The Second Bridge on Artificial Intelligence for Scholarly Communication (AAAI-26)

Contributions

https://doi.org/10.52825/ocp.v8i.3172

Published: 18 Dec. 2025

Synergistic AI Agents

Integrating Knowledge Graphs and Large Language Models for Scholarly Communication

Bharath Chand[1], Sanju Tiwari[2,4], Nandana Mihindukulasooriya[3], and Sören Auer[4]

[1]Department of Futures Studies, University of Kerala, India

[2]Sharda University, Delhi-NCR, India

[3]IBM Research, New York, USA

[4]TIB Leibniz Information Centre for Science and Technology and L3S Research Center, Leibniz University of Hannover, Germany

Abstract. Agentic AI is and emerging field of artificial intelligence and it has great impact on scholarly research. Agentic AI helps to handle large volume of information from vast corpora. Currently the Agentic AI systems depends on Large Language Models (LLM) for the tasks of information retrieval and reasoning. LLMs are very effective at Natural Language Understanding and the iterative reasoning. However, there exist some inherent limitations for LLMs, which pose challenges for Agentic AI. Provenance tracking, reasoning challenges, temporal staleness and context dilution are some examples. Incorporating Knowledge Graphs (KG) along with LLMs can mitigate these challenges, and can support deeps search in Agentic AI.

In this work, we are exploring the aspects of how KG is well suited for addressing these challenges, and how KG can complement LLMs in Agentic AI for scholarly research. Furthermore, we investigate the problem of frequency bias inherent in LLMs. Frequency bias distorts the outputs in LLMs by biasing towards the most frequent inputs. We examine how a KG integration can counteract this problem. Overall, through this work we aim to highlight the potential of Knowledge Graphs for Agentic AI in scholarly communication.

Keywords: Agentic AI, Knowledge Graph, Scholarly Research

1. Introduction

Traditional query-response paradigm is changing with the emergance of Agentic systems by following an active exploratory workflows that execute various tasks on its own [1]. Beyond generating text content, LLMs in this context are mainly participating in the plan-act-reflect loop of agents by performing reasoning and taking decisions autonomously[2].

Agentic systems have transformed the traditional query-response paradigm into active exploratory workflows [1]. In this context, LLMs are not only generating text but also

participating in the plan-act-reflect loop of agents, which includes performing reasoning and taking decisions on their own (autonomy) [2]. This shift has profound implications in transforming scholarly research. The research practices such as querying, screening, reading, and summarizing, etc when performed by humans are highly time consuming and are limited by human capacity and attention. Agentic AI promises to automate these workflows by ingesting large amounts of documents, formulating hypotheses on the fly, and returning structured outputs quickly. However, scholarly research imposes a standard that far exceeds consumer-level applications [3]. It demands a higher degree of assurance regarding provenance, reproducibility, peer review, insight, and so on. Current agentic systems with LLMs face some challenges in meeting this demand because of their inherent limitations. Some of these problems are highlighted in [3]. In this work, we examine how some of these problems can be addressed using Knowledge Graphs as a complement to LLMs in agentic deep search. The problem of frequency bias is also examined with the solution provided by Knowledge Graphs.
KGs and language models can complement and improve each other in several different aspects [4], creating a synergistic interaction where structured symbolic knowledge enhances the generative and reasoning capabilities of LLMs, and vice-versa. There is a large volume of KGs in the scholarly data domain already built, such as Open Research Knowledge Graph (ORKG) [5], Scholarly Data [6], and Scholarly Wikidata [7]. When Knowledge Graphs need to be incorporated with LLM-based agentic systems, a multi-agent approach is followed [8], [9], where LLM-based agents communicate with KG-based agents to facilitate their collaboration. Luo et. al. [10] presented a tutorial to integrate Large Language Models (LLMs) with Knowledge Graphs (KGs) to advance Artificial General Intelligence (AGI). This tutorial has covered, how KGs can enhance LLMs and LLMs can improve KG construction and completion along with the challenges and future research directions. Yang et. al., [11] presented a descriptive survey of Knowledge Graph–enhanced Pre-trained Language Models (KGPLMs) by categorizing existing approaches such as during-training, before-training, and post-training enhancement method. Cai et. al. [12] has proposed a comprehensive survey to explore the integration of LLMs and KGs by categorizing in three categories: LLM-enhanced KGs, KG enhanced LLMs and collaborative KGs and LLMs.
In this work, we are not focusing on the implementation of multi-agent systems; instead, we outline which KG techniques are appropriate for addressing the LLM-agents' problems. It is summarized in Figure 1, and in the following sections, we discuss these aspects in detail.

The main contributions of this paper are:

- A study that characterizes the challenges of existing Agentic AI systems in scholarly communications, especially with respect to knowledge-related aspects.
- Mapping of Knowledge Graph techniques that can mitigate the inherent limits of LLM-based agents for scholarly communication.

The rest of the paper is organized as follows: Section 2 highlights the challenges in LLMs, Section 3 discusses possible solutions to these challenges, and finally, Section 4 concludes the paper.

2. Agentic AI with LLM: Challenges for Agents in Scholarly Communication

In academic research, factual accuracy is crucial for any finding to be considered credible in scholarship and to withstand peer review and analysis. The probabilistic

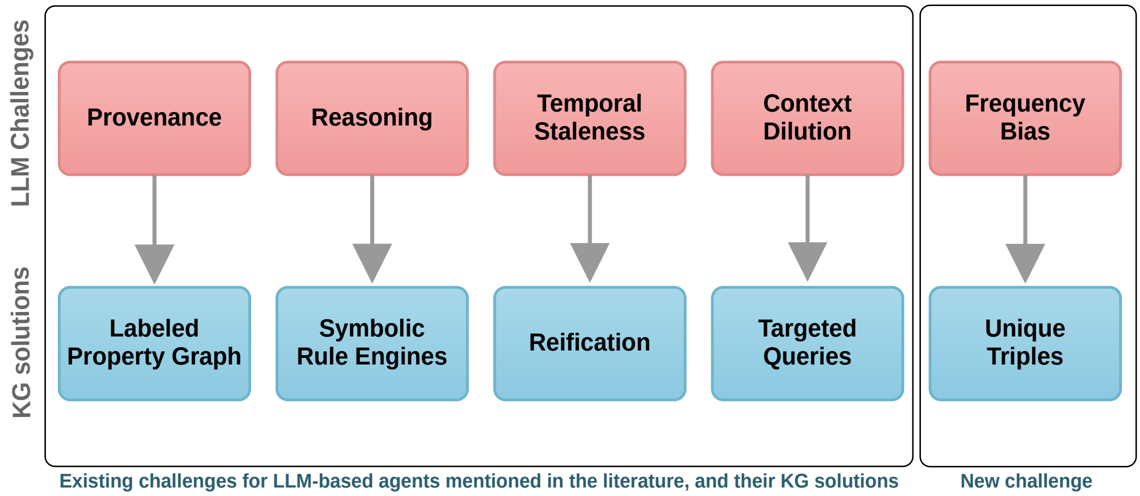

Figure 1. *The appropriate Knowledge Graph techniques that can be leveraged to apply to the identified problems associated with LLM-based agents.*

nature of LLMs, with a risk of hallucination, is fundamentally at odds with the demand for factual certainty. Standard practices, such as RAG, aim to enhance factual accuracy, but scholarly inquiry requires an even higher degree of assurance than that. An emerging extention of RAG - Graph RAG improves upon traditional RAG by supplying LLM agents with graph-structured context. However, in agentic systems, to track the provenance, an agent must maintain a trail that links LLM-generated statements to their exact source, using stable identifiers and specific reasoning steps to justify their inclusion.

Apart from the provenance problem, another challenge faced in agentic AI with LLM is associated with the Chain of Thought process. In agentic AI, LLMs are using a chain of thought process for reasoning, which is fundamentally sequential in nature. This constrains the ability of agents to perform reasoning with relational mapping or parallel execution. Modelling the reasoning process itself requires tree search or symbolic branching for causal relationships and dependencies.

Handling temporal details is also a challenge to be addressed, A statement or fact made during a particular period of time may not be true for another time. For example, The statement "Pluto is a Planet" is true for 2005. But "Pluto is a planet" is not true in 2007. It is a Dwarf Planet. Without structural frameworks to account for these temporal details, the answers can be wrong.

Another problem with LLM agents is Context Dilution. The performance may deteriorate when the context length of the content stretches beyond the limits. Scientific articles can contain systematic review appendices or complex figures, which may exceed the token limits of Language Models. In such cases, the ability to synthesize or deal with long documents is compromised.

The final challenge we have investigated is Frequency Bias, where hypotheses that appear more frequently lead to higher inference accuracy in LLM-based text generation [13], [14]. This leads to a bias towards the most frequently appearing predicates or concepts, resulting in the omission of critical facts that are rarely mentioned.

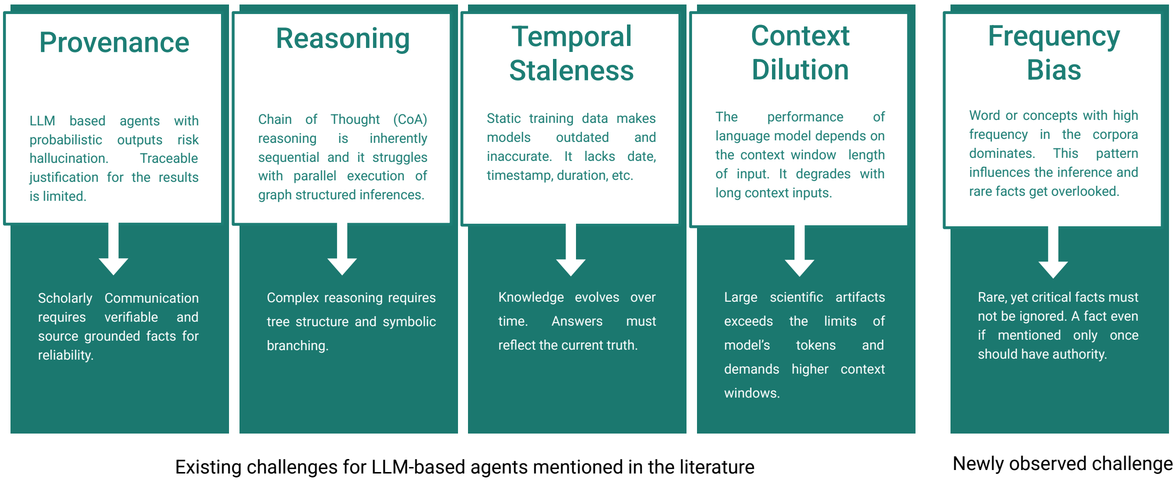

Figure 2. *The challenges associated with LLM-based Agentic AI for scholarly communication.*

Figure 2 summarizes these challenges we discussed above, and further In Section 3, we discuss the solutions of these challenges in detail. It points out which Knowledge Graph techniques are appropriate to mitigate these problems.

3. Knowledge Graph Infusion for Agentic AI

In this section, we examine how KG can complement LLMs for overcoming the limitations of Agentic AI in scholarly communication. Knowledge Graphs offers a wide variety of techniques that can be used to address these challenges and enhance performance of agents.

3.1 Provenance

Knowledge Graphs enhanced by labels for nodes and relationships are known as labeled Property Graphs (LPG). LPG holds these label properties in the form of key-value pairs and is a powerful Knowledge Graph with its rich and flexible structure. The research identifiers like DOI or URLS can also be represented as labels of LPG. When the agent works in either of its execution phases, it can trace this path from the graph. Adapting this framework will ensure that the agents are not hallucinating non existing content, but are actually giving outputs that are traceable within the graph.

Provenance from Knowledge Graphs can easily be facilitated through RDF and Ontologies, like PROV-O Ontology[15]. On the other hand, Labeled Property Graphs are used with its nodes and edges tagged with labels. For instance, Figure 3 shows an example of Labelled Property Graph. Here the node with label `:Author 1` have properties like `affiliation: "University X"`. And its relationship is labeled `:WROTE` to another node `:Paper`. The `:Paper` node is labelled with its `:title`. Relationships have label like, `year: 2022` also. This granularity makes LPGs well suited for modeling complex data by providing a machine-readable representation of knowledge, and to serve as the foundation of an agentic scholarly assistant.

When an Agentic AI interacts with the documents, its agents can create an audit trail by mapping the LPG nodes for the source of document. It can map sentences or data tables to citation relationships as well. This allows to trace back any final assertion

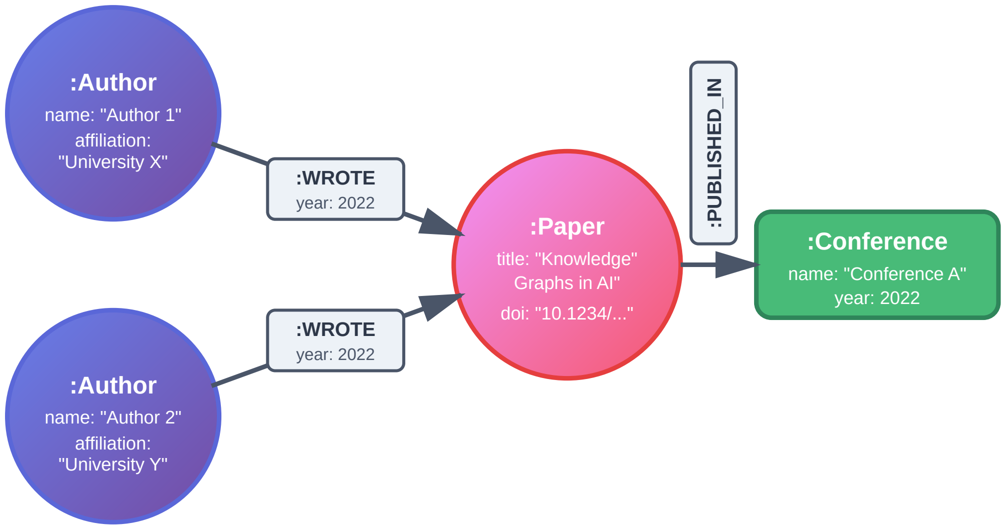

Figure 3. *Labeled Property Graph*

to its original source and serve as an evidence. This ensures scholarly admissibility and it mitigates the risks of hallucination and unvetted inputs.

3.2 Reasoning

Real world reasoning tasks like diagnosing systemic failures or planning with multiple interdependent variables are not linear, but are graphical in nature [16]. Symbolic AI uses tree searches to evaluate alternative hypotheses in parallel, and to understand the impact of a single node to the entire system, it maps out causal relationships and dependencies. Modelling these scenarios demands a structure that can explore branching possibilities. A linear Chain of Thought process might overlook it or simplify incorrectly. So, beyond the unstructured text, scientific agents leverage the graph structure of Knowledge Graphs to ensure the hypotheses are consistent with the known concepts [17].

When an agent queries or reasons over a KG for the next reasoning step, it is traversing a formally defined graph through the nodes and relations. This makes branching reasoning and tree exploration grounded in the Knowledge Graph structure. For example, suppose an agent needs to explore multiple hypotheses ("X causes Y", "X correlates with Y", "Z mediates X→Y"). In that case, a KG enables it to branch into each path through relationship links, rather than generating each hypothesis through text prediction. The causal relationships in graph edges naturally support dependency tracing, which also enables backtracking. However, the sequential Chain of Thought process is not suitable for backtracking.

Query languages like SPARQL, logical rule engines, and graph algorithms (such as shortest paths, community detection, and centrality measures) enable an agent to explore and verify relational knowledge programmatically. The agent can generate reasoning trees, prune invalid branches using KG constraints, and backtrack when contradictions appear. In this way, agents can integrate Symbolic AI and scientific reasoning workflows. With KG-assisted symbolic reasoning, agents can explore multiple reasoning paths (including multi-hop paths) simultaneously and backtrack when constraints are violated.

3.3 Temporal Staleness

The temporal aspect is essential for modelling dynamic domains in scientific discoveries, where the validity of information changes over time. Training or fine-tuning an LLM to incorporate every recent change is computationally expensive and slow. Knowledge Graphs can be used for storing temporal aspects, such as dates, durations, or timelines. By storing details like this separately from the LLM, only the KG needs to be updated to represent its corresponding time attribute when a fact changes, without needing to retrain or fine-tune the LLM.

Commonly, a Knowledge Graph stores information as triples (Subject-Predicate-Object). Because of this, standard KGs also have a limitation of appending timestamps beyond the triples [18]. To incorporate time, these triples can be extended with a temporal attribute with the technique of a "quad" or a "reified statement" especially on the KGs built on RDF standards, like the example shown in Figure 4. Here, Pluto was classified as a planet according to the fact in 2005. It is not true for 2007. In an RDF quad, both the triples are reified with their corresponding year. In this way, temporal aspects can be attributed in a manner decoupled from the core architecture of Large Language Models (LLMs). For more expressive details, it can incorporate temporal ontologies also [18].

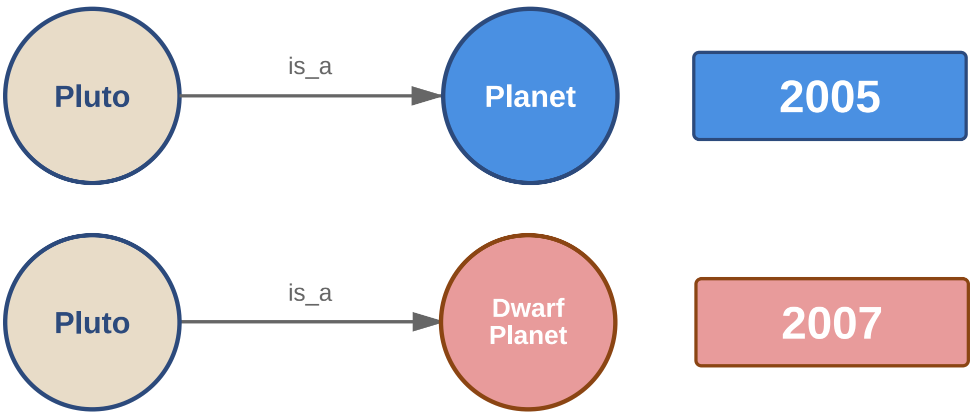

Figure 4. *A representation of quad reification in RDF with the year when a fact is valid.*

3.4 Context Dilution

Research artifacts are often very long and exceed the token limits of even the most advanced LLMs by means of their content. When these documents are processed in their entirety, it would require an excessively long sequence of tokens. As the input length increases, token-level representations compete for limited computational attention. It leads to a loss of focus on key contextual understanding. This is a process called context dilution. It reduces the model's ability to form meaningful insights or maintain coherence in long sections. This makes it particularly challenging for LLMs to handle literature that requires cross-referencing of results, interpreting detailed tables, or linking figures to corresponding text descriptions.

Instead of forcing the LLM to process the entire document text continuously, the KG can utilize the document's essential information as a structured network of nodes (entities and concepts) and edges (relationships between them). This allows the LLM-powered agent to reference this structured representation. By querying the KG for specific facts and relationships, the agent can reduce token usage and mitigate the degradation caused by context length. This approach enhances scalability by allowing selective referencing, such as querying specific nodes for relationships between variables, without reloading the entire content.

One another approach suggested by [19] to tackle this challenge in agentic systems is to employ a chain of agents (CoA). Yet, the performance can only be as good as the communication ability of agents and is affected by the latency between agents. Knowledge Graphs enable agents to selectively reference structured nodes and edges, rather than relying on the entire document each time, thereby improving scalability and reducing token-related degradation.

3.5 Frequency Bias Mitigation with Knowledge Graphs

As explained in the paper [14], the authors compute a frequency bias metric and they measure how frequently the verbal predicate in the hypothesis ("The ash contains iron") appears in a large corpus, compared to the predicate in the premise("The ash is rich with iron"). In the given example, the predicate "contains" is more common than "is rich with". The model is shown to rely on that frequency difference rather than any logical reason. From a general perspective, we can see that "the ash is rich with iron" is a more specific claim (richness) and "the ash contains iron" is a more general claim (containment).

As seen in the example mentioned above, many novel hypotheses or statements appear very rarely (perhaps only once) [20] in scholarly publications. Since the Knowledge Graphs are governed by a structured representation of relations (edges) and entities as unique triples, rather than text, agents can select based on structural attributes rather than how often particular words/concepts appear in a corpus. Thus, even the rarest predicates are preserved with full authority. The KG here acts as a structural anchor. Even a once-mentioned but highly authoritative fact remains accessible. Because the candidate query results stem from the KG, the agents can counter the frequency bias.

4. Conclusion

In this work, we discuss the applications of Knowledge Graphs in Agentic AI to complement LLMs. First, we identified four key limitations of LLM-based agents that align with the requirements for autonomous scientific assistance outlined by [3]. Additionally, we observed another issue related to frequency bias. We suggest Knowledge Graph as a complementary framework for LLM-based agents to mitigate these problems. As we have discussed, various Knowledge Graph techniques can be used to address the particular issues mentioned. Like, Labeled Property Graphs with the provenance tracking problem, Symbolic Rule Engines with the Reasoning Problem, Reification with Temporal Staleness, Targeted Queries with the Context Dilution, and Unique Triple representation of knowledge with the Frequency Bias, respectively. From the examples we discussed in this paper, it is clear that the agents can significantly benefit from the Knowledge Graph and its various features for improving performance, especially in scholarly communication.

Competing interests

The authors declare that they have no competing interests.

Author contributions

All authors have equal contribution.

References

[1] J. van den Akker and A. Siebes, "Enriching active databases with agent technology", in *International Workshop on Cooperative Information Agents*, Springer, 1997, pp. 116–125.

[2] J. Bailey, M. Georgeff, D. B. Kemp, D. Kinny, and K. Ramamohanarao, "Active databases and agent systems—a comparison", in *International Workshop on Rules in Database Systems*, Springer, 1995, pp. 342–356.

[3] M. Y. Jaradeh and S. Auer, "Deep research in the era of agentic ai: Requirements and limitations for scholarly research", in *CEUR Workshop Proceedings*, vol. 4065, 2025.

[4] H. Khorashadizadeh et al., "Research trends for the interplay between large language models and knowledge graphs", in *Proceedings of Workshops at the 50th International Conference on Very Large Data Bases, VLDB 2024, Guangzhou, China, August 26-30, 2024*, 2024.

[5] M. Y. Jaradeh et al., "Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge", in *Proceedings of the 10th international conference on knowledge capture*, 2019, pp. 243–246.

[6] A. G. Nuzzolese, A. L. Gentile, V. Presutti, and A. Gangemi, "Conference linked data: The scholarlydata project", in *International Semantic Web Conference*, Springer, 2016, pp. 150–158.

[7] N. Mihindukulasooriya, S. Tiwari, D. Dobriy, F. Å. Nielsen, T. R. Chhetri, and A. Polleres, "Scholarly wikidata: Population and exploration of conference data in wikidata using llms", in *International Conference on Knowledge Engineering and Knowledge Management*, Springer, 2024, pp. 243–259.

[8] X. Zhao et al., "Agentigraph: A multi-agent knowledge graph framework for interactive, domain-specific llm chatbots", *arXiv preprint arXiv:2508.02999*, 2025.

[9] M. Gridach, J. Nanavati, K. Z. E. Abidine, L. Mendes, and C. Mack, "Agentic ai for scientific discovery: A survey of progress, challenges, and future directions", *arXiv preprint arXiv:2503.08979*, 2025.

[10] L. Luo, C. Yang, E. Kharlamov, and S. Pan, "Integrating large language models and knowledge graphs for next-level agi", in *Companion Proceedings of the ACM on Web Conference 2025*, 2025, pp. 33–36.

[11] L. Yang, H. Chen, Z. Li, X. Ding, and X. Wu, "Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling", *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 7, pp. 3091–3110, 2024.

[12] L. Cai, Y. Kang, C. Yu, Y. Fu, H. Zhang, and Y. Zhao, "Bringing two worlds together: The convergence of large language models and knowledge graphs", in *2024 3rd International Conference on Automation, Robotics and Computer Engineering (ICARCE)*, IEEE, 2024, pp. 207–216.

[13] A. Amayuelas, J. Sain, S. Kaur, and C. Smiley, "Grounding llm reasoning with knowledge graphs", *arXiv preprint arXiv:2502.13247*, 2025.

[14] L. Cheng, Z. Wang, and M. Steedman, "Llms are frequency pattern learners in natural language inference", *arXiv preprint arXiv:2505.21011*, 2025.

[15] T. Lebo et al., "Prov-o: The prov ontology", 2013.

[16] Y. Yao, Z. Li, and H. Zhao, "Beyond chain-of-thought, effective graph-of-thought reasoning in language models", *arXiv preprint arXiv:2305.16582*, 2023.

[17] J. Wei et al., "From ai for science to agentic science: A survey on autonomous scientific discovery", *arXiv preprint arXiv:2508.14111*, 2025.

[18] F. Krause, T. Weller, and H. Paulheim, "On a generalized framework for time-aware knowledge graphs", in *Towards a Knowledge-Aware AI*, IOS Press, 2022, pp. 69–74.

[19] Y. Zhang, R. Sun, Y. Chen, T. Pfister, R. Zhang, and S. Arik, "Chain of agents: Large language models collaborating on long-context tasks", *Advances in Neural Information Processing Systems*, vol. 37, pp. 132 208–132 237, 2024.

[20] J. Wang, R. Veugelers, and P. Stephan, "Bias against novelty in science: A cautionary tale for users of bibliometric indicators", *Research Policy*, vol. 46, no. 8, pp. 1416–1436, 2017.