Synthesizing Scientific Literature with LLMs

Rachel So
rachel.so@4open.science

2025

1 Introduction/Background

The exponential growth of scientific literature has created significant challenges for researchers attempting to stay current with developments in their fields. Scientific discoveries often require synthesizing decades of research, a task that increasingly exceeds human information processing capabilities [22]. As the volume of scholarly articles continues to expand rapidly, conducting thorough literature analysis becomes increasingly time-consuming and difficult [7].

Large language models present a promising solution to these challenges. These AI systems, which derive their knowledge from statistical patterns, semantic relationships, and syntactical structures of language, have demonstrated impressive capabilities in various complex tasks including creative writing, translation, summarization, and code generation [42]. More recently, they have shown potential for advanced applications in scientific domains, particularly in performing scientific synthesis, inference, and explanation [42].

The role of LLMs in scientific research has become increasingly significant, with models like Llama, Gemini, and GPT-4 displaying profound capabilities in natural language understanding and generation [6]. Within scientific literature analysis specifically, applications such as literature summarization and knowledge extraction have seen practical deployments, enhancing researchers' productivity and broadening the scope of literature that can be synthesized and utilized [6].

By training on vast scientific literature, LLMs offer the potential to integrate noisy yet interrelated findings and potentially forecast novel results that might otherwise remain undiscovered [22]. This capability allows researchers to build upon the collective knowledge of their field more efficiently, addressing the fundamental challenge of information overload in modern scientific research [7].

Large language models (LLMs) have emerged as powerful tools for scientific literature synthesis, offering a solution to the challenge of processing vast amounts of research. These AI systems can potentially integrate findings across decades of scientific work, enhancing researchers' ability to analyze and build upon existing knowledge.

2 Challenges in Scientific Literature Analysis

The rapid expansion of scientific literature has created fundamental challenges for researchers attempting to analyze and build upon existing knowledge. This exponential growth makes it increasingly difficult for scientists to stay current with developments in their fields, turning in-depth literature analysis into a time-consuming and often overwhelming task [7] [28]. As research output continues to accelerate across disciplines, synthesizing knowledge has become a significant bottleneck in the scientific process.

A core challenge in scientific literature analysis is understanding the complex relationships between ideas presented in different papers. Research findings often have nuanced connections – they may be complementary, conflicting, or duplicative – and capturing these relationships requires sophisticated understanding of domain-specific concepts and methodologies [41]. Without explicit guidance, even advanced AI systems struggle to identify how different methods relate to specific tasks or how approaches might contrast with each other when applied to similar problems.

Current approaches to literature analysis using LLMs face several limitations. While these models can process individual papers effectively, they frequently fail to capture detailed relationships across large bodies of work [8]. This limitation becomes particularly pronounced when dealing with millions of potentially relevant facts and relationships, where unstructured approaches quickly become computationally prohibitive and financially impractical.

Another significant challenge is that existing LLM-based literature management approaches often overlook the rich structural and semantic relevance among scientific papers. This oversight limits their ability to accurately discern relationships between pieces of scientific knowledge and makes them susceptible to various types of hallucinations when attempting to synthesize information [40]. These limitations highlight the need for more sophisticated approaches that can better model the interconnected nature of scientific knowledge.

The biomedical domain represents a particularly challenging area for scientific literature analysis, given its highly specialized terminology and complex concepts. This has prompted the development of domain-specific language models like BioBERT [17] and BioGPT [21], which are specifically trained on biomedical corpora to better understand the nuanced language and relationships in this field [28].

The exponential growth of scientific literature has created significant barriers to comprehensive knowledge synthesis, with researchers struggling to navigate complex relationships between ideas and findings. LLMs face specific limitations in scientific literature analysis, including difficulty capturing nuanced relationships between papers and generating accurate structured representations of scientific knowledge.

3 Capabilities of LLMs for Scientific Literature Synthesis

Large language models have evolved into powerful tools for scientific literature synthesis, demonstrating advanced capabilities in processing and analyzing research publications. These AI systems derive their knowledge from statistical patterns, semantic relationships, and syntactical structures of language, enabling them to perform complex tasks including scientific synthesis, inference, and explanation [42]. Recent advances in models like Llama, Gemini, and GPT-4 have further expanded these capabilities, particularly in scientific literature analysis where applications such as literature summarization and knowledge extraction have seen practical deployments [6].

A key strength of LLMs in scientific literature synthesis is their ability to follow a range of instructions when given research articles as input. These models can extract information, summarize content, and answer questions about scientific papers, even when dealing with long input contexts such as entire research articles. Furthermore, they can structure their responses according to specific formats or schemas that support literature review aggregation or integration with augmented reading interfaces [30] [23].

LLMs have shown particular promise in multi-document summarization tasks, allowing researchers to synthesize information across multiple scientific abstracts or papers [11]. This capability addresses the challenge of information overload created by the exponential growth of scientific publications, helping researchers navigate large literature bodies more efficiently [29]. For example, models like Long T5 have proven effective for summarizing multiple documents in literature reviews, while BERT has been tailored specifically for summarizing scientific texts [3] [13] [39].

Another significant capability is hypothesis generation. LLMs trained on extensive datasets of scientific literature can recognize patterns and synthesize information across disciplines to propose novel scientific hypotheses [34] [33]. By leveraging their advanced natural language processing capabilities, these models can process and synthesize vast amounts of text to accelerate the analysis of scientific literature and generate new hypotheses for potential scientific discovery [38]. This ability to generate promising research ideas directly from existing literature has the potential to significantly streamline the traditionally time-consuming process of scientific research.

Domain-specific LLMs have further enhanced these capabilities for specialized scientific fields. Models like BioBERT and PubMedBERT focus on biomedical literature, while SciBERT covers a broader range of scientific disciplines [28] [17]. More recent models such as BioGPT have pushed the boundaries of scientific language modeling by incorporating advanced architectures and training techniques [21]. These specialized models excel at literature information retrieval, summarization, and question-answering within their respective domains, enabling more efficient navigation of scientific knowledge.

LLMs have also demonstrated capabilities in topic modeling for scientific literature. Approaches like PromptTopic leverage LLMs to extract topics at the sentence level from individual documents, then aggregate and condense these topics to provide coherent overviews of texts [3] [32]. Similarly, BERTopic has been integrated with models like Llama 2 for literature mining, allowing researchers to uncover intriguing and novel insights within broad research areas [14].

Through these diverse capabilities, LLMs are transforming how researchers interact with scientific literature, offering powerful tools for synthesizing knowledge across vast amounts of research and potentially accelerating scientific discovery.

> Large language models have demonstrated impressive capabilities in scientific literature synthesis, including summarization, knowledge extraction, hypothesis generation, and relationship identification across research papers. These models, especially domain-specific variants trained on scientific corpora, can process vast amounts of literature to help researchers navigate information overload and potentially generate novel scientific insights.

4 Applications and Systems

Several innovative systems have been developed to apply LLMs to scientific literature synthesis, each addressing specific challenges in research workflows:

SciDaSynth: A novel interactive system that enables researchers to efficiently build structured knowledge bases from scientific literature at scale. The system automatically creates data tables to organize and summarize knowledge through question-answering, while providing multi-level exploration capabilities that facilitate validation, correction, and refinement of generated content [35].

PyZoBot: An AI-driven platform that integrates Zotero's reference management capabilities with OpenAI's LLMs to streamline knowledge extraction and synthesis from extensive human-curated scientific literature databases [2].

SciQAG: A framework for automatically generating high-quality science question-answer pairs from large corpora of scientific literature using LLMs, supporting educational and research applications [31].

Step-by-Step Literature Survey Generation: A novel approach that leverages LLMs in a sequential prompting manner to generate scientific literature surveys. Given subjects and reference papers, the system guides LLMs to sequentially generate titles, abstracts, hierarchical headings, and main content, improving coherence while reducing API usage costs [16].

Scideator: A mixed-initiative tool for scientific ideation that extracts key facets (purposes, mechanisms, and evaluations) from user-provided papers and relevant literature. Users can interactively recombine these facets to synthesize inventive research ideas [26].

LLMs4Synthesis: A comprehensive framework designed to enhance open-source LLMs for generating scientific syntheses of comparable quality to those produced by larger proprietary models. The framework aims to enable automated and accurate integration of key insights from multiple scientific articles [12].

CORE-GPT: A question-answering platform that combines GPT-based language models with over 32 million full-text open access scientific articles. The system delivers evidence-based answers to questions with citations and links to referenced papers, increasing trustworthiness and reducing hallucinations [12] [25].

ByteScience: A non-profit cloud-based auto-finetuned LLM platform designed to extract structured scientific data and synthesize new scientific knowledge from vast scientific corpora [37].

Two-Step Literature Review Generation: An approach that outlines a plan for literature reviews and then executes steps in that plan to generate the actual review. Empirical studies indicate that these intermediate plans improve the quality of LLM-generated reviews compared to direct generation methods [1].

ARIA: A system that emulates a team of expert assistants to systematically replicate the human research workflow. ARIA autonomously searches, retrieves, and filters hundreds of papers, then synthesizes relevant literature into actionable research procedures [27].

These systems demonstrate how LLMs can be integrated into specialized applications to support different aspects of scientific literature synthesis, from initial discovery to knowledge extraction and integration.

Researchers have developed numerous specialized systems that leverage LLMs for scientific literature synthesis, including platforms for knowledge base creation, question generation, literature survey automation, and scientific ideation. These applications combine LLMs with structured approaches to assist researchers in extracting insights, generating research questions, automating literature reviews, and enhancing scientific discovery processes.

5 Approaches to Improving LLM Performance for Scientific Literature

The effectiveness of large language models for scientific literature synthesis depends significantly on specialized training and architectural improvements tailored to scientific domains. Several approaches have emerged to enhance LLM performance in this context, addressing their limitations when processing complex research materials.

Domain-specific pre-training represents a fundamental approach to improving LLM capabilities for scientific literature. Models like SciBERT demonstrated early success in this area by pre-training on large multi-domain corpora of scientific publications, addressing the challenge of limited high-quality labeled scientific data [4]. Building on this foundation, newer efforts like SciLit-LLM have further advanced domain adaptation by using continual pre-training on extensive academic texts, incorporating over 10 million academic papers while carefully balancing scientific and general knowledge to maintain broad capabilities [36].

Knowledge injection through carefully curated scientific corpora has emerged as another effective approach. Research suggests that LLMs benefit from training materials with qualities similar to exemplary textbooks: clarity, self-containment, instructiveness, and balance. Following this principle, some researchers have compiled substantial collections of high-quality scientific textbooks (over 73,000) and research papers (625,000) for targeted training, ensuring these materials meet copyright requirements [19].

The development of specialized frameworks represents a third approach to enhancing scientific synthesis capabilities. The LLMs4Synthesis framework, for example, aims to improve open-source LLMs to generate scientific syntheses comparable to those produced by larger proprietary models [12]. Such frameworks often build upon established language models like BERT [10] and integrate techniques for scientific content synthesis identified in previous research [25].

Researchers have also focused on improving structural understanding components that enhance scientific comprehension. One notable hypothesis suggests that understanding tables can significantly enhance LLM performance on scientific literature tasks by providing a more holistic understanding of research papers [15]. This approach recognizes the importance of non-textual elements in scientific communication and their role in conveying complex information.

To address the context window limitations of LLMs when processing lengthy scientific articles, retrieval-augmented generation (RAG) has emerged as a promising solution. This approach combines retrieval-based methods with generative models, enabling LLMs to access and process information beyond their limited context windows [9]. By first identifying relevant sections from a large corpus using vector search techniques and then feeding these sections into the LLM, RAG frameworks can handle larger volumes of text than traditional approaches [18].

Finally, structured knowledge representations offer a complementary approach to unstructured methods. When dealing with scientific corpora where millions of facts may influence an answer, unstructured approaches like standard RAG become computationally prohibitive and financially impractical. Structured representations enable more systematic analysis across entire corpora, capturing detailed relationships that might otherwise be missed [8].

Through these diverse approaches—domain-specific training, knowledge injection, specialized frameworks, structural understanding improvements, retrieval augmentation, and structured representations—researchers continue to enhance LLM performance for scientific literature synthesis, addressing the unique challenges presented by complex scientific content.

Researchers have developed multiple approaches to enhance LLM performance for scientific literature synthesis, including domain-specific training, knowledge injection, and structural improvements. Key strategies involve continual pre-training on scientific corpora, integration of table understanding capabilities, retrieval-augmented generation, and structured knowledge representations to overcome context limitations.

6 Integration with Data-Driven Research

The integration of LLMs with data-driven research approaches represents a powerful frontier in scientific discovery. While LLMs trained on vast scientific literature can potentially integrate interrelated findings to forecast novel results, their true potential emerges when combined with empirical data analysis [22]. This integration creates a synergistic relationship between literature-based insights and empirical observations, enhancing the overall research process.

A promising approach in this domain involves the combination of literature-based and data-driven hypothesis generation. Researchers have proposed frameworks where a literature-based hypothesis agent collaborates with a data-driven hypothesis agent to maintain and refine a shared pool of hypotheses [20]. This integrated approach ensures that generated hypotheses benefit from both data-driven adaptability and the grounding of existing scientific knowledge, creating a more robust foundation for scientific inquiry.

The ability of LLMs to generate scientific hypotheses represents a particularly valuable contribution to data-driven research. These models, trained on extensive scientific literature datasets, can recognize patterns and synthesize information across disciplines, proposing novel hypotheses that might not be immediately apparent to researchers [34] [33]. This capability can significantly streamline the traditionally time-consuming process of scientific research, accelerating the analysis of scientific literature and generating promising research ideas that can be tested through data-driven approaches [38].

Systems like SciDaSynth demonstrate how LLMs can bridge literature synthesis and data organization. This interactive system creates structured knowledge bases from scientific literature by automatically generating data tables that organize and summarize knowledge through question-answering approaches [35]. The multi-level exploration capabilities facilitate iterative validation and refinement, allowing researchers to effectively integrate insights from literature with their own data-driven investigations.

Platforms like ByteScience further exemplify this integration by providing cloud-based auto-finetuned LLM services designed specifically to extract structured scientific data and synthesize new scientific knowledge from vast scientific corpora [37]. These platforms enable researchers to move seamlessly between literature-based insights and structured data analysis, creating a more cohesive research workflow that leverages the strengths of both approaches.

The integration of LLMs with data-driven research not only accelerates hypothesis generation but also enhances the validation process. By grounding

LLM-generated hypotheses in empirical data, researchers can more effectively evaluate their plausibility and refine them through iterative testing. This creates a virtuous cycle where literature-informed hypotheses guide data collection and analysis, which in turn refines the understanding of the literature and suggests new avenues for exploration.

> Large language models are increasingly being integrated with data-driven research approaches to enhance scientific discovery processes. This integration combines the literature synthesis capabilities of LLMs with empirical data analysis to generate more robust hypotheses and accelerate knowledge creation in complex scientific domains.

7 Future Directions

As the field of LLM-based scientific literature synthesis continues to evolve, several important directions for future research are emerging. A particularly promising area is the development of more sophisticated approaches to capture the structural and semantic relationships between scientific papers and concepts. Current literature management approaches using LLMs often overlook these rich interconnections, limiting their ability to effectively discern relationships between pieces of scientific knowledge [40]. Addressing this limitation will be crucial for developing more effective scientific literature synthesis tools.

Another significant challenge that requires attention is the various types of hallucinations that LLMs produce when synthesizing scientific literature. These inaccuracies can significantly impact the reliability of generated insights and limit the practical utility of these systems in scientific research [40]. Future research will need to develop more robust methods for ensuring factual accuracy and proper attribution when generating syntheses of scientific literature.

The exponential growth of scientific literature presents a fundamental computational challenge that future systems will need to address. While unstructured approaches like retrieval augmented generation can identify relevant facts from large corpora, they become computationally prohibitive and financially impractical when millions of facts may influence an answer [8]. Structured representations offer a complementary approach that could enable more systematic analysis across entire corpora, suggesting that future systems may need to combine both structured and unstructured methods for optimal performance.

An emerging direction in this field involves developing specialized applications of LLMs beyond traditional literature analysis tasks. While applications like generating summaries, extracting insights, aiding in literature reviews, and enhancing knowledge integration have received significant attention, newer applications such as recommending research dimensions to obtain structured representations of research contributions represent an exciting frontier [24]. These novel applications could help researchers better understand the landscape of their fields and identify promising research directions.

The continued scaling of language models represents another important trend that will likely influence future developments in scientific literature synthesis. The dramatic improvements in task-agnostic, few-shot performance demonstrated by larger models like GPT-3 suggest that further scaling could yield additional benefits for scientific applications [24] [5]. However, these improvements will need to be balanced against the computational and environmental costs of training and deploying increasingly large models.

Future systems will likely also focus on enhanced integration with research workflows, moving beyond standalone literature analysis to support the entire scientific process from hypothesis generation to experimental design and result interpretation. This integration could significantly accelerate scientific discovery by providing researchers with more comprehensive support throughout their work .

As the field matures, addressing ethical considerations related to bias, attribution, and research integrity will become increasingly important. Ensuring that LLM-based literature synthesis tools promote fair and accurate representation of scientific contributions will be essential for their responsible deployment in research communities .

> Future research in LLM-based scientific literature synthesis will likely focus on developing more structured knowledge representations to improve relationship understanding between scientific concepts and reducing hallucinations. Addressing these challenges will require innovative approaches that combine semantic analysis with structured frameworks to better capture the complex interconnections in scientific knowledge.

Acknowledgements

Generative AI has been used to prepare this manuscript.

References

[1] Shubham Agarwal, Gaurav Sahu, Abhay Puri, I. Laradji, K. Dvijotham, Jason Stanley, Laurent Charlin, and Christopher Pal. LitLLMs, LLMs for Literature Review: Are we there yet?, 2024.

[2] S. Alshammari, Lama Basalelah, Walaa Abu Rukbah, Ali Alsuhibani, and D. Wijesinghe. PyZoBot: A Platform for Conversational Information Extraction and Synthesis from Curated Zotero Reference Libraries through Advanced Retrieval-Augmented Generation. *arXiv.org*, 2024.

[3] Ibrahim Al Azher, V. D. R. Seethi, Akhil Pandey Akella, and Hamed Alhoori. LimTopic: LLM-based Topic Modeling and Text Summarization for Analyzing Scientific Articles limitations. *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*, 2024.

[4] Iz Beltagy, Kyle Lo, and Arman Cohan. SciBERT: A Pretrained Language Model for Scientific Text. *Conference on Empirical Methods in Natural Language Processing*, 2019.

[5] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, J. Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, T. Henighan, R. Child, A. Ramesh, Daniel M. Ziegler, Jeff Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Ma teusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, I. Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. *Neural Information Processing Systems*, 2020.

[6] Hengxing Cai, Xiaochen Cai, Junhan Chang, Sihang Li, Lin Yao, Changxin Wang, Zhifeng Gao, Hongshuai Wang, Yongge Li, Mujie Lin, Shuwen Yang, Jiankun Wang, Yuqi Yin, Yaqi Li, Linfeng Zhang, and Guolin Ke. SciAssess: Benchmarking LLM Proficiency in Scientific Literature Analysis. *arXiv.org*, 2024.

[7] Hengxing Cai, Xiaochen Cai, Shuwen Yang, Jiankun Wang, Lin Yao, Zhifeng Gao, Junhan Chang, Sihang Li, Mingjun Xu, Changxin Wang, Hongshuai Wang, Yongge Li, Mujie Lin, Yaqi Li, Yuqi Yin, Linfeng Zhang, and Guolin Ke. Uni-SMART: Universal Science Multimodal Analysis and Research Transformer. *arXiv.org*, 2024.

[8] Abhipsha Das, Nicholas Lourie, Siavash Golkar, and Mariel Pettee. What's In Your Field? Mapping Scientific Research with Knowledge Graphs and Large Language Models, 2025.

[9] F. Delgado-Chaves, Matthew J. Jennings, Antonio Atalaia, Justus Wolff, R. Horváth, Zeinab M. Mamdouh, J. Baumbach, and L. Baumbach. Transforming literature screening: The emerging role of large language models in systematic reviews. *Proceedings of the National Academy of Sciences of the United States of America*, 2025.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *North American Chapter of the Association for Computational Linguistics*, 2019.

[11] Julia Evans, Jennifer D'Souza, and Sören Auer. Large Language Models as Evaluators for Scientific Synthesis. *Conference on Natural Language Processing*, 2024.

[12] Hamed Babaei Giglou, Jennifer D'Souza, and Sören Auer. LLMs4Synthesis: Leveraging Large Language Models for Scientific Synthesis. *Proceedings of the 24th ACM/IEEE Joint Conference on Digital Libraries*, 2024.

[13] Yash Gupta, Pawan Sasanka Ammanamanchi, Shikha Bordia, Arjun Manoharan, Deepak Mittal, Ramakanth Pasunuru, Manish Shrivastava, M. Singh, Mohit Bansal, and P. Jyothi. The Effect of Pretraining on Extractive Summarization for Scientific Documents. *SDP*, 2021.

[14] Boshko Koloski, Nada Lavrac, B. Cestnik, S. Pollak, Blaž Škrlj, and Andrej Kastrin. AHAM: Adapt, Help, Ask, Model - Harvesting LLMs for literature mining. *arXiv.org*, 2023.

[15] Buse Sibel Korkmaz and Antonio Del Rio Chanona. Integrating Table Representations into Large Language Models for Improved Scholarly Document Comprehension. *SDP*, 2024.

[16] Yuxuan Lai, Yupeng Wu, Yidan Wang, Wenpeng Hu, and Chen Zheng. Instruct Large Language Models to Generate Scientific Literature Survey Step by Step. *Natural Language Processing and Chinese Computing*, 2024.

[17] Jinhyuk Lee, WonJin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinform.*, 2019.

[18] Patrick Lewis, Ethan Perez, Aleksandara Piktus, F. Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Kuttler, M. Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Neural Information Processing Systems*, 2020.

[19] Sihang Li, Jian Huang, Jiaxi Zhuang, Yaorui Shi, Xiaochen Cai, Mingjun Xu, Xiang Wang, Linfeng Zhang, Guolin Ke, and Hengxing Cai. SciLitLLM: How to Adapt LLMs for Scientific Literature Understanding. *arXiv.org*, 2024.

[20] Haokun Liu, Yangqiaoyu Zhou, Mingxuan Li, Chenfei Yuan, and Chenhao Tan. Literature Meets Data: A Synergistic Approach to Hypothesis Generation. *arXiv.org*, 2024.

[21] Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. BioGPT: Generative Pre-trained Transformer for Biomedical Text Generation and Mining. *Briefings Bioinform.*, 2022.

[22] Xiaoliang Luo, Akilles Rechardt, Guangzhi Sun, Kevin K. Nejad, Felipe Y´a˜nez, Bati Yilmaz, Kangjoo Lee, Alexandra O. Cohen, Valentina Borghesani, Anton Pashkov, D. Marinazzo, Jonathan Nicholas, Alessandro Salatiello, Ilia Sucholutsky, Pasquale Minervini, Sepehr Razavi, Roberta Rocca, Elkhan Yusifov, Tereza Okalova, Nianlong Gu, Martin Ferianc, Mikail Khona, Kaustubh R. Patil, Pui-Shee Lee, Rui Mata, Nicholas E. Myers, Jennifer K Bizley, Sebastian Musslick, I. Bilgin, Guiomar Niso, Justin M. Ales, Michael Gaebler, Apurva Ratan Murty, Chloe M. Hall, Jessica Dafflon, Sherry Dongqi Bao, and Bradley C. Love. Large language

models surpass human experts in predicting neuroscience results. *Nature Human Behaviour*, 2024.

[23] I. Marshall and Byron C. Wallace. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic Reviews*, 2019.

[24] Vladyslav Nechakhin, Jennifer D'Souza, and Steffen Eger. Evaluating Large Language Models for Structured Science Summarization in the Open Research Knowledge Graph. *Inf.*, 2024.

[25] David Pride, Matteo Cancellieri, and Petr Knoth. CORE-GPT: Combining Open Access research and large language models for credible, trustworthy question answering. *International Conference on Theory and Practice of Digital Libraries*, 2023.

[26] Marissa Radensky, Simra Shahid, Raymond Fok, Pao Siangliulue, Tom Hope, and Daniel S. Weld. Scideator: Human-LLM Scientific Idea Generation Grounded in Research-Paper Facet Recombination. *arXiv.org*, 2024.

[27] Joaquin Ramirez-Medina, Mohammadmehdi Ataei, and Alidad Amirfazli. Accelerating Scientific Research Through a Multi-LLM Framework, 2025.

[28] Chandan K. Reddy and Parshin Shojaee. Towards Scientific Discovery with Generative AI: Progress, Opportunities, and Challenges. *arXiv.org*, 2024.

[29] Gautam Kishore Shahi and Oliver Hummel. On the Effectiveness of Large Language Models in Automating Categorization of Scientific Texts. *arXiv.org*, 2025.

[30] David Wadden, Kejian Shi, Jacob Daniel Morrison, Aakanksha Naik, Shruti Singh, Nitzan Barzilay, Kyle Lo, Tom Hope, Luca Soldaini, Shannon Zejiang Shen, Doug Downey, Hanna Hajishirzi, and Arman Cohan. SciRIFF: A Resource to Enhance Language Model Instruction-Following over Scientific Literature. *arXiv.org*, 2024.

[31] Yuwei Wan, Yixuan Liu, Aswathy Ajith, Clara Grazian, B. Hoex, Wenjie Zhang, Chunyu Kit, Tong Xie, and Ian Foster. SciQAG: A Framework for Auto-Generated Science Question Answering Dataset with Fine-grained Evaluation, 2024.

[32] Han Wang, Nirmalendu Prakash, N. Hoang, Ming Shan Hee, Usman Naseem, and Roy Ka-Wei Lee. Prompting Large Language Models for Topic Modeling. *BigData Congress [Services Society]*, 2023.

[33] Qingyun Wang, Doug Downey, Heng Ji, and Tom Hope. SciMON: Scientific Inspiration Machines Optimized for Novelty. *Annual Meeting of the Association for Computational Linguistics*, 2023.

[34] Wenxiao Wang, Lihui Gu, Liye Zhang, Yunxiang Luo, Yi Dai, Chen Shen, Liang Xie, Binbin Lin, Xiaofei He, and Jieping Ye. SciPIP: An LLM-based Scientific Paper Idea Proposer. *arXiv.org*, 2024.

[35] Xingbo Wang, S. Huey, Rui Sheng, Saurabh Mehta, and Fei Wang. SciDaSynth: Interactive Structured Knowledge Extraction and Synthesis from Scientific Literature with Large Language Model. *arXiv.org*, 2024.

[36] Dayong Wu, Jiaqi Li, Baoxin Wang, Honghong Zhao, Siyuan Xue, Yanjie Yang, Zhijun Chang, Rui Zhang, Li Qian, Bo Wang, Shijin Wang, Zhixiong Zhang, and Guoping Hu. SparkRA: A Retrieval-Augmented Knowledge Service System Based on Spark Large Language Model. *Conference on Empirical Methods in Natural Language Processing*, 2024.

[37] Tong Xie, Hanzhi Zhang, Shaozhou Wang, Yuwei Wan, Imran Razzak, Chunyu Kit, Wenjie Zhang, and B. Hoex. ByteScience: Bridging Unstructured Scientific Literature and Structured Data with Auto Fine-tuned Large Language Model in Token Granularity. *2024 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2024.

[38] Guangzhi Xiong, Eric Xie, Amir Hassan Shariatmadari, Sikun Guo, Stefan Bekiranov, and Aidong Zhang. Improving Scientific Hypothesis Generation with Knowledge Grounded Large Language Models. *arXiv.org*, 2024.

[39] Benjamin Yu. Evaluating Pre-Trained Language Models on Multi-Document Summarization for Literature Reviews. *SDP*, 2022.

[40] Jiasheng Zhang, Jialin Chen, Ali Maatouk, Ngoc Bui, Qianqian Xie, L. Tassiulas, Jie Shao, Hua Xu, and Rex Ying. LitFM: A Retrieval Augmented Structure-aware Foundation Model For Citation Graphs. *arXiv.org*, 2024.

[41] Zhi Zhang, Yan Liu, Sheng hua Zhong, Gong Chen, Yu Yang, and Jiannong Cao. Mixture of Knowledge Minigraph Agents for Literature Review Generation, 2024.

[42] Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, A. T. Nguyen, Lauren T. May, Geoffrey I. Webb, and Shirui Pan. Large Language Models for Scientific Synthesis, Inference and Explanation. *arXiv.org*, 2023.

Author Biography

Rachel So is an AI scientist. She focuses on the impact of artificial intelligence on the scientific process and academic publishing. Her work bridges traditional concerns about authorship ethics with emerging questions about the role of AI in knowledge production. Rachel aims to develop frameworks that maintain research integrity while acknowledging the growing presence of AI in academic workflows.