

透過檢索增強生成技術提升學習成效：

應用於大學計算機概論課程的教育聊天機器人

*Enhancing Learning Outcomes through Retrieval-Augmented Generation:
An Educational Chatbot for Computer Science Courses*

ABSTRSCT

本研究旨在設計並開發一款基於檢索增強生成(RAG)技術的教育聊天機器人，並將其應用於大學計算機概論課程，以提供個性化且高效的學習支持。系統整合歷屆考試題目與教學資源，並採用語義檢索與內容生成技術，確保在學習場景中的實用性、互動性與精確性。研究針對南部某大學 98 名一年級學生，進行為期四週的系統測試，旨在測量系統在語義完整性、內容準確性及學習支持功能方面的表現。進一步探討 Warmth Perception、Competence Perception、Post-Recovery Satisfaction、Re-Use Intentions 等心理變項。研究結果顯示，這些心理變項之間呈現顯著正向關係，反映了情感支持與系統效能的提升用戶滿意度與參與度方面的重要性。然而，根據相依樣本 t 檢定結果，系統在使用前後的系統可用性量表(SUS)未達顯著差異。揭示了系統在滿足使用者需求方面仍需進一步改進。反映出系統設計與使用者需求間仍存在優化空間。未來可進一步優化系統設計，滿足學習者多樣化需求，提升教育科技於高等教育場域的應用潛力。

INDRODUCTION

聊天機器人技術在過去六十年間實現了顯著進步，這得益於自然語言處理(NLP)和機器學習(ML)的快速發展，以及大型語言模型(LLMs)的廣泛應用。聊天機器人是一種基於文本的數位對話代理，通過自然語言處理技術與用戶進行互動(Wirtz et al., 2018)。相比於傳統資訊系統，聊天機器人在互動性和智慧化方面具有優勢(Maedche et al., 2019)。這些特性使其成為企業提升服務效率與提供個性化體驗的重要工具。隨著聊天機器人技術的成熟，其應用範圍已迅速擴展至服務業、旅遊業、醫療保健及教育等多個領域(Van et al., 2020)。從學術角度看，研究重點集中於數位代理的設計元素(Kull et al., 2021)、用戶對數位互動的反應(Mozafari et al., 2021)以及用戶對系統失敗的應對策略(Sands et al., 2022)。在處理敏感議題方面，聊天機器人表現出巨大的潛力。例如，性騷擾問題對全球數百萬人造成影響，全球約 35% 的女性在其一生中經歷過性伴侶或非伴侶暴力(Vakayil et al., 2024)。受害者往往因為羞恥或擔心受到評判而不願求助，這可能進一步導致心理健康問題(Karami et al., 2021)。在這種情況下，聊天機器人作為匿名支持工具，能為受害者提供關鍵資訊和心理安慰(Følstad et al., 2021)。

根據技術架構，聊天機器人可分為以下兩類：(1)基於規則的聊天機器人：此類系統依賴於固定的對話路徑，基於明確的規則或決策樹實現互動，適用於簡單的任務操作。Bauer et al. (2020)設計了一款針對性騷擾倖存者的規則型聊天機器人，該系統在識別騷擾案件方面具有 98% 以上的準確率，但在處理特定類型的騷擾問題時仍存在挑戰。(2)神經型聊天機器人：神經型系統基於深度學習模型，能直接從輸入文本生成高質量回應，適用於處理複雜對話場景。Socatiyanurak et al. (2021)設計的 LawU 是一款專為泰國地區設計的神經型聊天機器人，

基於 Llama-2 模型，專注於為性騷擾倖存者提供法律支持。儘管取得了顯著技術進步，聊天機器人仍然面臨不少挑戰。回應失敗，如無法正確理解或生成適當答案，常引發用戶的不滿與失望(Brendel et al., 2020; Seeger & Heinzl, 2021)。近 75% 的用戶認為聊天機器人無法正確處理其需求，而約 30% 的用戶表示一次不良體驗即可影響其未來的使用意願(Weiler et al., 2022)。這反映了處理自然語言輸入的固有複雜性，尤其是在面對非結構化或不可預測輸入時，顯示出對恢復策略的迫切需求(Choi et al., 2021)。大型語言模型憑藉其對龐大數據的訓練，在語言細微差別與情感線索的理解方面取得了突破性進展，這對於應用於創傷情境尤為重要(Li et al., 2023)。檢索增強生成(RAG)技術的引入進一步提升了聊天機器人的能力，使其能基於可信來源提供一致且準確的回應。然而，評估這些系統需要大量標註數據，而這些數據的生成耗時且成本高昂。理想的評估數據集需涵蓋廣泛知識領域並具備多元答案格式(Lewis et al., 2021)。Wang (2022)強調，問題生成技術可通過邏輯推理、常識推理等方法提升對話系統的準確性與多樣性。同時，Yuan et al. (2022)提出了一種基於上下文 n-gram 相似度的篩選技術，能有效移除不相關問題，進一步提高數據集的質量。

本研究旨在開發一款基於 RAG 技術的教育聊天機器人，將歷屆計算機概論考古題嵌入系統，並應用於大一計算機概論課程中。該系統通過檢索增強生成技術(RAG)提升檢索效率與內容生成的準確性，提供個性化且高效的學習輔助功能。同時，本研究將深入探討教育聊天機器人在學習過程中的心理層面影響，包括對學生恢復後滿意度、信任感和再利用意圖的影響。研究的核心目標在於揭示基於 RAG 的教育聊天機器人在學術應用中的價值，並分析以下三個關鍵心理層面的關係。我們的研究問題如下：Yuan et al. (2022)

- (1) 基於 RAG 聊天機器人檢索效率與生成內容準確性是否顯著影響學生的恢復後滿意度？
- (2) 基於 RAG 的聊天機器人對學生的信任感是否具有顯著影響？
- (3) 溫暖感知(Warmth Perception)、能力感知(Competence Perception)、恢復後滿意度(Post-Recovery Satisfaction)與再利用意圖(Re-Use Intentions)之間是否存在顯著相關？

RELATED WORK

聊天機器人在學習動機與參與度提升中的角色

現代教育聊天機器人整合了多媒體內容、即時反饋與遊戲化設計(Haristiani & Rifa, 2020)，並且逐步引入增強現實(AR)與虛擬現實(VR)技術，以提升學習的沉浸感與互動性。這些設計特性顯著提高了學習者的動機與參與度。遊戲化元素的引入(特別是任務導向與即時獎勵機制)增強了學習者的內在驅動力與持續投入。Yang (2022)指出，教師認為聊天機器人可以有效處理日常教學任務，釋放教師精力以專注於更具挑戰性的課堂活動。學生則認為即時反饋功能幫助他們深入理解學習內容，特別是在語言學習這類需要高頻率練習的領域。表明聊天機器人的設計需更加聚焦於學習者的需求，以最大化其教育效能。Yin et al. (2021)比較了使用聊天機器人支持的實驗組與未使用聊天機器人支持的控制組在本科生中的表現，發現兩組在學習成績上無顯著差異。然而，與聊天機器人互動的學生展現出更高的學習動機，顯示出聊天機器人在激發學習興趣方面的潛能。暗示學習者對學習資源的接觸頻率及主動性可能是其學習成果的重要決定因素，而非僅僅依賴學習模式的改變。另一方面，Hsu et al. (2023)指

出，聊天機器人能減少學習者的焦慮與認知負擔，營造出更具吸引力的學習環境。透過即時回應與個性化的支持，聊天機器人提供了類似於人際互動的學習體驗，進一步促進了學習者的參與感與專注力。然而，Fryer et al. (2019)強調，學生對聊天機器人的新奇感可能會隨著時間減弱，導致長期參與度的下降。這種效應揭示了設計具持久吸引力的互動元素對於教育聊天機器人的重要性。Pérez et al. (2020)將教育聊天機器人分為以下兩類：(1)服務導向型聊天機器人：提供招生、入學和圖書館服務等支援。(2)教師導向型聊天機器人：作為課堂助手，生成知識、提高學生參與度並提供智能反饋(Vázquez-Cano et al., 2021)。此外，這些工具亦協助高等教育機構優化現有資源配置，減少人力成本，同時開發創新教育服務(Hien et al., 2018)。教育聊天機器人結合人工智慧與自然語言處理技術，可進行文本或語音形式的互動式教學(Pérez et al., 2020)。FAQ 型聊天機器人(Han & Lee, 2022)、模仿人機交流的早期 NLP 程式 ELIZA(Natale, 2019)，以及 MOOC 平台中促進學習者互動的虛擬代理(Tegos et al., 2019)。這些系統已在多樣化的學習場景中被廣泛接受。聊天機器人能在教師不便隨時提供指導時，模擬人類對話，幫助學生回顧學習內容(Göschlberger & Brandstetter, 2019)。不僅能提升學生的學業成就與自我效能，還促進了自適應學習能力(Fadhil & Villafiorita, 2017)。更重要的是，這些工具作為師生之間的橋樑，鼓勵學生以自己的學習節奏進行探索，並促進不願在傳統教室環境中提問的學生積極參與(Verleger & Pembridge, 2018)。

在語言學習領域，聊天機器人整合到語言學習應用中的趨勢在近年來迅速增加(Kohnke, 2023)。聊天機器人可以為語言學習者提供補充資源和即時幫助，例如訪問在線詞典、數位材料和即時社交媒體。聊天機器人可以整合語音識別和合成技術(R&S)，並根據個別學習者的具體需求進行定制，從而在學習過程中提供個性化的反饋和適應性的支持(Jeon et al., 2023)，其主要目的是讓學習者在數位環境中參與模擬對話，提供互動式練習以鍛煉發音、詞彙和語法，實時檢測和糾正錯誤，並根據個別學習者的需求調整教學內容。聊天機器人作為虛擬語言導師的廣泛可及性，無論時間和空間限制如何，都相對於人類教師而言是一個重大優勢。研究顯示，人工智慧聊天機器人(AIC)在語法(Kharis et al., 2022)、詞彙(Ajisoko, 2020)、寫作(Lin, et al., 2020)和對話(Hakim & Rima, 2022)等方面促進顯著改進。AIC 能即時糾正語言錯誤，提供精準的語言支持，使學習者能快速識別並改進語言薄弱環節。這些功能不僅提升了語言學習效率，也為學習者提供了多樣化的練習機會。然而，受限於依賴大數據進行學習與預測，其在處理不同口音、語境與語言細微差異時可能出現誤解或錯誤。Pham et al. (2018)探討了行動應用語言學習(MALL)中聊天機器人的角色，指出其易用性與實用性獲得了用戶的高度評價。透過移動設備接觸學習資源的便利性，使得學習不再受限於時間與空間，進一步提高了學習資源的普及性與可及性。Chen et al. (2020)研究顯示聊天機器人對詞彙學習的輔助效果顯著，並通過技術接受模型(TAM)驗證了其對學習者滿意度的正向影響。Yang (2022)研究指出，教師認解決等高層次的教育目標。然而，儘管教師認為聊天機器人能有效減輕教學負擔，其在處為聊天機器人能有效減輕教學負擔，讓他們能將更多精力集中於批判性思維、問題理語言細節上的不足仍是教學中需克服的挑戰。Vera (2023)指出，聊天機器人在處理複雜句子結構、上下文含義與非標準語言時存在局限性。為了克服這一限制，設計者需加入更靈活的語境調整功能，並採用更豐富的語料庫訓練模型。Rapp et al. (2021)指出，聊天機器人的人性化互動有限，可能使學生感到疏離，進而影響學習效果。揭示了在技術規劃中融入更自然語言生成與情感計算的必要性，以減少學習過程中的冷漠感。

情緒與動機在聊天機器人輔助學習中的互動關係

情感在學習過程中扮演著至關重要的角色，深刻影響學生的認知過程、注意力和信息處理(Pekrun, 1992)。情感已被證明會引導不同的行動及其動機意圖(Scarantino, 2014)。在教育聊天機器人互動的情境中，需特別關注負向情感對工作記憶的影響。工作記憶是用於保留與操作信息以執行各種心智能力的記憶系統(Boon, 2011)。此外，這些情感可能會中斷並重新組織學習目標，使學習者的注意力從正在進行的任務轉移到其他目標(Losenno et al., 2020)。這強調了在教育聊天機器人互動中理解與管理誘發情感的重要性，以創造支持性與有效的學習體驗。誘發情感受到聊天機器人回應、使用者的個人行為以及聊天機器人與人類互動整體體驗的影響(Jin & Youn, 2023)。興趣、愉悅和好奇心等正向情感與更高的學習動機、更好的信息記憶以及增強的問題解決能力密切相關。相對地，挫折、無聊和焦慮等負向情感阻礙學習過程，導致學習動機下降和參與度降低。Ramirez-Arellano et al. (2018)的研究指出，動機與情感之間具有因果互動關係。低動機可能導致學習任務的缺乏啟動或中斷，而高學習焦慮可能會影響內在與外在學習動機(Wang et al., 2022)。其他研究也在不同情境中證實了這種關係。例如，Tze et al. (2016)的一項元分析發現，無聊對中學與大學學生的動機有顯著影響。而 Pekrun et al. (2011)揭示，憤怒、焦慮與羞愧等負向情感會在上課、學習或參加考試時降低內在動機。為了應對情感對學習的影響，教育中的聊天機器人逐漸融合情感智能，能夠辨識並適當回應使用者的情緒狀態(Ehtesham-Ul-Haque et al., 2024)。儘管關於教育聊天機器人及其對學習動機影響的研究不斷增多，但對雙重情感(ambivalent emotions)對學習動態影響的理解仍存在顯著的研究空白(Kiuru et al., 2020)。雙重情感指在互動過程中同時體驗正向和負向情感(Yin et al., 2024)。例如，學習者在與教育聊天機器人互動時，可能會同時感受到獲得即時回饋的興奮和難以理解概念的挫敗。Tulis & Fulmer(2013)指出，動機與情感顯著影響學習者在學術環境中的參與度、持續性和成功。當學生感到有能力、認為行動與結果之間存在穩定聯繫、重視學科內容並體驗到正向情感時，他們更可能被激勵。相對地，負向情感會導致學習者的脫離。創建有利的學習環境並有效管理情感可增強學習的持久性和資源管理能力(Boekaerts, 2010)。情感敏感的教學能抑制恐懼和憤怒等負向情感，促進快樂與掌握感等正向情感，從而創造更支持的學習環境(Astleitner, 2000)。

而回饋(feedback)在學習中具有多種用途，包括自我評估、支持學習、評分、澄清期望、檢測錯誤以及提高動機(Hattie & Timperley, 2007)。自然語言處理(NLP)領域的進展使教育聊天機器人具備了適應性回饋功能。學生現在可以針對既定標準，獲得即時且全面的指導，並接受量身定制的改進策略。Chuah & Kabilan (2021)定義了教育聊天機器人情境中的感知(perception)，指用戶對其與聊天機器人互動的態度與看法，包括感知的實用性、接受度及潛在興趣。了解用戶對聊天機器人的觀點，包括其接受度與偏好，是不可或缺的。同樣地，滿意度被描述為用戶在與聊天機器人互動中，其需求與期望被滿足的程度，涵蓋語言及設計層面。Chocarro et al. (2023)強調聊天機器人的主動性及用戶個體特徵在塑造學生滿意度中的重要性。儘管聊天機器人日益流行，但針對教育聊天機器人互動中情感作用的研究仍相對有限(Deng & Yu, 2023)。多數情感聊天機器人的研究集中在健康與福祉領域(Stević et al., 2023; Tudor Car et al., 2020)。Kuhail et al. (2023)回顧了教育聊天機器人，分析了其涵蓋的平台、設

計原則、功能、互動風格及限制等多方面內容。他們發現大多數教育聊天機器人以網絡為導向，涵蓋計算機科學、語言、教育、工程和數學等領域。這些聊天機器人通常作為指導代理存在，遵循既定路徑，有些則使用個性化學習。然而，這些研究未直接探討互動中的情感反應。**Okonkwo & Ade-Ibijola (2021)**強調了聊天機器人在教育中的多種優勢，包括無縫的內容整合、快速的信息訪問、動機與參與度的提升、多用戶支持及即時協助。然而，這些優勢未涉及教育聊天機器人互動中的情感與動機層面。考慮到情感與動機之間的密切關係，進一步研究此領域的必要性尤為突出。**Guo et al. (2023)**探索了一種新穎的課堂辯論方法，利用聊天機器人互動來促進論辯性對話。該方法分為三個階段：學生與聊天機器人互動以激發創意、小組內部對生成的想法進行討論，並參與與其他小組的辯論。這些過程受到聊天機器人互動的見解與情感的影響。該研究評估了學生在行為、認知和情感三個層面的參與度。數據來源包括聊天機器人互動記錄、聊天日誌、音頻錄音及學生的反思筆記。研究發現，學生對將聊天機器人互動融入辯論準備的態度普遍正向。這種積極情緒源於聊天機器人激發創新想法的能力、促進辯論的獨特方式，以及通過互動引發的情感，這些情感營造了一種放鬆且富有成效的學習氛圍。**Jasin et al. (2023)**開發了自動化問答聊天機器人，用於幫助在線化學學生。該聊天機器人結合了同步溝通與教師即時性技術，並以情感化方式設計，旨在通過建立類人連接來增強學生的學習。在一項試點研究中，新加坡一所大學的 12 名在線化學學生參與了質性訪談與自我報告數據收集。主題分析顯示了多樣的情感與行為結果，包括在解答複雜問題或提供資源指引時，學生信心與熟練度的增加。然而，當聊天機器人給出錯誤提示或誤解問題時，學生會產生負向情感，暴露了其局限性。這項研究突出了以情感為核心的聊天機器人改善在線學習的潛力，特別是強調準確的理解與回應。

從對話中檢索：RAG 對聊天機器人語境理解的提升

檢索增強生成(Retrieval-Augmented Generation、RAG)是一種先進的自然語言處理(NLP)技術，結合了檢索與生成技術，以提升文本生成任務的性能。概念首次由 **Lewis, et al.(2020)**提出，並在其重要論文"Retrieval-augmented generation for knowledge-intensive NLP tasks"中闡述，主要應用於 Seq2Seq 模型。RAG 有效地將參數化記憶(模型參數)與非參數化的外部記憶結合，以實現更好的結果。RAG 通過檢索相關文檔片段來補充大型語言模型(LLMs)的外部知識，以解決其在特定領域或知識密集型任務中的局限性(**Kandpal et al., 2023**)。透過語義相似性計算，RAG 增強了生成內容的事實準確性，有效減少了 LLM 常見的幻覺(hallucination)現象(**Zhang et al., 2023**)。RAG 與 LLM 的結合促進了其在現實應用中的廣泛採用，成為推進 聊天機器人技術和增強 LLM 現實應用適用性的關鍵技術(**Arora et al., 2023**)。

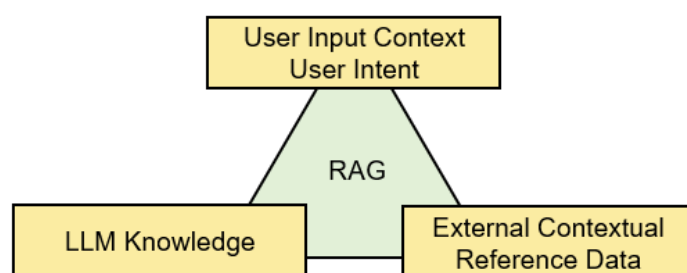


Fig 1 : LLMs and Data Management

RAG 的發展與 Transformer 架構和預訓練模型(Pre-training Models、PTMs)興起密切相關。在早期階段，RAG 專注於通過 PTMs 引入外部知識以增強語言模型(Arora et al., 2023)。ChatGPT 的出現進一步加速了 RAG 的發展。隨著強大的上下文學習(In-Context Learning、ICL)能力的引入，研究重點轉向利用 RAG 在推理階段提供更好的信息，以應對更複雜和知識密集型的任務(Lewis et al., 2020)。隨著研究進展，RAG 開始結合微調技術(Fine-tuning)，超越僅僅局限於推理階段的增強。這些進展被分為三種主要範式：Naive RAG、Advanced RAG 和 Modular RAG，其中 Naive RAG 是最早的方法學，遵循「檢索-閱讀」(Retrieve Read)框架，包括索引、檢索和生成(Ma et al., 2023)。過程從索引開始，將 PDF、HTML 等各種格式的原始數據清理並分段為更小的片段，編碼為向量表示並存儲在向量數據庫中。在檢索階段，RAG 將用戶查詢轉換為向量表示，計算相似度分數，檢索出最相關的片段，這些片段隨後用於生成回答。然而 Naive RAG 仍面臨以下挑戰：(1)準確性和召回率的困難，可能導致檢索到不相關或遺漏的信息。(2)可能出現幻覺、不相關或偏見的回答，影響質量和可靠性。(3)檢索信息的整合可能導致輸出冗餘或不連貫(Ma et al., 2023)。

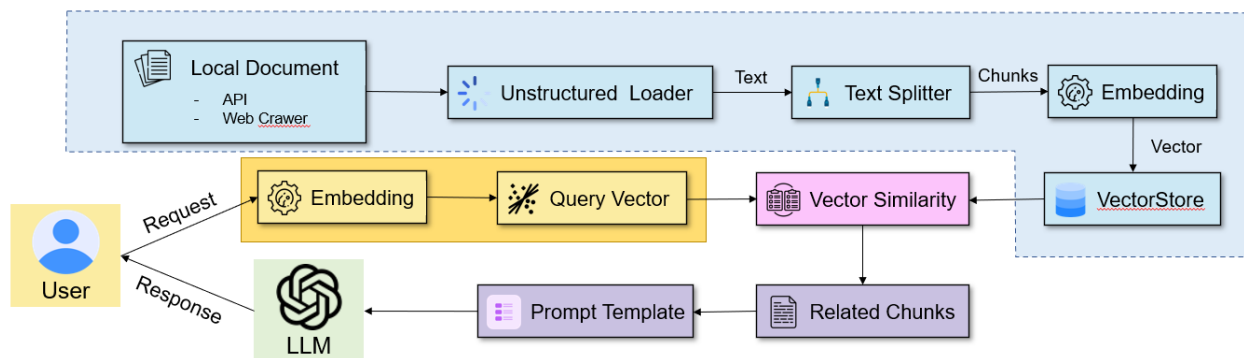


Fig 2：RAG 架構示意圖

高效的檢索(Retrieval)是 RAG 的關鍵，因為檢索來源和粒度會顯著影響生成內容。檢索包括(1)檢索來源：包括非結構化數據(文本、維基百科)、半結構化數據(PDF)和結構化數據(知識圖譜)。最近的趨勢還包括使用 LLM 生成的內容進行檢索(Zhang et al., 2023)。(2)檢索粒度：粗粒度單位(例如文檔)提供更多上下文，但可能引入噪聲；而細粒度單位(例如句子)提高了精確性，但增加了計算負擔(Teja, 2023)。(3)索引優化：通過分塊策略、元數據附加和層次結構或知識圖譜索引增強檢索質量(Teja, 2023)。索引階段的質量直接決定了後續檢索和生成的精確性。有效的索引可以減少檢索階段的不確定性，提高系統的整體表現。嵌入(Embedding)通過計算查詢和文檔片段的語義相似性來實現檢索。常用方法包括稀疏編碼器(例如 BM25)和密集檢索器(基於 BERT)。主要進展包括：(1)混合檢索：結合稀疏和密集嵌入以捕獲互補特徵，增強魯棒性(Robustness)(Lewis et al., 2020)。此技術特別適用於處理包含稀有實體或具體專業術語的查詢，有助於系統應對多樣化的應用場景。(2)微調嵌入模型：針對特定領域(像是醫療、法律)調整嵌入或通過監督學習使檢索器與 LLM 對齊。這種方法能夠提高系統在專業領域的適配性，並增強對特定上下文的理解能力。在檢索之後，有效的生成(Generation)涉及精煉檢索內容和微調 LLM：(1)上下文策劃：過濾冗餘信息以避免干擾 LLM，通過重排序和上下文壓縮等技術優先處理相關信息(Liu et al., 2024)。(2)重排序：重

排序策略不僅有助於突出最相關的內容，還能通過過濾噪聲進一步提升生成的精確性 (Zhuang et al., 2023)。(3)微調 LLM：使模型輸出與特定領域需求或用戶偏好對齊，利用強化學習或知識蒸餾技術(Zhuang et al., 2023)。同時對生成模型進行定向微調還可以實現個性化響應，滿足多樣化的用戶需求。RAG 正在不斷發展，迭代、遞歸和自適應檢索方法逐漸受到關注。迭代檢索通過重複搜索建立全面的知識庫，而遞歸檢索則細化查詢以提高相關性。這些技術不僅增強了系統對複雜查詢的處理能力，也為探索新的應用場景提供了可能性。此外，自適應檢索方法允許 LLM 動態決定檢索的時機和內容，提高效率和準確性。這些方法通過引入主動判斷機制，賦予模型更高的靈活性和決策能力，有助於提升系統在多樣化場景中的表現。

METHODOLOGY

研究流程

本研究旨在設計並實施一套基於 RAG 技術的聊天機器人，並聚焦於系統的實驗過程，旨在檢測學生在實際使用系統過程中的認知變化、情緒反應及行為意圖，並透過科學化的量化與質化分析，深入探討系統使用後的學習滿意度、能力感知以及再使用意圖等心理層面的影響。研究對象為台灣南部大學工程相關科系共 98 位一年級學生，以大一必修計算機概論進行實驗，為期 4 周進行。為了深入了解學生在學習過程中的心理反應，實驗設計特別聚焦於三個核心心理成分的測量：溫暖感知(學生對系統親和力及支持感的評估)、能力感知(學生對系統產出內容的專業性與可信度的主觀認知)，以及學習後滿意度(學生在使用系統後的整體情緒反饋與體驗)。這些心理成分的測量將透過標準化的量表進行評估，例如 SUS(System Usability Scale)量表及自編的學習滿意度問卷，在數據分析階段，本研究將採用量化分析方法，確保結果的嚴謹性與解釋力。在量化分析方面，透過信度分析、相依樣本 t 檢定、縣性回歸與 Pearson 相關分析探討系統使用過程中的檢索效率、內容準確性與學生心理成分間的相互關聯性，進一步驗證系統對於學習信心、溫暖感知及再使用意圖的影響程度。最後，在結果討論與結論推動階段，本研究將依據分析結果進行綜合性討論，並著重於以下幾個重點：系統在提升學習效能方面的具體成效、學生心理感知層面的變化趨勢，以及系統可用性對學習動機與再使用意圖的影響。本研究的最終目標在於提出具體而可行的建議，為未來教育輔助工具的設計與應用提供實證依據，並進一步探討如何優化學習者的使用體驗與學習成效，促進教育科技在大學教學場域中的落地與深化應用。

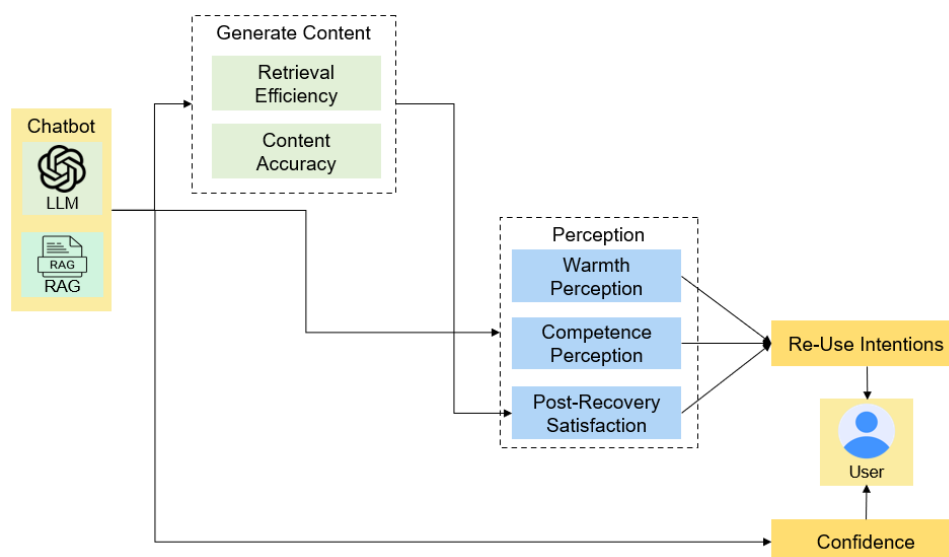
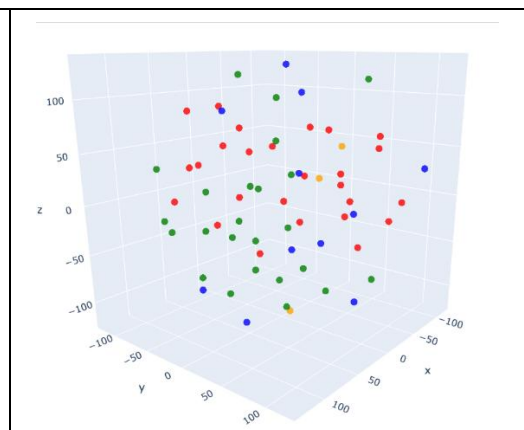
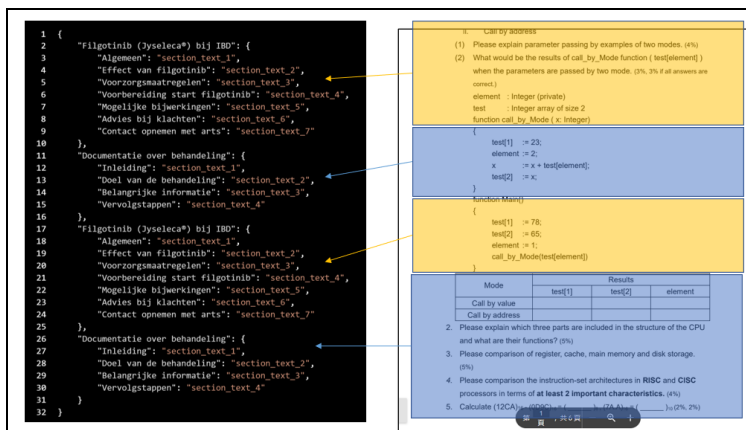


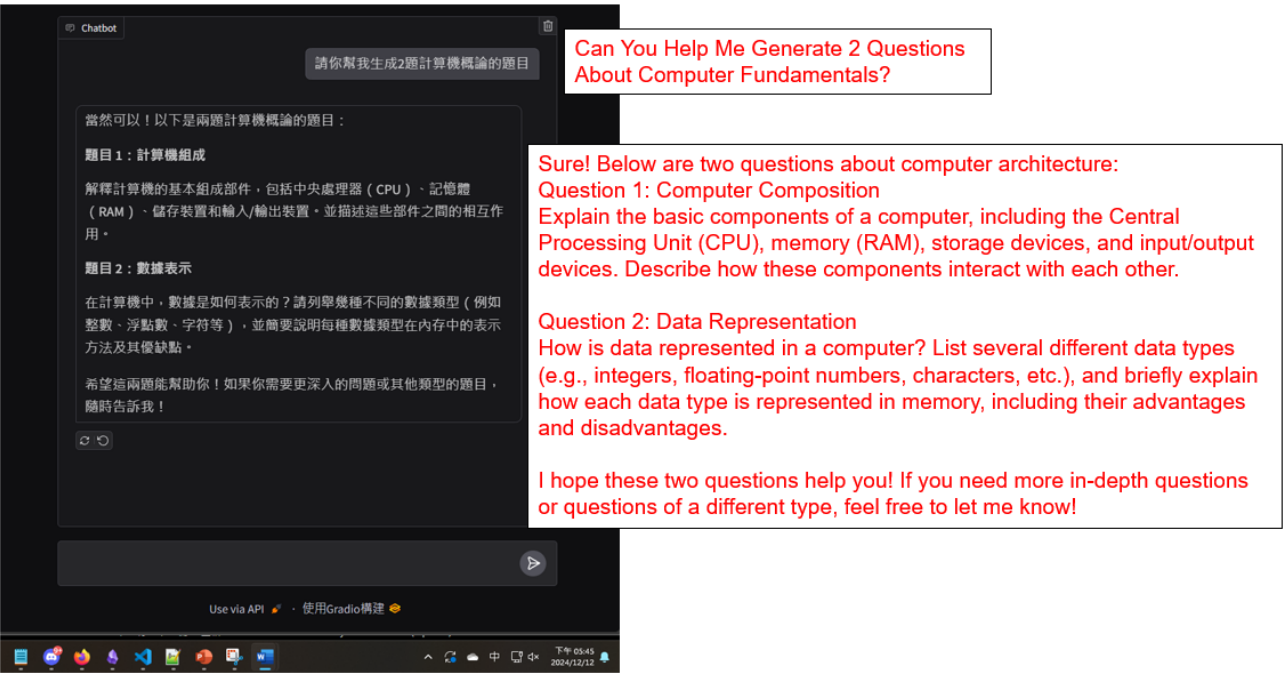
Fig 3：研究架構圖

系統架構

本研究旨在設計與實現一套基於檢索增強生成技術(RAG)的教育聊天機器人系統，專為大學計算機概論課程的輔助學習設計。該系統嵌入歷屆考古題，結合語義檢索與生成技術，為學生提供個性化的學習支持，顯著提升學習效率與效果。同時，本系統以實際教育場景需求為基礎，重點強調技術的應用性與學生的互動體驗，確保能解決教學中的痛點問題。整體架構分為數據處理模組、檢索模組、生成模組與用戶交互模組，每個模組均依照特定功能進行設計，並在系統中相輔相成。數據處理模組負責高效整理並轉換大量教育資源，檢索模組則著重於準確定位學生所需的內容，生成模組以智能化回應支持學習深度，用戶交互模組則提供直觀、易用的界面，實現全方位的學習支援，從而打造一個高效且互動的教學輔助工具。在數據蒐集階段，本研究以過去五年的計算機概論考古題、教學講義、課堂錄音及其他學習資源為核心數據集。為了實現高效檢索與語義一致性，我們採用 LangChain 中的 DirectoryLoader 與 TextLoader，將所有文本數據轉換為內存文檔並附加相關元數據(如文件類型、來源等)。為提升檢索效率與上下文完整性，我們使用 CharacterTextSplitter 將文檔分割為長度為 2000 字符的文本塊，並設置 400 字符的重疊區間，確保語義連貫性並避免信息缺失。檢索模組使用 OpenAI 的嵌入模型(OpenAIEmbeddings)，將處理後的文本塊轉換為高維向量，並存儲於 Chroma 向量數據庫中。這種嵌入技術允許系統通過語義相似性檢索相關內容，同時具備高效檢索能力與持久化存儲特性。為進一步提升檢索準確性，我們對 retriever 進行優化，包括引入動態篩選策略以提高語義相關性的檢索結果。



生成模組採用 OpenAI 的 GPT 模型作為核心生成引擎，並結合 LangChain 的 ChatOpenAI 工具進行實現。該模組通過調整 GPT 模型參數，實現內容生成的語氣、風格與專業性需求。此外，我們採用 ConversationBufferMemory 記錄對話歷史，確保多輪對話的上下文一致性，並通過 ConversationalRetrievalChain 整合檢索與生成模組。此架構確保系統生成的答案既具備高度相關性，又能滿足用戶的上下文需求。



為了提升用戶體驗，我們選用 Gradio 構建友好的交互界面，使用戶能夠方便地輸入問題並獲取回應。Gradio 的實時交互特性支持多輪對話模式，進一步增強了系統的智能化表現。同時，界面設計注重簡潔與直觀性，方便學生快速上手使用。上述模組通過嚴密的技術整合，構建了完整的 RAG 型教育聊天機器人系統。數據處理模組負責數據預處理與分割，檢索模組通過高維向量實現語義檢索，生成模組則提供與檢索結果高度相關的內容生成，而用戶交互模組確保了流暢的使用體驗。這一整體架構的設計與實施，旨在提供精確、高效且具互動性的學習支持，為學生帶來嶄新的學習體驗。

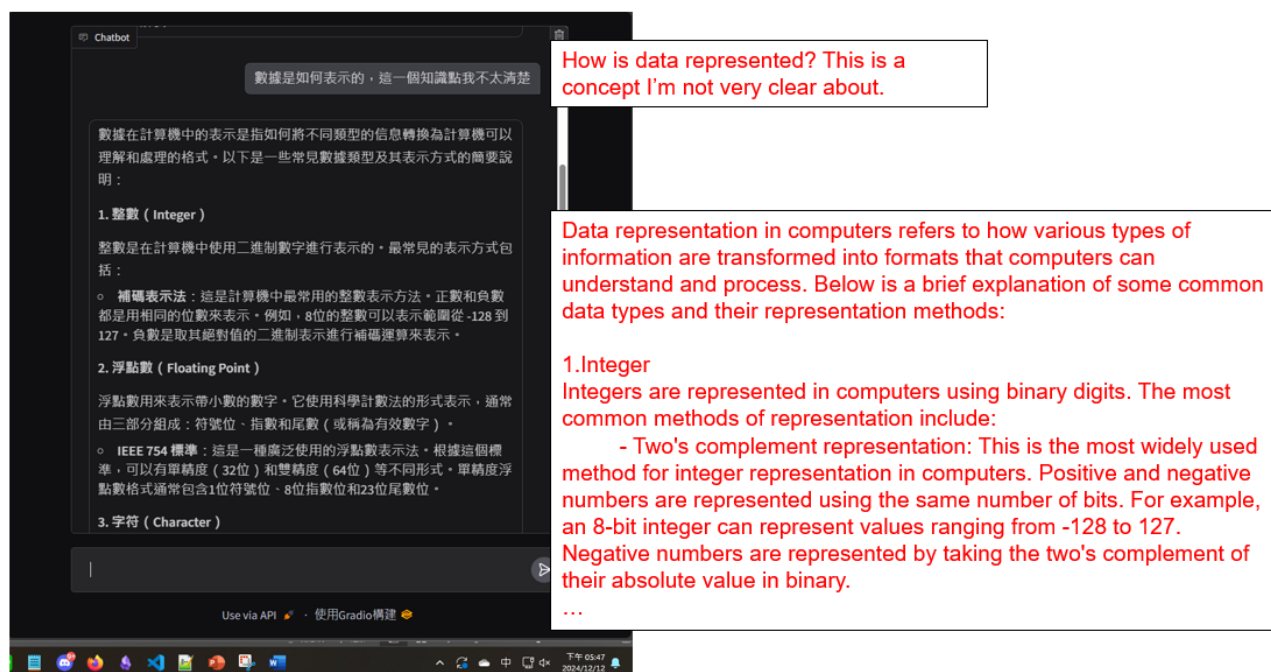


Fig 7：維持上下文記憶

研究工具

對於研究工具，我們選用 Haupt et al(2023)研究之量表進行修改，並稱呼為聊天機器人交互體驗量表(Chatbot Interaction Experience Scale, CIES)。量表一共包含五個維度，每個維度共 3 題。溫暖感知(Warmth Perception)用於衡量受試者用戶對聊天機器人在情感層面的感知，特別是其是否具備親和力、友善和關懷的特質。能力感知(Competence Perception)評估用戶對聊天機器人在執行任務和解決問題時的專業性與效率的感知。恢復後滿意度(Post-Recovery Satisfaction)評估用戶在經歷服務問題後，對聊天機器人恢復處理的滿意程度。再利用意圖(Re-Use Intentions)衡量用戶在未來是否願意再次使用聊天機器人服務的意圖。檢索效率(Retrieval Efficiency)評估聊天機器人提供檢索功能的速度和準確性。生成內容準確性(Content Accuracy)評估聊天機器人生成的答案是否準確、相關且有用。本研究針對聊天機器人交互體驗量表進行可信度分析，分析結果顯示，其 Cronbach' s Alpha 內部一致性係數為 0.96，依據 Nunnally(1978)之標準，具可接受之信度。

Table 1：聊天機器人交互體驗量表(Chatbot Interaction Experience Scale, CIES)

Item	Description
WP1	I think the chatbot is warm.
WP2	I think the chatbot is kind.
WP3	I think the chatbot is caring.
CP1	Do you think the chatbot is competent?
CP2	Do you think the chatbot is effective?
CP3	Do you think the chatbot is efficient?
RS1	I feel satisfied with the chatbot's service.

RS2	I find the chatbot's service pleasant.
RS3	I find the chatbot's service good.
US1	I will use this chatbot again.
US2	I intend to use the chatbot regularly.
US3	I would use the chatbot for learning computer science.
RE1	The chatbot's response speed meets my needs.
RE2	I can quickly find relevant answers using the chatbot.
RE3	The chatbot provides accurate and consistent responses.
CA1	The chatbot provides accurate answers.
CA2	The chatbot's generated responses are highly relevant to the questions.
CA3	The chatbot's responses align with the learning materials.

而為了有效評估本系統本身的使用性，我們選用 SUS 量表(System Usability Scale)修改並作為系統測量依據。量表由 Brooke(1996)提出，至今已成為國際上最廣泛使用的使用性評估工具之一。該量表因其簡單、易用且具高度可靠性，廣泛應用於軟硬體、網站及各類系統的使用者體驗(User Experience, UX)評估中，尤其是在學術研究、商業開發和使用測試等多個領域，具有重要的參考價值。本研究針對 SUS 系統量表進行可信度分析，分析結果顯示，其 Cronbach' s Alpha 內部一致性係數為 0.83；具可接受之信度。

Table 2：SUS 系統量表

Item	Description
SUS1	I think I would frequently use this chatbot to assist my computer science learning.
SUS2	I find this chatbot very easy to use.
SUS3	I think the chatbot integrates search and output functions well.
SUS4	I believe most students can quickly learn how to use this chatbot.
SUS5	I feel confident when using this chatbot.
SUS6	I think technical support is needed to effectively use this chatbot.
SUS7	I need to learn many skills before I can use this chatbot.
SUS8	I find the chatbot's design overly complex.
SUS9	I think the chatbot's design lacks consistency.
SUS10	I think the chatbot's operation is cumbersome.

RESULT

首先針對 Retrieval Efficiency、Content Accuracy 對應 Post-Recovery Satisfaction 之關係。結果如表所示：在考慮其他變項下，Retrieval Efficiency 與 Post-Recovery Satisfaction 有顯著之關係($\beta=0.23$, $p=0.03$)。Retrieval Efficiency 得分愈高，其 Post-Recovery Satisfaction 表現也愈好。Content Accuracy 與 Post-Recovery Satisfaction 也有顯著之關係($\beta=0.45$, $p=0.00$)。Content Accuracy 得分愈高，其 Post-Recovery Satisfaction 表現也愈好。當系統能夠快速且高效地檢

索到使用者所需的內容時，用戶的滿意度將顯著提高。這一現象可部分歸因於受試者在遭遇問題時，對於本研究開發之聊天機器人抱有高度期待。當受試者在緊急或高壓情境中，檢索效率的優劣直接影響他們的壓力水平與情感體驗。高效的檢索不僅縮短了等待時間，還能有效降低受試者的挫敗感與不滿，從而改善其整體互動體驗。這進一步強調了優化檢索技術的重要性，尤其是對於具有高頻交互需求的應用場景，如教育輔助與醫療諮詢等領域。此外系統生成內容的準確性在影響受試者滿意度方面具有更為關鍵的作用。即使檢索效率高，但若提供的內容不具備準確性或相關性，則受試者滿意度依然會大幅降低。內容準確性不僅與受試者的直接需求有關，還影響受試者對系統整體信任感的評估。特別是在信息過載的環境中，準確的信息能幫助受試者快速做出決策，減少認知負擔。受試者對信息質量的期望也反映了現代數位環境中對精準性和專業性的重視。這一結果可能源於用戶在面對大量冗餘或錯誤信息時，對於系統的依賴性更高，因此內容準確性成為衡量系統效能的核心指標之一。整體模型可以解釋 61% 的變異項($R^2=0.61$, $\text{Adj } R^2=0.38$)，意味該結果有效解釋用戶滿意度的變異性。這說明系統的設計策略應兼顧這兩個因素，實現檢索與內容生成的協同優化。特別是針對不同使用者群體的需求，可以進一步開發適配性高的檢索與生成機制，以提升模型的實用性與應用範圍。

Table 3：Retrieval Efficiency、Content Accuracy 對應 Post-Recovery Satisfaction 之線性回歸

	Post-Recovery Satisfaction				
	B	SE B	β	T	p
Retrieval Efficiency	0.23	0.10	0.23	2.24	0.03
Content Accuracy	0.42	0.09	0.45	4.49	0.00
R^2	.61 ^a				
Adj R^2	0.38				
F	28.91				
df	(2, 96)				

在針對學生對於 RAG 的聊天機器人之信任感，我們透過以相依樣本 t 檢定分析發現，使用前與使用後並無顯著差異($t=-0.30$, $p=0.76$)。我們認為這可能反映了幾個潛在的原因：(1)與干預的設計效能有關，我們認為研究開發之聊天機器人未能顯著改變學生的體驗或滿意度，因此在統計檢定中可能無法檢測到使用前後的顯著差異。SUS 量表的分數顯示出使用前後變化極小，說明干預過程中系統設計可能未能充分滿足學生在功能性或體驗上的需求。(2)研究對象也可能影響結果。如果研究對象的使用習慣或背景相對一致，統計檢定可能缺乏足夠的檢測力。即使干預對部分學生可能產生了影響，但整體的影響未能在樣本層面上達到顯著。(3)學生的學習動機或使用態度可能影響結果。如果學生對於聊天機器人的接受度較低，或僅因實驗要求而被動使用系統，這樣的互動很難產生積極的滿意度改變。此外，部分學生可能對新技術抱有較高期待，但實際體驗未能符合這些期望，從而導致分數未出現顯著變化。(4)聊天機器人的設計可能偏離學生實際的學習需求。例如，如果系統過於強調檢索效率與技術精確度，但未能提供情感支持或互動樂趣，這可能導致系統的整體體驗不足，從而未能有效提升使用者的滿意度。

Table 4：使用前與使用後之相依樣本 t 檢定

score	Mean(SD)		df	T	p
	pre-use	post-use			
SUS	2.83(1.17)	2.86(1.12)	100	-0.30	0.76

最後，為了瞭解 Warmth Perception、Competence Perception、Post-Recovery Satisfaction 對應 Re-Use Intentions 之對應關係。我們採取 Pearson 相關係數差異 t 檢定，結果顯示三者皆與 Re-Use Intentions 呈現顯著的正向關係。Warmth Perception 與 Re-Use Intentions 之間的正向關係($r(96)=0.68, p<0.01$)反映了系統在人性化與親和力上的表現對使用者未來意圖的重要性。Warmth Perception 主要關注使用者對系統友善、關懷與情感支持的主觀評價，這類親和性設計在教育環境中特別重要。當使用者感受到系統能夠提供溫暖與支持的互動體驗時，他們的情緒體驗得到改善，進而提升系統的可接受度與滿意度，促使他們願意再次使用。這一現象也呼應了情感在學習過程中扮演的關鍵角色，正向情感不僅增強了使用者的學習投入，還能降低因技術操作所帶來的焦慮與排斥感。其次，Competence Perception 與 Re-Use Intentions 之間的顯著正向關係($r(96)=0.74, p<0.01$)顯示出系統專業性與效能對使用者持續使用意圖的關鍵影響。Competence Perception 指的是使用者對系統在任務完成、問題解決等方面的專業性與效率的評估。在本研究中，當聊天機器人能夠提供準確且高效的回應時，使用者對系統的信任感大幅提升，進而增強他們再次使用的意願。進一步強調了系統效能在學習工具設計中的核心地位，特別是對於需要即時資訊檢索與準確內容生成的學習環境。若系統表現出專業性，使用者會將其視為可靠的學習夥伴，從而增加使用意圖。最後，Post-Recovery Satisfaction 與 Re-Use Intentions 之間的正向關係($r(96)=0.64, p<0.01$)則揭示了使用者在遭遇問題後對系統恢復處理的滿意度對再使用意圖的重要性。Post-Recovery Satisfaction 反映了使用者在系統故障或服務中斷後，對系統恢復處理過程的評價。當系統能迅速且有效地解決問題，並滿足使用者的需求時，使用者的整體體驗得以改善，進而提升對系統的信任與滿意度，這也成為使用者再次選擇該系統的關鍵驅動力。結果說明系統不僅需在常態使用中展現效能，還需具備有效的問題恢復與處理能力，以確保使用者在面臨困難時仍能獲得良好的體驗，進而強化再使用意圖。

	WP1	CP1	RS1	US1
WP1	-			
CP1	.648**	-		
RS1	.690**	.767**	-	
US1	.683**	.745**	.645**	-

**．相關性在 0.01 層級上顯著（雙尾）。

CONCLUSION

本研究發了一款基於 RAG 技術的教育聊天機器人，並專注於提升大學計算機概論課程的學習效果。系統包含數據處理、語義檢索、內容生成等系統架構的整合，並展現高效的檢索能力及內容生成功能，為學生提供了即時且個性化的學習支援。在分析結果上也進一步驗證

Warmth Perception、Competence Perception、Post-Recovery Satisfaction 與 Re-Use Intentions 之間的影響。研究顯示，情感設計與功能性表現同樣重要，對於增強系統的接受度及使用意圖具有積極作用。此外，檢索效率與內容準確性在影響用戶滿意度方面也扮演了重要角色，強調系統在資料檢索與內容生成上的優化潛力。然而，使用前後的系統可用性測試未達顯著差異，反映出系統仍需針對使用者需求進行進一步優化。例如，系統設計可更聚焦於學生的互動體驗與學習需求，加入更靈活的回應機制及增強情感連結的功能，以提升整體學習滿意度。此外，未來研究可考量更長期的介入時間及多樣化的學習場景，驗證系統在不同學習情境下的效果。此外，系統還應著眼於提高穩定性與可用性，尤其是在面對多樣化用戶需求與非結構化數據時，進一步整合語境學習與個性化模型調整功能，以確保能夠應對更複雜的學習場景與多變的使用需求。透過不斷的迭代與用戶反饋循環，這款教育聊天機器人有望在學習輔助工具領域創造更多價值。

研究雖初步驗證了 RAG 技術在教育聊天機器人應用中的潛力，但研究結果也存在可進一步探討的面向。像是擴展研究範疇，將系統應用於其他學科或跨學科領域，檢驗其在不同學習內容與目標下的適用性與效能。此外，系統可加入更先進的自然語言處理技術，如語境強化模型或強化學習機制，以進一步提升回應的精確性與互動的自然度。在情感整合方面，未來可探索如何設計更具情感敏感度的聊天機器人，使其能夠即時辨識並回應使用者的情緒狀態，創造更加貼近人類互動的學習體驗。情感設計不僅可提升學習者的參與度，還能促進其學習動機與學習成效，對於情境性教學場景尤為重要。此外，針對長期使用的影響，未來研究可透過縱向研究設計，分析學生的學習動機、情感反應及學習成果的變化趨勢，進一步評估系統的持續性效益。特別是在多輪對話場景中，透過分析系統在長期互動中的穩定性與一致性，將有助於揭示教育聊天機器人在學習輔助中的最大化價值。透過技術優化與設計創新，教育聊天機器人將能成為提升學習效果與滿足使用者需求的重要工具。未來研究不僅應專注於技術層面的改進，也應考慮如何在實際教育情境中結合創新應用，從而促進教育科技的長期可持續發展，實現智慧教育的全新可能性。

REVERENCE

- Ajisoko, P. (2020). The use of Duolingo apps to improve English vocabulary learning. *International Journal of Emerging Technologies in Learning (iJET)*, 15(7), 149-155.
- Arora, D., Kini, A., Chowdhury, S. R., Natarajan, N., Sinha, G., & Sharma, A. (2023). Gar-meets-rag paradigm for zero-shot information retrieval. *arXiv preprint arXiv:2310.20158*.
- Astleitner, H. (2000). Designing emotionally sound instruction: The FEASP-approach. *Instructional science*, 28, 169-198.
- Bauer, T., Devrim, E., Glazunov, M., Jaramillo, W. L., Mohan, B., & Spanakis, G. (2020). #MeTooMaastricht: Building a chatbot to assist survivors of sexual harassment. In *Machine Learning and Knowledge Discovery in Databases: International Workshops of ECML PKDD 2019, Würzburg, Germany, September 16–20, 2019, Proceedings, Part I* (pp. 503-521). Springer International Publishing.
- Boekaerts, M. (2010). The crucial role of motivation and emotion in classroom learning. *The nature*

of learning: Using research to inspire practice, 91-111.

- Boon, H. (2011). Raising the bar: Ethics education for quality teachers. *Australian Journal of Teacher Education (Online)*, 36(7), 104-121.
- Brendel, A. B., Greve, M., Diederich, S., Bührke, J., & Kolbe, L. M. (2020). You are an Idiot!-How Conversational Agent Communication Patterns Influence Frustration and Harassment. In *AMCIS*.
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry*, 189(194), 4-7.
- Chen, H. L., Vicki Widarso, G., & Sutrisno, H. (2020). A chatbot for learning Chinese: Learning achievement and technology acceptance. *Journal of Educational Computing Research*, 58(6), 1161-1189.
- Chocarro, R., Cortiñas, M., & Marcos-Matás, G. (2023). Teachers' attitudes towards chatbots in education: a technology acceptance model approach considering the effect of social language, bot proactiveness, and users' characteristics. *Educational Studies*, 49(2), 295-313.
- Choi, S., Mattila, A. S., & Bolton, L. E. (2021). To err is human (-oid): how do consumers react to robot service failure and recovery?. *Journal of Service Research*, 24(3), 354-371.
- Chuah, K. M., & Kabilan, M. (2021). Teachers' views on the use of chatbots to support English language teaching in a mobile environment. *International Journal of Emerging Technologies in Learning (iJET)*, 16(20), 223-237.
- Deng, X., & Yu, Z. (2023). A meta-analysis and systematic review of the effect of chatbot technology use in sustainable education. *Sustainability*, 15(4), 2940.
- Ehtesham-Ul-Haque, M., D'Rozario, J., Adnin, R., Utshaw, F. T., Tasneem, F., Shefa, I. J., & Al Islam, A. A. (2024). EmoBot: Artificial emotion generation through an emotional chatbot during general-purpose conversations. *Cognitive Systems Research*, 83, 101168.
- Fadhil, A., & Villafiorita, A. (2017, July). An adaptive learning with gamification & conversational UIs: The rise of CiboPoliBot. In *Adjunct publication of the 25th conference on user modeling, adaptation and personalization* (pp. 408-412).
- Følstad, A., Araujo, T., Law, E. L. C., Brandtzaeg, P. B., Papadopoulos, S., Reis, L., ... & Luger, E. (2021). Future directions for chatbot research: an interdisciplinary research agenda. *Computing*, 103(12), 2915-2942.
- Fryer, L. K., Nakao, K., & Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in human Behavior*, 93, 279-289.
- Göschlberger, B., & Brandstetter, C. (2019). Conversational AI for Corporate e-learning [Conference session]. In *Twenty-first International Conference on Information Integration and Web-based Applications and Services*. New York, NY: Association for Computing Machinery.
- Guo, K., Zhong, Y., Li, D., & Chu, S. K. W. (2024). Investigating students' engagement in chatbot-supported classroom debates. *Interactive Learning Environments*, 32(9), 4917-4933.
- Hakim, R., & Rima, R. (2022). Chatting with AI Chatbots applications to improve English communication skill. *Journal of English Language Studies*, 7(1), 121-130.
- Han, S., & Lee, M. K. (2022). FAQ chatbot and inclusive learning in massive open online courses.

Computers & Education, 179, 104395.

- Haristiani, N., & Rifa'i, M. M. (2020). Combining chatbot and social media: Enhancing personal learning environment (PLE) in language learning. *Indonesian Journal of Science and Technology*, 5(3), 487-506.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112.
- Haupt, M., Rozumowski, A., Freidank, J., & Haas, A. (2023). Seeking empathy or suggesting a solution? Effects of chatbot messages on service failure recovery. *Electronic Markets*, 33(1), 56.
- Hien, H. T., Cuong, P. N., Nam, L. N. H., Nhung, H. L. T. K., & Thang, L. D. (2018, December). Intelligent assistants in higher-education environments: the FIT-EBot, a chatbot for administrative and learning support. In *Proceedings of the 9th International Symposium on Information and Communication Technology* (pp. 69-76).
- Hsu, M. H., Chen, P. S., & Yu, C. S. (2023). Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments*, 31(7), 4297-4308.
- Jasin, J., Ng, H. T., Atmosukarto, I., Iyer, P., Osman, F., Wong, P. Y. K., ... & Cheow, W. S. (2023). The implementation of chatbot-mediated immediacy for synchronous communication in an online chemistry course. *Education and Information Technologies*, 28(8), 10665-10690.
- Jeon, J., Lee, S., & Choi, S. (2024). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*, 32(8), 4613-4631.
- Jin, S. V., & Youn, S. (2023). Social presence and imagery processing as predictors of chatbot continuance intention in human-AI-interaction. *International Journal of Human-Computer Interaction*, 39(9), 1874-1886.
- Kandpal, N., Deng, H., Roberts, A., Wallace, E., & Raffel, C. (2023, July). Large language models struggle to learn long-tail knowledge. In *International Conference on Machine Learning* (pp. 15696-15707). PMLR.
- Karami, A., Spinel, M. Y., White, C. N., Ford, K., & Swan, S. (2021). A systematic literature review of sexual harassment studies with text mining. *Sustainability*, 13(12), 6589.
- Kharis, M., Schön, S., Hidayat, E., Ardiansyah, R., & Ebner, M. (2022). Mobile Gramabot: Development of a chatbot app for interactive German grammar learning. *International Journal of Emerging Technologies in Learning (iJET)*, 17(14), 52-63.
- Kiuru, N., Spinath, B., Clem, A. L., Eklund, K., Ahonen, T., & Hirvonen, R. (2020). The dynamics of motivation, emotion, and task performance in simulated achievement situations. *Learning and Individual Differences*, 80, 101873.
- Kohnke, L. (2023). L2 learners' perceptions of a chatbot as a potential independent language learning tool. *International Journal of Mobile Learning and Organisation*, 17(1-2), 214-226.
- Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1), 973-1018.
- Kull, A. J., Romero, M., & Monahan, L. (2021). How may I help you? Driving brand engagement

- through the warmth of an initial chatbot message. *Journal of business research*, 135, 840-850.
- Lewis, P., Wu, Y., Liu, L., Minervini, P., Küttler, H., Piktus, A., ... & Riedel, S. (2021). PAQ: 65 million probably-asked questions and what you can do with them. *Transactions of the Association for Computational Linguistics*, 9, 1098-1115.
- Li, C., Wang, J., Zhang, Y., Zhu, K., Hou, W., Lian, J., ... & Xie, X. (2023). Large language models understand and can be enhanced by emotional stimuli. *arXiv preprint arXiv:2307.11760*.
- Lin, M. P. C., & Chang, D. (2020). Enhancing post-secondary writers' writing skills with a chatbot. *Journal of Educational Technology & Society*, 23(1), 78-92.
- Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12, 157-173.
- Losenno, K. M., Muis, K. R., Munzar, B., Denton, C. A., & Perry, N. E. (2020). The dynamic roles of cognitive reappraisal and self-regulated learning during mathematics problem solving: A mixed methods investigation. *Contemporary Educational Psychology*, 61, 101869.
- Ma, X., Gong, Y., He, P., Zhao, H., & Duan, N. (2023). Query rewriting for retrieval-augmented large language models. *arXiv preprint arXiv:2305.14283*.
- Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., ... & Söllner, M. (2019). AI-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering*, 61, 535-544.
- Mozafari, N., Hammerschmidt, M., & Weiger, W. (2021, September). That's so embarrassing! When not to design for social presence in human-chatbot interactions. In *Proceedings of the International Conference on Information Systems*.
- Natale, S. (2019). If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA. *new media & society*, 21(3), 712-728.
- Nunnally, J. C. (1978). *Psychometric Theory* 2nd edition (New York: McGraw).
- Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033.
- Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Applied psychology*, 41(4), 359-376.
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, 36(1), 36-48.
- Pérez, J. Q., Daradoumis, T., & Puig, J. M. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Computer Applications in Engineering Education*, 28(6), 1549-1565.
- Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018, November). Chatbot as an intelligent personal assistant for mobile language learning. In *Proceedings of the 2018 2nd International Conference on Education and E-Learning* (pp. 16-21).
- Ramirez-Arellano, A., Acosta-Gonzaga, E., Bory-Reyes, J., & Hernández-Simón, L. M. (2018). Factors affecting student learning performance: A causal model in higher blended education.

Journal of Computer Assisted Learning, 34(6), 807-815.

- Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151, 102630.
- Sands, S., Campbell, C., Plangger, K., & Pitt, L. (2022). Buffer bots: The role of virtual service agents in mitigating negative effects when service fails. *Psychology & Marketing*, 39(11), 2039-2054.
- Scarantino, A. (2014). The motivational theory of emotions.
- Seeger, A. M., & Heinzl, A. (2021, June). Chatbots often fail! Can anthropomorphic design mitigate trust loss in conversational agents for customer service?. In ECIS.
- Socatiyanurak, V., Klangpornkun, N., Munthuli, A., Phienphanich, P., Kovudhikulrungsri, L., Saksakulkunakorn, N., ... & Tantibundhit, C. (2021). Law-u: Legal guidance through artificial intelligence chatbot for sexual violence victims and survivors. *IEEE Access*, 9, 131440-131461.
- Stević, I., Vukmirović, D., Vujović, V., & Marinković, V. (2023, May). Artificial Intelligence Chatbots and Conversational Agents—An Overview of Clinical Studies in Health Care. In *International KES Conference on Innovation in Medicine and Healthcare* (pp. 44-52). Singapore: Springer Nature Singapore.
- Tegos, S., Psathas, G., Tsiatsos, T., & Demetriadis, S. N. (2019). Designing Conversational Agent Interventions that Support Collaborative Chat Activities in MOOCs. In *EMOOCs-WIP* (pp. 66-71).
- Teja, R. (2023). Evaluating the ideal chunk size for a rag system using llamaindex. *LLAMAi*, [Online]. Available: <https://www.llamaindex.ai/blog/evaluating-the-ideal-chunk-size-for-a-rag-system-using-llamaindex-6207e5d3fec5>, 30, 31.
- Tudor Car, L., Dhinakaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y. L., & Atun, R. (2020). Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*, 22(8), e17158.
- Tulis, M., & Fulmer, S. M. (2013). Students' motivational and emotional experiences and their relationship to persistence during academic challenge in mathematics and reading. *Learning and Individual Differences*, 27, 35-46.
- Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: A meta-analysis. *Educational Psychology Review*, 28(1), 119-144.
- Vakayil, S., Juliet, D. S., & Vakayil, S. (2024, April). RAG-Based LLM Chatbot Using Llama-2. In *2024 7th International Conference on Devices, Circuits and Systems (ICDCS)* (pp. 1-5). IEEE.
- Van Pinxteren, M. M., Pluymaekers, M., & Lemmink, J. G. (2020). Human-like communication in conversational agents: a literature review and research agenda. *Journal of Service Management*, 31(2), 203-225.
- Han Vera, F. (2023). Integrating Artificial Intelligence (AI) in the EFL classroom: Benefits and challenges. *Transformar*, 4(2), 66-77
- Verleger, M., & Pembridge, J. (2018, October). A pilot study integrating an AI-driven chatbot in an

- introductory programming course. In 2018 IEEE frontiers in education conference (FIE) (pp. 1-4). IEEE.
- Wang, Z. (2022). Modern question answering datasets and benchmarks: A survey. arXiv preprint arXiv:2206.15030.
- Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information–Inoculation messages against conversational agents’ response failures. *Electronic markets*, 32(1), 239-258.
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. *Journal of Service Management*, 29(5), 907-931.
- Yang, J. (2022). Perceptions of preservice teachers on AI chatbots in English education. *International Journal of Internet, Broadcasting and Communication*, 14(1), 44-52.
- Yin, J., Goh, T. T., & Hu, Y. (2024). Interactions with educational chatbots: the impact of induced emotions and students’ learning motivation. *International Journal of Educational Technology in Higher Education*, 21(1), 47.
- Yin, J., Goh, T. T., Yang, B., & Xiaobin, Y. (2021). Conversation technology with micro-learning: The impact of chatbot-based learning on students’ learning motivation and performance. *Journal of Educational Computing Research*, 59(1), 154-177.
- Yuan, X., Wang, T., Wang, Y. H., Fine, E., Abdelghani, R., Lucas, P., ... & Oudeyer, P. Y. (2022). Selecting better samples from pre-trained LLMs: A case study on question generation. arXiv preprint arXiv:2209.11000.
- Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., ... & Shi, S. (2023). Siren's song in the AI ocean: a survey on hallucination in large language models. arXiv preprint arXiv:2309.01219.
- Zhuang, S., Liu, B., Koopman, B., & Zuccon, G. (2023). Open-source large language models are strong zero-shot query likelihood models for document ranking. arXiv preprint arXiv:2310.13243.
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33, 9459-9474.