

# 智慧圖書館：基於 GPT-4 的智慧館員

## Smart Library: Intelligent Librarian Powered by GPT-4

謝翰誼

Han-Yi Hsieh

東海大學軟體工程與技術中心

Software Engineering & Technology Center, Tunghai University

葛芝辰

Chih-Chen Ko

東海大學數位創新碩士學位學程

Digital Master Program of Digital Innovation, Tunghai University

余霖鍇

Lin-Kai Yu

東海大學資訊工程學系

Computer Science Department, Tunghai University

陳廷恩

Ting-En Chen

東海大學資訊工程學系

Computer Science Department, Tunghai University

周忠信 (通訊作者)

Jung-Sing Jwo

Email: jwo@thu.edu.tw

東海大學資訊工程學系、數位創新碩士學位學程、軟體工程與技術

中心教授

Professor, Computer Science Department, Tunghai University

李央晴

Yang-Ching Li

國立公共資訊圖書館科員

Officer, National Library of Public Information

張卉君

Hui-Chun Chang

國立公共資訊圖書館科員

Officer, National Library of Public Information

馬湘萍

Hsiang-Ping Ma

國立公共資訊圖書館館長

Director General, National Library of Public Information

### 【摘要 Abstract】

圖書館隨著時代及科技的進步，其服務與功能持續不斷進化。圖書館的參考服務隨著科技的進步有了巨大的轉變。紙本印刷形式的參考資源逐漸轉變為數位形式，而參考諮詢服務經歷了傳統的面對面諮詢、電子郵件、網路聊天室即時諮詢，隨著人工智慧技術的快速崛起，圖書館進入智慧化服務的新階段，以文字或語音互動的聊天機器人開啟了智慧參考服務的先河。而 AI 技術在圖書館中的應用，不僅提升了資料檢索和資訊推薦的精準度，還開創了新的服務模式，如個性化閱讀推薦、智慧分類和知識圖譜構建等。近年來，大語言模型逐漸成熟，ChatGPT-4 的出現為圖書館帶來革命性影響。大語言模型具備強大的自然語言理解和生成能力，可大幅提升服務效率，滿足現代讀者多樣化和個性化的需求。本研究的目標是探討基於 GPT-4 的智慧館員「曉書」，強調圖書館服務能力和讀者體驗。曉書是具備語音溝通能力的數位雙生，能模仿真實館員的行為和語言風格，提供自然和人性化的互動。智慧館員「曉書」自 2023 年 12 月 28 日於國立公共資訊圖書館啟用以來，「曉書」運用大語言模型提供高效準確的資訊回應，並通過豐富的人設設計增強與讀者的互動體驗。未來，本研究將繼續完善智慧館員「曉書」的功能和服務，重點包括資料庫擴充與更新、

情感識別與回應、個性化推薦、多語言支持、反應速度提升及使用者回饋機制。透過不斷的技術創新和服務優化，智慧館員「曉書」將能夠為讀者提供更有效率、更智慧、更貼近讀者需求的圖書館服務，成為智慧圖書館不可或缺的重要一環。

With the progression of technology, library services and functions have continuously evolved. The reference services in libraries have undergone significant transformations alongside technological advancements. Traditional paper-based reference resource are gradually shifting to digital formats. Reference consultation services have evolved from face-to-face inquiries and email consultations to real-time online chat services. With the rapid rise of artificial intelligence technology, libraries have entered a new stage of intelligent services, pioneering smart reference services through text or voice-interactive chatbots. The application of AI technology in libraries has not only enhanced the accuracy of data retrieval and information recommendation but also initiated new service models, such as personalized reading recommendations, intelligent classification, and knowledge graph construction. In recent years, large language models have gradually matured, and the emergence of ChatGPT-4 has brought revolutionary impacts to libraries. Large language models possess powerful natural language understanding and generation capabilities, significantly improving service efficiency and meeting the diverse and personalized needs of modern readers. This study aims to explore the intelligent librarian "Xiaoshu" based on GPT-4, emphasizing the enhancement of library service capabilities and reader experience. Xiaoshu, a digital twin with voice communication abilities, can mimic the behavior and language style of real librarians, providing natural and human-like interactions. Since its implementation on December 28, 2023, at the National Public Information Library, Xiaoshu has utilized large language models to offer efficient and accurate information responses, enhancing

reader interaction through rich character design. Looking ahead, this study will continue to refine the functionalities and services of the intelligent librarian. Key areas of focus include database expansion and updates, emotion recognition and response, personalized recommendations, multilingual support, response time improvement, and user feedback mechanisms. Through continuous technological innovation and service optimization, the intelligent librarian aims to provide readers with more efficient, intelligent, and reader-centric library services, becoming an indispensable part of the smart library ecosystem.

### 【關鍵詞 Keywords】

智慧圖書館、大語言模型、生成式人工智慧、智慧館員、參考諮詢服務  
Smart library, Large language model, Generative AI, Intelligent librarian  
Reference services

## 壹、背景與動機

阮甘納桑 (S.R. Ranganathan) 的圖書館五律，是圖書館學的基本原則，概括了圖書館的核心職能和目標 (Ranganathan, 1931)。五律中，「圖書館是一個不斷成長的有機體」，充分說明了圖書館隨著時代的轉變，其服務與功能持續不斷進化的特質。早期的圖書館，如古代亞歷山大圖書館，主要集中在收藏和保護手稿和書籍，以供學者和研究者查閱。隨著印刷術的發明，圖書館逐漸向公眾開放，成為普及知識的重要場所。在這個過程中，Ranganathan 的五律強調了讀者和圖書之間的互動，促使圖書館在服務品質和效率上不斷進化。

自 20 世紀下半葉開始，圖書館先後經歷了資訊化、網際網路以及行動智慧手機普及化等三個關鍵進化階段。首先，資訊化階段始於 20 世紀下半葉，當時電腦技術開始應用於圖書館管理中。圖書館首先實現了館藏書目數位化，建立了線上公用目錄 (OPAC)，取代傳統的卡片目錄 (林呈潢, 1995, 頁 2220-2221)。這一變革使得讀者能夠更

快捷檢索所需書籍和資料，提高了館藏資源的利用率。此外，圖書館也開始引入電子資料庫和電子期刊，為讀者提供更豐富的數位資源。資訊化的進程，大大提高了圖書館的管理效率和服務品質。

進入網際網路階段，圖書館服務進一步擴展，實現了線上資源共享與遠端服務。網際網路的普及使得圖書館能夠將數位化的館藏資源上傳到網路上，讀者可以通過圖書館的網站或數位圖書館平台，隨時隨地取得電子書、期刊、資料庫等資源。這一階段的標誌性發展，包括數位圖書館的建立、線上公用查詢目錄（Web OPAC）的推廣，以及各類數位資源的互聯互通等（Rao, 2001, p25-26）。讀者的查詢、借閱和資料下載變得更加便捷，圖書館服務的覆蓋範圍和影響力更是顯著提升。

到了行動網路成熟以及智慧手機的普及，圖書館服務進入了移動互聯時代（梁鴻栩，2009，頁 77-78；柯皓仁，2012，頁 188）。行動智慧手機為圖書館提供了全新的服務平台。通過開發行動應用 app，圖書館能夠向讀者提供更加靈活和個性化服務。讀者可以使用手機隨時隨地進行圖書查詢、預約、續借等操作，並接收到圖書館的最新通知和活動資訊。圖書館更是應用智慧手機，整合掃碼借還書、電子書閱讀、線上諮詢問答、個性化推薦等功能，優化讀者使用體驗。

從圖書館的服務來看，參考服務是讀者服務重要的一環，而諮詢服務，則是影響參考服務成效的最重要工作之一。圖書館的參考服務在於：1)回答讀者問題；2)幫助讀者尋找資料(張淳淳，2003，頁 288)。

參考服務最早起源於 1876 年。格林（Samuel Green）於 1876 年在美国麻薩諸塞州伍斯特公共圖書館(Worcester Public Library)建立第一個正式的參考服務。他強調對於需要資訊的人提供協助的重要性，在此之前，圖書館員通常將自己視為檔案的保管人，而不是主動提供資訊的人（Katz, 1982, p.4）。

參考服務發展的初期，館員以圖書館卡片目錄及紙本參考工具書，例如紙本書目、索引或百科全書、字詞典等參考資源提供參考諮詢服

務，諮詢的方式也侷限於面對面或書信問答。在這半世紀以來，隨著科技的進步，參考資源與參考諮詢服務也發生了巨大的轉變。

在參考資源方面，隨著資訊科技的演進，紙本印刷形式的參考資源，例如卡片目錄、紙本參考工具書，也逐漸轉變為數位形式，例如線上公用目錄、線上資料庫、光碟資料庫及網路資源等；在參考諮詢服務方面，除了傳統的面對面、書信諮詢，隨著數位時代的來臨也逐漸走向數位參考諮詢服務（張淳淳，2003，頁 13、25、295-296；莊健國，2004，頁 46-47）。

數位參考諮詢服務最早期的形式可從傳真和電話參考服務談起，繼而利用電子郵件和網頁表單傳遞問題和解答（蘇小鳳，2005，頁 13），隨後亦發展出網路聊天參考服務及合作參考服務，館員以線上聊天室即時回答讀者問題，並透過各個合作圖書館參考館員的合作輪班制度，讓參考服務延伸為 24 小時服務（莊健國，2004，頁 76-77）。

近年來，人工智慧（Artificial Intelligence, AI）的興起，再次改變了參考服務的形式。圖書館紛紛投入聊天機器人的開發，例如大學圖書館嘗試建置以文字互動為主的參考諮詢聊天機器人（陳宜琳，2019）。國立公共資訊圖書館（國資圖）也於 2019 年導入智慧語音服務，民眾可透過口語化介面，查詢國資圖各項服務資訊與館藏，並且以實體智慧服務（圖書館內設置資訊便利站）及線上智慧服務（官方網站客服）兩種形式並行提供諮詢服務（陳志銘，2022，頁 235-237）。

AI 技術在圖書館中的應用，不僅能夠提升資料檢索和資訊推薦的精準度，同時也開創了新的服務模式。例如，AI 可以通過大數據分析和機器學習技術等，分析讀者的借閱行為和需求，進而提供個性化的閱讀推薦等（Cho & Han, 2019, April；Sivasankari et al., 2024）。此外，AI 技術在圖書館資源管理中的應用，如智慧分類、標籤生成和知識圖譜構建等，都可大大提高圖書館的營運效率和服務品質。AI 技術的應用，還能有助於圖書館進行館藏數位化和文獻保存，例如透過影像識別和自然語言處理等技術，實現歷史文獻的數位化保存和自動化

管理。

近年來 AI 領域中的大語言模型(Large Language Models, LLMs) 逐漸成熟，而基於 GPT(Generative Pre-trained Transformer) 的 ChatGPT 聊天生成式預訓練轉換器的橫空出世，更為圖書館帶來了革命性的影響 (Zhao et al., 2023; Yang et al., 2024)。LLMs 具備強大的自然語言理解和生成能力，可以處理大量的文本數據，提供準確的問答服務，並生成高品質的文本內容。而這樣的能力，是過去圖書館的服務機器人所無法企及。因此，運用 LLMs 發展智慧圖書館館員，是圖書館智慧化的必然趨勢。透過 LLMs 發展的智慧館員除精通各國語言外，上知天文下知地理且熟讀百科全書，同時又能通曉圖書館所有服務項目。因此在回應眾多讀者需求上，除可大幅提升服務效率外，更能滿足現代讀者多樣化和個性化的需求，真正實現 Ranganathan 所提出的圖書館理想。

本研究的主要目標是開發一個基於 GPT-4 的智慧圖書館館員，而智慧館員成功的關鍵因素，除了服務能力外，使用者體驗 (User Experience, UX) 也至關重要 (Lew, & Schumacher, 2020; Zheng, et al., 2022, April)。圖書館作為知識和資訊的提供者，其主要目標是滿足讀者的需求，提供高品質的服務。因此虛擬圖書館員在能力上，必須能夠運用書籍資料庫，提供作者與書籍的相關介紹與推薦；根據館務服務規定（如辦理借書證等），回覆讀者詢問；能夠與圖書館內現有系統，如書籍管理系統等，實現整合並即時回覆書籍所在位置。

從讀者體驗設計角度來看，智慧圖書館員必須具備自然和人性化的互動設計，以有助於減少讀者對機器的陌生感和排斥感。因此當智慧圖書館館員的行為和語言風格與真實館員相似時，讀者才會更容易接受和信任此類 AI 服務。本研究的虛擬館員，實為一真實館員的數位雙生 (digital twin) (Barricelli, Casiraghi, & Fogli, 2019; Iakovides et al., 2022, October)。數位雙生是指用數位技術創建的一個虛擬實體，精確模擬其物理對象的特徵和行為。此數位雙生具備語音溝通能力，

亦即能聽得懂讀者、同時也能以語音回覆，而更重要的是在對話過程中，具備一如常人的說話型態。透過專注於 AI 的讀者體驗設計，圖書館方能夠實現從傳統服務模式朝向智慧化服務模式之過程。這不僅是技術上的創新，更是圖書館在現代社會中發揮核心作用的重要保障。

綜合上述，本研究旨在開發一個運用大語言模型 GPT-4 的虛擬館員。此虛擬館員，是一個支持語音互動的虛擬助理之智慧系統。除了能夠模仿真實圖書館員的行為，如回答查詢、提供資訊外，尚可透過對話進行書籍推薦。虛擬館員同時也是知識百科，能夠回答讀者對知識的詢問。透過對虛擬館員的研究與發展，預期為圖書館達到以下三個效益：

- 一、提升讀者體驗：智慧館員可滿足當代讀者迅速、方便且智慧化的服務需求，從而優化圖書館的服務品質。讀者更可從中獲得 AI 時代的體驗，並提升圖書館新魅力。
- 二、提高工作效益：面對未來人力資源短缺趨勢，智慧館員可以協助真實館員提供服務。除了解決人力不足問題外，館員也才有更多時間致力於提供更高品質的服務項目與進行策展等工作。
- 三、創建未來圖書館形象：透過 AI 等數位科技的融入，新時代的圖書館五律勢必重塑。同時未來圖書館在與讀者的互動上，也將與過去全然不同。此外，本研究成果可供其他公共圖書館借鑒和應用。

## 貳、大語言模型文獻回顧

大語言模型（Large Language Models, LLMs）的發展是人工智慧領域的一個重要進程，它們在自然語言處理（Natural Language Processing, NLP）中展現了強大的能力。LLMs 的發展可以追溯到早期的統計語言模型，並在近年來隨著深度學習（deep learning）技術的進步而取得了顯著的突破。以下是 LLMs 發展的幾個關鍵階段。

早期在 LLMs 出現之前的語言模型研究，主要基於統計方法。N-gram

模型是早期最具代表性的模型之一，通過統計固定長度的詞序列的出現概率來進行文本生成和預測 (Cavnar, & Trenkle, 1994, April)。此類模型雖然簡單，但存在諸多限制，如無法有效捕捉長距依賴關係，且隨著 N 值增大，計算量和數據需求也會劇增。

2000 年初，隨著神經網路技術的發展，研究者開始將其應用於語言建模。Bengio 等人提出了基於神經網路的語言模型，這一模型能夠通過學習詞嵌入 (word embeddings) 來捕捉語言中的語義資訊，並通過多層感知機進行預測，這一模型開創了神經語言建模的先河 (Bengio, Ducharme, & Vincent, 2000)。而隨後的進展，如循環神經網路 (Recurrent neural network, RNN) 和長短期記憶網路 (Long Short-Term Memory, LSTM) 在處理序列數據方面展現了強大的能力。RNN 能夠處理變長序列的數據 (Tomáš Mikolov et al., 2010)，LSTM 則通過設計特定的記憶單元來緩解 RNN 在長序列中存在的梯度消失問題 (Sundermeyer, Schlüter & Ney, 2012)。

2017 年 Vaswani 等人提出的 Transformer 模型，引入注意力機制，徹底改變了自然語言處理的格局 (Vaswani, 2017)。Transformer 模型能夠並行處理序列數據，極大化提高了訓練效率和模型性能。其核心在於自注意力機制，能夠動態地關注序列中不同位置的單詞，捕捉長距離依賴。Transformer 模型，能夠高效率地處理長序列數據，成為現代 LLMs 的基礎。隨著計算能力和數據量的增加，研究者開始訓練大規模預訓練模型，這些模型在大量文本數據上進行預訓練 (pre-trained)，然後在特定任務上進行微調。OpenAI 的 GPT (Generative Pre-trained Transformer) 系列模型和 Google 的 BERT (Bidirectional Encoder Representations from Transformers) 模型是這一領域的代表 (Floridi, & Chiriatti, 2020; Ravichandiran, 2021)。在 LLMs 的發展中，GPT 是目前最大且最具影響力的語言模型之一。其展示了前所未有的生成和理解自然語言的能力，成功證明了模型規模與性能之間的密切關係，並引發了廣泛的應用和研究。

雖然大語言模型在自然語言處理領域取得了顯著的進展，但它們仍然面臨許多挑戰和問題（Kaddour et al., 2023）：

### 一、幻想與不準確的內容生成

LLMs 有時會生成看似合理但實際上不正確的內容，這種現象被稱為“幻覺”（hallucination）。這是因為模型基於訓練數據進行語言生成，但無法理解或驗證事實，可能會編造不存在的細節或錯誤的內容。

### 二、偏見與歧視

LLMs 的訓練數據來自於網際網路的大量文本，這些數據中可能包含社會偏見和歧視。模型會無意識地學習並再現這些偏見，導致生成的文本中出現種族、性別、年齡等方面的偏見。

### 三、缺乏上下文理解

儘管 LLMs 能夠生成連貫的內容，但它們對於更大背景和上下文的理解能力有限。模型往往只基於當前的上下文視窗進行推理，無法長期保持對話的連續性和一致性。

### 四、效率與成本

訓練和運行 LLMs 需要大量的算力資源，這不僅昂貴而且對環境有較大影響。模型的計算效率和能源消耗是目前的一大挑戰。

### 五、安全與濫用風險

LLMs 可以被濫用來生成虛假內容、垃圾內容或進行社交工程攻擊。這些模型的強大生成能力如果被不法分子利用，可能會造成嚴重的安全問題。

## 參、智慧館員架構設計

在構建基於 GPT-4 技術的智慧圖書館館員時，除了需要應對前一節所述的有關大型語言模型可能出現的問題，如幻覺等，還需在設計上充分考慮讀者體驗的人機互動設計。圖 1 展示了虛擬智慧館員的

3D 設計示意圖。智慧館員的實體設計借用了最早出現在倫敦街頭的紅色公共電話亭這一空間意象，象徵人類可以在此空間內與智慧館員進行直接對話。智慧館員名為“曉書”，寓意「通曉圖書，智慧諮詢」，其外觀如圖 3.1 中的人像所示，是館內某位真實館員的 AI 數位雙生。該人像亦由 AI 生成，但以漫畫風格展現，旨在避免“恐怖谷(Uncanny Valley)理論”的影響，同時保持對讀者的吸引力 (Wang, Lilienfeld, & Rochat, 2015)。此外，電話亭內還設有桌面，桌面上配備一塊液晶面板，用於與智慧館員互動時顯示對話內容，或在虛擬智慧館員需要回覆大量資訊、展示網頁或 QR Code 等情況下使用。



圖 1 智慧館員實體設計示意圖

圖 2 則是本研究的智慧館員系統架構設計圖，其左半部為智慧館員的前端系統、右半部為後台服務，分別說明如下：

### 一、前端系統

該系統由五個主要單元組成，旨在支持系統與讀者之間的人機互動介面。其中“曉書播放”單元負責在系統與讀者對話時，即時以動畫方式呈現智慧館員的面部表情。語音輸入與語音輸出兩個單元則分別負責系統與讀者對話時的聲音收錄與播放。在對話過程中，收錄的

聲音經由前端系統呼叫後台的語音轉文字 STT(Speech to Text)服務，將類比聲音識別成文字，供後端計算使用；同時，透過文字轉語音 TTS (Text to Speech) 技術，將文字轉化成類比聲音，並模仿真實館員的聲音，通過前端的語音輸出喇叭播出。這三個單元的主要目的是創造出讀者在與虛擬館員互動時，具備類似與真實館員互動的讀者體驗。

此外，“桌面呈現”單元用來展示前端系統自後台系統接收到的所有內容，包括雙方對話、推薦書籍資訊、書籍館藏位置的 QR Code、熱門書籍資訊、新書資訊網頁、館務服務說明、交通資訊網頁及附近位置的 Google 地圖等。此單元的顯示面板位於圖 1 中的桌面上。此設計旨在幫助智慧館員曉書在對話時維持在 20 秒內，其餘內容則在桌面上呈現，避免對話過於冗長而導致讀者不耐煩。

最後，前端系統中還包含一個“性別年齡層識別”單元。此單元隱藏於電話亭前端的鏡頭內，通過 AI 人物識別模型，記錄目前與虛擬館員互動的讀者性別與年齡層。這些紀錄資料不包含個人隱私，例如人臉識別，僅包含性別與年齡層資訊，其目的是為了提供改善未來智慧館員的參考依據。

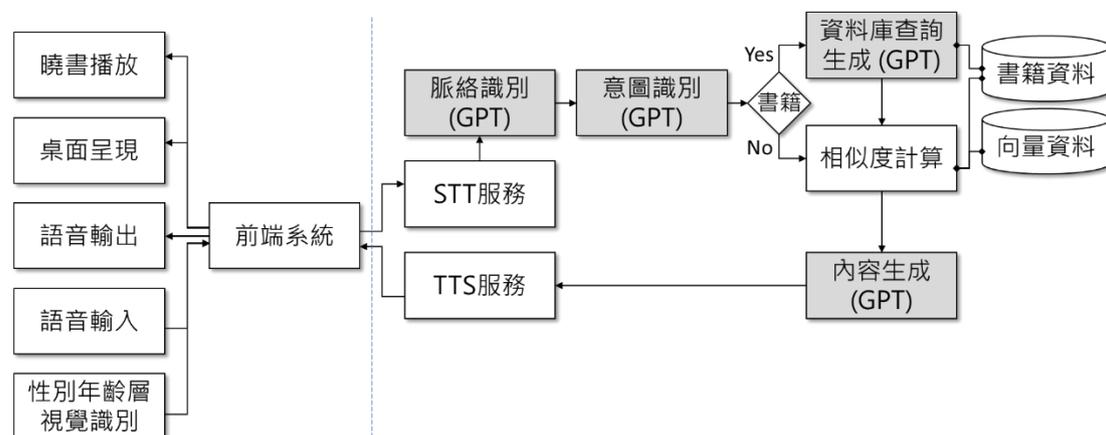


圖 2 智慧館員系統架構設計圖

## 二、後台服務

作為支持前端系統的服務請求，首要任務是提供語音與文字的轉換。本系統採用微軟公司 Azure 上的 STT (Speech to Text) 與 TTS

(Text to Speech) 服務。其優點在於可以用真人聲音進行訓練，TTS 服務再根據訓練後的模型生成類似該人物的聲音。此外，該系統能自動辨識讀者使用中文或英文進行對話，並據此生成相應的回應內容。

當讀者的對話被轉換成文字後，智慧館員首先必須判斷此對話是否與前一對話內容相關，即是否存在對話脈絡關係。雖然 ChatGPT 在脈絡對話上表現不錯，但對於一個提供公共服務的虛擬館員而言，面對不同的讀者或同一讀者但詢問主題隨時變換的情況下，識別新脈絡的起點及結束顯然不同於 ChatGPT 這種個人對話服務。因此虛擬智慧館員必須具備自主識別此次對話是否存在脈絡關係，或是需要重新建立新的脈絡。

在創建初期，本研究以國資圖原有的智慧語音問答集和館員修訂之各項館務問答集，以及預先導入的 16 萬筆書目資料，整合進入智慧館員曉書生成回應來源的資料庫。為避免智慧館員在不同問題上答非所問或產生幻覺，本系統加入了意圖識別服務。此服務主要目的是在面對某些問題的回應涉及正確與否時，例如書籍相關資訊、人物、館務等問題；另外，某些問題雖非圖書館業務，但需要回應正確資訊，例如氣象、交通、圖書館周邊等問題。面對這些類型的意圖，智慧館員只能使用系統內部的資料庫或預先指定的外部網站來生成回應，以確保其正確性。

由於圖書館藏書眾多，讀者有時會尋求書籍推薦、有時詢問某本書的內容、有時詢問作者的其他著作、有時想知道書籍位置等。因此，智慧館員需能夠自行進入書目資料庫或與圖書館現有系統整合，而這正是後台“資料庫查詢生成”單元的主要任務。

為了快速從讀者的問題中比對出資料庫中最相關的內容，本系統採用了“檢索增強生成”(Retrieval-Augmented Generation, RAG) 方法 (Lewis et al., 2020)。RAG 是一種結合檢索方法和生成模型的自然語言處理技術，其結構包括檢索器和生成器，透過整合這兩個組件來生成更具連貫性和上下文相關的內容。檢索器的基本原理是透過將讀

者提出的問題與資料庫中的內容進行比對，從而找到最相關的內容。這一過程通常包括兩個主要步驟：首先，將使用者的查詢與內容轉換為特定的檢索向量形式，並根據兩者之間的相似度計算選擇最相關的內容返回。最後，將所有取得的內容與讀者問題一併提供給 GPT，由 GPT 生成最終的回應內容。

#### 肆、實際應用與成果評估

系統完成開發後，為確保智慧館員的服務水平達標，圖書館設計了一套涵蓋書籍、館務、服務及智慧館員「曉書」的人設等共計 168 道測試題目。經測試，智慧館員在此測試集中達到了 83.97% 的正確率。錯誤的主要原因集中在書籍名稱、推薦或作者相關資訊方面。進一步分析表明，部分錯誤源於書籍資訊不足，另一些則是由於 RAG 方法的局限性所致。在對上述問題進行進一步修正後，系統的正確率已超過預設的目標（80%）。於是，智慧館員「曉書」於 2023 年 12 月 28 日在國資圖正式啟用。圖 3 展示了曉書在國資圖的實際建置圖。



圖 3 國立公共資訊圖書智慧館員—曉書

圖 4 展示了智慧館員「曉書」自 2023 年 12 月 28 日開始服務至 2024 年 5 月 12 日期間的每日訪問量。其中，藍色表示工作日，橘色表示假日，空白處則為休館日。整體而言，假日的使用量顯著高於工作日，且訪問量大多在 200 次以上。此外，除了極少數日期外，工作日的訪問量也多數達到 50 次以上。

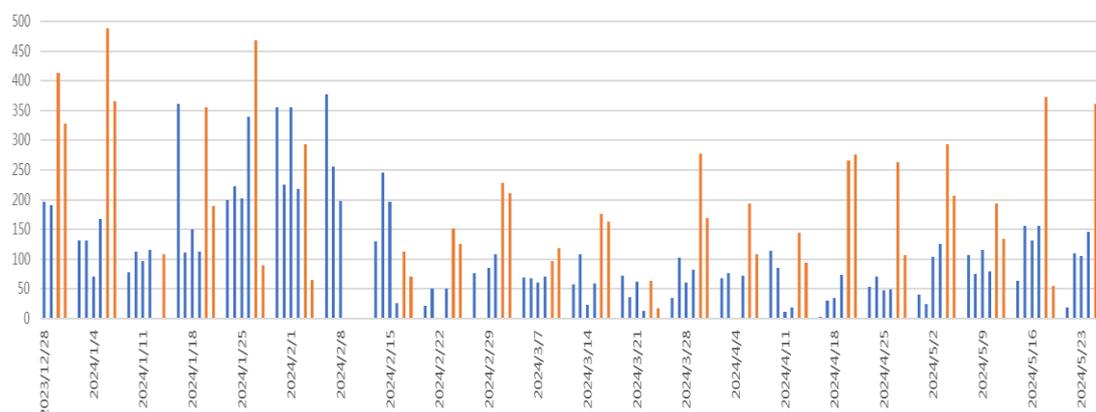


圖 4 虛擬館員「曉書」每日訪問量

表 1 展示了智慧館員「曉書」自開始服務後，所接受到各類問題的歷史紀錄。最左側的灰色表格顯示的是 2024 年 1 月至 2 月的統計數據，中間的藍色表格是 2024 年 3 月至 4 月的統計數據，最右側的綠色表格則顯示了 2024 年 5 月的統計數據，也就是本文撰寫前一個月的數據。從表中可以看出，讀者與智慧館員「曉書」對話的主題大致可分為八類：館務、書籍、即時資訊、人設、人物、政治、倫理和閒聊。其中，館務與書籍問題是圖書館的主要服務業務；即時資訊包括館外資訊如氣象、交通，以及新書和熱門書籍等查詢服務；人設問題指對智慧館員「曉書」的身分、年齡、嗜好等方面的好奇詢問；人物問題則涉及政府或重要單位的現職者資訊；政治問題涉及黨派、選舉等議題；倫理問題包括隱私、安全、誹謗、辱罵、虛假訊息或違反法律等相關議題；最後，閒聊則涵蓋了其他七種主題之外的各類提問。

	佔比		佔比		佔比
閒聊	54.12%	閒聊	34.20%	閒聊	26.26%
館務詢問	15.64%	館務詢問	20.56%	館務詢問	19.15%
人設詢問	12.38%	人設詢問	16.95%	人設詢問	27.92%
書籍詢問	11.23%	書籍詢問	17.82%	書籍詢問	18.27%
即時資訊	3.34%	即時資訊	6.04%	即時資訊	4.32%
倫理	1.59%	倫理	2.69%	倫理	2.42%
人物詢問	1.50%	人物詢問	1.00%	人物詢問	1.56%
政治	0.20%	政治	0.74%	政治	0.09%
Total	100.00%	Total	100.00%	Total	100.00%
2024/01-02		2024/03-04		2024/05	

表 1 智慧館員「曉書」問題歷史紀錄分類與占比

首先，從表 1 中可以發現，閒聊類問題在建置後初期兩個月內是最多的提問類別，甚至超過了書籍與館務問題的總和。同時，關於「曉書」人設的提問也高於書籍詢問。剩下的其他四類問題總和僅占比 6.63%。進一步分析紀錄發現，閒聊類提問內容豐富多樣，這證明了基於大語言模型 GPT 發展的智慧館員「曉書」在使用者體驗上對讀者確實更具吸引力。其原因歸納如下：

- 一、自然對話能力：GPT 模型具有強大的自然語言處理能力，能夠理解和生成流暢且具上下文關係的對話，使讀者感受到更自然和人性化的互動。
- 二、多樣化回應：由於 GPT 模型能處理多種話題，讀者可以就任何感興趣的話題進行提問，從而增加了互動的樂趣和參與度。這在初期大量閒聊類提問中得到了體現。
- 三、即時反應：智慧館員「曉書」能夠即時回應讀者的問題，提供快速而準確的資訊，提升了服務效率，滿足了現代讀者對即時性和便利性的需求。
- 四、廣泛知識庫：基於 GPT 模型，智慧館員「曉書」能夠訪問並利用龐大的知識庫，提供多領域、多層次的資訊內容，滿足讀者的多樣化需求。
- 五、個性化互動：對智慧館員「曉書」人設的高頻提問表明，讀者對虛擬館員本身產生了興趣，這反映了虛擬館員在與讀者建立情感聯繫方面的潛力。這種個性化的互動能增加讀者的使用黏著度和

忠誠度。

有關「曉書」人設的設計，在初期發展階段並未投入太多關注。然而，從表 1 中可以發現，有關「曉書」人設的提問在這五個月內明顯增長。因此，本研究進一步豐富了「曉書」的人設內容，包括她喜歡的飲食、書籍等，以改善讀者與智慧館員互動時的使用者體驗。此舉的目的在於增強智慧館員的親和力和人性化特質，從而提高讀者的參與度和滿意度。透過提供更詳細和個性化的智慧館員「曉書」背景設定，讀者能夠對智慧館員產生更深的情感聯繫，並激發更多的互動興趣。這些改進措施不僅豐富了虛擬智慧館員的角色形象，還提升了整體服務品質，進一步確保了智慧館員在實際應用中的有效性和可持續性。

另外，從表 1 中還可以觀察到，館務與書籍的提問總和在後三個月內逐步超過了初期占比超過 50% 的閒聊類問題。這可能顯示出讀者在服務初期主要因為好奇而與智慧館員「曉書」互動。然而，隨著使用情境的回歸正常，館務與書籍相關的詢問才成為主要需求。這一現象表明，智慧館員「曉書」在初期吸引了大量讀者的注意，並透過多樣化的互動內容成功激發了讀者的興趣。隨著時間推移，讀者逐漸將智慧館員「曉書」視為一個可靠的資源，主要用來解決實際的館務與書籍相關問題。這也說明了智慧館員能夠有效地適應並滿足讀者的核心需求，進一步提升了其在圖書館服務中的價值與地位。

## 伍、結論與未來展望

本研究的主要目標是開發一個基於 GPT-4 的智慧館員「曉書」，而智慧館員成功的關鍵因素，除了回答內容的正確性外，讀者體驗也至關重要。從讀者體驗設計角度來看，本研究的智慧館員在外型與聲音上模仿真實館員，做為真實館員的數位雙生，此數位雙生具備語音溝通能力，亦即能聽得懂讀者的問題、同時也能以語音回覆，更重要的是在對話過程中具備一如常人的說話型態，提供更人性化的使用者體驗。透過專注於生成式 AI 的問答正確率和讀者體驗設計，作為臺

灣第一個採用生成式人工智慧的智慧館員「曉書」，在圖書館智慧化諮詢服務上產生了劃時代的意義。

智慧館員「曉書」自 2023 年 12 月 28 日正式於國資圖啟用以來，初期大量的閒聊類提問顯示出讀者對新技術的高度興趣。隨著時間推移，館務與書籍相關的提問逐漸成為主要需求，這表明讀者逐漸意識到並且願意理解智慧館員在實際圖書館服務中的價值。「曉書」在設計上結合了自然語言處理技術和 RAG 方法，不僅提供了高效準確的資訊回應，還通過豐富的人設設計增強了與讀者的互動體驗。透過分析讀者的互動數據，本研究不斷調整和改進系統，以提升服務的精確性和使用者滿意度。

本研究成果在創建未來圖書館方面，勢必重塑新時代對圖書館諮詢服務的認識。同時，未來圖書館在與讀者的互動上，也將與過去全然不同。此外，本研究成果可供公共圖書館借鑒和應用。

對於智慧館員「曉書」的功能和服務，建議未來優化及改善重點如下：

- 一、資料庫擴充與更新：持續擴充和更新圖書館的書目和館務資料庫，確保智慧館員「曉書」能夠提供最新、最全的書籍和館務資訊。
- 二、情感識別與回應：研究並引入情感識別技術，使智慧館員「曉書」能夠更好地理解讀者的情感狀態，提供更具人性化的回應。
- 三、個性化推薦：利用機器學習技術，根據讀者的歷史互動和偏好，提供更加個性化的書籍推薦和資訊服務，提升讀者的使用體驗。
- 四、多語言支持：進一步提升智慧館員「曉書」的多語言能力，滿足不同語言背景讀者的需求，促進國際化服務的拓展。
- 五、反應速度提升：受限於 GPT，目前對讀者提問的反應時間有時過長，建立私有專屬的 LLMs 是未來方向之一。
- 六、使用者回饋機制：建立健全的使用者回饋機制，及時收集並分析讀者的意見和建議，作為系統改進的重要依據。

鑒於未來人工智慧和機器學習技術將繼續提升和發展，圖書館能

更準確地預測讀者需求、提供個性化的服務和資源推薦。未來運用 GPT-4 模型結合其它人工智慧模型等新技術，可為圖書館提供包括個別化服務、資訊尋找、閱讀體驗增強、數據分析和服務改進，以及雲端和數位化服務。透過不斷的技術創新和服務優化，採用 GPT-4 的智慧館員「曉書」將能夠在未來為讀者提供具效率、更智慧、更貼近讀者需求的圖書館服務，成為未來圖書館不可或缺的重要一環。

### 【參考書目】

- 林呈潢 (1995)。線上公用目錄。在圖書館學與資訊科學大辭典 (下冊)。臺北市：漢美。
- 柯皓仁 (2012)。圖書館資訊科技與應用。在國家圖書館編，中華民國一〇一年圖書館年鑑。臺北市；編者。
- 張淳淳 (2003)。參考資源與服務。新北市：國立空中大學。
- 梁鴻翔 (2009)。淺談 M-Libraries 行動通訊圖書館服務。中華民國圖書館學會會訊，17(2)，77-81。
- 莊健國 (2004)。圖書館數位合作參考服務的理論與實務。臺北市：文華圖書館管理資訊。
- 陳志銘 (2022)。圖書館資訊科技與應用。在國家圖書館編，中華民國一一〇年圖書館年鑑。臺北市；編者。
- 陳宜琳 (2019)。國立臺灣師範大學圖書館參考諮詢機器人建置與評估。未出版之碩士論文，國立臺灣師範大學圖書資訊學研究所，臺北市。
- 蘇小鳳 (2005)。即時數位參考諮詢服務。臺北市：文華圖書館管理。
- Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). *A survey on digital twin: Definitions, characteristics, applications, and design implications. IEEE access, 7, 167653-167671.* New York : IEEE.

- Bengio, Y., Ducharme, R., Vincent, P. & Jauvin, C. (2000). A neural probabilistic language model. In T. Leen, T. Dietterich & V. Tresp (Ed.), *Advances in neural information processing systems*, 13.
- Cavnar, W. B., & Trenkle, J. M. (1994, April). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval* (Vol. 161175, p. 14). Ann Arbor, Michigan: Environmental Research Institute of Michigan.
- Cho, E., & Han, M. (2019, April). AI powered book recommendation system. In *Proceedings of the 2019 ACM Southeast Conference* (pp. 230-232) New York: Association for Computing Machinery.
- Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. In Mariarosaria Taddeo (Ed.), *Minds and Machines*, 30, 681-694.
- Iakovides, N., Lazarou, A., Kyriakou, P., & Aristidou, A. (2022, October). *Virtual library in the concept of digital twin*. In *2022 International Conference on Interactive Media, Smart Systems and Emerging Technologies (IMET)* (pp. 1-8). New York: IEEE.
- Katz, W. A. (1982). *Introduction to reference work (4th ed)*. New York: McGraw-Hill.
- Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). *Challenges and applications of large language models*. arXiv preprint arXiv:2307.10169.
- Lew, G., & Schumacher, R. M. (2020). *AI and UX: Why artificial intelligence needs user experience*. New York: Apress.
- 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. New York: Curran Associates.
- Ranganathan, S. R. (1931). *The five laws of library science*. Madras, India: Madras Library Association and London: Edward Goldston.

- Rao, K. N., & Babu, K. H. (2001). Role of librarian in Internet and World Wide Web environment. *Informing Science*, 4, 25.
- Ravichandiran, S. (2021). *Getting Started with Google BERT: Build and train state-of-the-art natural language processing models using BERT*. Birmingham, UK : Packt Publishing.
- Sivasankari, R., Suriya, S., Sindhu, S., Devi, J. S., & Dhilipan, J. (2024). AI-Powered Recommendation Systems and Resource Discovery for Library Management. In Iman Magdy Khamis (Ed.), *Applications of Artificial Intelligence in Libraries* (pp. 223-244). Pennsylvania : IGI Global.
- Sundermeyer, M., Schlüter, R., & Ney, H. (2012, September). Lstm neural networks for language modeling. In *Interspeech* (Vol. 2012, pp. 194-197).
- Tomáš Mikolov, Martin Karafiát, Lukáš Burget, Jan Černocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Eleventh annual conference of the international speech communication association*, 2010.
- Vaswani, A. (2017). *Attention is all you need*. arXiv preprint arXiv:1706.03762.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention is all you need*. In T. Leen and T. Dietterich & V. Tresp (Ed.), *Advances in neural information processing systems*, 30.
- Wang, S., Lilienfeld, S. O., & Rochat, P. (2015). The uncanny valley: Existence and explanations. *Review of General Psychology*, 19(4), 393-407.
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., & Hu, X. (2024). *Harnessing the power of llms in practice: A survey on chatgpt and beyond*. In Jian Pei (Ed.), *ACM Transactions on Knowledge Discovery from Data*, 18(6), 1-32. New York : Association for

Computing Machinery.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., & Wen, J. R. (2023). *A survey of large language models*. Retrieved from <https://arxiv.org/abs/2303.18223>.

Zheng, Q., Tang, Y., Liu, Y., Liu, W., & Huang, Y. (2022, April). *UX research on conversational human-AI interaction: A literature review of the ACM digital library*. In Simone Barbosa & Cliff Lampe (Ed.), *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1-24). New York : Association for Computing Machinery

# Smart Library: Intelligent Librarian Powered by GPT-4

Han-Yi Hsieh

Software Engineering & Technology Center, Tunghai University

Chih-Chen Ko

Master Program of Digital Innovation, Tunghai University

Lin-Kai Yu

Computer Science Department, Tunghai University

Ting-En Chen

Computer Science Department, Tunghai University

Jung-Sing Jwo (Corresponding author)

Email: jwo@thu.edu.tw

Professor, Computer Science Department, Tunghai University

Yang-Ching Li

Officer, National Library of Public Information

Hui-Chun Chang

Officer, National Library of Public Information

Hsiang-Ping Ma

Director General, National Library of Public Information

## Abstract

With the progression of technology, library services and functions have continuously evolved. The reference services in libraries have undergone significant transformations alongside technological

advancements. Traditional paper-based reference resources are gradually shifting to digital formats. Reference consultation services have evolved from face-to-face inquiries and email consultations to real-time online chatbot services.

With the rapid rise of artificial intelligence technology, libraries have entered a new level of intelligent services, such as smart reference services through text or voice-interactive chatbots. The application of AI technology in libraries has not only enhanced the accuracy of data retrieval and information recommendation but also initiated new service models, such as personalized reading recommendations, intelligent classification, and knowledge graph construction.

In recent years, large language models become matured, and the emergence of ChatGPT-4 has brought revolutionary impacts to libraries. Large language models possess powerful natural language understanding and generation capabilities, significantly improving service efficiency and meeting the diverse and personalized needs of modern readers.

This study aims to explore the intelligent librarian "Xiaoshu" based on GPT-4, emphasizing the enhancement of library service capabilities and reader experience. Xiaoshu, a digital twin with voice communication abilities, can mimic the behavior and language style of real librarian, providing natural and human-like interactions. Since its implementation on December 28, 2023, at the National Public Information Library, Taiwan, Xiaoshu has utilized large language models to offer efficient and accurate information responses, enhancing reader interaction through rich character design.

Looking ahead, this study will continue to refine the functionalities and services of the intelligent librarian. Key areas of focus include database expansion and updates, emotion recognition and response, personalized

recommendations, multilingual support, response time improvement, and user feedback mechanisms. Through continuous technological innovation and service optimization, the intelligent librarian aims to provide readers with more efficient, intelligent, and reader-centric library services, becoming an indispensable part of the smart library ecosystem.

### **Keywords**

Smart library, Large language model, Generative AI, Intelligent librarian, Reference services

## ***SUMMARY***

### **Background and Motivation**

As technology advances, library services and functions have continuously evolved. The reference services in libraries have undergone significant transformations alongside technological advancements. Traditional paper-based reference resources are gradually shifting to digital forms. Reference consultation services have evolved from face-to-face inquiries and email consultations to real-time online chatbot services.

With the rapid rise of artificial intelligence technology, libraries have entered a new level of intelligent services, such as smart reference services through text or voice-interactive chatbots. The application of AI technology in libraries has not only enhanced the accuracy of data retrieval and information recommendation but also initiated new service models, such as personalized reading recommendations, intelligent classification, and knowledge graph construction.

In recent years, large language models become matured, and the emergence of ChatGPT-4 has brought revolutionary impacts to libraries. Large language models possess powerful natural language understanding and generation capabilities, significantly improving service efficiency and meeting the diverse and personalized needs of modern readers.

This study aims to develop an intelligent librarian based on GPT-4. This voice-interactive virtual assistant is designed to simulate a real librarian, providing services such as answering questions, recommending books, and delivering information. The goal of this research is to enhance user experience and improve library services through the development of this intelligent librarian.

## **Architecture Design**

The design of the intelligent librarian is composed of front-end and back-end components. The front-end component primarily supports human-computer interaction (HCI) between the system and the users, including the animations of the intelligent librarian Xiaoshu, the processing of the voice dialogues between intelligent librarian and users, and displaying the content and information generated by the back-end component.

For the back-end component, it utilizes Microsoft's Azure STT (Speech to Text) and TTS (Text to Speech) services to support voice and text conversion. Once the dialogue is converted, back-end component determines whether there is a contextual relationship between the dialogues or if a new context needs to be established. To prevent irrelevant responses, back-end component implements an intent recognition feature that determines whether the response should be generated from the internal database or via designated external sources. When the intention is to query book-related information, back-end component implements another

feature called database query generation. It automatically generates SQL code to query book database based on the user's intention.

### **Practical Application and Outcome Evaluation**

From December 28, 2023, to May 12, 2024, the intelligent librarian received more than 200 visits per-day. The types of conversations between readers and the intelligent librarian Xiaoshu can be broadly classified into eight categories, namely, library services, book inquiries, live information, Xiaoshu personal information, people, politics, ethics, and casual chat. Initially, dialogues related to the casual category were predominant. However, as the usage became more established, library service requests gradually exceeded the casual type dialogues. This trend indicates that the intelligent librarian has effectively adapted to and fulfilled the evolving needs of the readers.

### **Conclusions and Recommendations for Future Work**

The primary goal of this study is to develop an intelligent librarian, Xiaoshu based on GPT-4. The success of the intelligent librarian hinges not only on the accuracy of its responses but also on the quality of the user experience. Xiaoshu is designed to mimic a real librarian in both appearance and voice, while emphasizing the accuracy of generative AI responses and user experience design, with the aim of providing a more humanized interaction.

Since its implementation, readers have gradually come to recognize and appreciate the value of the intelligent librarian in real library services. Future improvements will focus on several directions, including books database expansion and updates, emotion recognition and response, personalized recommendations, and multilingual support. It is anticipated

that the future intelligent librarian Xiaoshu can deliver more efficient, intelligent, and user-centric library services.

### ***ROMANIZED & TRANSLATED REFERENCE FOR ORIGINAL TEXT***

張淳淳 (2003)。參考資源與服務。新北市:國立空中大學。

【Chang, Chun-Chun (2003). *Tsan Kao Zih Yuan Yu Fu Wu*. New Taipei: National Open University.】

莊健國 (2004)。圖書館數位合作參考服務的理論與實務。臺北市：文華圖書館管理資訊。

【Zhuang, Jian-guo (2004). *Tu Shu Guan Shu Wei He Zuo Tsan Kao Fu Wu De Li Lun Yu Shih Wu*. Taipei : Mandarin Library & Information Service.(in Chinese)】

陳宜琳 (2019)。國立臺灣師範大學圖書館參考諮詢機器人建置與評估。未出版之碩士論文，國立臺灣師範大學圖書資訊學研究所，臺北市。

【Chen, Yi-Lin (2019). Development and Evaluation of a Chatbot for Reference Service in National Taiwan Normal University Library (Unpublished master's thesis). National Taiwan Normal University, Taipei. (in Chinese)】

陳志銘 (2022)。圖書館資訊科技與應用。在國家圖書館編，中華民國一一〇年圖書館年鑑(頁 235-237)。臺北市；編者。

【Chen, Chih-Ming (2022). *Tu Shu Guan Zih Syun Ke Ji Yu Ying Yong*. In National Central Library (Ed.), *Yearbook of Librarianship in Taiwan 2021* (pp. 235-237). Taipei: National Central Library. (in Chinese)】

Cho, E., & Han, M. (2019, April). AI powered book recommendation system. In *Proceedings of the 2019 ACM Southeast Conference* (pp. 230-232) New York : Association for Computing Machinery.

- 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. New York : Curran Associates.
- Sivasankari, R., Suriya, S., Sindhu, S., Devi, J. S., & Dhilipan, J. (2024). AI-Powered Recommendation Systems and Resource Discovery for Library Management. In Iman Magdy Khamis (Ed.), *Applications of Artificial Intelligence in Libraries* (pp. 223-244). Pennsylvania : IGI Global.
- Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., & Hu, X. (2024). *Harnessing the power of llms in practice: A survey on chatgpt and beyond*. In Jian Pei (Ed.), *ACM Transactions on Knowledge Discovery from Data*, 18(6), 1-32. New York : Association for Computing Machinery.
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., & Wen, J. R. (2023). *A survey of large language models*. Retrieved from <https://arxiv.org/abs/2303.18223>.