

# 研究計畫書

基於 RA-LLMs 架構下的古漢語知識問答系  
統之構建

研究學生：臺灣中原大學 資訊管理學系 本科三年級 潘驄杰

指導教授：臺灣中原大學 資訊管理學系 副教授 賴錦慧

## 目錄

一、摘要.....	1
二、研究動機與研究問題 .....	1
(一) 研究動機.....	1
(二) 研究問題.....	2
三、文獻回顧與探討 .....	2
(一)、古漢語資訊處理技術的研究現狀與方法 .....	2
(二) 檢索增強生成技術 (RAG) 與檢索增強大語言模型 (RAG-LLMs) 之基本原理.....	4
四、研究方法及步驟 .....	6
(一)、研究方法.....	6
(二)、研究步驟.....	7
五、預期結果.....	10
六、參考文獻.....	11

# 研究計畫內容

## 一、摘要

本研究計畫旨在開發基於檢索增強生成大語言模型 (Retrieval-Augmented Large Language Models, RA-LLMs) 訓練策略的古漢語知識問答系統，以解決學生在古漢語學習中面臨的挑戰，特別是在理解典故、歷史事件和文化思想等需要較多背景知識方面的困難，並滿足學生的即時性學習需求。研究方法採用「LLMs First」的序列訓練策略，集中升級檢索模組 (Dense Passage Retriever, DPR)，並與生成模型 (GPT-4o) 無縫結合，實現高精準度的知識生成與查詢。本研究預期開發一款能即時回答古漢語知識問題的智慧問答系統，以提升學生學習效率，促進國文教育數位化。同時，這一系統將檢驗 RA-LLMs 技術在古漢語處理中的應用效果，並為教育科技與數位人文領域提供創新解決方案。

## 二、研究動機與研究問題

### (一) 研究動機

古漢語知識作為中華文化的核心組成部分，蘊含豐富的文學、歷史、哲學與藝術內涵，是理解中華文化精髓的重要途徑。然而，古漢語學習對學生來說挑戰重重，尤其是在語文考試或課堂學習中，常涉及典故、歷史事件或文化思想等專業知識，這些內容需要具備深厚的背景知識與對文化內涵的理解能力。以學生群體(如初中生、高中生及大學生)為例，他們在學習過程中常面臨以下三大難題：

- **背景知識不足**：面對文言文中的歷史典故或哲學思想，由於缺乏相關背景知識，學生難以準確理解或解釋文本內容。
- **學習資源受限**：傳統的學習工具（如字典和參考書）檢索效率低，網路搜尋結果往往資訊分散且結構化不足；教師輔導則受限於時間與資源，難以滿足即時學習需求。
- **生成模型的局限**：雖然如 GPT-4 等大語言模型在生成文本方面頗具潛力，但其生成古漢語學術內容時，常因缺乏專業背景知識而出現偏差或錯誤，無法完全滿足學生學術需求。

有鑒於上述挑戰，學生需要一款即時、高效且專業的智慧學習輔助工具，來幫助他們解決古漢語學習中的困難，提升學習效率與成效。因此，本文提出基於檢索增強生成技術 (RAG) 的學習輔助系統，結合檢索模組 (Retriever)

與生成模組 (Generator) 的優勢，在生成答案前主動檢索相關專業知識，確保回答的準確性與可靠性。為進一步提升系統效能，本研究引入檢索增強生成大語言模型訓練 (RA-LLMs Training)，透過優化檢索器訓練，精準提取古漢語文本所需的背景資料，增強生成模組的專業性與結構化表現。本研究的最終目標是開發一款智慧輔助工具，能即時解決學生在古漢語學習中遇到的知識難點，並促進國文學習的便利性與成效。同時，該工具有潛力推動教育數位化進程，並成為數位人文研究中的創新應用。

## (二) 研究問題

基於以上研究動機，本研究聚焦於以下核心問題，力求通過技術創新解決學生在學習古漢語時面臨的背景知識不足與即時性需求的挑戰，並驗證工具在提升學習效率的成效。因此，核心研究問題在於：

- 基於檢索增強生成大語言模型 (RA-LLMs) 的古漢語知識問答系統，是否可以有效說明學生克服古漢語學習中背景知識不足與即時性學習需求的挑戰，從而提升學習效率？

## 三、文獻回顧與探討

### (一)、古漢語資訊處理技術的研究現狀與方法

#### 1. 古漢語資訊處理技術研究

古漢語資訊處理技術與專業大語言模型的研究，不僅是古籍文本數位化處理的關鍵支柱，也是推動數位人文研究與文化傳承的重要力量。其核心目的也是解決古籍整理、語言、國文教育及數位元人文研究中的文本處理問題。由於古籍文本多數缺乏標點符號或分詞標注，這些技術的目標在於提升文本解析的精確性與自動化程度。因此，本部分先從句讀標點、分詞與詞性標注、文白翻譯等技術方向之方面，探討古漢語智慧處理的研究現狀與未來發展。

#### (1) 句讀標點

首先，古籍文本大多無標點符號，因此句讀標點的核心目的是為缺乏標點的古漢語文本添加停頓和結構標記，從而提高文本的可讀性和研究效率。因此，有研究提出基於深層語言模型 (如 BERT) 結合條件隨機場 (CRF) 與卷積神經網路 (CNN) 的方法，自動化處理句讀並顯著提高斷句準確率 (胡韜奮 et al., 2021)。而有研究者利用雙向迴圈神經網路 (Bi-GRU) 的斷句模型也進一步增強了對上下文的理解能力 (王博立 et al., 2017)。也有研究基於深度學習，以及 BERT 模型，實現句讀與標點的一體化處理，不僅解決了斷句與標點串列任務中的錯誤傳遞問

題，並提升了整體準確性 (俞敬松 et al., 2019) (袁義國 et al., 2022)。

## (2) 分詞與詞性標注

其次，在古漢語中，詞彙的意義高度依賴上下文理解的語言現象，源于其單音節詞主導的特性和靈活的語法結構。以「國家」與「國」、「家」二字為例：作為雙音合成詞的「國家」整體指向現代意義的國家概念，但單獨使用「國」或「家」時，其意涵會隨語境擴展或收縮。例如，在獨立語境中，「國」可指諸侯封地（如「齊國」）、政權（如「治國」）或都城（如「國門」）；而「家」既可指家庭（如「治家齊國」），也能表示學術流派（如「儒家」），此現象反映了古漢語的詞類活用與多義性，但也使機器自動分詞難以準確識別詞彙邊界，因此分詞技術旨在解決古漢語文本的詞彙切分問題。

因此，在古漢語分詞領域的研究中，有學者提出自動分詞的詞彙處理與量化統計的分詞方法，並驗證了古代漢語中雙音詞比例隨時間穩定增長的趨勢 (邱冰 & 皇甫娟, 2008)。針對特定文本的應用，有研究透過條件隨機場模型進行《孟子》語料的自動分詞，並引入小句 F 值評估指標，提升了實驗效果的可靠性和準確性 (梁社會 & 陳小荷, 2013)。此外，亦有研究結合先秦典籍專用詞表與條件隨機場模型，顯著提升分詞準確率至 97.47%，在古典文獻處理中表現出色 (黃水清 et al., 2015)。為進一步優化模型性能，一體化的 BiLSTM-CRF 模型被提出，實現了分詞、詞性標注與斷句的聯合處理，同時有效解決傳統分步處理中多級錯誤傳遞的問題 (程寧 et al., 2020)。

## (3) 文白翻譯

再來，文白翻譯的意義在於將古漢語轉化為現代漢語，從而降低閱讀門檻並促進文化普及。有研究者先構建了目前最大規模的古今漢語平行語料庫，並基於 Transformer 模型進行翻譯實驗，獲得了較高的 BLEU 評分 (Liu et al., 2020)。隨後，亦有研究者設計了 Erya 翻譯框架，採用雙音節替換與雙遮罩 Transformer 模型，改善了零樣本與微調場景中的翻譯效果，此框架在流暢性與準確性上超越前者研究構建的模型 (Guo et al., 2023)。

然而，古漢語的處理面臨「低資源、富知識」的挑戰。「低資源」即現有的標注資料稀缺且品質有限，例如大多數古籍文本缺少標點符號；而「富知識」指蘊含豐富的文化典故與歷史知識，這意味著為了充分理解和解析古籍文本，模型需要具備深厚的背景知識 (李紳 et al., 2024)。這對於語言資源建設和模型研究帶來了極大困難，在於如何在資源匱乏的情況下提升模型的效能，以及如何有效地將豐富的背景知識轉化為可用於模型訓練的結構化資料。

這就需要通過大型語言模型的多工學習能力與小樣本學習特性解決以上問題。多工學習能力，即模型可以同時執行句讀、典故識別、文白翻譯等多種任務，

而小樣本學習特性，即是在有限語料的條件下，通過少量標注資料微調以提升處理精度。因此，大語言模型的建構，亦為古漢語資訊處理的趨勢。

## 2. 古漢語專業大語言模型的探索與應用

針對古漢語的大語言模型，首先「荀子」(Xunzi-Qwen-Chat)<sup>①</sup>是其中一個重要的專業化探索。「荀子」模型基於 Qwen-7B-Chat 的架構，經過針對《四庫全書》等古漢語語料的微調，旨在提升古籍處理與語義分析的專業表現。而另外「九思」模型則結合多模態技術，聚焦于古籍文本的語意解讀與圖文融合場景，展現了在多層次語義理解中的潛力，然而此模型尚未公開使用或開源(劉金柱 et al.)。

另外，關於古漢語語言模型的評估方法，有學者提出了 AC-EVAL 評估基準為古漢語理解能力提供了重要的測試框架，涵蓋短文本理解、長文本處理及歷史知識掌握三方面的能力評估 (Wei et al., 2024)。

實驗顯示，「荀子」模型在歷史知識任務中表現尚可(正確率 60.20%)，但在長文本理解(僅 30.87%)和少樣本學習上表現不如同規模的通用模型(如 Qwen-7B-Chat 和 ERNIE-Bot 4.0)。這凸顯出該專業化模型在應對多樣化任務時的適應性不足，特別是在長文本處理及連貫性分析方面，中文通用大模型(如 GLM 和 ERNIE 系列)依然保持領先地位。

然而，另一個針對古漢語的「AI 太炎」模型，通過微調結合領域知識與僅使用 1.8B 參數量，便在句讀與文白翻譯等任務中已達到甚至超越人類基線的效果，顯示了大語言模型在古漢語處理中仍具有一定的潛力(李紳 et al., 2024)。

雖然大語言模型在語言理解和生成上表現優異，但依然存在一些難以克服的內在缺陷。例如，幻覺現象(Hallucination)，即模型可能會生產出不準確或無中生有的內容；知識過時，即模型內部的知識隨著時間推移可能會失效，特別是在面對需要最新資訊的任務時表現不佳；再訓練與微調需要較高成本，且需要較高的算力要求。面對這些問題，檢索增強生成技術(RAG)成為另一個重要解決方案。RAG 方法通過外部知識檢索補充模型的內部知識，能有效提升生成文本的準確性和相關性，很適合於古漢語回答此較需要專業化答覆之領域。

### (二) 檢索增強生成技術(RAG)與檢索增強大語言模型(RA-LLMs)之基本原理

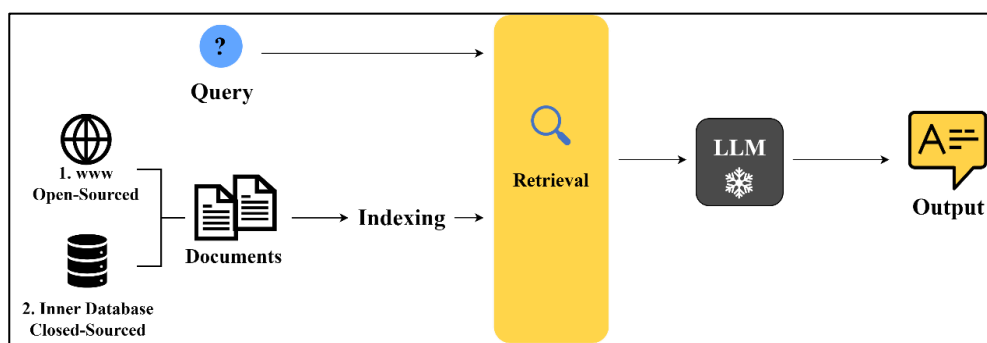
檢索增強生成技術(Retrieval-Augmented Generation, RAG)是一種結合檢索系統與生成式模型的方法，通過檢索外部知識庫中的相關資訊，輔助大型語言模型(LLMs)生成更加準確且上下文相關內容。而 RA-LLMs 則為 RAG 的改良方法，其在訓練階段融入檢索機制，使模型能夠更深入地理解並利用外部知識。

---

① 參見：<https://github.com/Xunzi-LLM-of-Chinese-classics/XunziALLM>。

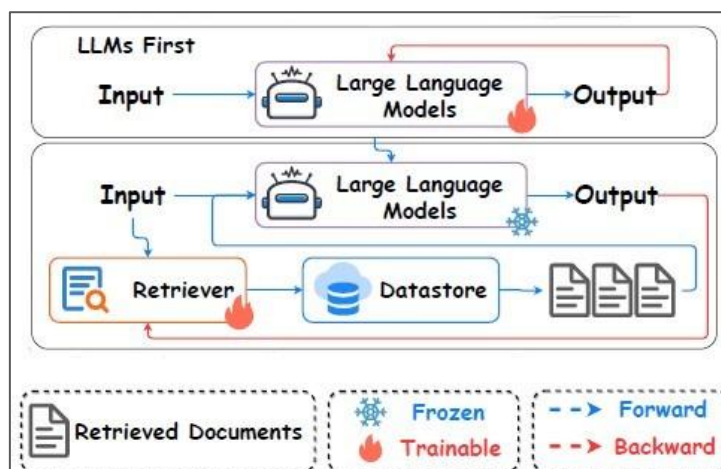
首先，RAG (Retrieval-Augmented Generation) 的基礎運行框架流程可分為四個主要步驟：索引建立、檢索、增強與生成。在索引建立階段，系統會對網路開來源資料或內部資料庫的檔進行索引編制。接著進入檢索階段，系統會根據使用者的查詢，從外部資料庫中提取相關的文本資料。

隨後，這些檢索到的文本會被整合至語言模型的輸入中 (通常語言模型直接使用，可不再微調)，作為增強資訊的來源。最後，語言模型基於這些增強資訊生成回應。此技術不僅有效解決了大型語言模型 (LLMs) 因訓練資料靜態或不足而可能產生的「幻覺」(hallucination) 問題，更允許系統即時更新知識庫，以應對動態變化的需求(Lewis et al., 2021)。



圖一、RAG 基礎運行框架 (LLM 下的雪花代表可不須再行微調或訓練)

其次，「檢索增強大語言模型訓練」(RetrievalAugmented Large Language Models: RA-LLMs)則將檢索機制融入模型訓練，提升效能與外部知識利用能力，欲解決傳統 RAG 方法上，檢索器在知識引用模糊或不精準的問題。模型通過多模組架構 (檢索與生成模組) 學習有效選擇與使用檢索資料，增強知識密集型任務表現，並應對跨領域或特定領域 (如醫療、教育、企業) 的挑戰。此外，RA-LLMs 動態整合最新資訊，減少錯誤或過時內容生成，提升模型可靠性與實用性 (Fan et al., 2024)。



圖二、RA-LLMs Sequence Training 中 LLMs 優先訓練之運行框架

如以上圖二所示，以 Sequence Training 中「LLMs 優先訓練運行框架」為例，說明 RA-LLMs 的運作流程。此框架分為上下兩層：上層是大型語言模型(LLMs)，下層是檢索器 (Retriever)。由於這裡採用的是「LLMs 優先」的設定，首先由上層的 LLMs 開始運作。具體過程如下：

### 1. 訓練 LLMs (上層)

上層的 LLMs 先經過訓練 (一般來說，LLMs 通常由其他團隊事先訓練完成)，完成訓練後，這些模型的參數將被鎖定，稱為「Frozen」，無法再進行進一步的調整。這樣的設置使得 LLMs 能在特定任務上表現良好，但也造成了隨著時間推移，LLMs 的知識可能變得陳舊。

### 2. 檢索器的動態調整 (下層)

為了彌補 LLMs 知識可能過時的問題，下層的檢索器 (Retriever) 會根據 LLMs 的輸出結果再進行訓練、進行動態調整。這樣，檢索器能提供更加符合當前情境需求的資訊，提升檢索結果的品質。

### 3. 協同作用

在這個架構中，檢索器和 LLMs 互相配合，實現「檢索 + 生成」的結合。透過檢索器即時更新的資訊，生成內容的準確性和相關性得到了提升。同時，這也確保了知識能夠動態更新，以應對變化中的需求(Fan et al., 2024)。

此外，隨著技術的不斷發展，研究者也開始探索如 Graph-RAG 等變體，在檢索來源 (Retrieval Sources) 上利用結構化知識圖譜進一步提升檢索和生成的精準性 (Peng et al., 2024)。

綜上所述，在古漢語綜合知識問答等需高度文獻佐證的應用場景中，結合檢索增強生成 (Retrieval-Augmented Generation, RAG) 的即時知識調用能力與大型語言模型 (Large Language Models, LLMs) 的強大語言理解與生成能力，基於 RA-LLMs (檢索增強大型語言模型) 的架構尤為適合。特別是在 Sequence Training 中採用「LLMs 優先訓練」的運行框架，該方法以生成模組 (LLMs) 作為核心，確保其生成能力穩定後，再重點訓練、優化檢索模組 (Retriever)，有效彌補模型內部知識過時或不足的問題。

## 四、研究方法及步驟

### (一)、研究方法

在知識密集型任務中，尤其是古漢語綜合知識問答這類需要高準確性和即時性的應用場景，檢索增強生成 (Retrieval-Augmented Generation, RAG) 技術的結

合為傳統大語言模型 (LLMs) 提供了有效的解決方案。然而，由於 RAG 中檢索模組 (Retriever) 與生成模組 (Generator) 的協同運作需要精確對接，訓練策略的選擇至關重要。

因此，在 RA-LLMs (檢索增強生成大語言模型) 的訓練過程中，採用 LLMs 優先的序列訓練方法顯得尤為重要。LLMs 優先的序列訓練方法的核心邏輯是在訓練過程中優先確保生成模組的穩定性，將 LLMs 的參數固定 (Frozen)，並重點優化檢索模組 (Retriever)。這種策略帶來以下幾點關鍵優勢：

1. **提升訓練效率**：避免對生成模組重複微調，顯著降低了算力與資源需求。
2. **保障生成穩定性**：優先完成 LLMs 的訓練，確保其生成語言的流暢性與結構完整性不受後續檢索模組調整的影響。
3. **動態優化檢索能力**：將訓練重點放在檢索模組上，有助於實現即時的知識補充與精準的外部資料引用，彌補 LLMs 知識的靜態化問題。

基於上述考慮，本研究設計了一個適應古漢語綜合知識問答的 RA-LLMs 訓練框架，並採用 LLMs 優先的序列訓練方法，以開發出高效、準確、專業的智慧輔助系統。

## (二)、研究步驟

基於以上方法，研究步驟包含以下關鍵部分：

### 1. 知識庫構建

以中學、高中、大學中文系教科書內容、大學入學考試國文題目及其網路公開詳解為主要資料來源，並引入古籍等公開資料以擴展語料範圍。資料經過分段處理，每段約 100 至 300 字 (Passage)，確保語義處理時的上下文完整性與檢索結果的準確性。

最後，透過稠密向量編碼技術 (如 BERT 或 Contriever 模型)，將文本段落轉換為高維向量，並構建基於 FAISS 的高效向量索引系統，支援快速檢索。所有這些過程均在 Azure 平臺上完成，確保分散式處理和資料安全性。

### 2. 檢索模組訓練

檢索模組的訓練設計採用 Dense Passage Retriever (DPR)，是一種基於雙塔編碼器 (Bi-Encoder) 的檢索架構，能有效抓取查詢與段落之間的深層語義關係，大幅提升檢索的準確性與效率 (Karpukhin et al., 2020)。檢索模組訓練資料來源包括人工標注的問題與答案對以及 GPT-4o 自動生成的問題變體，進一步擴充

語料多樣性。透過對比學習 (Contrastive Learning) 方法，最大化正樣本的語義相似性並最小化負樣本的相似性，進行模組訓練。最終優化採用交叉熵損失函數 (Cross-Entropy Loss)，確保檢索模組具備高精確度與語義敏感性。此訓練過程預計利用 Azure 提供的 GPU 叢集進行大規模分散式訓練，加速模型訓練。

### 3. 檢索與生成整合

針對生成需求，本研究設計了適配 GPT-4o 的提示工程 (Prompt Engineering)，將檢索結果與使用者查詢相結合形成生成輸入。而後，在 Azure 平臺上部署研究環境，進行多組檢索生成模式的對比測試，評估檢索模組對生成效能的貢獻。

### 4. 系統效能評估

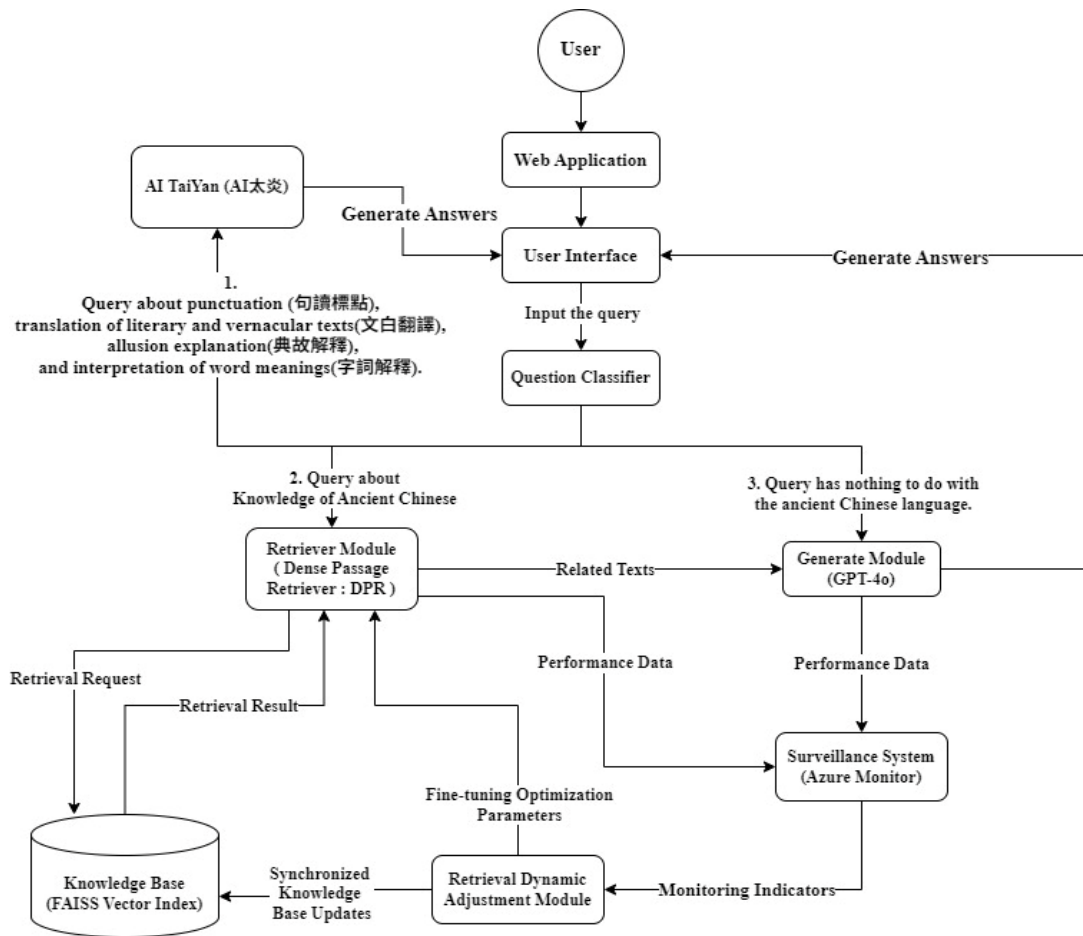
在評估系統效能時，本研究將著重於檢索模組與生成模組的綜合表現。檢索模組的準確性將透過平均精度 (MAP) 與 P@k 等指標進行評估，並分析多樣性指標以確保檢索結果涵蓋廣泛內容。

生成模組的評估則主要基於 Wei 等人 (2024) 提出的 AC-EVAL 評估基準，利用提示詞工程，測試系統在短文本理解、長文本處理及歷史知識掌握等方面的能力。為確保生成內容評估的公正性與可靠性，本研究將參考該研究的評估方法 (李紳 et al., 2024)。邀請兩位中文研究所的碩士或博士生擔任評估員。針對每條生成結果 (包含模型與人類的作答)，我們將提供三組匿名且隨機排序的答案，以確保評估員無法得知答案的來源。評估員將被提供參考答案，並允許查閱相關資料，以評估作答的準確性。評分標準如下：

- **1 分**：答案正確、精準，且有助於理解。
- **0.5 分**：答案接近正確，但存在部分問題或不清楚之處。
- **0 分**：答案錯誤、離譜，可能誤導他人。

為確保評估員對評分標準的理解一致，我們將首先進行試評估與討論，然後再進行正式評估。最終的人工評估結果將分為「準確率」與「嚴格準確率」兩項指標，兩者滿分皆為 100%。「嚴格準確率」代表答案完全正確 (得 1 分) 的比例，而「準確率」則代表答案完全正確與部分正確 (得 0.5 分) 的比例之總和。

最終，基於以上研究方法與研究步驟的設計，我們可以設計如以下圖三之系統架構圖，搭配架構圖之解說，以更加直觀瞭解此系統之實際架構：



圖三、基於 RA-LLMs 之古漢語知識問答系統架構圖

此架構基於檢索增強生成技術 (Retrieval-Augmented Generation, RAG)，結合大型語言模型 (GPT-4o) 與檢索模組 (Dense Passage Retriever, DPR)，以提供高效、準確的古漢語知識解答。系統運行流程始於使用者查詢，用戶通過網頁應用輸入問題，經由「問題分類器」分析後進行處理三類狀況。第一、若問題屬於古漢語處理相關任務 (如句讀標點、文白翻譯、典故解釋、字詞解釋)，則交給協力網站「AI 太炎」處理相對應任務，而後返回結果；第二、若問題就是屬於古漢語知識問題，則經檢索模組從知識庫中提取相關文本，並傳遞給生成模組回答；第三、若為其他類型問題，則直接由生成模組生成答案。知識庫以 FAISS 向量索引構建。同時，系統以 Azure Monitor 即時監控檢索與生成模組效能，收集性能資料以優化檢索模組參數和實現知識庫更新，確保系統能動態適應使用者需求。此架構利用問題分類器的智慧分流功能結合檢索與生成模組的優勢，不僅確保回答的專業性與準確性，還兼具靈活性與高效性，以應用于教育與文化傳承等場景。

## 五、預期結果

本研究致力於開發一款基於檢索增強大語言模型（Retrieval-Augmented Large Language Models, RA-LLMs）的古漢語知識問答系統，使用者能透過網頁端簡單的聊天畫面(例如 ChatGPT)即時解答古漢語知識性問題，以及句讀標點、文白翻譯、典故解釋、字詞解析任務。該系統整合了檢索模組（Dense Passage Retriever, DPR）與生成模組（GPT-4o），通過 Azure 平臺的監控系統（Azure Monitor）動態微調優化檢索模組參數，同時實現知識庫的即時更新，並輔以問題分類器實現查詢分流，有效提升系統的回應準確性和專業性。此檢索增強生成模組的結合，能改善生成模型對古漢語背景知識的掌握，避免產生偏誤資訊。

透過此系統的開發與應用，該工具將有效提升學生理解國文學習中的難點，促進國文教育的數位化發展，為教育科技領域提供創新解決方案。計畫之結果期望能夠驗證該研究計畫之問題，即**基於 RA-LLMs 的古漢語知識問答系統，在多大程度上能夠幫助學生克服古漢語學習中的困難、是否有效解決即時性學習需求，以提升其國文學習效率**。此外，此計畫將檢驗 RA-LLM 技術在古漢語綜合知識問答上的應用效果，並探索此類輔助工具在教育場景中的長期應用潛力。

## 六、參考文獻

- Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.-S., & Li, Q. (2024). A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. <https://doi.org/10.48550/arXiv.2405.06211>
- Guo, G., Yang, J., Lu, F., Qin, J., Tang, T., & Zhao, W. X. (2023). Towards Effective Ancient Chinese Translation: Dataset, Model, and Evaluation. <https://doi.org/10.48550/arXiv.2308.00240>
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-t. (2020). Dense Passage Retrieval for Open-Domain Question Answering. <https://doi.org/10.48550/arXiv.2004.04906>
- Lai, C.-H., & Tseng, K.-C. (2022). Applying Deep Learning Models to Analyze Users' Aspects, Sentiment, and Semantic Features for Product Recommendation. *Applied Sciences*, 12(4), 2118. <https://doi.org/10.3390/app12042118>
- 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. (2021). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. <https://doi.org/10.48550/arXiv.2005.11401>
- Liu, D., Lv, J., Yang, K., & Qu, Q. (2020). Ancient-Modern Chinese Translation with a Large Training Dataset. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 19(1), 1-13. <https://doi.org/10.1145/3325887>
- Peng, B., Zhu, Y., Liu, Y., Bo, X., Shi, H., Hong, C., Zhang, Y., & Tang, S. (2024). Graph Retrieval-Augmented Generation: A Survey. <https://doi.org/10.48550/arXiv.2408.08921>
- Wei, Y., Xu, Y., Wei, X., Yang, S., Zhu, Y., Li, Y., Liu, D., & Wu, B. (2024). AC-EVAL: Evaluating Ancient Chinese Language Understanding in Large Language Models. <https://doi.org/10.48550/arXiv.2403.06574>
- 王博立, 史曉東, & 蘇勁松. (2017). 一種基於迴圈神經網路的古文斷句方法. *北京大學學報自然科學版*, 53(2), 255-261. <https://doi.org/10.13209/j.0479-8023.2017.032>
- 劉金柱, 王錦繡, 羅捷春, 李志芳, 袁方, 餘靜靜, 龔丹, 謝雨霏, 羅婉滢, 鄭蘇楠, 陳曠心, 賀心雨, 張潤哲, 夏婉婷, 謝佳延, 呂佳源, 呂萍, 余樂妍, 鄭詩銘, ... 劉根輝. AI 九思：用大語言模型煥新古漢語之美. <https://chinaxiv.org/abs/202501.00212>
- 李紳, 胡勍奮, & 王立軍. (2024). 古漢語大語言模型的構建及應用研究. *語言戰*

- 略研究, 9(05), 22-33. <https://doi.org/10.19689/j.cnki.cn10-1361/h.20240502>
- 邱冰, & 皇甫娟. (2008). 基於中文資訊處理的古代漢語分詞研究. *微電腦資訊*, 24(24), 100-102.
- 俞敬松, 魏一, & 張永偉. (2019). 基於 BERT 的古文斷句研究與應用. *中文資訊學報*, 33(11), 57-63.
- 胡韜奮, 李紳, & 諸雨辰. (2021). 基於深層語言模型的古漢語知識表示及自動斷句研究. *中文資訊學報*, 35(4), 8-15.
- 袁義國, 李斌, 馮敏萱, 賀勝, & 王東波. (2022). 基於深度學習的古籍文本自動斷句與標點一體化研究. *圖書情報工作*, 66(22), 134-141. <https://doi.org/10.13266/j.issn.0252-3116.2022.22.012>
- 梁社會, & 陳小荷. (2013). 先秦文獻《孟子》自動分詞方法研究. *南京師範大學文學院學報*(03), 175-182.
- 黃水清, 王東波, & 何琳. (2015). 以《漢學引得叢刊》為領域詞表的先秦典籍自動分詞探討. *圖書情報工作*(11), 127-133.
- 程甯, 李斌, 葛四嘉, 郝星月, & 馮敏萱. (2020). 基於 BiLSTM-CRF 的古漢語自動斷句與詞法分析一體化研究. *中文資訊學報*, 34(4), 1-9.