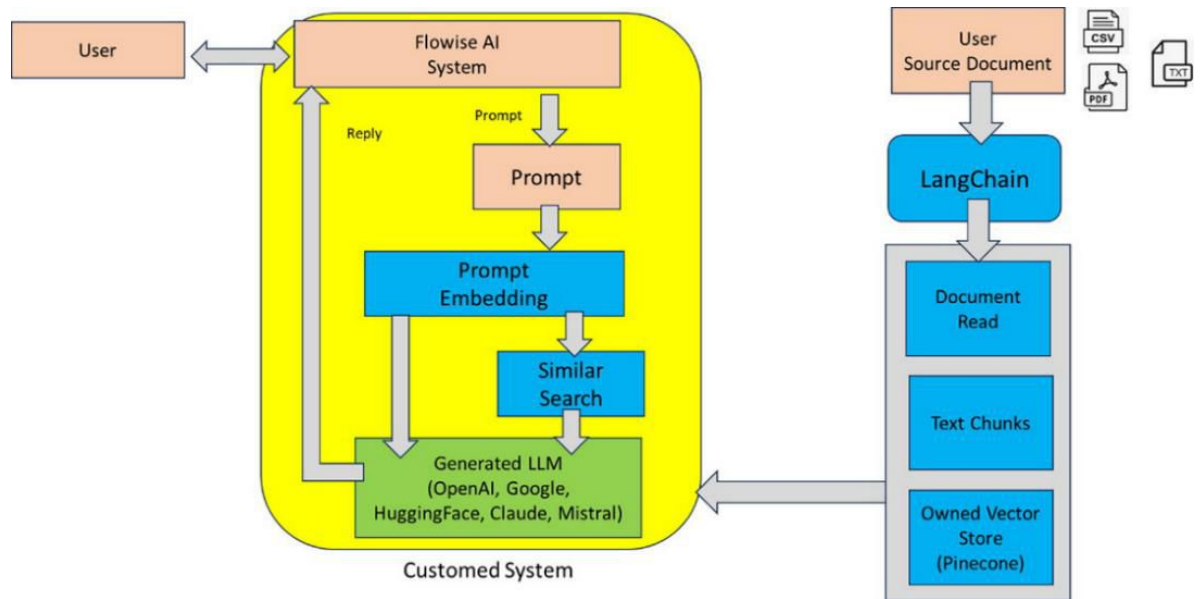


圖 4-1-1 Langchain 進行 RAG 串接的架構 圖片來源：CIO Taiwan(2024)

二、系統實現

1. 系統架構

本計畫所開發的系統架構(圖七)可分為資料處理層(Data Layer)、應用層(Application Layer)及使用者介面層(User Interface Layer)。系統架構概述如下：

- (1) 資料處理層：負責擷取招生簡章、各校審查資料準備指引，將 HTML、PDF 等格式的非結構化數據解析為可存取的結構化資料，並將所有處理後的數據將儲存於資料庫中。
- (2) 應用層(Application Layer)：負責資訊檢索、資料整合及生成建議。根據學生輸入的查詢條件，從資料庫檢索目標校系的招生需求與審查標準，並透過 LangChain 結合 RAG 技術，依據查詢結果產出備審資料撰寫建議。
- (3) 使用者介面層(User Interface Layer)：提供直觀介面，學生可查詢校系需求、上傳個人學習歷程並獲取生成建議；輔導老師則可查詢各校系的招生資訊與審查標準，以協助學生申請。

用戶層	1.學生介面:提供查詢科系需求、上傳資料、查看建議等功能。 2.管理員介面:用於更新數據。
應用層	1.資料檢索模組：從資料庫中查詢相關科系的需求。 2.資料整合模組：將多來源需求進行統一處理。 3.建議生成模組：基於科系需求，生成撰寫建議。
數據層	1.結構化數據庫：存儲學校、科系需求的文字內容。 2.非結構化數據庫：存儲學生上傳的證書和圖片（如 AWS S3）。

圖 4-2-1 系統架構示意圖

2. 技術實現

(1) 資料處理層(Data Layer)

資料處理層負責自動化擷取與解析各大學招生簡章、審查標準及備審資料規範。系統採用 PyMuPDF、pdfplumber 等 PDF 解析工具，針對非純文字 PDF 則結合 Tesseract OCR 技術進行影像文字辨識，確保各類型文件皆能高效擷取內容。解析後，透過正規表達式(Regex)與自然語言處理(NLP)技術，自動抽取學校名稱、科系、文件要求、繳交期限、評分標準等關鍵欄位。所有資訊經標準化轉換為結構化格式(如 JSON 或關聯式資料表)，並存入 PostgreSQL 資料庫，資料表設計兼顧一致性與擴充性，方便後續檢索、分析與串接向量資料庫。每筆資料均標註來源、處理時間與信心指標，確保可追溯性與品質控管。

(2) 應用層(Application Layer)

應用層核心在於資訊檢索、資料整合與個人化建議生成。系統根據學生輸入的查詢條件，從資料庫檢索目標校系的招生需求與審查標準，並結合 LangChain 平台與 RAG(Retrieval-Augmented Generation)技術，將查詢結果與用戶輸入一併向量化，通過 FAISS 等向量資料庫進行相似度檢索，提取與查詢高度相關的文檔片段。這些檢索結果與查詢上下文共同作為輸入，傳遞至 OpenAI GPT-4 或 Google Gemini 等大型語言模型，進行自回歸文本生成。系統設計多層 Prompt 模板，確保生成內容具備結構性、針對性與真實性，並可根據需求自動產生表格、心智圖等多元內容。應用層亦負責用戶管理、權限控管、歷史紀錄、配額查詢等功能，確保系統安全與穩定運作。

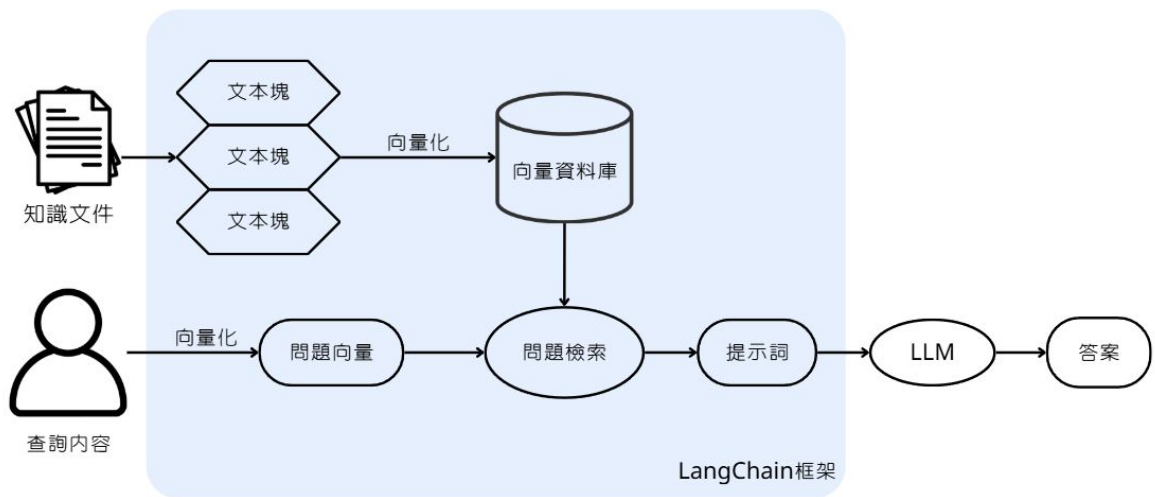


圖 4-2-2 RAG 應用流程圖

(3) 使用者介面層(User Interface Layer)

介面層以 Flask 結合 Jinja2 模板引擎實現，設計簡潔、操作直觀。學生可於介面查詢校系需求、上傳個人學習歷程並獲取自動生成的備審建議；輔導老師則可查詢各校系招生資訊與審查標準，協助學生申請。主要頁面包括登入、註冊、備審查詢、段落生成、會員管理等，並支援即時回饋、段落重新生成、內容精簡、心智圖生成、Word 檔匯出等多元功能。所有操作流程皆有明確提示，並集中於 profile 頁面進行管理，提升用戶體驗與資料管理便利性。前端與後端透過 API 進行資料交換，確保資料安全與即時性。

伍、實驗結果與系統展示

一、系統設計與功能實現

1. 備審資料指引查詢模組

本系統設計了備審資料指引查詢模組，目的是協助學生快速獲取各大學、各學系的備審資料準備重點與官方建議。使用者可於網頁介面選擇學校與學系，系統即自動查詢後端資料庫，回傳該系所的審查重點、準備指引、學系特色等結構化資訊。這些資料來源於自動化文件解析與結構化建檔，確保內容一致且具備專業性。查詢結果頁面支援表格、分段重點顯示，並可收藏、下載，提升資訊取得效率。後端 API 設計上，所有查詢請求皆經過權限驗證，確保資料安全。此模組大幅降低學生自行搜尋與比對各校系備審要求的時間與困難，提升備審準備的精準度與便利性。

學校： 學系：

☒ 需繳交資料 ☐ 學系特色 ☐ 準備建議

A: 修課紀錄

B: 書面報告

C: 實作作品

G: 社團活動經驗

- 1.社團活動經驗：參採校內、外社團活動(與語文、數學、科技等任一領域相關者尤佳)參與證明，並說明參與心得。

H: 幹部經驗

圖 5-1-1 備審準備指引查詢介面

學校： 學系：

☒ 需繳交資料 ☒ 學系特色 ☐ 準備建議

一、學系特色

教師專長涵蓋資訊技術、計量決策、資訊管理等領域。

以大數據分析及人工智慧應用為研究發展主軸。

特色應用領域包含商業智慧、生醫資訊等。

開設人工智慧及資安跨領域課程。

提供業界實習機會，高比例學生至知名企業實習。

提供雙聯計畫、海外移地教學等多元國際交流資源。

二、培養重點

學習與資訊科技能力。

溝通表達與合作能力。

自我精進與問題解決能力。

具備大數據分析及資安防護能力。

強化學生實作能力，提升專業競爭力。

培養宏觀國際視野，強化國際移動力。

圖 5-1-2 學系特色與培養重點列點呈現

需繳交資料

學系特色

準備建議

一、適合強調的能力或特質

學習與資訊科技能力：強調在語文、數學、科技領域的學習成果，展現對資訊領域的興趣。

溝通表達與合作能力：說明參與社團、擔任幹部等經驗，體現團隊合作和溝通能力。

自我精進與問題解決能力：展示參與競賽、專案等經驗，說明如何解決問題、持續學習。

國際視野：若有參與國際交流活動，強調國際移動力及文化適應能力。

二、建議呈現方式與撰寫方向

課程學習成果：選擇語文、數學或科技相關課程報告或作品，連結到資訊管理領域的應用。

多元表現：參與與資訊科技相關的社團活動、競賽或營隊，說明學習心得與收穫。

學習歷程自述：反思高中學習經驗，說明重要成就、挑戰及對自身影響，展現對資管的熱忱。

未來學習計畫：結合系所特色，說明大學四年學習規劃及畢業後生涯規劃，展現明確目標。

經濟不利背景：若屬經濟不利，可闡述對資訊科技的濃厚興趣及克服困難的具體行動。

三、可避免的常見問題

避免過於空泛的描述，需具體事例佐證，例如參加過的競賽、專案等。

避免抄襲網路範本，應展現個人獨特性，說明個人特質如何與資管系契合。

避免過度強調學業成績，應著重於學習過程中的收穫與成長，展現學習熱忱。

避免未來規劃與系所關聯性不足，應深入了解系所特色，提出具體的學習計畫。

避免篇幅過長，應精簡扼要地呈現重點，突出個人優勢。

圖 5-1-3 各學系備審資料準備建議

2. 備審段落生成模組

系統結合 OpenAI GPT-4o-mini 與 Google Gemini-2.0-Flash 等大型語言模型，設計多層提示詞(prompt engineering)，根據用戶輸入、學系需求與資料庫內容，動態生成個人化、針對性強的備審段落。為提升生成內容的專業性與多樣性，系統導入 RAG(Retrieval-Augmented Generation)技術，將本地向量資料庫(FAISS)中相似學系、歷屆優秀範例作為輔助上下文，與用戶輸入一同組合進入 LLM，顯著提升生成內容的真實性與針對性。生成結果可即時預覽、重新生成，並支援 Word 檔下載與歷史紀錄管理，方便學生反覆修訂。

欲生成段落

多元表現綜整

請選擇要生成的段落

課程學習成果

多元表現綜整

學習歷程反思

學習動機

未來計畫

圖 5-1-4 配合台灣大學升學制度，各段落有不同的生成器

12

G - 社團活動經驗

1. 社團活動經驗：參採校內、外社團活動(與語文、數學、科技等任一領域相關者尤佳)參與證明，並說明參與心得。

請輸入社團活動經驗相關內容...

0/300 字

圖 5-1-5 提供引導式介面，引導學生分項填寫自身資料

開始識別 **清除全部**

✓ 已處理 6 個文件 (內容已自動整理)

統整內容：

- * AMC 10 參與證明 (2020)
- * 基隆市立中正國中 服務員證書：2020小小樂手_管樂體驗營 隊輔，服務時數16小時 (2020/1/21-1/22)
- * 臺北醫學大學 推廣教育課程班結訓證明書：2020 全國青少年暑假醫學營-現代醫學營 (2020/8/18-2020/8/21)
- * 2020 全國科學探究競賽：全球暖化加劇台灣地震？榮獲高中職組入選 (與尚承嬭、薛可柔同學共同參與)
- * 流行音樂歌唱 結業證書
- * PADI Scuba Dive 體驗證明

圖 5-1-6 學生可透過上傳 PDF、JPG 等檔案，由系統自動整理成對應資料

語氣風格

請選擇風格

正式

自然

簡潔

詳細

自訂: 開朗活潑有朝氣

自訂: 活潑

自訂: 測試

自訂: 我的語氣

圖 5-1-7 因應個人化生成，設計多種語氣供學生選擇喜好風格

新增自訂風格

✕

風格名稱

請輸入風格名稱

語氣

例：正式、中性、活潑

描述文字的整體語氣風格

偏好詞彙

用逗號分隔，例：專業, 清晰

希望優先使用的詞彙，用逗號分隔

圖 5-1-8 除了預設風格以外，學生亦可自定義屬於自己的風格

生成結果

編輯心智圖

```

1 mindmap
2   root((N-多元表現綜整心得))
3     branch1[團隊合作能力]
4       sub1[熱音社鼓手:我在做這件事]
5       sub2[排球社學球員:之前我曾經]
6     branch2[組織與管理能力]
7       sub3[總務股長經驗:我很喜歡食]
8       sub4[小老師角色:希望我可以召]
9     branch3[創新與探索精神]
10      sub5[數學競賽參與:之前我曾經]

```

更新心智圖

心智圖預覽

放大查看

圖 5-1-9 生成心智圖結果，即時透過模態框預覽

大標題

重新生成

和諧的團隊合作

段落 1 - 內容

重新生成

團隊合作對我來說就像一首和諧的樂曲，每個音符都不可或缺。我在熱音社擔任鼓手時，深刻體會到這一點。一開始，我常常只顧著自己打得開心，忽略了和其他樂器的配合，導致整體音樂聽起來有些雜亂。經過不斷的練習和溝通，我開始學會聆聽其他樂器的聲音，調整自己的節奏和力度，讓鼓聲能夠更好地融入樂曲中。我才發現團隊合作的真諦在於互相配合、互相成就，讓每個人的優點都能夠充分發揮，最終呈現出最棒的音樂。

圖 5-1-10 生成結果預覽

二、技術應用與成效分析

1. Skeleton of Thought (SoT)與多階段生成流程設計

為提升生成內容的層次性與一致性，本系統導入 Skeleton of Thought (以下簡稱 SoT)骨架生成技術，使模型能在生成前先建立「思維結構骨架」(Skeleton)，再依序展開細部內容。此設計理念源自人類寫作過程中「先組織後表達」的邏輯結構，能有效降低大型語言模型在長篇生成時常見的語意漂移(semantic drift)與重複生成問題。其中，在整體生成流程共分為四個階段，分別為：

(1) 釐清任務與骨架生成(Clarifier / SoT Outliner)

在多重 LLM 串聯架構中，澄清與骨架生成階段是整個生成流程的起點，其核心任務在於「釐清輸入語意、補足缺漏資訊、建立思維骨架」。由於使用者在實際輸入時常僅提供零散的個人資料(如學習動機、活動經驗、比賽成果)，若直接交由主模型生成，容易導致段落主題偏移、內容不連貫或邏輯鬆散。因此，本階段透過 GPT-4o-mini 執行預處理，在正式內容生成前先進行語意澄清(Semantic Clarification)與骨架構建(Skeleton Generation)，確保主模型能於清晰、結構化的語境中生成文本。具體來說，共會執行：

A. 澄清階段(Clarifier)

系統在接收到使用者輸入(user inputs)和目標段落代碼(section code)後，會優先執行欄位完整性檢查，以判斷是否遺漏關鍵資訊，例如是否包含反思內容或具體描述與學系相關的動機；一旦檢測到關鍵欄位不足(如使用者僅輸入「我對電腦有興趣」)，系統便會在介面上提示使用者補充必要內容(例如：請補充描述具體事件、課程或專案經驗)以引導使用者補充回答。若使用者未進行補答，系統仍會以現有輸入為基礎產出較保守的草稿，並在輸出的說明中明確註記資料不完整可能導致的推論限制；這種處理方式能有效避免系統在缺少事實依據時自動產生高自信但無事實支撐的內容，從而降低實際運作中的幻覺率(Hallucination Rate)。

表 5-2-1 使用澄清階段與未使用澄清階段版本比較表

評估指標	未使用澄清階段	使用澄清階段	改善幅度
語意錯位率(Semantic Misalignment)	9.1%	4.7%	↓ 48.4%
段落主題一致性(Topic Consistency)	84%	94%	+10.0%
邏輯性與結構得分*(Logic & Structure)	17.16 / 20	18.55 / 20	+8.1%
內容完整度*(Content Completeness)	15.83 / 20	17.95 / 20	+13.4%

資料來源：本研究實驗數據(n=324)，2025 年 3 月內部測試。

*說明：邏輯性與結構得分與內容完整度評分方式使用 GPT-4o-mini、Gemini-2.0-Flash 兩種模型自評(Self-Evaluation)與跨模型互評(Cross-Model Evaluation)成績平均，每項滿分 20 分。其中邏輯性與結構得分評分標準為評估段落間邏輯連貫、語意層次清晰度與結構合理性；內容完整度評分標準為衡量生成段落是否充分回應題意並涵蓋必要要素。

B. Skeleton of Thought (SoT) 骨架構建

完成任務釐清與資訊補全後，系統會啟動骨架生成程序，依照 SoT 方法產出結構化骨架，其中內容包含：主題(main title)、子主題(subheading)寫作意圖(intent)與段落要點草稿(content outline)，作為主模型逐段生成的提示模板。以「學習歷程反思」為例，系統會產出類似下圖之骨架，表示在主題與子題層級已先行確定邏輯方向與關鍵論點；主生成階段即以此骨架說明逐一指示模型依序展開內容，從而降低偏題、重複敘述與語意漂移的機率。

```
● JSON

{
  "main_title": "探索資訊管理的樂趣與挑戰",
  "paragraphs": [
    {
      "sub_heading": "音樂與團隊的和諧",
      "intent": "引出對資訊管理的興趣，強調團隊合作的重要性",
      "content_outline": "在熱音社擔任鼓手的經驗中，我學會了如何與隊友協作，讓音樂 更和諧，這讓我意識到團隊合作在資訊管理中的關鍵角色。"
    },
    {
      "sub_heading": "管理資源的挑戰",
      "intent": "展示在組織與管理方面的能力，與資訊管理的關聯",
      "content_outline": "作為總務股長，我負責班上的帳務，這讓我學會了如何有效管理 資源，這種能力在資訊管理中同樣重要，讓我能夠更好地處理數據與資源。"
    },
    {
      "sub_heading": "數學思維的探索之旅",
      "intent": "反思參加數學競賽的經驗，展望未來的學習目標",
      "content_outline": "參加AMC 10數學競賽讓我挑戰自己的數學思維，增進了解決問題 的能力，這種探索精神將引導我在資訊管理的學習中不斷前進。"
    }
  ]
}
```

圖 5-2-1 學習歷程反思骨架生成範例

C. Prompt 正規化與語意一致化

在澄清任務與 SoT 骨架生成階段完成後，系統隨即進入「Prompt 正規化與語意一致化」程序，其核心目的在於讓所有進入生成階段的資料能維持一致的格式與語意結構，確保模型能確切理解生成需求，避免因指令誤解而產生不穩定的結果。此步驟會整合多方資料，包括前端的使用者輸入(如 user inputs、section code、語氣風格、偏好用詞與避免詞等偏好設定)以及後端經 OCR 模組辨識、文字濃縮模組萃取重點的檔案內容。接著，由後端的 Prompt Formatter 模組將通過驗證的欄位填入固定格式的 Prompt 模板中，同時執行欄位白名單控制、長度與結構限制、以及字元清洗與轉義等三項安全與一致性控制，以防止提示詞注入攻擊與語意混亂。

標準化後的 Prompt 會形成結構統一的輸入(範例如：「請根據 {section code}(所屬系所：{department})與下列重點，生成一段約 150-200 字的學習歷程反思：重點={content outline}；風格={tone}；避免詞={avoid words}。」)，此輸入可供主生成階段(Main Writer)重複使用，從而確保整份文件在語氣與邏輯上的一致性。

(2) 主生成階段(Main Writer)

在主生成階段，系統以 SoT 產生的段落概要骨架為驅動，並採取「分段同步生成 x 多模型交互」的策略，大幅降低文件生成時間，在兼顧生成速度與內容品質的前提下，產出具邏輯層次、語氣一致且貼合使用者個人特質的完整段落。

主生成階段具體流程如下：系統先把骨架中待展開的段落分組(每段視為一組)，以非同步/並行呼叫(asynchronous parallel generation)方式同時向初稿模型發出請求。在取得所有段落初稿後，系統再將其交由第二階段的 Refiner 模型進行語氣與風格的統一化處理，最後由 Composer 模組負責合併各段落並執行全局語意檢查，確保文本在邏輯、風格與格式上保持一致。此流程具備以下關鍵特性：

A. 分段同步生成(Parallelization)

系統將多個段落同時發出請求，而非逐段順序處理，使總生成時間趨近於「單段落生成時間」而非「所有段落時間的總和」。

實驗中，依順序生成一份完整備審文件約需 150 秒，採用分段並行後，文件生成平均耗時降至 20-25 秒，成功將整體延遲縮短約 83%。

B. 多模型交互(Multi-Model Chaining)

主生成階段採取「快速初稿 + 高質潤飾」的雙階段策略：初稿模型由 Gemini-2.0-Flash 模型負責生成具結構與要點的段落內容；Refiner 模型則採用 GPT-4o-mini 模型，著重於語氣調整、術語一致性校正與局部事實性檢查。

藉由此設計在效能與品質間取得平衡，先以快速模型完成內容骨架，再以高精度模型進行細部強化，實現「耗時短、品質高」的最終文本。

表 5-2-2 主生成階段不同生成策略之比較

比較項目	順序生成	分段並行+多模型串聯	改善幅度
完整生成平均時間(s)	153.6	24.8	↓ 83.9%
語句重複率	28.8%	22.4%	↓ 22.2%
邏輯性與結構得分*	17.16 / 20	18.55 / 20	+8.1%

*說明：邏輯性與結構得分評分方式使用 GPT-4o-mini、Gemini-2.0-Flash 兩種模型自評(Self-Evaluation)與跨模型互評(Cross-Model Evaluation)成績平均，每項滿分 20 分。邏輯性與結構得分評分標準為評估段落間邏輯連貫、語意層次清晰度與結構合理性。

(3) 結構整合階段(Composer)

在完成各段文字的生成與初步風格微調後，系統會進入「結構整合階段(Composer)」，此階段的主要任務是將所有段落初稿整合成具一致語氣與格式的完整文件，確保輸出結果邏輯銜接自然、用詞符合要求。

A. 結構統整(Structural Assembly)

首先根據 SoT 階段建立的階層式骨架，將主生成階段所產生的多段 JSON 段落輸出，依據骨架順序進行重新排列與組合，並插入必要的過渡句，以維持段

落之間的邏輯延續與語意流暢。同時，亦會依據骨架的順序自動補上缺漏的段落或調整段落順序，確保整體文件結構與原始 SoT 思維框架一致。

B. 語意檢查與風格一致性調整(Semantic Validation & Consistency Tuning)

在整合階段，系統會對所有段落進行語意層級的比對與校正。主要檢查項目包括：段落間是否出現重複內容或相互矛盾的陳述；主題轉換是否平順、有無突兀斷層；人稱代詞與時態是否保持一致；用詞風格是否延續前文設定(如正式、自然、溫暖)。若系統偵測到語意斷層，Composer 會自動生成過渡句或提示交由 Refiner 再次校正，以確保整篇文件的語意連續性與敘事一致性。

2. RAG 技術應用與資料檢索優化

在多階段 LLM 生成架構中，雖然 SoT 能有效強化文本的邏輯與層次，但若生成內容缺乏真實資料支撐，仍可能出現事實性錯誤或「幻覺」(hallucination)。為確保系統生成的備審文件兼具事實依據性與資料針對性，本研究導入 RAG 技術，在生成階段之前，利用檢索層從資料庫中選出該校系備審準備指引中最相關的段落撰寫提示內容，再與使用者輸入結合形成完整的上下文提示(context-enhanced prompt)。

(1) RAG 檢索與生成流程

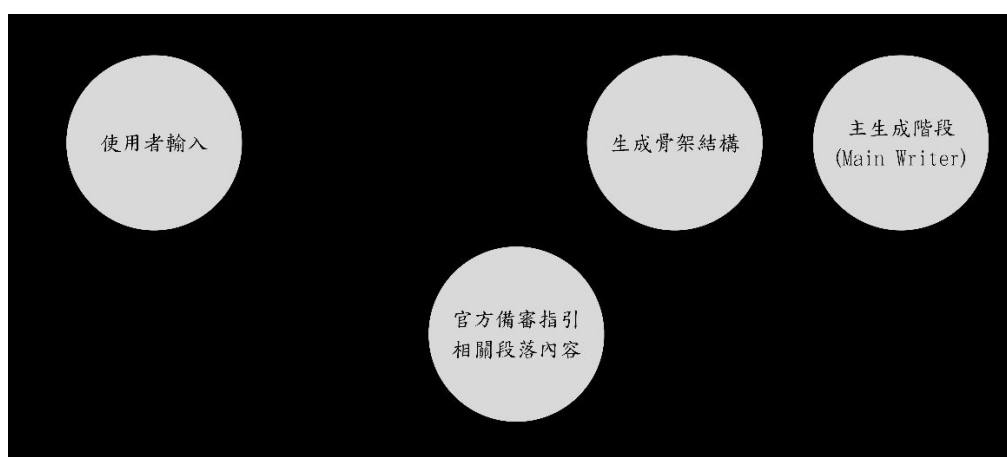


圖 5-2-2 RAG 檢索在生成流程上的應用

- A. 查詢向量化(Query Embedding)：使用者輸入的內容與區段名稱 (例如 “learning motivation”) 會被轉換為查詢向量 q ，並與資料庫中預存的文件向量 $d_1 \dots d_n$ 計算相似度。
- B. 檢索(Retrieval)：系統採用 BM25 演算法進行檢索，並根據分數擷取前 5 筆最相關段落(Top-5 retrieval)。
- C. 上下文組合(Context Assembly)：
檢索結果會與使用者輸入(user inputs)共同嵌入至 prompt 模板，形成「使用者輸入+官方資料」的混合輸入，範例 Prompt 如下：「根據以下資料與使用者經驗，撰寫一段學習動機：使用者輸入：{user inputs} 參考資料：{retrieved contexts}」

(2) 檢索效果與優化策略

為提升 RAG 層的檢索精度表現，本研究進行了向量模型與索引參數的測試：

表 5-2-3 向量模型表現比較

組合	向量模型	Top-k	命中率(Hit@5)	幻覺率(Hallucination Rate)
A	SBERT-Base	5	86%	15.2%
B	BGE-M3	5	94%	4.5%
C	E5-Base	5	92%	6.3%

結果顯示，使用 BGE-M3 模型並採 Top-5 檢索時，在命中率與幻覺率之間取得最佳平衡，幻覺率自 15.2%降至 4.5%，內容一致性由 82%提升至 93%。

3. 個性化生成與評估指標分析

在備審資料撰寫情境中，雖然結構完整與內容真實是必要條件，但真正能凸顯個人特色的關鍵在於語氣與風格。因此本系統導入「風格自訂(Style Customization)」機制，讓使用者能夠明確定義生成文本的語氣、偏好用詞與寫作傾向，使最終輸出不再只是通用模板，而能展現個人化表達與風格一致性。

其中，自定義風格控制的核心挑戰在於「讓模型懂得使用者的偏好卻不破壞原始語意」。為達成此目的，本系統採用三層自定義風格控制策略：

表 5-2-4 自定義風格控制策略

控制層級	技術方式	目的
Prompt 層	在生成模板中插入 tone、preferred words、avoid words 等參數	明確引導模型語氣方向
語體分析層	由樣本文字建立 Style Vector，映射句型與語調特徵	對齊使用者既有文風
Refiner 層	二次微調，統一語氣、調整句法、消除語意突兀	保持整體風格一致性

從實際測試結果中觀察到，經過上述三層控制後，生成文本在風格穩定性與自然度上均有明顯提升，顯示此個性化生成模組設計能有效套用使用者自定義風格，並透過 Refiner 模型使生成結果在整體語感上保持流暢連貫、風格一致。

4. 整體成效

測試結果顯示，系統能有效協助學生釐清各學系審查重點，並產生貼近需求的備審段落，顯著提升備審資料準備效率與內容品質。

段落 4		
學科競賽經驗		
事實描述	學習心得	能力反思
參加 2020 年 American Mathematics Competition 10	準備過程中，我學會系統性分析問題，並運用數學知識找到解決方案。	培養了邏輯思維和解決問題的能力，對未來學習資訊科技非常重要。
參加 2020 全國科學探究競賽，主題為全球暖化與台灣地震，榮獲高中職組入選	從文獻蒐集到報告撰寫，我學會運用科學方法探究複雜議題。	提升了資料分析能力，與資管系的大數據分析方向高度相關。

圖 5-2-3 表格生成結果

段落 6

重新生成

除了學科學習，我也積極參與社團活動，以培養溝通表達與合作能力。在排球社和熱音社中，我學會了如何與不同個性的同學合作，朝著共同目標努力。雖然只是社員，但我總是積極參與練習，並在團隊需要時提供協助。擔任總務股長和地理/公民小老師的經驗，更讓我深刻體會到溝通協調的重要性。作為總務股長，我負責班級經費管理，需與同學溝通收費事宜，並向老師報告財務狀況。擔任小老師則要求我將複雜知識整理成易於理解的內容，並向同學講解。

段落 7

重新生成

幹部經驗		
事實描述	學習心得	能力反思
擔任總務股長	學會與同學溝通收費事宜，並向老師報告財務狀況。	培養了組織能力和溝通協調能力，這在未來職場上非常重要。
擔任地理/公民小老師	學會將複雜知識整理成易理解的內容，並向同學講解。	提升了解釋能力和溝通表達能力，對未來在團隊中分享知識有幫助。

圖 5-2-4 多元成果表現段落中，幹部經驗部分生成結果

段落 3

重新生成

🕒 段落重新生成中...

圖 5-2-5 段落亦提供重新生成功能，學生可將不滿意之段落重新生成

陸、結論

本研究透過開發「備審資料自動生成系統」，初探生成式 AI 在升學備審文件製作中的輔助應用，成功整合網頁爬蟲、RAG 增強檢索、多重 LLM 串聯與 SoT 骨架生成技術，實現一站式招生資訊查詢與個人化文本生成功能。系統不僅解決了傳統備審準備中資訊分散、資料整合困難與內容生成不一致的痛點，還為學生提供針對性強的撰寫建議，提升申請文件的邏輯性、事實性與個性化程度；同時輔助老師快速查詢審查標準，優化輔導流程。

實驗結果證實，系統在 324 組樣本測試中，導入 SoT 與 RAG 後，「邏輯性與結構」得分平

均達 18.55/20，提升約 11%；幻覺率降至 4.5%，內容一致性達 93%；生成耗時控制在 30.17 秒內，證明多階段生成策略在教育應用場域的有效性。此成果不僅驗證生成式 AI 在個人申請輔導的潛力，還突顯 RAG 技術在減緩模型幻覺與提升事實依據方面的貢獻，為台灣多元入學制度提供實用工具。

然而，本研究仍存在限制，如資料擷取受限於網站結構變動，可能需定期維護爬蟲模組；此外，生成內容雖經人工評分驗證，但尚未進行大規模使用者實測，未來可擴充 A/B 測試與回饋機制。展望未來，可進一步整合多模態輸入(如語音轉文字或圖像分析)，並探索跨學系推薦功能，以擴大系統在教育輔導的應用範圍。總體而言，本計畫證實生成式 AI 能有效輔助備審文件製作，促進教育公平與效率，為後續研究奠定基礎。

柒、參考文獻

1. 江宜芷,&林子斌.(2020).當「大學選才」遇見「高中育才」：大學多元入學考招改革的批判論述分析.*教育研究與發展期刊*,16(1),1–34.
2. 李浩仲,李文傑,&連賢明.(2016).多錢入學？從政大學生組成看多元入學.*經濟論文*,44(2),207–250.
3. 陶宏麟,&吳澤玫.(2019).從效率與公平評估臺灣的大學入學制度改革.*人文及社會科學集刊*,31(3),385–426.
4. 葉玉玲,&丁一顧.(2021).我國大學多元入學方案之回顧與省思.*臺灣教育評論月刊*,10(4),1–9.
5. Allard,M.(2019).WhatisaTransformer?*Medium*.Retrievedfrom<https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04>
6. Bengio,Y.,Courville,A.,&Vincent,P.(2013).Representationlearning:Areviewandnewperspectives.*IEEETransactionsonPatternAnalysisandMachineIntelligence*,35(8),1798–1828.<https://doi.org/10.1109/TPAMI.2013.50>
7. Devlin,J.&Chang,M.(2018).BERT:Pre-trainingofdeepbidirectionaltransformersforlanguageunderstanding.*arXiv*.
8. Gao,Y.&Xiong,Y.(2024).Retrieval-augmentedgenerationforlargelanguagemodels:Asurvey.*arXiv*.
9. Han,Y.,Liu,C.,&Wang,P.(2023).Acomprehensivesurveyonvectordatabase:Storageandretrievaltechniques,challenges[Preprint].*arXiv*.<https://doi.org/10.48550/arXiv.2310.11703>
10. Hou,Y.(2024).Researchontheapplicationofbigdataineducation.*CommunicationsinHumanitiesResearch*.
11. Lan,Y.,Li,X.,Du,H.,Lu,X.,Gao,M.,Qian,W.,&Zhou,A.(2024).Surveyofnaturallanguageprocessingforducation:Taxonomy,systematicreview,andfuturetrends[Preprint].*arXiv*.<https://arxiv.org/abs/2401.07518>
12. LangChainDevelopers.(2023).LangChain:Aframeworkfordevelopingapplicationspoweredbylanguage models.Retrievedfrom<https://python.langchain.com>
13. Lewis,P.,Perez,E.,Piktus,A.,Petroni,F.,Karpukhin,V.,Goyal,N.,Küttler,H.,Lewis,M.,Yih,W.-t.,Rocktäschel,T.,Riedel,S.,&Kiela,D.(2020).Retrieval-augmentedgenerationforknowledge-intensiveNLPtasks.In*Proceedingsofthe34thInternationalConferenceonNeuralInformationProcessingSystems(NeurIPS)*(pp.5998–6008).Vancouver,Canada.

14. OpenAI.(2023).GPT-4technicalreport.Retrievedfrom<https://api.semanticscholar.org/CorpusID:257532815>
15. Pintrich,P.R.(2004).Aconceptualframeworkforassessingmotivationandself-regulatedlearningincollegestudents.*EducationalPsychologyReview*, 16(4),385–407.<https://doi.org/10.1007/s10648-004-0006-x>
16. Tang,Y.,Zhang,R.,Guo,J.,deRijke,M.,Liu,S.,Wang,S.,Yin,D.,&Cheng,X.(2025).Generativeretrievalforbooksearch[Preprint].*arXiv*.<https://doi.org/10.48550/arXiv.2501.11034>
17. Vaswani,A.,Shazeer,N.,Parmar,N.,Uszkoreit,J.,Jones,L.,Gomez,A.N.,Kaiser,Ł.,&Polosukhin,I.(2017).Attentionisallyouneed.In*AdvancesinNeuralInformationProcessingSystems(NeurIPS)*(pp.5998–6008).
18. Yasunaga,M.,Aghajanyan,A.,Shi,W.,James,R.,Leskovec,J.,Liang,P.,Lewis,M.,Zettlemoyer,L.,&Yih,W.-t.(2023).Retrieval-augmentedmultimodallanguagemodeling.In*ICML2023*.*arXiv*.<https://doi.org/10.48550/arXiv.2211.12561>

附錄A. 專題工作內容

本專題依照系統模組與研發流程進行分工，核心工作涵蓋了資料處理、系統後端建置、前端介面開發，以及最重要的 RAG 檢索與多階段生成式 AI 模組設計。

一、資料蒐集與處理模組

- 1.資料獲取：蒐集各大學校系的招生簡章、審查標準及備審資料準備指引。
- 2.自動化擷取：設計爬蟲流程，整合 Scrapy、Selenium 等工具，高效自動化擷取線上資料。
- 3.非結構化資料解析：針對 PDF 或掃描文件，我們導入 Tesseract OCR 進行文字識別，並執行資料正規化與關鍵欄位抽取。
- 4.資料庫建置：建立結構化資料庫(採用 PostgreSQL)，規劃校系資訊、審查重點、文件類型等核心欄位，確保資料品質。

二、後端與系統架構開發

- 1.系統架構設計：規劃資料處理層、核心應用層與使用者介面層的整體分層架構。
- 2.後端 API 實作：開發核心服務 API，包含招生資訊查詢、備審內容生成、以及使用者權限管理功能。
- 3.向量檢索系統串接：串接 FAISS 向量資料庫，實作文件切分、向量化儲存，以及高效的相似度檢索流程。
- 4.RAG 流程整合：導入 LangChain 作為核心框架，將檢索(Retrieval)與生成(Generation)結合成完整的 RAG Pipeline。

三、生成式 AI 模組與多階段生成流程

- 1.骨架思維(SoT)導入：設計 Skeleton of Thought (SoT)流程，要求 LLM 先規劃內容骨架，確保結構完整性。
- 2.四階段 LLM 串聯：建構一套精細的生成流程：Clarifier(釐清需求)→Outliner(生成大綱)→Main Writer(主體撰寫)→Composer(潤飾與校對)。
- 3.多模型協同運作：採用異質模型合作策略，利用 Gemini Flash 快速產出初稿，再由 GPT-4o-mini 進行內容優化與精煉(Refine)。
- 4.參數控制機制：建立精準的參數設定介面，讓使用者可以控制語氣風格、自定義風格，以及設定偏好詞彙與避免詞彙。

四、前端介面設計與互動功能

- 1.技術選用：使用 Flask 搭配 Jinja2 模板引擎完成網頁頁面。
- 2.介面設計：區分設計「學生端」與「輔導老師端」兩種查詢及操作介面。
- 3.功能實現：完成備審段落生成、歷史紀錄查詢、檔案上傳解析、以及輔助性的心智圖生成等核心功能。
- 4.使用者體驗優化：針對表格資料進行整理，並優化提示訊息設計與使用者操作流程(UI/UX)，提升系統易用性。

附錄B. 專題心得與建議

一、 專題執行心得

本次專題讓我完整體驗了從底層資料處理到頂層 AI 應用模組的開發過程，並帶來幾項關鍵認知：

- 1.資料處理是最大挑戰：由於各校招生文件格式極度不一致，執行 OCR、正規化與欄位抽取佔據了大量時間，這是系統建置初期最具挑戰性的部分。
- 2.RAG 品質的關鍵：在生成階段，若提示工程(Prompt Engineering)不足，模型極易產生幻覺 (Hallucination)。因此，SoT 骨架與 Clarifier 流程的設計對於穩定生成內容的邏輯性與準確性至關重要。
- 3.效率優化：透過多模型協作與分段並行生成，我們有效降低了生成延遲，並更深入理解了大型語言模型 (LLM) 部署與架構優化的瓶頸所在。
- 4.實務應用價值：本專題不僅在技術層面(向量資料庫、檢索演算法)有所深入，更讓我們貼近理解學生和輔導老師在實際應用 AI 工具時的迫切需求。

二、 對未來改進的建議

為了使系統更具彈性與市場競爭力，提出以下改進方向：

- 1.強化爬蟲維護性：招生網頁格式變動頻繁，未來應導入自動錯誤偵測，或探索利用 LLM 輔助資料抽取，提升爬蟲的健壯性。
- 2.擴充資料集範圍：目前資料以部分校系為主，應逐步擴展至全國大學與科系，以提升系統的覆蓋率與實用性。
- 3.加入多模態功能：擴大資料輸入範圍，未來可支援學習歷程封面圖像、影片或自錄簡介等，以處理更多元的資料格式。
- 4.大規模使用者實測：應透過 A/B 測試，廣泛收集實際高中生的回饋，據此持續改善介面與 AI 生成內容的品質。
- 5.探索跨校系推薦服務 (DSS)：依據學生的特質與歷程資料，開發「適配度分析」或學系推薦模型，將系統服務延伸至更深層次的決策輔助系統 (DSS) 領域。