Avances en Interacción Humano-Computadora, 2024, Volumen 9, No. 1, pp. 21-25
DOI: 10.47756/aihc.y9i1.141

RESEARCH PAPER

A Study of LLM-Powered Student Query Support

Cecilia Delgado-Solorzano, Elias Tzoc, Suzanne Rook, Christopher Vinson, Carlos Toxtli

published: 30 November 2024

Abstract

In this paper, we explore the use of Large Language Models (LLMs) to help students improve their information-seeking skills while encouraging the use of references to aid library literacy efforts. This study aims to expand the reach of library support by introducing an approach that leverages the capabilities of LLMs and well-structured prompts. Our approach begins with surveying the current changes students have faced in the last two years concerning their study habits and how they search for information. We subsequently propose a multi-step system prompt, referred as prompting architecture, for foundational and instructed LLMs. The proposed prompt architecture powers a web application named LibRef. We explore the adaptability of the prompting architecture to different information retrieval needs by refining search prompts and providing academic references. A field experiment is conducted using LibRef in academic settings. Our results suggest that the use of LibRef enhances students' academic information-seeking experience. Our research underscores the potential of prompting architectures in procedural refinement of academic queries from students. We believe our findings can provide valuable insights on the current capabilities of LLMs for instructing students to provide more targeted prompts as well as incentivize the use of references.

Keywords:

Learning; Large Language Models; User Study.

1 Introduction

The advent of the digital age has brought about a seismic shift in the way academic libraries operate and serve their patrons. As an increasing number of libraries continue to digitize their collections and develop online platforms, the challenge of effectively aiding students in scanning these digital resources grows correspondingly. Although technology has made it simpler to obtain large amounts of information, it has also made it challenging to filter, find and retrieve high-quality, relevant content. [3]. This has resulted in a critical need for new solutions to enhance the user experience within the digital library ecosystem.

This paper proposes an intersection of Information Science (LIS) and Artificial Intelligence (AI) as a potential solution to this challenge, with a particular focus on the use of multi-step system prompts, referred as prompting architectures in large language models (LLMs) for student queries in academic topics. Prompting architectures can create dynamic and interactive systems that have the potential to execute workflows directed by LLM-based decisions. This can help students to find academic references and refine search prompts for improved search results. A good understanding of how students' study habits and information-gathering techniques evolve is necessary to customize these systems to meet their demands.

We start our research by examining students' modifications to their study habits and information-gathering techniques over the past two years. Understanding these pain points forms the foundation for our proposed solution, allowing us to tailor the prompting architecture to the users' needs. We delve into the proposed system's design, implementation, and operation, emphasizing its adaptability to different use cases. While this work is still ongoing, preliminary insights suggest a promising potential for prompting architectures in LLMs to promote library objectives. Our ultimate goal is to expand the reach of library support, providing students with a tool that not only aids in their academic information search but also helps them to self-assess their information-seeking skills. By doing so, we aim to improve the overall user experience within the academic library ecosystem.

The explorations outlined in this paper can offer insights for various stakeholders - academic institutions looking to enhance their library services, library practitioners seeking to understand and leverage AI in their work, and AI researchers exploring applications of LLMs. Through this study, we aim to simplify the way for a more interactive and user-friendly academic library experience in the digital age.

2 Related Work

As our work lies at the intersection of LIS and AI, our literature review encompasses both these domains, primarily focusing on the use of AI in information-searching services, large language models, and prompting systems.

Use of technologies in Libraries. In the field of LIS, researchers have begun to explore the potential of AI technologies. Sin and Kim reviewed the use of the Internet in libraries, focusing on its potential benefits, the use of it by users and the presence of libraries online. Their works show the interest of libraries in providing the most actualized way of distributing the information and adaptability that the libraries are looking for [7]. A recent study by Omame addressed the implementation of AI-based chatbots in libraries,

Delgado-Solorzano, C.., Tzoc E., Rook S., Vinson C., Toxtli C.
Clemson University
Clemson, USA
cecilid@clemson.edu, etzoc@clemson.edu, rook@clemson.edu, vinsonc@clemson.edu, ctoxtli@clemson.edu

which shows some overlap with our research. However, it does not explore the use of prompting systems in LLMs [5].

Large Language Models. The application of large language models, such as GPT-4, in various domains has been widely explored. Brown et al. introduced GPT-3, discussing its language tasks and performance [2]. A subsequent study by Bender et al. examined the risks and challenges related to using LLMs, providing valuable insights for our work since it is crucial to explore how the information is obtained and used by an LLM [1]. However, to our knowledge, the application of LLMs in academic libraries remains relatively unexplored.

Prompting Systems. Finally, the study of prompting systems has seen significant interest in recent years. A survey by Zamfirescu-Pereira et al. offered a broad overview of different prompting techniques, which forms a fundamental basis for our work [8]. Reynolds et al. discussed the use of dynamic prompting in LLMs, an approach we have adopted for our proposed architecture [6]. Our work stands apart in its application of a well-structured prompting system in LLMs for academic search, which aims to solve unique challenges students face when interacting with AI digital resources. This study expands on the work of previous researchers and introduces an approach in the LIS field.

3 LibRef Description

The proposed system incorporates a Large Language Model (LLM) and a prompting architecture. This allows for conversational interaction with students, guiding them through digital library resource search, access, and utilization. LibRef employs an LLM similar to GPT-4 Turbo, trained on diverse and extensive text corpora to generate human-like responses to queries. The prompting architecture guides LLM responses, yielding answers, references, and learning resources. It addresses students' unique difficulties when using Language Models for study materials, ensuring a comprehensive and instructive support system. The architecture encompasses various steps, including search, validation, and navigational prompts, each serving a specific function.

System operation begins with a student query, which is parsed through the prompting architecture and processed by the LLM. The system provides a response and reference for each query. A feedback loop allows for continuous improvement during a session, with previous prompts appended to the conversation.

Interface Design. The next stage is the actual implementation of the system after it has been designed. This design involves creating the system itself, which combines the prompting architecture and the LLM seamlessly. As an intermediary, the prompting architecture directs user interactions in response to their inquiries. The prompting architecture acts as a mediator and a guide, pointing the LLM toward clarifying strategies for better prompt creation. In other words, it goes over the previously provided prompts as part of the context, looking for differences between past and new prompts and updating references. The overall aspect of the application is shown in Figure 1.

The design has two parts. The first part and sub-parts described in this paragraph are shown in Figure 2 with the respective numbers. It has a "presentation area" (number 1) with text followed by an input-text-box area with its separate button; in this part, users write their first prompt, and when users press the button, the input-textbox and button. The next area is activated with the copy of the first prompt (number 2), the corresponding result (number 3), and references (number 4).

In response to the needs articulated by librarian users, the system is poised to incorporate a feature outlined as Requirement 1, focusing on a space named "History Area of Research" see Figure 3. This user-driven requirement emphasizes the implementation of a carousel/section that keeps track of the record of all previous users' requests to quickly identify the student's progress within the web page interface.

The designated "History Area of Research" is a dedicated space to retain and organize information gathered during searches. This includes storing the retrieved lesson Figure 3 number 2, recommended prompts Figure 3 number 3, references Figure 3 number 4, and staged results Figure 3 number 5. The rationale behind this feature is rooted in the user's practical need to compare and maintain order among the diverse pieces of information sought during the research process. This requirement is derived from the opinions of librarian users who interact with the system firsthand.

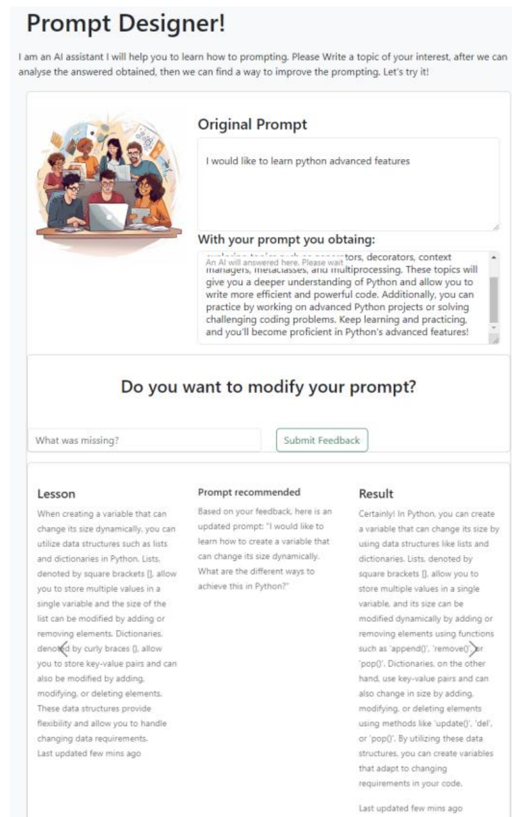


Fig. 1. General aspect of the web application

Their observations highlight how crucial it is to set aside a specific area for gathering and referencing data to facilitate an orderly research process. A fit criterion has been established to evaluate the successful implementation of this requirement. Using a carousel is deemed to keep information from the exact search together, enhancing the user's ability to navigate and comprehend the results of their information-seeking process. Importantly, this requirement stands independently without dependencies or conflicts with other system functionalities, ensuring seamless.

Before deployment, a pilot study is conducted to ensure system readiness for academic settings. This approach, employing an LLM and structured prompting architecture, aims to enhance students' academic information search efficiency and effectiveness. It also offers potential applications for validating literary references and refining search prompts. The system's originality lies in the integration of multi-step prompts specifically designed for challenges faced by students in academic libraries, potentially transforming the user experience within the educational library ecosystem.

4 Methodology

This study recruited 31 participants via Prolific, residing in the USA, with English as their first language, holding a bachelor's degree or higher, and ranging from 18 to 59 years old. Participants have diverse knowledge areas, including Computer Science, Business, Health, Finance, Engineering, History, and others. A mixed-method approach was employed, comprising a pre-survey, user-experience activities, and a post-survey. The pre-survey examined changes in students' study habits and information-gathering techniques over the past two years, focusing on the transition to online activities and LLM incorporation. It also inquired about reference use and LLM experience

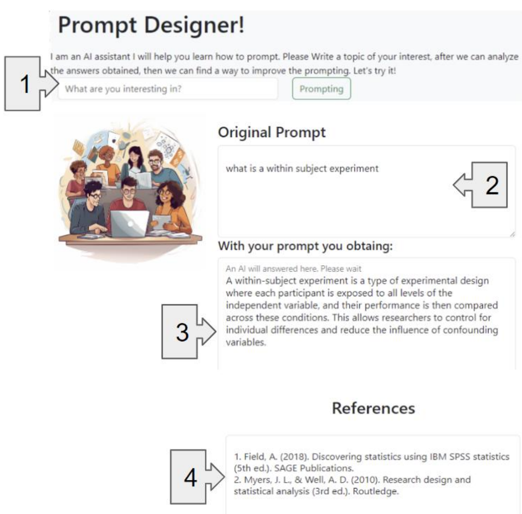


Fig. 2. First part of the web application interaction. #1: input textbox for the first interaction, #2: area where the first prompt is copied, #3 the result that users obtain with the prompt, and #4 corresponding references.

In brainstorming sessions, we generate three scenarios for the user-experience activities interacting with the interface. The first one is any topic of interest to the participant. We expect that participants will choose a public domain topic so they will not pay attention to the references, or the topic will not need references. The second one is about vitamin K and its positive effects. This topic is more specific without diffusion; we expect the user to pay attention to the references. The third one is to generate a prompt by looking for species that could live in different environments. This topic is more general than the previous one, and we want the participants to be more specific in their prompts. Then, the topics are general to particular; with this dynamic, we want to analyze how often the participants give feedback before obtaining a result containing the information they desire and the references the system provides. The experiment of 1 hour duration involved participants who were compensated with 12 dollars per hour. They engaged with the interface, initiating their information-seeking journey through several steps. To commence, they formulated a prompt, setting the stage for their exploration. After retrieving information, a critical evaluation ensued to assess the depth of detail acquired. Participants smoothly transitioned to the next step if deemed insufficient, generating a follow-up prompt to fill in the input textbox "What was missing?" Figure 3 number 1.

The system played an integral role in this interactive process, generating a lesson to refine prompt generation. Subsequently, participants carefully reviewed this feedback, ensuring a better outcome for their subsequent inquiries. Once armed with comprehensive information, participants were responsible for determining the completeness of the gathered data and checked for essential components, including references.

After acquiring information, participants proceeded to the subsequent phase of the activities. We ask them to compile their findings with a report showing their exploration findings. This report will help us explore if they use references. Participants provided valuable feedback on the interface's functionality in the final step, assessing whether its components operated seamlessly or if any issues warranted attention. This iterative and systematic approach ensured a thorough exploration and utilization of the interface, with participants actively shaping and refining their information-seeking experience.

In the post-survey, we ask how the system gives references, lessons, results, and prompt recommendations. Additionally, we ask about the general experience using the interface based on Lewis [4].

All participants reported that they thought the information they received was reliable. With this interaction, the system provides a lesson, suggested prompt, references, and a new result. In our evaluation of the lesson, we explore three aspects using a 5-point Likert scale: overall rating (1 indicating "Very poor" and 5 "Excellent"), satisfaction (1- "Very unsatisfied," 5-"Very satisfied") and perceived usefulness (1-"Strongly disagree" 5 "strongly agree").

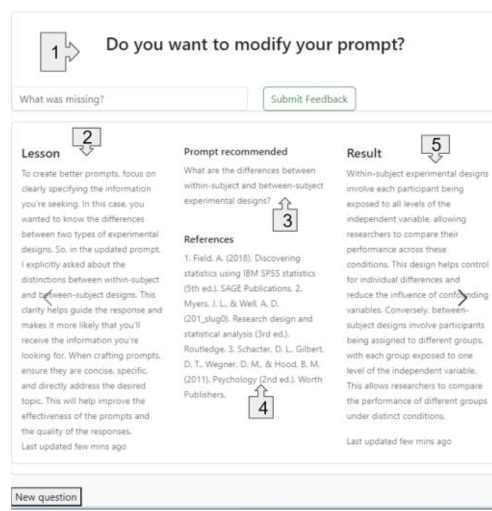


Fig. 3. (1) Feedback area, below it, is found the Requirement 1 "History Area of Research." which contains a (2) lesson, a (3) prompt recommended, (4) references and a (5) new result

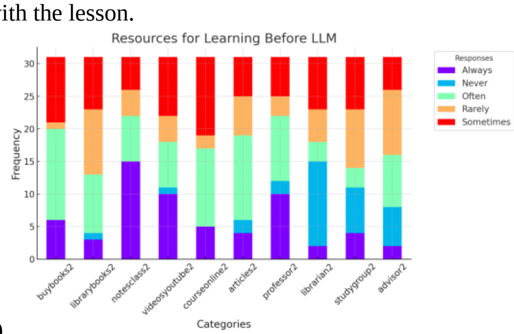
The final step involves an analysis of the collected data to evaluate the effectiveness of the prompting architecture. Key metrics such as system response accuracy, user satisfaction, and overall performance are scrutinized. These insights serve as the foundation for subsequent refinements to the system, initiating an iterative improvement process that ensures continuous enhancement and adaptation to user needs. Our approach combines user-centered design principles with experimental evaluation to develop a solution that is both useful to students and effective in its operation, ensuring it is grounded in real user needs and backed by empirical data.

5 Results

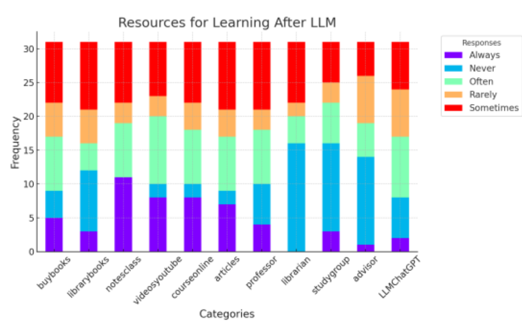
In response to the multiple-option question about how participants use references, the presurvey results revealed that 77.4% of participants reported using Google Scholar as their primary reference tool, while 54.8% used library web pages. The primary purposes for obtaining references were researching and writing papers and homework (22.6%). Figure 4a shows the resources used before the incorporation of LLMs, and Figure 4b illustrates the resources used after.

Participants reported that the system provided references with DOI numbers and in APA style, which contributed to their trust in the references' authenticity. Of the participants, 51.6% reported checking the references and finding them, 6.5% checked but did not find them, and 41% did not check the references. We asked participants how many times they gave feedback; the distribution is shown in Figure 5; the most common was two times.

One activity was to do a report with the information they obtained; the objective was to know if the participants used references; we found that 67% of participants used them, and the rest did not. Participants interact with the system by prompting and giving feedback. The average rating for the lesson was 4, indicating a strong positive reception, suggesting that participants found the lesson to be good. This is mirrored in the satisfaction score, which averaged 4.09, meaning participants were generally very satisfied with the lesson.



a)



b)

Fig. 4. Variety of resources used to learn by users (a) Before the LLM existed and (b) after the LLM appeared.

Similarly, the lesson's perceived usefulness scored an average of 4.06, reflecting its excellence in meeting participants' learning needs and expectations. We also asked how they 1" strongly disagree" or 5 "strongly agree" that the lesson clearly explains how to improve the prompts; the average was 4.29, signifying that participants predominantly felt positive about the clearness of the lesson.

Likewise, we ask participants to evaluate the prompt suggested by the system. They rate on average 4.09, satisfaction with an average of 4.16 and usefulness with 4.35, indicating a high level of approval in all three domains. In this interaction, the system generated 816 references; the minimum was one reference, the maximum was six references, and the maximum frequency was three references. The distribution can be observed in Figure 6. One activity was to do a report with the information they obtained; the objective was to know if the participants used references; we found that 67% of participants used them, and the rest did not.

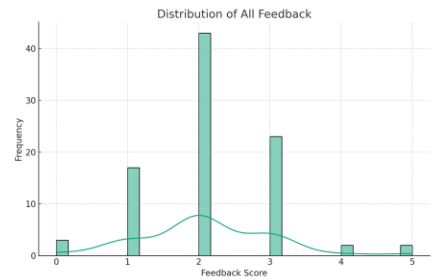


Fig. 5. Distribution of the number of feedback given by users.

6 Discussion

The results of this study provide valuable insights into the effectiveness of the LibRef system and its impact on students' information-seeking behavior in academic contexts. The high adoption rate of Google Scholar among participants highlights the prevalence of online search tools in academic research, while the substantial use of library web pages indicates that traditional academic resources remain relevant. The LibRef system's ability to provide references with DOI numbers and in APA style addresses a crucial need in academic research, potentially bridging the gap between general search engines and specialized academic resources. The system's performance in reference provision is promising, with a majority of participants successfully finding and verifying the provided references. However, a significant portion of participants did not check references, indicating a need for further encouragement of reference verification in future iterations of the system.

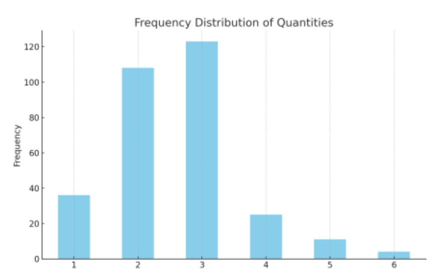


Fig. 6. Distribution of quantity of references given by system

User engagement with the system, as evidenced by the frequency of feedback, demonstrates that participants actively interacted with and refined their queries. This level of engagement suggests that the prompting architecture successfully encouraged the iterative improvement of search strategies. The high ratings for system-generated lessons and prompt suggestions indicate that the LibRef system effectively supports learning and skill development in information seeking. The high satisfaction and usefulness scores further corroborate the system's value in enhancing the academic research experience. The system's ability to generate an appropriate number of references per query demonstrates its capacity to provide comprehensive yet manageable source lists for academic purposes.

This feature could significantly streamline the research process for students. While these results are encouraging, they also highlight areas for improvement, such as increasing reference verification rates. Future research should explore the system's adaptability to different academic disciplines and its potential for integration with existing library systems, as well as its long-term impact on students' information literacy skills.

7 Limitations and future work

The limited generalizability of the study's findings is one of its primary limitations. Although the selected sample size was adequate to observe the use of the system in this exploratory investigation, larger samples in subsequent studies would offer higher statistical power and strengthen the validity of the findings. The conclusions' validity may have been impacted by systematic mistakes produced by this sample bias. As a result, consideration should be taken when interpreting the data, and more study with representative and meaningful samples is required to validate the conclusions.

Future research should aim to include larger and more diverse samples to enhance the generalizability of the findings. Additionally, replicating the study using alternative methodologies could further validate the findings and provide deeper insights into the studied phenomena.

8 Conclusion

This study examined the use of large language models with prompting architectures for academic information seeking. The LibRef system effectively provided references, enhancing user trust. Most participants successfully found and verified the provided references. System-generated lessons and prompts received high ratings for quality, satisfaction, and usefulness. Results indicate that the LLM-powered LibRef system enhanced students' academic information-seeking experience. Areas for improvement include increasing reference verification rates. Future research should focus on external reference validation mechanisms and the system's long-term impact on information literacy skills. This study provides a foundation for integrating AI-powered tools in academic libraries, potentially transforming student interactions with digital resources.

9 References

[1] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 610–623.

[2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems 33 (2020), 1877–1901.

[3] Kun Huang, Xiaoyu Wang, Shichao Luo, Qiuping Su, and Lei Li. 2022. What difficulties did the college students encountered in information seeking during the COVID-19 pandemic? Data and Information Management 6, 2 (2022), 100005.

[4] James R Lewis. 1995. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human-Computer Interaction 7, 1 (1995), 57–78.

[5] Isaiah Michael Omame and Juliet C Alex-Nmecha. 2020. Artificial intelligence in libraries. In Managing and adapting library information services for future users. IGI Global, 120–144.

[6] Laria Reynolds and Kyle McDonell. 2021. Prompt programming for large language models: Beyond the few-shot paradigm. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. 1–7.

[7] Sei-Ching Joanna Sin and Kyung-Sun Kim. 2008. Use and non-use of public libraries in the information age: A logistic regression analysis of household characteristics and library services variables. Library & information science research 30, 3 (2008), 207–215.

[8] JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why Johnny can't prompt: how non-AI experts try (and fail) to design LLM prompts. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 1–21.

10 Appendix: Prompting Architecture

Your main goal is to help me improve my skills on how to give you instructions more effectively. You are required to provide a response based on my initial prompt, then ask me for my feedback on your result, and then, based on my feedback, suggest a more effective prompt. You MUST always follow the following steps:

SET_SYSTEM_RULES: You will get a set of rules to operate. GET_PROMPT: You will get my prompt and jump to GIVE_RESPONSE. GENERATE_RESPONSE: You will give a response to my prompt. Focus on only providing a result without requesting extra information. GET_FEEDBACK: You will receive my feedback on your result. GENERATE_UPDATED_PROMPT: You will generate a more effective prompt based on my feedback. Ensure to integrate any previous updated prompt. Then go to GIVE_LESSON. GIVE_LESSON: You will provide me with a lesson on how to improve my prompt creation skills. Then go to GIVE_NEW_RESPONSE. GIVE_NEW_RESPONSE: You will use the GIVE_UPDATED_PROMPT prompt that you provided and generate a new response. Then jump to GET_FEEDBACK. GIVE_FINAL_LESSON: In case I do not have more feedback, you will provide me with an integrated lesson that covers all your previous advice.

Begin! SET_SYSTEM_RULES: You are a system implemented in a University so your output must comply with all the integrity rules that academic institutions enforce. GET_PROMPT: