DLSU Research Congress 2025
De La Salle University, Manila, Philippines
June 26 to 28, 2025

Application of Generative AI for Contextual and Exploratory Search in a Library Discovery Service

Luis Ezra D. Cruz
The Libraries, De La Salle University
luis.cruz@dlsu.edu.ph

Abstract: The increasing volume of scholarly output and interdisciplinary research has complicated researchers' ability to locate, assess, and connect with relevant materials. This makes it harder for scholars to identify key contributions, often leading to information overload and the risk of overlooking significant studies. Meanwhile, integrating artificial intelligence (AI) in library discovery services offers benefits like enhanced query comprehension and accuracy. To address the limitations of traditional keyword searching that library users might experience, a university library in the Philippines adopted Primo Research Assistant in January 2025, integrated into the Primo VE discovery platform. This generative AI tool allows users to pose natural language queries and provides structured summaries rather than simple lists of search results. This paper presents a preliminary evaluation of Primo Research Assistant, using a heuristic analysis based on Jakob Nielsen's 10 Usability Heuristics. The evaluation reveals the tool's potential to improve academic discovery through AI-powered summarization and natural language interaction, but also points to areas needing improvement, including query interpretation, error feedback, exploratory refinement, and support for advanced user needs. Addressing these issues is crucial for the tool to effectively meet the diverse needs of academic researchers.

Key Words: Generative AI, Library Discovery Services, Usability Evaluation, Information Retrieval

1. INTRODUCTION

The increasing volume of scholarly output and the expansion of interdisciplinary research have complicated researchers' ability to locate, assess, and connect relevant materials. This sharp rise in published research makes it harder for scholars to identify key contributions in their fields of interest (Khalid & Wu, 2020). Researchers are frequently overloaded with sources that are hard to sort through due to the sheer number and diversity of topics. While traditional keyword-based search engines and discovery services are effective at retrieving specific resources, they are often inadequate for exploratory searches, especially when users lack familiarity with terminology or context (Oviedo-Garcia, 2016; Tang, 2024). In these cases, search results often lack relevance, causing frustration. The absence of advanced filtering and ranking based on relevance and quality can yield thousands, even millions, of irrelevant results (Schölvinck et al., 2024). Despite the rapid growth of indexed content in search engines, databases, and digital libraries, the lack of consistent structure leads to numerous irrelevant search results, hindering the location of pertinent research (Khan, Afzal, & Qadir, 2012). Consequently, researchers often face information overload, risking the inclusion of less significant studies. Without automated quality

and relevance assessment, manual sorting of vast data volumes is inefficient and error-prone (Oviedo-Garcia, 2016).

Meanwhile, integrating AI in library discovery services offers numerous benefits that enhance both user experience and operational efficiency. AI technologies such as Natural Language Processing and Machine Learning enhance the comprehension of user queries, making search results more relevant and accurate (Chemulwo & Sirorei, 2021; Shamsitdinova et al., 2024). AI can interpret the context of queries, reducing the ambiguity and imprecision often associated with traditional search methods (Shamsitdinova et al., 2024). In addition, AI-powered recommendation systems provide personalized content suggestions based on user behavior and preferences, increasing user engagement and satisfaction (Sivasankari, 2024).

To address the limitations of conventional keyword searching, particularly in contextual and exploratory inquiries, an academic library in the Philippines adopted Primo Research Assistant in January 2025. This generative AI tool was integrated into the Primo VE discovery platform, a discovery service by Ex Libris. Its primary advantage is its ability to process natural language queries, moving beyond simple keyword searching to understand a user's intent and context. Leveraging generative AI, the tool synthesizes information from the Central Discovery Index (CDI) to provide structured summaries rather than simple lists of search results (Ex Libris, 2025). It also supports exploratory search by suggesting related research areas and concepts that users may not initially consider.

Primo Research Assistant was adopted following an internal evaluation within the academic library. After the evaluation, the tool was made available to the wider academic community. Users can access it within the Primo VE interface, where it appears as a persistent menu tab. The underlying platform, Primo VE, is a subscription service provided by Ex Libris, a library technology firm under Clarivate, a global provider of academic data and analytics. This paper presents a preliminary evaluation of Primo Research Assistant, based on a heuristic analysis conducted by three expert evaluators.

2. METHODOLOGY

This study employed heuristic evaluation, utilizing Jakob Nielsen's 10 Usability Heuristics (Nielsen Norman Group, 2024) (Table 1), to assess the interface and functionality of Primo Research Assistant, a conversational search support feature integrated into a Primo VE discovery platform within an academic library in the Philippines. Heuristic evaluation is a widely accepted method for identifying usability issues in interactive systems (Figueroa et al., 2019; Gonzalez-Holland et al., 2017; Sanchez et al., 2023), and Nielsen's framework has proven effective in detecting a broad range of interface design and interaction problems (Gonzalez-Holland et al., 2017; Hancock et al., 2020). Compared to other usability techniques, heuristic evaluation allows for the early identification of issues with relatively low resource demands, making it particularly suited to library systems evaluation (Figueroa et al., 2019).

Table 1. Nielsen's 10 Usability Heuristics

Heuristic	Description
1. Visibility of system status	Keep users informed about what is going on through timely and appropriate feedback.
2. Match between system and the real world	Use language, concepts, and conventions familiar to the user rather than system-oriented terms.
3. User control and freedom	Allow users to undo and redo actions and easily navigate without getting stuck.
4. Consistency and standards	Follow platform conventions and maintain consistency throughout the interface.
5. Error prevention	Design to prevent problems from occurring in the first place.
6. Recognition rather than recall	Minimize user memory load by making objects, actions, and options visible.

Heuristic	Description
7. Flexibility and efficiency of use	Allow both novice and expert users to tailor frequent actions and shortcuts.
8. Aesthetic and minimalist design	Avoid irrelevant or rarely needed information; keep interface clean and focused.
9. Help users recognize, diagnose, and recover from errors	Provide clear, constructive error messages and guidance for recovery.
10. Help and documentation	Offer easily accessible help and documentation to assist users when needed.

To strengthen the validity and breadth of the evaluation, three expert evaluators participated: (1) the author, which is a librarian specializing in library technologies and systems, (2) a subject librarian from a different academic institution in the Philippines, and (3) a reference librarian from a third academic library. All evaluators had professional experience in academic reference services, discovery platforms, and interface evaluation, providing a well-rounded and practitioner-informed perspective.

Over a four-week period (February 3–28, 2025), each evaluator independently interacted with Primo Research Assistant using a structured set of natural language queries representative of academic research behavior across disciplines. These queries were designed to reflect a wide range of academic research tasks. Exploratory or broad-topic queries involved open-ended questions aimed at gathering general background information or conceptual overviews, such as "What are the major theories of climate change?" Factual and bibliographic lookups focused on specific requests for factual details or complete citations to support precise information needs, for example, "Give the complete citation of *Pedagogy of the Oppressed* by Paulo Freire." Analytical and multi-layered prompts consisted of complex questions requiring comparison, synthesis, or evaluation across multiple concepts, such as "Compare the effects of monetary policy in emerging and developed economies." Technical and field-specific questions involved inquiries using specialized terminology or concepts from particular disciplines, for instance, "Explain the difference between CRISPR-Cas9 and RNA interference in gene editing." Finally, instructional or process-oriented prompts addressed practical requests related to research workflows or academic writing, including "How do I format tables and figures in APA 7th edition?"

The system's interface, behavior, and responses were evaluated against Nielsen's ten usability heuristics (Nielsen Norman Group, 2024): visibility of system status, match between system and the real world, user control and freedom, consistency and standards, error prevention, recognition rather than recall, flexibility and efficiency of use, aesthetic and minimalist design, error recovery, and help/documentation.

Evaluator feedback was analyzed qualitatively through a structured, three-step process. First, all observations were organized by Nielsen's ten heuristics, noting both positive features (observed strengths) and usability violations (issues). Second, each usability issue was categorized by type (e.g., navigation, language, system feedback) and assigned a severity rating, minor, moderate, or critical, based on its potential impact on user experience. Third, findings were compared across the three evaluators to identify recurring themes, shared concerns, and points of divergence. This process produced a detailed heuristic evaluation table, which includes the specific heuristic evaluated, observed strengths, usability issues identified, and severity ratings. This comprehensive synthesis enabled a holistic understanding of the tool's usability and informed practical recommendations for interface and interaction improvements.

3. RESULTS AND DISCUSSION

The heuristic evaluation of Primo Research Assistant identified key strengths and areas for improvement in supporting contextual and exploratory academic information retrieval. Table 2 summarizes the findings according to Jakob Nielsen's 10 Usability Heuristics. The table presents three columns: the specific heuristic evaluated, observed strengths in the system, and usability issues identified, including the assigned severity rating

reflecting the potential impact of each issue on the user experience.

Table 2. Summary of Nielsen Heuristic Evaluation of Primo Research Assistant

Heuristic	Observed Strengths	Usability Issues Identified
1. Visibility of system status	Generally good feedback during query processing, keeping users informed that the system is working.	Occasional delays in response generation causing uncertainty about system activity. Severity: Minor
2. Match between system and real world	Use of academic language in summaries aligned with user expectations and aided readability.	Inconsistent interpretation of complex or ambiguous queries; sometimes missing important contextual details. Severity: Moderate
3. User control and freedom	Easy query revision and navigation through results.	Lack of intuitive options for refining or redirecting exploratory searches limits deeper information discovery. Severity: Minor
4. Consistency and standards	Interface mostly adheres to standard web conventions, aiding ease of use.	Minor inconsistencies in terminology across interface elements that may affect clarity. Severity: Minor
5. Recognitior rather than recall	Clearly labeled summaries and well-presented key information facilitate quick comprehension.	No significant reported. Severity: None
6. Error prevention	N/A	No explicit guidance on how to structure natural language queries to prevent errors. Severity: Moderate
7. Error recovery	N/A	Vague error messages without constructive suggestions hinder users from correcting queries. Severity: Moderate
8. Flexibility and efficiency of use	Supports natural language queries, accommodating various experience levels.	Lack of advanced search options or controls to adjust summary depth and scope limits affects experienced users' efficiency. Severity: Minor
9. Aesthetic and minimalist design	Clean, focused interface that minimizes distractions.	No significant reported. Severity: None
10. Help and documentatior	N/A	Absence of comprehensive help or user guidance/ Severity: Moderate

3.1 Visibility of System Status

The system generally maintained good visibility of status during query processing, providing users with timely indications that their input was being processed. This feedback is essential in academic research settings to assure users that their queries are being addressed. However, occasional delays in generating responses, particularly for broad or complex queries, resulted in periods where users lacked confirmation that the system was still active. This issue, rated as minor in severity, can reduce user confidence and may discourage further interaction or query reformulation.

3.2 Match Between System and the Real World

The match between the system and the real world was supported by the use of academic language in generated summaries, which aligned with user expectations and improved comprehension. Despite this, inconsistencies appeared in the system's handling of complex or ambiguous queries. For example, when asked to summarize the impact of social media on political polarization in Southeast Asia, the response focused on the general effects of social media without addressing the specific regional and political context. Such limitations, considered moderate in severity, suggest the system's understanding of user intent requires further refinement to meet complex academic inquiry demands.

3.3 User Control and Freedom

User control and freedom were evident in the ease of query revision and navigation through results. However, the system lacked clear, intuitive options for refining or redirecting exploratory searches based on initial responses, limiting users' ability to pursue deeper or alternative lines of inquiry. A more prominent placement of recommended questions can be done to improve exploratory searches.

3.4 Consistency and Standards

For consistency and standards, the interface largely conformed to established web conventions, which supports usability through familiar design elements. Nonetheless, minor inconsistencies in terminology across interface components were observed, which could confuse users and detract from overall clarity.

3.5 Recognition Rather Than Recall

A notable strength was the system's support for recognition rather than recall. Summaries were clearly labeled, and key information was well presented, facilitating quick comprehension without requiring users to remember prior steps or commands.

3.6 Error Prevention and Recovery

The evaluation uncovered shortcomings in error prevention and recovery. The system did not provide clear guidance or examples to help users formulate effective natural language queries. When errors occurred, feedback was vague and uninformative, such as messages stating the system could not understand the query without suggesting ways to correct it. This was identified as a moderate issue that somehow impairs user experience.

3.7 Flexibility and Efficiency of Use

The interface has demonstrated flexibility by accommodating users with varying levels of search expertise through conversational query input. However, it lacked advanced search options and offered limited controls to adjust the depth and scope of the summaries provided, which limits efficiency for expert users who require precise control over their search parameters and outputs.

3.8 Aesthetic and Minimalist Design

The minimalist interface design supported focused attention on results, minimizing distractions.

3.9 Help and Documentation

Help features and documentation were sparse and not clearly visible, potentially hindering users' onboarding and troubleshooting. The absence of comprehensive support materials may limit users' ability to fully utilize the tool's capabilities and resolve issues independently.

4. CONCLUSION

Primo Research Assistant showed strong potential in enhancing academic discovery through its natural language query interface and automated summarization features. One of its key strengths lies in the intuitive design that aligns with common web conventions, facilitating ease of use across users with varying levels of expertise. The system effectively maintains visibility of its status during query processing, providing timely feedback that helps users understand when their input is being handled. This aspect is critical in academic contexts where researchers expect responsive interaction while exploring complex topics. Additionally, the use of clear academic language in the generated summaries meets the expectations of scholarly users and supports comprehension, reducing cognitive load by presenting information in a familiar and concise manner. The emphasis on recognition over recall through well-labeled summaries and clearly presented key information further aids users in quickly grasping relevant content without needing to remember previous steps or commands.

Despite these strengths, the evaluation highlights important areas requiring improvement. The system's interpretation of complex and ambiguous queries occasionally falls short, resulting in outputs that do not fully capture essential contextual or disciplinary nuances. This limitation restricts the tool's utility for more detailed or specialized research questions. Moreover, the absence of clear guidance or constructive feedback when queries fail hampers users' ability to refine their input effectively, which can cause frustration and reduce confidence in the tool. The lack of intuitive options for refining or redirecting exploratory searches limits deeper engagement with the information and inhibits the natural progression of academic inquiry based on emerging insights. For more experienced users, the absence of advanced customization options to control the depth and scope of generated summaries constrains efficiency and precision, key factors in rigorous academic research. Lastly, the minimal help documentation presents a barrier to onboarding and effective troubleshooting, which may impact the tool's wider adoption in diverse research environments.

In conclusion, while Primo Research Assistant offers a promising approach to academic search with its user-friendly interface and well-crafted summarization, addressing its limitations is essential. Enhancing query understanding, providing clearer error feedback, supporting more flexible exploratory search, and expanding user customization and support resources will be critical steps in ensuring the system meets the varied and demanding needs of academic researchers. These improvements will help establish Primo Research Assistant as a valuable, reliable, and efficient tool in scholarly research workflows across disciplines.

REFERENCES

Chemulwo, M. J., & Sirorei, E. C. (2021). Managing and adapting library information services for future users: Applying artificial intelligence in libraries. In I. Management Association (Ed.), *Research anthology on collaboration, digital services, and resource management for the sustainability of libraries* (pp. 644-659). IGI Global. https://doi.org/10.4018/978-1-7998-8051-6.ch037

Ex Libris. (2025). *Primo research assistant*. Ex Libris Group. https://knowledge.exlibrisgroup.com/Primo/Product_Materials/001_Next_Discovery_Experience_(NDE)/Primo_Research_Assistant

Figueroa, I., Jiménez, C., Allende-Cid, H., & Leger, P. (2019). Developing usability heuristics with PROMETHEUS: A case study in virtual learning environments. *Computer Standards & Interfaces*, *65*, 132-142. https://doi.org/10.1016/j.csi.2019.03.003

Gonzalez-Holland, E., Whitmer, D., Moralez, L., & Mouloua, M. (2017). Examination of the use of Nielsen's 10 usability heuristics & outlooks for the future. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 1472-1475. https://doi.org/10.1177/1541931213601853

Hancock, G. M., Anvari, S. S., Mok, N. B., Ayvazyan, A., Bai, X., McCoy, K. M., Nare, M. T., Mather,

G. P., Machado, C. L., Chompff, R. M. E., McBride, A. S., & Morales, N. (2021). A heuristic evaluation of usability for environmental control units' eye-tracking interfaces at Veterans Affairs Spinal Cord Injuries and Disorders Centers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *64*(1), 741-745. https://doi.org/10.1177/1071181320641171

Khalid, S., & Wu, S. (2020). Supporting scholarly search by query expansion and citation analysis. *Engineering, Technology & Applied Science Research*, *10*(4), 6102-6108. https://doi.org/10.48084/etasr.3655

Khan, S. A., Afzal, M. T., & Qadir, M. A. (2012). Exploiting title-keywords relation to develop concept classifier for scientific community. In R. Benlamri (Ed.), *Networked digital technologies. NDT 2012* (Communications in Computer and Information Science, Vol. 294, pp. 388-400). Springer. https://doi.org/10.1007/978-3-642-30567-2_37

Nielsen, J. (2024, January 30). *10 usability heuristics for user interface design*. Nielsen Norman Group. https://www.nngroup.com/articles/ten-usability-heuristics/

Oviedo-García, M. Á. (2016). Ex ante evaluation of interdisciplinary research projects: A literature review. *Social Science Information*, *55*(4), 568-588. https://doi.org/10.1177/0539018416658147

Sánchez, D., Arcos, D., Pantoja Y., W. L., & Collazos, C. A. (2023). Heuristics for interface design in telemedicine systems. In *Proceedings of the VIII Iberoamerican Conference of Human-Computer Interaction (HCI 2022)*, October 13-15, 2022, La Habana, Cuba. *CEUR Workshop Proceedings*, *3570*, Paper 5. CEUR-WS.org.

Schölvinck, A.-F., Uygun-Tunç, D., Lakens, D., Vaesen, K., & Hessels, L. K. (2024). How qualitative criteria can improve the assessment process of interdisciplinary research proposals. *Research Evaluation*, *33*, rvae049. https://doi.org/10.1093/reseval/rvae049

Shamsitdinova, M., Khashimova, D., Niyazova, N., Nasirova, U., & Khikmatov, N. (2024). Harnessing AI for enhanced searching in digital libraries: Transforming research practices. *Indian Journal of Information Sources and Services*, *14*(3), 102-109. https://doi.org/10.51983/ijiss-2024.14.3.14

Sivasankari, R., Suriya, S., Sindhu, S., Shyamala Devi, J., & Dhilipan, J. (2024). AI-powered recommendation systems and resource discovery for library management. In I. Khamis (Ed.), *Applications of artificial intelligence in libraries* (pp. 223-244). IGI Global. https://doi.org/10.4018/979-8-3693-1573-6.ch009

Tang, C. (2024). Design methodology for an academic resource-oriented search engine. In *Proceedings of SPIE 13289, International Conference on Optical Communication and Optoelectronic Technology (OCOT 2024)*, *1328908*. https://doi.org/10.1117/12.3049221

Tory, M., & Moller, T. (2005). Evaluating visualizations: Do expert reviews work? *IEEE Computer Graphics and Applications*, *25*(5), 8-11. https://doi.org/10.1109/MCG.2005.102