Conversational Exploration of Literature Landscape with LitChat

Mingyu Huang[1] , **Shasha Zhou**[2] , **Yuxuan Chen**[3] and **Ke Li**[2]

[1]College of Computer Science and Engineering, University of Electronic Science and Technology of China, Xiyuan Avenue 2006, Chengdu, China

[2]Department of Computer Science, University of Exeter, EX4 4QF, Exeter, United Kingdom

[3]James Watt School of Engineering, University of Glasgow, G12 8QQ, Glasgow, United Kingdom

m.huang.gla@outlook.com, {sz484, k.li}@exeter.ac.uk, 2959902c@student.gla.ac.uk

Abstract

We are living in an era of "big literature", where the volume of digital scientific publications is growing exponentially. While offering new opportunities, this also poses challenges for understanding literature landscapes, as traditional manual reviewing is no longer feasible. Recent large language models (LLMs) have shown strong capabilities for literature comprehension, yet they are incapable of offering "comprehensive, objective, open and transparent" views desired by systematic reviews due to their limited context windows and trust issues like hallucinations. Here we present LitChat, an end-to-end, interactive and conversational literature agent that augments LLM agents with data-driven discovery tools to facilitate literature exploration. LitChat automatically interprets user queries, retrieves relevant sources, constructs knowledge graphs, and employs diverse data-mining techniques to generate evidence-based insights addressing user needs. We illustrate the effectiveness of LitChat via a case study on AI4Health, highlighting its capacity to quickly navigate the users through large-scale literature landscape with data-based evidence that is otherwise infeasible with traditional means.

1 Introduction

The past decade has witnessed an unprecedented expansion of scientific literature on many topics. This deluge of digital data on publications offers unprecedented opportunities for scholars and practitioners to explore patterns characterizing the structure and evolution of the underlying literature landscapes while simultaneously confronting them with new challenges. Scholars often develop systematic literature reviews (SLRs) [Kitchenham et al., 2009] to obtain comprehensive overviews of the relevant topics by manually reviewing a set of relevant publications. However, the exponentially increasing volume of literature means that the provision of "comprehensive, objective, open and transparent" (COOT) principles is no longer feasible by traditional means. For example, the number of publications on the applications of artificial intelligence in healthcare has reached 100,000 in 2023, which is far beyond the capacity of human experts to review.

Large language models (LLMs) like ChatGPT and Gemini have now emerged as transformative tools that promise to revolutionize scientific literature exploration and understanding. Autosurvey [Wang et al., 2024] is a recent example of such a system that uses LLMs to automatically generate literature reviews. Another system, PaperQA2 [Skarlinski et al., 2024], leverages LLMs and retrieval-augmented generation (RAG) [Gao et al., 2023] to answer questions about scientific publications. While they have shown some promise, they also fail to provide COOT views of literature landscapes. For example, due to the limited context windows of LLMs, it is infeasible to feed the entire literature corpus, which may contain thousands of papers, to the LLMs. This is in contrast to the "comprehensive" principle of SLR. Similarly, as probabilistic models, LLMs function as black boxes and are prone to hallucinations [Ji et al., 2023], which may lead to incorrect or misleading information that violates the "objective" and "transparent" principles.

Long before the emergence of LLMs, scholars have explored the use of data-driven methods to generate insights that can guide literature exploration. For example, Schmallenbach et al. leverages topic modeling to explore latent themes in large corpora of literature on the use of artificial intelligence in life science research. Jeong et al. uses community detection techniques to explore key research topics in authentication. Our recent work [Huang and Li, 2024] combines these with author collaboration network analysis to provide a holistic view on the research landscape of multi-objective optimization research. These methods could provide quantitative evidence to facilitate SLR under the COOT principles.

Inspired by this, here we present LitChat, an end-to-end, interactive and conversational literature agent that combines LLM agents and data-driven discovery tools to facilitate literature exploration. LitChat can automatically interpret user queries, design search queries, retrieves relevant sources from online databases, constructs bibliographic knowledge graphs, and employs diverse data-mining techniques to deliver evidence-based insights addressing user needs. It is able to perform explorations of different granularities, from providing an overview of the literature landscape to drilling down into the details of individual papers or researchers. We demonstrate the effectiveness of LitChat through a case study on AI4Health, highlighting its capacity to guide users through a timely, large-scale and rapidly evolving research landscape.
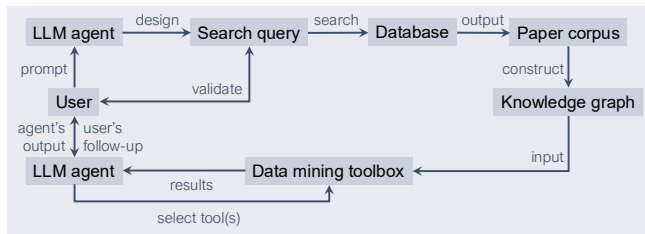
Figure 1: Schematic overview of LitChat.

2 LitChat

A schematic overview of LitChat is depicted in Figure 1 on page 2. Overall, LitChat consists of two LLM agents, both of which can interact with users in a conversational manner. The first agent (upper) takes user queries and devises boolean search queries to retrieve relevant publications from online databases. The collected papers along with metadata are then used to construct a bibliographic knowledge graph (BKG). The second agent (lower) then selects appropriate data-mining techniques to explore the obtained BKG and generate responses to user queries based on the analysis results.

At a higher level, the whole system is connected to a frontend built using JavaScript and React, which allows users to interact through a web interface. This frontend is connected to a backend using a Python Flask web app. The backend holds the LLM agents, the BKG (via Neo4j graph database), the paper metadata (via SQLite databases), and the data mining library. Below we delineate each component of LitChat.

2.1 Query Design and Data Acquisition

Initial query design. When a user enters a query, e.g., "I want to know what healthcare tasks researchers have explored using LLMs for" (see Figure 2 on page 3), the first LLM agent identifies the key research domains involved and synthesizes them into a boolean search query that can be used on established databases like Web of Science, Scopus, or Semantic Scholar. Benefiting from the extensive training on large-scale corpora, LLMs have shown human-level performance in query formulation [Wang *et al.*, 2023]. To further ensure the quality of the designed query, 10 in-context-learning examples are provided to the LLM agent. These examples are extracted from established SLRs from diverse domains in which domain experts explicitly describe the search queries they used for literature search.

Note on the choice of LLMs: Unlike other LLM-based literature agents like Autosurvey and GraphRAG, which prompt LLMs to scrutinize the entire text corpus before reaching the final outputs, LitChat only uses LLMs to directly interact with the users for essential communications like discussing the query, explaining the results, and answering questions. This makes LitChat significantly more "token-efficient" and thus more affordable for broader applications. We thereby choose GPT-4o to enhance performance, though any other LLMs can also be used.

Interactive refinement. The initial query will then be sent to the user for confirmation. The user can either approve it or provide feedback to refine, e.g., by adding or removing keywords. In the latter case, the LLM agent will update the query accordingly and ask for confirmation again. This process continues until the user is satisfied with the query.

Database search and metadata. Once the query is confirmed, the LLM agent forwards it to an online database via API calls to retrieve relevant publications. Here we use the WoS API for this purpose, which is one of the most comprehensive literature databases and a common choice for SLR. The WoS API returns the matched publications in JSON format, which consists of rich curated metadata such as abstract, authors, venues, publication date, citations. The agent stores this information in a SQLite database for further processing.

Note on full paper data: Due to copyright issues, LitChat typically only retrieves paper metadata from online databases, unless the full paper is open access. Yet, a user is able to upload PDF documents to LitChat for further analysis.

2.2 Bibliographic Knowledge Graph Construction

Once the paper metadata is collected, LitChat parses it to construct a BKG, which essentially contains all the information regarding the underlying literature landscape. Notably, this graph is heterogeneous, consisting of different types of entities including papers, authors, venues, keywords, institutions, etc., as well as complex relationships among them like citation, co-authorship, and keyword co-occurrence, etc. Paper abstracts are also integrated as part of node attributes, and we generate embeddings for them using the `voyage-3-large`[1] model, a top performer on the MTEB leaderboard. The final BKG provides the basis for all the subsequent data mining and response generation tasks. Upon construction, it will be stored in a Neo4j graph database to allow for efficient querying and analysis.

2.3 Data Mining and Response Generation

After obtaining the BKG, the second LLM agent, also based on GPT-4o, would take another look at the user query and select one or more data mining approaches from a predefined library that can be helpful for answering the query. While an exhaustive list of data mining techniques is not feasible here, some important ones include:

- **Topics and keywords:** Applying topic modeling [Grootendorst, 2022] to uncover latent research themes and their temporal evolution in the BKG; Traverse keyword co-occurrence networks [Miao *et al.*, 2022] to identify interdisciplinary research topics.

- **Papers and citations:** Exploiting the topology of paper citation network to identify or predict impactful works [Weis and Jacobson, 2021] and recommend similar papers [Boyack and Klavans, 2010].

- **Authors and research groups:** Exploring author collaboration networks to identify active researchers and groups with network metrics [Palla *et al.*, 2007].

- **Scientific discovery:** Learning hidden patterns embedded in the BKG to infer meaningful research ideas

---
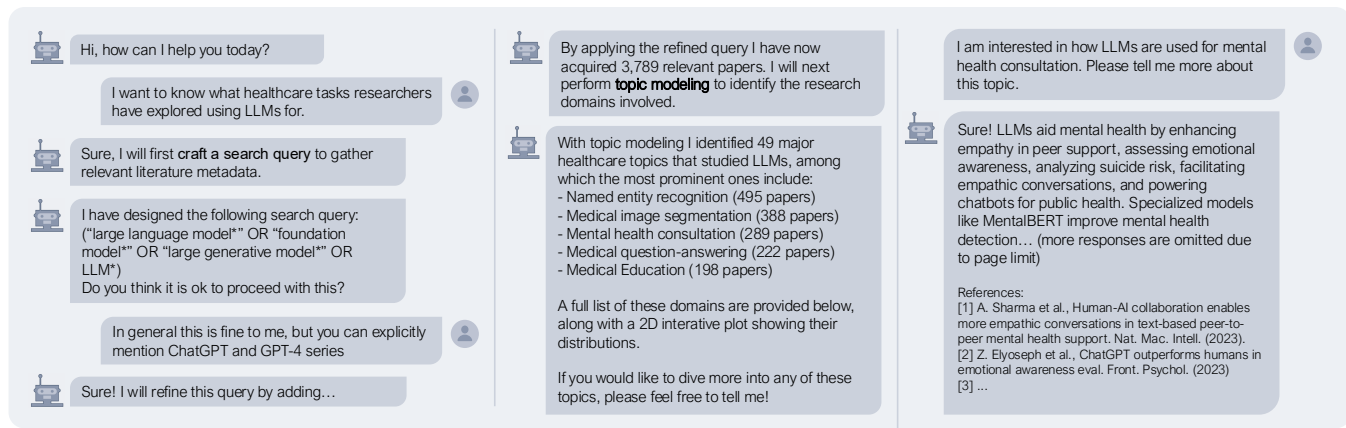[1]https://blog.voyageai.com/2025/01/07/voyage-3-large/

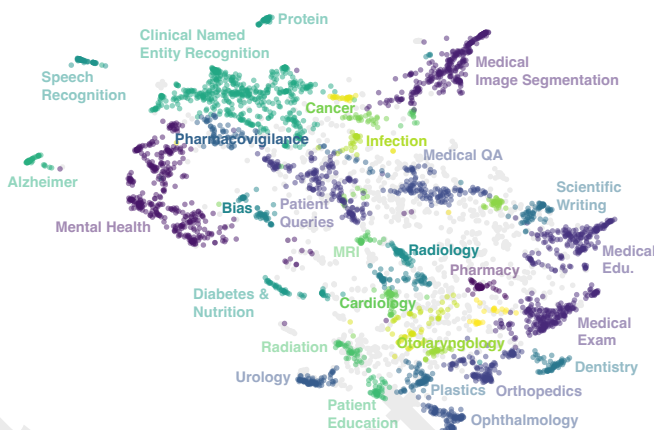Figure 2: An example of LitChat in action.

Figure 3: 2D visualization of the research landscape of LLM applications in healthcare provided by LitChat. Each point represents a paper, colored by the research topic as identified by topic modeling. Light gray points indicate outliers that are not assigned to any topic.

that transcend existing individual knowledge and cross-domain boundaries [Krenn *et al.*, 2023; Bai *et al.*, 2024].

After LitChat executes a tool, it composes the results of the operations into a natural language response that it returns to the user. LitChat generates these responses by filling in templates associated with each operation based on the results. The responses also include sufficient context to understand the results and opportunities for following up. In addition to these advanced data-driven tools, LitChat also supports basic tasks like question-answering and summarization based on RAG techniques just as many other literature chatbots.

3 Case Study on AI4Health

To demonstrate the effectiveness of LitChat, here we present a case study on the AI4Health domain. Specifically, as shown in Figure 2 on page 3, the user first asks "I want to know what healthcare tasks researchers have explored using LLMs for?" As one can imagine, answering this requires a comprehensive audit of the relevant literature, which is beyond the capacity of human experts. To address this, LitChat first creates a candidate search query and engages with the user to refine it to allow for a comprehensive coverage of relevant literature. The final query is then sent to the Web of Science database, from which 3,700+ papers are retrieved and synthesized into a BKG. Given the large volume of papers, traditional LLM-based systems would struggle to process them all, and this is where LitChat's data-driven tools come into play.

Since the user is interested in an overview of certain research topics, LitChat chooses topic_modeling to analyze the BKG. This function clusters the retrieved papers into different topics based on their semantic proximity (e.g., in the upper left region are topics related to brain and recognition, e.g., "Alzheimer", "Mental Health", and "Speech Recognition"). LitChat then generates response by highlighting the most prominent topics, and also provides a 2D visualization of the literature landscape (see Figure 3 on page 3). Note that in the response, LitChat is able to explicitly indicate the size of each topic as determined by the topic model, which conforms to the "objective" and "transparent" principles of SLR.

When the user asks for more details on a specific topic (here "mental health"), LitChat can use the topic tool to recommend representative papers, and generate topic summaries based on them. This conversation can continue to drill down into the details of one specific paper in this topic, historical trends of the topic, or the authors involved, etc., and each response would also be grounded on data-based evidence.

4 Conclusion

In this paper we introduced LitChat, an interactive LLM agent that is augmented with data-driven discovery tools to aid the exploration of large-scale literature landscapes. With a case study on AI4Health, we demonstrated how LitChat can help users to quickly navigate the terrain of LLM applications in healthcare research in an end-to-end fashion, which is otherwise infeasible with traditional manual reviewing or pure LLM-based systems. We hope that LitChat opens up new possibilities for scholars and practitioners to understand scientific literature in this era of literature explosion.

Acknowledgements

Contribution Statement

Mingyu Huang and Shasha Zhou contributed equally to this work. Mingyu Huang and Shasha Zhou conceptualized the work, designed the experiments and wrote the manuscript. Mingyu Huang developed the KG construction and bibliometric analysis tools. Shasha Zhou integrated LLMs for analysis and developed the web interface. Yuxuan Chen assisted in constructing the keyword network as part of the bibliographic KG. Ke Li supervised the whole project.

References

[Bai *et al.*, 2024] Jiaru Bai, Sebastian Mosbach, Connor J. Taylor, Dogancan Karan, Kok Foong Lee, Simon D. Rihm, Jethro Akroyd, Alexei A. Lapkin, and Markus Kraft. A dynamic knowledge graph approach to distributed self-driving laboratories. *Nat. Commun.*, 15(1):462, 2024.

[Boyack and Klavans, 2010] Kevin W. Boyack and Richard Klavans. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *J. Assoc. Inf. Sci. Technol.*, 61(12):2389–2404, 2010.

[Gao *et al.*, 2023] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *CoRR*, abs/2312.10997, 2023.

[Grootendorst, 2022] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based TF-IDF procedure. *CoRR*, abs/2203.05794, 2022.

[Huang and Li, 2024] Mingyu Huang and Ke Li. A survey of decomposition-based evolutionary multi-objective optimization: Part II - A data science perspective. *CoRR*, abs/2404.14228, 2024.

[Jeong *et al.*, 2023] Jongkil Jay Jeong, Yevhen Zolotavkin, and Robin Doss. Examining the current status and emerging trends in continuous authentication technologies through citation network analysis. *ACM Comput. Surv.*, 55(6):122:1–122:31, 2023.

[Ji *et al.*, 2023] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Yejin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM Comput. Surv.*, 55(12):248:1–248:38, 2023.

[Kitchenham *et al.*, 2009] Barbara A. Kitchenham, Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen G. Linkman. Systematic literature reviews in software engineering - A systematic literature review. *Inf. Softw. Technol.*, 51(1):7–15, 2009.

[Krenn *et al.*, 2023] Mario Krenn, Lorenzo Buffoni, Bruno Coutinho, Sagi Eppel, Jacob Gates Foster, Andrew Gritsevskiy, Harlin Lee, Yichao Lu, João P. Moutinho, Nima Sanjabi, Rishi Sonthalia, Ngoc Mai Tran, Francisco Valente, Yangxinyu Xie, Rose Yu, and Michael Kopp. Forecasting the future of artificial intelligence with machine learning-based link prediction in an exponentially growing knowledge network. *Nat. Mac. Intell.*, 5(11):1326–1335, 2023.

[Miao *et al.*, 2022] Lili Miao, Dakota Murray, Woo-Sung Jung, Vincent Larivière, Cassidy R. Sugimoto, and Yong-Yeol Ahn. The latent structure of global scientific development. *Nat. Hum. Behav.*, 6(9):1206–1217, 2022.

[Palla *et al.*, 2007] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, 2007.

[Schmallenbach *et al.*, 2024] Leo Schmallenbach, Till W. Bärnighausen, and Marc J. Lerchenmueller. The global geography of artificial intelligence in life science research. *Nat. Commun.*, 15(1):7527, 2024.

[Skarlinski *et al.*, 2024] Michael D. Skarlinski, Sam Cox, Jon M. Laurent, James D. Braza, Michaela M. Hinks, Michael J. Hammerling, Manvitha Ponnapati, Samuel G. Rodriques, and Andrew D. White. Language agents achieve superhuman synthesis of scientific knowledge. *CoRR*, abs/2409.13740, 2024.

[Wang *et al.*, 2023] Shuai Wang, Harrisen Scells, Bevan Koopman, and Guido Zuccon. Can chatgpt write a good boolean query for systematic review literature search? In *SIGIR'23: Proc. of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1426–1436. ACM, 2023.

[Wang *et al.*, 2024] Yidong Wang, Qi Guo, Wenjin Yao, Hongbo Zhang, Xin Zhang, Zhen Wu, Meishan Zhang, Xinyu Dai, Min zhang, Qingsong Wen, Wei Ye, Shikun Zhang, and Yue Zhang. Autosurvey: Large language models can automatically write surveys. In *NeurIPS'25: Proc. of the 38th Annual Conference on Neural Information Processing Systems*, 2024.

[Weis and Jacobson, 2021] James W. Weis and Joseph M. Jacobson. Learning on knowledge graph dynamics provides an early warning of impactful research. *Nat. Biotechnol.*, 39(10):1300–1307, 2021.