https://doi.org/10.51408/issi2025_048

From Search to Recommendation: Using an LLM to Assess the Usefulness of Academic Articles

Frans van der Sluis[1], Alesia Zuccala[2], Haakon Lund[3]

[1] *frans@hum.ku.dk*, [2]*a.zuccala@hum.ku.dk*, [3]*hl@hum.ku.dk*
Department of Communication, University of Copenhagen, Karen Blixens Vej 8, 2300, Copenhagen (Denmark)

Abstract

The goal of this paper is to explore what an AI-powered LLM can do to help academics/scientists organize, classify, summarize, and make recommendations concerning the relevance of reference articles for the preparation of a literature review. Literature reviewing is a core task in academia, which requires systematic planning and thinking, but today, enormous amounts of information make this process onerous. Many scholars are familiar with research management tools like Endnote, Zotero and Mendeley; however, the advent of LLMs means that new potentialities are on the horizon. We investigate one LLM's ability to make synthesized judgements about a set of article abstracts retrieved from Scopus (n=194), to prepare a literature review for one 'case paper'. Our finding was that its selecting and filtering capabilities were not quantitatively impressive, though qualitatively, it produced many useful recommendations. Here, we describe the kind of inferences the LLM can make about scientific relevance and discuss the potential of LLMs in utilizing academic literature.

Introduction

As the amount of scientific literature published each year increases, it becomes harder to keep up to date with current information and use it when writing a manuscript. Despite this challenge, literature search and reviewing are core skills that an academic needs to situate and contextualize new work. For the uninitiated, Onwuegbuzie and Frels (2015) have produced a guidebook, titled: S*even Steps to A Comprehensive Literature Review*. Within this 'seven step' approach, the first five constitute an "exploratory phase," followed by an 'interpretive phase,' then finally the seventh 'communicative phase.' The exploratory phase alone involves: 1) establishing a research question, 2) initiating a systematic search, 3) storing, and organizing the information, selecting, and then 'deselecting' information based on an established set of criteria – i.e., to choose or not to choose a source.

Early on, this may have involved basic note cards, but today, the average graduate student does not have to sit amongst notes and papers "Piled high" and "Deep" (PhD) just to produce a comprehensive literature review. Today, software tools like Endnote, Mendeley, and Zotero, make this process much easier. All three tools are useful for storing and organizing references, keeping user notes, inserting citations into a manuscript, and automatically formatting bibliographies. An added benefit of Mendeley and Zotero, is that both possess capabilities as reference finders. For example, the Mendeley 'suggest' feature, implements several different recommendation algorithms (i.e., collaborative & content-based filtering; popularity-based & trend-based models) to help academics "discover new research"

based on [their libraries, search behaviors], and general short-term and long-term interests" (Wordpress, 2015).

Many academics are familiar with Mendeley (Zaug et al., 2011), but now AI-powered Large Language Models (LLM) are inspiring researchers to investigate how useful they are at producing textual summaries (Ahmed Antu et al., 2023, Cai et al., 2024; Jin et al., 2024; Nechakhin,2024), as well as feedback (Liang et al., 2024). Experts are positive about the range of applications and potential impact that LLM will have on the higher education system (Pearson, 2024; Luo et al., 2025). Still, they advise academics to maintain skills in critical thinking, problem solving, and ethical decision-making (Fetcher et al., 2023; Watson et al., 2025).

In this paper we look to an LLM both as a tool for filtering relevant research articles and for recommending how useful the articles are for preparing a literature review. We address this specifically by retrieving document abstracts from Scopus and prompting an LLM to sort and contextualize them for their potential value as references, beyond mere keyword or topical relatedness. This is challenging, since current academic search engines (e.g., Scopus) already deliver relevant results based on extensive queries and keyword-based retrieval, making it difficult to improve significantly upon their effectiveness. In contrast to existing recommender systems, our case serves a well-defined need for academic literature in relation to a paper in progress. To ensure accessibility and scalability, we use an open-source LLM and consumer-grade GPU.

Can an LLM help with literature review? Our aim is to answer this question in the context of academic search (Christou et al., 2024), with the added goal of extending earlier work (Azzopardi & Van Der Sluis, 2024; Van der Sluis & Azzopardi, 2025). Specifically, we examine how an LLM can estimate and detail the relevance and usefulness of scientific article abstracts for writing a 'case paper,' which builds on and follows from that earlier research.

Related work

A key issue in academic search is the subjective and multifaceted nature of relevance (Christou et al., 2024; Jordan & Tsai, 2024). Search engines like Google Scholar rely heavily on ranking algorithms that prioritize citation counts and the presence of search terms (Beel & Gipp, 2009; Mallapaty, 2024), but these methods are neither transparent nor comprehensive. This reliance on citation counts reinforces biases such as the "Matthew effect," where already-cited works gain disproportionate visibility, while less-cited but potentially valuable contributions are relegated to the "long tail" of academic literature (Gould, 2009). It also means that search engines may be misconstrued as informants in knowledge production, rather than inert sources of information. This results in a system that favors established viewpoints and overlooks innovative or niche research, limiting the diversity of knowledge accessible to researchers.

Current approaches to relevance evaluation, including binary and graded judgments, focus primarily on topicality or algorithmic matching, often failing to address a user's specific goals/needs or context (Borlund, 2003; Saracevic, 2007). While graded judgments offer a more nuanced assessment, by assigning degrees of relevance, they

remain centered on query-content relationships and fall short of addressing the practical value of information in specific tasks (Cole et al., 2009; Van der Sluis et al., 2010). The reliance on these relevance-based judgments underpins traditional search engine algorithms like Google Scholar, which conceptualize relevance as relatedness rather than usefulness. Judging usefulness, however, needs more information than can typically be captured in a query or easily evaluated on a search engine index (Cole et al., 2009).

Recent advancements in language models have led to the use of LLMs for judging the relevance and ranking of research papers (Luo et al., 2025). These developments are part of a broader suite of Retrieval-Augmented Generation (RAG) technologies, where LLMs interact with traditional search engines and indexes to ground their outputs in external, up-to-date knowledge sources (Argawal et al., 2022; Huang & Huang, 2024). RAG enables extensive, semantic queries that represent full abstracts when searching a database. Here, LLMs are used for query expansion by extending abstracts with related terms and pseudo-references, leveraging information available in the corpus (Shi et al., 2023). Additionally, LLMs assist in relevance estimation and re-ranking, using both supervised and zero-shot methods to reorder search results based on their conceptual fit with an abstract (Argawal et al., 2024; Hou et al., 2023). Their ability to query by abstract enhances literature exploration by understanding context beyond simple keyword matching, allowing for more precise, user-specific retrieval.

Despite these advancements, the primary focus of LLMs in academic research has remained on generation rather than retrieval, particularly in summarization and literature review writing (Pearson, 2024; Luo et al., 2025). Existing systems, such as AutoCite (Wang et al., 2021) and BACO (Ge et al., 2021), generate structured citation texts by leveraging citation networks and textual data to produce contextually relevant citation texts. Similarly, hierarchical clustering techniques in RAG-based models enhance literature reviews by structuring research fields. These systems excel at summarization and organization, enabling automated literature review writing. While advances in sentence-based planning and contextual summarization have refined the automated presentation of prior work, no existing system explicitly supports ideation and writing by helping authors strategically select and integrate references. In this work, we take a step before fully automated writing, exploring whether an LLM can assist authors in assessing a reference's contribution to their own work.

Method

Instruments and Equipment

The Gemma2 language model was used[1], an open-source large language model (LLM) developed by Google, which features 27 billion parameters. The model was

[1] URL: https://huggingface.co/bartowski/gemma-2-27b-it-GGUF
Model file: gemma-2-27b-it-Q6_K_L.gguf

instruction-trained and employed a recommended quantization level of 6 bits (Team et al., 2024). Gemma2 represents a trend towards smaller yet high-performing models, designed for open exploration, fine-tuning, and testing in diverse applications[2]. Instruction-tuned models usually follow a system-user-assistant prompt structure. Gemma2 omits the system role but supports an assistant role for examples; however, this was deliberately omitted to focus on user-directed instructions.

Inferencing was performed using a consumer-grade Nvidia RTX 4090 GPU, equipped with 24 GB of VRAM. The model's context window was set to 2024 tokens to fit within the available VRAM. This limits the number of tokens that can be included in a single prompt, restricting the number of abstracts that can be supplied simultaneously. LlamaCPP, a foundational API for LLMs, was used to structure prompts and ensure compatibility with the Gemma2 model.

Procedure and Materials

The procedure had two phases (see Figure 1). First, we iteratively refined a Scopus query to identify search results relevant to the case paper's topics: information seeking and green consumption (Azzopardi & Van Der Sluis, 2024; Van der Sluis & Azzopardi, 2025). Scopus, a comprehensive database of academic literature (Mallapaty, 2024), provides detailed results, including titles, abstracts, authors, and other metadata. Standard keyword selection and refinement practices focused on the query while limiting the results list's size. An abstract of the case paper informed query development, with synonyms generated using ChatGPT 4o and selectively added to avoid overexpanding the results list. This process ensured a highly relevant set of abstracts. The results are available on Github (https://github.com/fsluis/scopus-llm-review). In total, 194 abstracts were obtained. The final Scopus query was:

TITLE-ABS-KEY(("search behavior" OR "search behaviour" OR "information seeking" OR "web search" OR "information evaluation" OR "information retrieval" OR "consumer search behavior" OR "green complexities" OR "search on information" OR "greenwashing" OR "green washing" OR "information barriers" OR "knowledge barriers") AND ("responsible consumption" OR "sustainable consumption" OR "green consumerism" OR "conscious consumer" OR "ecological consumer" OR "environmentally sustainable" OR "eco-conscious" OR "ethical consumerism" OR "ethical consumer" OR "socially responsible purchasing" OR "sustainable behavior" OR "sustainable behaviours" OR "sustainable decision making" OR "eco-friendly decision-making" OR "consumer decision making" OR "ethical decision making" OR "purchase decision making" OR "green consumption" OR "green shopping" OR "consumption gap") AND NOT ("infrastructure" OR "enterprise" OR "corporate"))

Final Scopus query

[2] For an informal benchmark, visit https://dubesor.de/benchtable

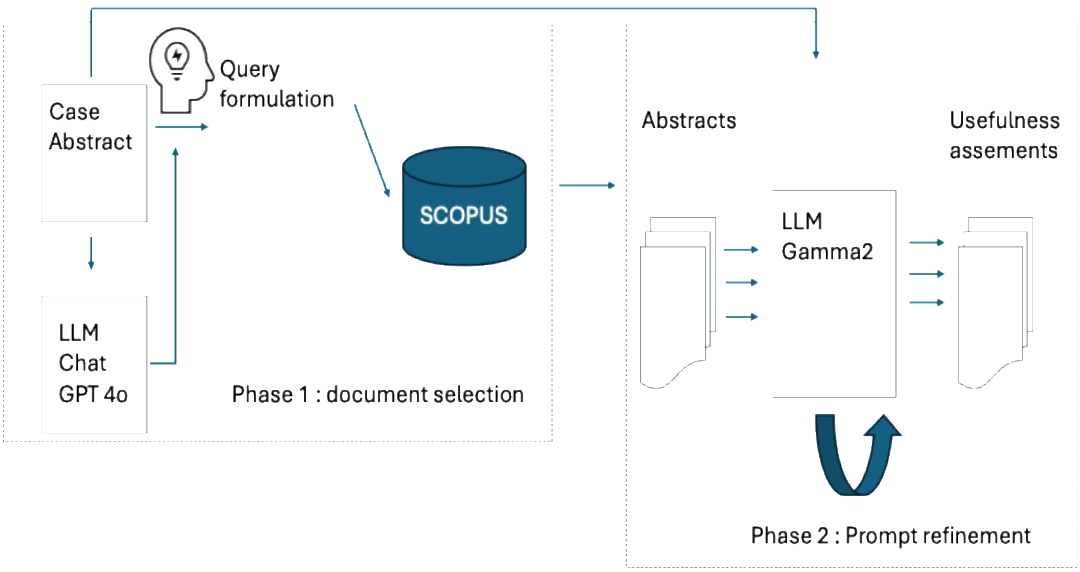

Figure 1. Document selection and prompt refinement. Phase 1 includes an LLM to support manual refinement of a Scopus query. Phase 2 includes an LLM for automated usefulness assessments of article abstracts retrieved from Scopus.

Table 1. Prompt used with Gemma 2.

Section	Prompt
1	I want you to evaluate whether an abstract of a reference paper is relevant to a paper I'm writing. I'll give you details of both my paper and the reference paper.
2	My paper: Abstract: {my_abstract} Reference paper: Title: {title} Abstract: {abstract}
3	I am particularly interested in knowing whether a paper relates to either of: a) Information seeking: Studies of information seeking and sustainable or responsible consumption, including information seeking challenges experienced by consumers; b) Information availability: Studies showing the influence of information availability or barriers on responsible or sustainable consumer behavior; c) Asymmetries: Studies showing the existence of information asymmetries between market players and consumers, such as through greenwashing practices; d) Sustainability: Studies showing the importance of sustainable practices, but are not directly relevant to my study; e) Other: There might be other categories of relations. Do feel free to add / interpret new types of relations.

4	Be critical when estimating relevance. If it is not about sustainability or responsible consumption, it is not relevant.
5	Does the reference paper seem relevant? If yes, how can it be utilized in my research? Answer in a structured way: Relevance: Yes, possibly, no Relation: Seeking, availability, asymmetries, sustainability, other Utilization: Explain how this paper can be utilized in my research

Results

Here we analyze the responses received from the LLM, both quantitatively and qualitatively. The focus is purely on the end-product of our exploration. The complete set of responses is available on GitHub (https://github.com/fsluis/scopus-llm-review).

Quantitative results

The LLM did not reduce the amount of text needed for an evaluation from a quantitative perspective. The average text length of the full-text assessments is 1463.51 characters, while abstracts, at 1490.07 characters on average, are close to the same length. Of the 194 references analyzed, 41 were identified as relevant, while 129 were labeled 'possibly' relevant. With only 24 reference articles identified as not relevant, the tool did not drastically reduce what was considered useful as a final collection. In fact, it left a substantial amount of work for the authors to manually sort through and identify relevant versus non-relevant articles. The usefulness of the LLM tool was, from a quantitative perspective, limited.

Qualitative results

Qualitatively, we observed certain features of LLM responses that made it overall a powerful tool. Here, we highlight features of responses identified through initial observations and subsequent collaborative discussions amongst the authors. Quotes were selected randomly, though with a preference for brevity. The LLM responses were consistent in focus and style, suggesting that a sample of responses provides a reliable representation of the overall output.

Some of the full-text responses focused primarily on *relevance relations* specified in Section 3 of the prompt, expanding on the structured (quantitative) responses:

(1) *"**Information Availability:** The reference paper directly addresses how retailers can influence consumer behavior by making eco-friendly products more visible and accessible. This aligns with your focus on "information barriers" and how accessible information can drive prioritization of responsible aspects. You can cite this paper to support your argument that improving information availability is crucial for fostering responsible consumption."*

[DOI: 10.1016/j.jretconser.2016.07.008, Relevance: yes, Relation: availability, asymmetries]

(2) *"**Greenwashing as an Information Asymmetry:** The reference paper highlights how greenwashing practices create an information asymmetry between companies and consumers. Companies may present their products as more environmentally friendly than they actually are, misleading consumers who lack complete information."*

[DOI: 10.1108/YC-10-2020-1241, Relevance: yes, Relation: asymmetries]

These examples demonstrate that the LLM typically explains the relevance relation observed. Additional information is included from the reference paper abstract, as seen in Quote 2, and in certain cases about potential use cases, as seen in Quote 1. More often, the usage-part of the response expands on the initial relevance relations, by highlighting other potential uses:

(3) *"**Connecting to Your Research:** You can use this paper to support your argument about the importance of **information availability and transparency** in shaping responsible consumption. The reference paper demonstrates how a lack of accurate information (due to greenwashing) can hinder consumers' ability to make ethical choices, even if they have initial intentions to consume responsibly."*

[DOI: 10.1108/YC-10-2020-1241, Relevance: yes, Relation: asymmetries]

(4) *"**Establishing a baseline:** The paper's findings on the increasing prevalence of online information seeking for major financial decisions could be used to highlight the growing importance of online information for consumers in general. You could then argue that this trend extends to responsible consumption decisions, even if the paper itself doesn't directly address sustainability."*

[DOI: 10.1080/16522354.2007.11073450, Relevance: possibly, Relation: information seeking]

Quote 3 corresponds to previously presented Quote 2 [i.e., DOI: 10.1108/YC-10-2020-1241]. Here, the LLM's response directly supports a core argument of the case paper by identifying how greenwashing hinders sustainable consumption. Quote 4 indirectly supports a broader, non-core argument, concerning the importance of information seeking in consumer decision-making. Both illustrate the LLM's ability to expand on initial relevance assessments and describe how each paper might potentially be used, as requested in Section 5 of the prompt.

The LLM also identified several 'other' relationships, peripheral to the case paper. 'Other' relationships were less common (only 16 occurrences with a positive relevance annotation) and typically came with more extensive full-text responses:

(5) *"**Information Seeking in Niche Communities:** The paper highlights how online drug forums serve as platforms for information exchange about a specific product (MXP). This could be drawn upon to illustrate how consumers in niche markets, even those involving potentially harmful products, actively seek information to inform their purchasing decisions. You could discuss this*

as a parallel to consumers seeking information about ethically sourced or sustainable products."

[DOI: 10.1080/02791072.2014.974002, Relevance: possibly, Relation: other]

(6) *"**Decision-Making Strategies**: The paper identifies decision-making strategies used by consumers in omni-channel contexts. Understanding these strategies could be helpful in contextualizing how information about responsible consumption influences consumer choices. For example, if consumers primarily rely on "weighted additive rule" (WADD), it suggests they weigh different product attributes, potentially including ethical considerations, in a systematic manner."*

[EIC: 2-s2.0-85126569618, Relevance: possibly, Relation: other]

Quote 5 focuses on the consumption of a synthetic drug (MXP) with potential personal harm rather than environmental harm. Quote 6 focuses on fashion buying, suggesting that similar heuristics may apply to sustainable considerations. Both examples provide interesting ways to enrich our case paper, methodologically and/or contextually.

Sometimes the LLM provided a few creative (but good!) suggestions for alternative papers to use; papers that did not have any initial relevance relation to the case paper, but may be useful when writing the implications or discussion sections of the case paper:

(7) *"**Expanding Your Discussion:** You could discuss the reference paper findings in the context of your own research on information barriers. provides a concrete example of how information asymmetries can act as barrier to responsible consumption. You could also explore how you proposed solutions (improved information systems, decision-support tools could help mitigate the negative effects of greenwashing and empowe consumers to make more informed decisions."*

[DOI: 10.1108/YC-10-2020-1241, Relevance: yes, Relation: asymmetries

(8) *"**Exploring interventions:** The study's use of retro-cueing to increas information availability could inspire you to explore potential intervention or design strategies that make ethical information more salient and accessibl to consumers."*

[DOI: 10.1016/j.jbusres.2019.08.012, Relevance: possibly, Relation: availability

Both of these examples point at possible use cases beyond the initial relevance relations prompted. It expands the scope of literature considered, surfacing sources that might otherwise be overlooked.

And finally, in addition to reference papers with potentially creative uses, the LLM responded with critical notes, related to Section 4 of the prompt:

(9) *"Remember to critically evaluate the reference paper's methodology and findings in the context of your own research."*

[DOI: 10.1080/17543266.2024.2343934, Relevance: yes, Relation: Asymmetries, Sustainability]

*"**Important Note:** While the reference paper offers relevant insights, it's crucial to acknowledge the differences in context. Health information seeking and responsible consumption involve distinct decision-making processes and motivations. Therefore, avoid directly applying the findings to your research without careful consideration and contextualization."*

[DOI: 10.1016/j.hlpt.2024.100842, Relevance: possibly, Relation: information seeking]

These critical notes appeared frequently. Quote 9 reminds us that abstracts alone are insufficient to assess a reference paper's merits. Quote 10, which was more common, cautions against over-generalizing a reference paper's findings to the case paper's context. We found these critical notes to be well-grounded and comment on this further in the Discussion section.

Overall, these 10 quotes highlight the LLM's strength in contextualizing reference abstracts and presenting structured, clear assessments. Clear headings enabled quick scanning of reference papers, while the structured and concise format made it easier to evaluate papers on their potential usefulness. By going beyond relevance to provide actionable suggestions, some responses guided the incorporation of references into the case paper, helping refine arguments and expand its scope and implications.

Discussion

This work positions LLMs as a transformative tool in literature reviews by addressing two key contributions. First, it demonstrates how LLMs fulfill the longstanding ambition of implementing usefulness as a core relevance concept, moving beyond traditional binary or graded relevance judgments to actionable insights. By structuring responses with relevance labels and task-specific suggestions, LLMs bridge the gap between search engine outputs and the practical support of ideation and writing processes. Second, it extends the scope of Retrieval-Augmented Generation (RAG) approaches, showing that information retrieval not only enhances text generation, but also that LLMs can augment traditional article-based approaches. By connecting relevance to usefulness, LLMs unify these two paradigms, advancing both the practical application of retrieved items and raising the possibility of generation-augmented retrieval (GAR), where LLMs become part of the retrieval process.

Our results show that LLM-generated assessments add significant value beyond reference abstracts by helping researchers interpret diverse and dispersed details. By consolidating information into structured insights, LLMs assist in evaluating both relevance and usefulness in relation to a researcher's work. This streamlines the literature review process in two key ways: saving time when sifting through large volumes of references and supporting writing through creative ideas and recommendations for integrating citations. For researchers with limited time or resources, LLMs running on consumer-grade hardware provide a scalable and efficient alternative to traditional methods. However, these findings are based on a

sole case study and reflect the authors' perspectives, which may limit their generalizability. Even though both the authors and intended readership are well-positioned to judge the examples presented in this case study, it remains an open question as to whether these conclusions hold across different authors, disciplines, or research contexts.

These findings suggest a broader role for LLMs in the literature review cycle. By mitigating biases introduced by citedness-based rankings in search engines like Google Scholar (Mallapaty, 2024), LLMs can delve into less-explored references, potentially democratizing the academic literature (Fecher et al., 2023). By easing access to lesser-cited but valuable works, LLMs could even out the long tail of underused articles and give smaller, lesser-known studies a better chance of being cited. This contributes to a more equitable distribution of academic attention and resources.

Despite these advantages, quality control remains a critical limitation. The risk of misuse, where LLMs might shortcut the review process without proper validation, underscores the need for robust quality mechanisms. In a landscape where LLMs increasingly support both the reading and writing of academic literature (Fecher et al., 2023), the emphasis on peer review and expert judgment is heightened. This is especially vital given the proliferation of non-peer-reviewed repositories like arXiv and the potential for errors to propagate, echoing concerns seen in the replication crisis within other disciplines (Open Science Collaboration, 2015). As reliance on LLMs grows, the importance of quality control (Van der Sluis, 2022; Van der Sluis et al., 2024) cannot be overstated.

Future work

To generalize the current findings, future work could consider repeating the presented approach across different case papers and disciplines. Establishing ground truth labels would allow for more formal evaluation using retrieval metrics such as precision and recall, while also enabling comparative testing against existing tools such as Scopus rankings or Zotero Suggest. This could help quantify the practical advantages of LLMs in literature review workflows, beyond the currently highlighted qualitative strengths such as interpretability and perceived usefulness.

Future work could also extend the technical contributions of this study by testing different LLMs and refining prompt design. In addition, automated querying, developed and researched as part of the RAG suite, presents promising opportunities for academic literature search. LLMs can support query drafting, refinement, and synonym generation for complex academic search engines like Scopus. While the current study focuses on interpreting retrieved abstracts, future systems could integrate both querying and evaluation in a single LLM workflow.

Nevertheless, the computational demands and environmental footprint of LLMs warrant continued investigation. Developing efficient workflows for consumer-grade hardware could broaden access and promote more sustainable and responsible deployment of these tools in academic research. Addressing these challenges alongside optimizing consumer-grade hardware use offers a dual opportunity: advancing LLM capabilities for academic purposes and promoting their responsible,

sustainable deployment. These efforts would support more equitable access to research tools, reinforcing the democratization of academic practices.

References

Ahmed Antu, S., Chen, H., & Richards, C. K. (2023). Using LLM (Large Language Model) to improve efficiency in literature review for undergraduate research. In *Proceedings of the Workshop on Empowering Education with LLMs – the Next-Gen Interface and Content Generation*, Tokyo, Japan, 7 July 2023, 8–16.

Agarwal, S., Laradji, I. H., Charlin, L., & Pal, C. (2024). LitLLM: A toolkit for scientific literature review. *arXiv*, doi:10.48550/arXiv.2402.01788.

Aytar, A. Y., Kilic, K., & Kaya, K. (2024). A retrieval-augmented generation framework for academic literature navigation in data science. *arXiv*, doi:10.48550/arXiv.2412.15404.

Azzopardi, L., & Van der Sluis, F. (2024). Seeking socially responsible consumers: Exploring the intention–search–behaviour gap. In P. Clough, M. Harvey, & F. Hopfgartner (Eds.), *Proceedings of the 2024 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '24)* (pp. 153–164), Sheffield, UK, March 10–14, 2024. ACM. doi:10.1145/3627508.3638324

Van der Sluis, F., & Azzopardi, L. (2025). Search changes consumers' minds: How recognizing gaps in understanding drives ethical choices. In D. McKay & D. Oard (Eds.), *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25)*, Melbourne, Australia, March 24–28, 2025. ACM. doi:10.1145/3698204.3716456

Beel, J., & Gipp, B. (2009). Google Scholar's ranking algorithm: The impact of citation counts (an empirical study). In *Proceedings of the 2009 Third International Conference on Research Challenges in Information Science (RCIS)* (pp. 439–446). Fez, Morocco, April 22–24, 2009. IEEE. doi:10.1109/RCIS.2009.5089308

Borlund, P. (2003). The concept of relevance in IR. *Journal of the American Society for Information Science and Technology*, 54(10), 913–925. doi:10.1002/asi.10286

Cai, H.; Cai, X.; Chang, J.; Li, S.; Yao, L.;Wang, C.; Gao, Z.; Li, Y.; Lin, M.; Yang, S.; et al. (2024)

SciAssess: Benchmarking LLM Proficiency in Scientific Literature Analysis. arXiv, doi: 10.48550/arXiv.2403.01976.

Christou, E., Parmaxi, A., & Zaphiris, P. (2025). A systematic exploration of scoping and mapping literature reviews. *Universal Access in the Information Society*, 24(1), 941–951. doi:10.1007/s10209-024-01120-3

Cole, M., Liu, J., Bierig, R., Gwizdka, J., Liu, C., Zhang, J., & Zhang, X. (2009). Usefulness as the criterion for evaluation of interactive information retrieval. In *Proceedings of the 2009 HCIR Workshop on Human–Computer Interaction and Information Retrieval* (pp. 1–4), Cambridge, MA, USA, October 23, 2009.

Fecher, B., Hebing, M., Laufer, M., Pohle, J., & Sofsky, F. (2025). Friend or foe? Exploring the implications of large language models on the science system. *AI & Society*, 40(2), 447–459. doi:10.1007/s00146-023-01791-1

Ge, Y., Dinh, L., Liu, X., Su, J., Lu, Z., Wang, A., & Diesner, J. (2021). BACO: A background knowledge- and content-based framework for citing sentence generation. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL/IJCNLP 2021)* (pp. 1466–1478). Online, August 1–6, 2021. Association for Computational Linguistics. doi:10.18653/v1/2021.acl-long.116

Gould, T. H. P. (2009). The future of academic publishing: Application of the long-tail theory. *Publishing Research Quarterly*, 25(4), 232–245. doi:10.1007/s12109-009-9134-y

Huang, Y., & Huang, J. (2024). A Survey on Retrieval-Augmented Text Generation for Large Language Models. *ArXiv*. doi: doi:10.48550/arXiv.2404.10981

Hou, Y.,Zhang, J.,Lin, Z., Lu, H., Xie, R., McAuley, J., Zhao, W.X. (2023) Large Language Models are zero-shot rankers for recommender systems. *ArXiv*, doi:10.48550/arXiv.2305.08845

Jin, H.; Zhang, Y.; Meng, D.; Wang, J.; Tan, J. (2024). A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods. *arXiv*, doi: doi:10.48550/arXiv.2403.02901.

Jordan, K., & Tsai, S. P. (2024). Ranking 'by relevance' in academic literature searches: Prevalence, definitions, and implications. *Postdigital Science and Education*. doi:10.1007/s42438-024-00530-z

Latif, E.; Fang, L.; Ma, P.; Zhai, X. (2023). Knowledge distillation of llm for education. *arXiv*, doi: 10.48550/arXiv.2312.15842.

Liang, W.; Zhang, Y.; Cao, H.; Wang, B.; Ding, D.; Yang, X.; Vodrahalli, K.; He, S.; Smith, D.; Yin, Y.; et al. (2023). Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *arXiv*, doi:10.48550/arXiv.2310.01783.

Luo, Z., Yang, Z., Xu, Z., Yang, W., & Du, X. (2025). LLM4SR: A Survey on Large Language Models for Scientific Research. *ArXiv*. doi:10.48550/arXiv.2501.04306.

Mallapaty, S. (2024). Can Google Scholar survive the AI revolution? *Nature*, 635(8040), 797–798. doi:10.1038/d41586-024-03746-y

Nechakhin, V., D'Souza, J., & Eger, S. (2024). Evaluating large language models for structured science summarization in the open research knowledge graph. *Information(Switzerland)*, *15(6)*, 328. doi:10.3390/info15060328

Onwuegbuzie, A. J., & Frels, R. K. (2015). *Seven steps to a comprehensive literature review: A multimodal and cultural approach*. SAGE Publications.

Open Science Collaboration. (2015). *Estimating the reproducibility of psychological science. Science*, *349*(6251), aac4716. doi:10.1126/science.aac4716

Pearson, H. (2024). Can AI review the scientific literature—and figure out what it all means? Nature, 635(8038), 276–278. doi:10.1038/d41586-024-03676-9

Saracevic, T. (2007). Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. *Journal of the American Society for Information Science and Technology*, 58(13), 1915–1933. doi:10.1002/asi.20682.

Shi, Z., Gao, S., Zhang, Z., Chen, X., Chen, Z., Ren, P., & Ren, Z. (2023). Towards a unified framework for reference retrieval and related work generation. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 5785–5799). Singapore: Association for Computational Linguistics. doi:10.18653/v1/2023.findings-emnlp.385.

Team, G., Riviere, M., Pathak, S., Sessa, P. G., Hardin, C., Bhupatiraju, S., Hussenot, L., et al. (2024). Gemma 2: Improving Open Language Models at a Practical Size. *arXiv*. doi: doi:10.48550/arXiv.2408.00118.

Van der Sluis, F., van den Broek, E. L., & van Dijk, B. (2010). Information Retrieval eXperience (IRX): Towards a human-centered personalized model of relevance. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2010)* (pp. 322–325). Toronto, Canada, August 31 – September 3, 2010. IEEE. doi:10.1109/WI-IAT.2010.222

Van der Sluis, F. (2022). *A conversationalist approach to information quality in information interaction and retrieval*. In *CHIIR'22 Workshop on Information Quality in Information Interaction and Retrieval*. doi:10.48550/arxiv.2210.07296

Van der Sluis, F., Faure, J., & Homnual, S. P. (2024). An empirical exploration of the subjectivity problem of information qualities. *Journal of the Association for Information Science and Technology*, 75(7), 829–843. doi:10.1002/asi.24884.

Watson, S., Brezovec, E., & Romic, J. (2025). The role of generative AI in academic and scientific authorship: An autopoietic perspective. *AI & Society, 1-11*. doi:10.1007/s00146-024-02174-w

Wang, Q., Xiong, Y., Zhang, Y., Zhang, J., & Zhu, Y. (2021). AutoCite: Multi-modal representation fusion for contextual citation generation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21)* (pp. 788–796). Virtual Event, Israel, March 8–12, 2021. ACM. doi:10.1145/3437963.3441739

Wordpress (2015). A practical guide to building recommender systems. From Algorithms to Product. Mendeley Suggest. Available at: https://buildingrecommenders.wordpress.com/2015/11/13/mendeley-suggest/.

Zaugg, H., West, R. E., Tateishi, I., & Randall, D. L. (2011). Mendeley: Creating communities of scholarly inquiry through research collaboration. *TechTrends*, 55(1), 32–36. doi:10.1007/s11528-011-0467-y