Towards AI-Supported Research: a Vision of the TIB AIssistant

Sören Auer[1,2], Allard Oelen[1], Mohamad Yaser Jaradeh[2,1], Mutahira Khalid[1], Farhana Keya[1], Sasi Kiran Gaddipati[1], Jennifer D'Souza[1], Lorenz Schlüter[3], Amirreza Alasti[3], Gollam Rabby[2], Azanzi Jiomekong[4,1] and Oliver Karras[1]

[1]*TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany*

[2]*L3S Research Center, Leibniz University of Hannover, Hannover, Germany*

[3]*Leibniz University of Hannover, Hannover, Germany*

[4]*University of Yaounde 1, Yaounde, Cameroon*

Abstract

The rapid advancements in Generative AI and Large Language Models promise to transform the way research is conducted, potentially offering unprecedented opportunities to augment scholarly workflows. However, effectively integrating AI into research remains a challenge due to varying domain requirements, limited AI literacy, the complexity of coordinating tools and agents, and the unclear accuracy of Generative AI in research. We present the vision of the TIB AIssistant, a domain-agnostic human-machine collaborative platform designed to support researchers across disciplines in scientific discovery, with AI assistants supporting tasks across the research life cycle. The platform offers modular components — including prompt and tool libraries, a shared data store, and a flexible orchestration framework — that collectively facilitate ideation, literature analysis, methodology development, data analysis, and scholarly writing. We describe the conceptual framework, system architecture, and implementation of an early prototype that demonstrates the feasibility and potential impact of our approach.

Keywords

AI-Supported Research, LLMs for Science, Scholarly AI Platform, Scholarly Research Assistant

1. Introduction

The developments of Generative AI, and specifically Large Language Models (LLMs), have had a significant impact in many areas of our society [1] already. Additionally, in the scientific domain, LLMs are increasingly utilized, for example, to assist researchers with academic writing [2]. LLMs are used across a wide variety of scholarly domains, such as medicine in life sciences [3], social sciences [4], chemistry [5], law [6], and coding in computer science [7].

While many approaches have been proposed and demonstrated, it remains challenging for researchers to get started with LLMs in their field. The ability of individual researchers to optimally leverage AI for their research strongly depends on their AI literacy, i.e., their ability to evaluate, communicate with, and collaborate using AI [8]. AI can be used to support domain-independent tasks, such as finding related work, assisting with paper authoring, and proofreading, as well as for domain-specific tasks, including supporting methodologies, implementations, or evaluations. While the possibilities are virtually unlimited, the benefits that a single researcher gains from AI-assisted research heavily depend on the user and the specifics of her research work, which determine the required prompts and the tasks that can be outsourced to the LLM. Prompt engineering is a crucial aspect for effective LLM
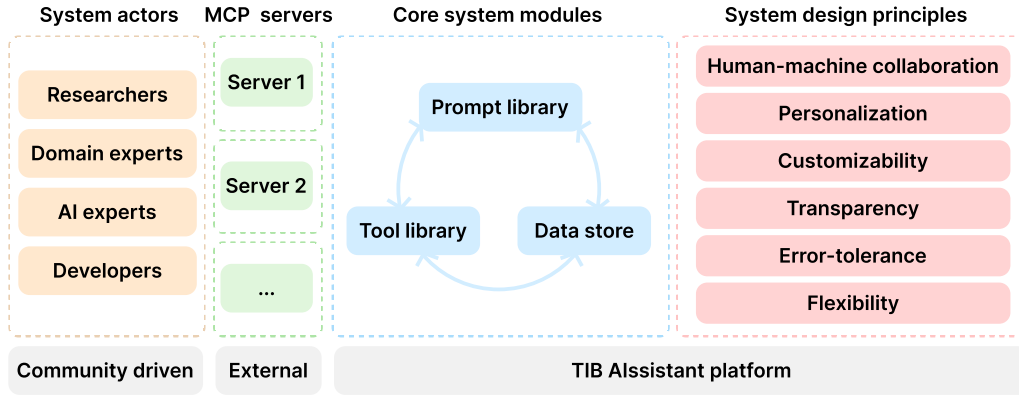
Figure 1: Proposed framework of AI-assisted research, highlighting system actors, external MCP (Model-Context Protocol) servers consisting of collections of tools, core system modules, and the system's design principles.

usage [9], but can be a bottleneck for non-AI experts [10]. Even if researchers possess the necessary skills to operate LLMs, research work relies on diverse tools designed to support specific tasks, such as data analysis in computing environments like R Studio or digital libraries that support knowledge discovery. The effective integration of Generative AI with the diverse tools used in research remains a challenge. For example, the integration of an LLM-based assistant with digital libraries can provide additional context via Retrieval-Augmented Generation [11] or by calling external services using Tool Callings [12].

Based on these considerations, we identified the following challenges for AI-assisted research:

- **Challenge 1**: lacking domain-specific AI literacy to leverage AI for research tasks effectively.
- **Challenge 2**: the skill to effectively engineer prompts and context injection.
- **Challenge 3**: leveraging existing tools and services into AI workflows and providing appropriate user interfaces for them.
- **Challenge 4**: technical capability to organize and orchestrate different AI agents to accomplish a single task.

In this work, we present our vision for an AI-supported, domain-agnostic research platform, named TIB AIssistant. Figure 1 depicts the conceptual framework of our approach. The platform serves as a central repository for scholarly AI agents and their corresponding prompts. Additionally, the platform provides the scholarly tools necessary to accomplish research tasks. To serve researchers across various domains, we argue that the flexibility of the system is crucial for accommodating the diverse requirements and use cases arising from diverse research work. Figure 2 illustrates an example research life cycle within the TIB AIssistant. The user begins with using the TIB AIssistant to generate ideas, proceeds through the different phases of the life cycle (supported by various agents), and utilizes external tools as needed. The data store stores results from different agents and makes them available to other agents.

2. Vision of the TIB AIssistant

Our vision for the TIB AIssistant is to empower researchers through an AI-supported, human-centered, and domain-agnostic platform that redefines the conduct of scholarly research. We envision a collaborative research environment where humans and machines co-create knowledge. Rather than aiming for full automation, the TIB AIssistant centers on *human-machine collaboration*, enabling researchers to retain control, orchestrate processes, and critically evaluate AI-generated results throughout the research life cycle.

At the core of this vision is a *flexible, modular, and transparent infrastructure* that facilitates AI integration without imposing rigid workflows. The TIB AIssistant is conceived as a lightweight yet

powerful research hub where customizable AI agents, scholarly tools, and curated prompts work together seamlessly. Each element of the system — ranging from prompt libraries to external tool integrations — is designed to be interoperable, extensible, and openly accessible.

We aim to lower the barrier to AI adoption in academia by addressing the four key challenges faced by researchers: i) understanding the scope of AI in domain-specific tasks, ii) developing effective prompts and contextual inputs, iii) integrating external scholarly tools into AI workflows, iv) and coordinating diverse AI agents to perform complex research processes.

Inspired by principles from Integrated Development Environment (IDE) interfaces, the TIB AIssistant provides a research-friendly environment where users can initiate ideation and literature exploration, formulate research questions, iterate on methodologies, analyze and synthesize results, and ultimately author and refine scholarly publications. This vision is grounded in a set of foundational design principles: *Personalization, Customizability, Trustworthiness and Transparency, Error-tolerance*, and *Open science and community engagement*. Ultimately, the TIB AIssistant aspires to become a central AI hub for scholarly research, not just a tool, but an evolving community-driven platform that transforms how research is conceptualized, conducted, and communicated in the age of generative AI.

3. Related Work

The use of LLMs to support scholarly activities is a growing field. Existing approaches can be broadly categorized into single task assistance and those that offer a more integrated, multi-task framework.

3.1. Single Task Assistants

A significant body of work focuses on leveraging LLMs to streamline specific, often labor-intensive, components of the research process. For instance, the STORM approach [13] provides a systematic approach to the research and pre-writing stages, which are the core of a literature review. Similarly, ResearchAgent [14] is an LLM-powered system focused on research idea generation by defining problems, proposing methods, and designing experiments. The system utilizes collaborative LLM-powered reviewing agents to refine these ideas iteratively based on feedback. Another specialized application of LLM agents is simulating complex scholarly interactions. The AgentReview framework [15], for example, utilizes LLM-based agents to simulate the entire peer-review process, allowing for the study of its dynamics, including reviewer bias and the influence of author identity. Also, there are domain-specific tools such as *Name2SMILES* (for converting molecule names to SMILES), *ReactionPlanner* (for multi-step synthesis planning), *PatentCheck* (for checking compound patent status), and *SafetySummary* (for retrieving safety information) from the ChemCrow platform [5], for research-related tasks in chemistry.

3.2. Multi-Task Assistants

Moving beyond single-task applications, a growing number of initiatives aim to create more comprehensive, multi-faceted research assistants. These systems often integrate several capabilities to support researchers throughout their workflow. Paper Copilot [16], for example, functions as a personalized academic assistant that maintains a real-time updated database of research papers. It can derive a user's research profile, analyze the latest trending topics, and provide advisory services, thereby combining multiple support functions into a single system. Furthermore, there is The AI Scientist [17], a framework designed for fully automated, open-ended scientific discovery. This system represents a significant step towards end-to-end automation by autonomously performing a sequence of research tasks. Starting with ideation based on existing literature, followed by experimentation, and finally, the paper authoring. This results in a complete manuscript in LaTeX, which LLM agents also review.

While these approaches demonstrate the potential of full automation, their rigid, pipeline-driven nature highlights several challenges that our vision for the TIB AIssistant aims to address. The emphasis on a fully autonomous process limits the role of the researcher, contrasting with our core principle of human-machine collaboration, where the human expert orchestrates and validates each step. The fixed
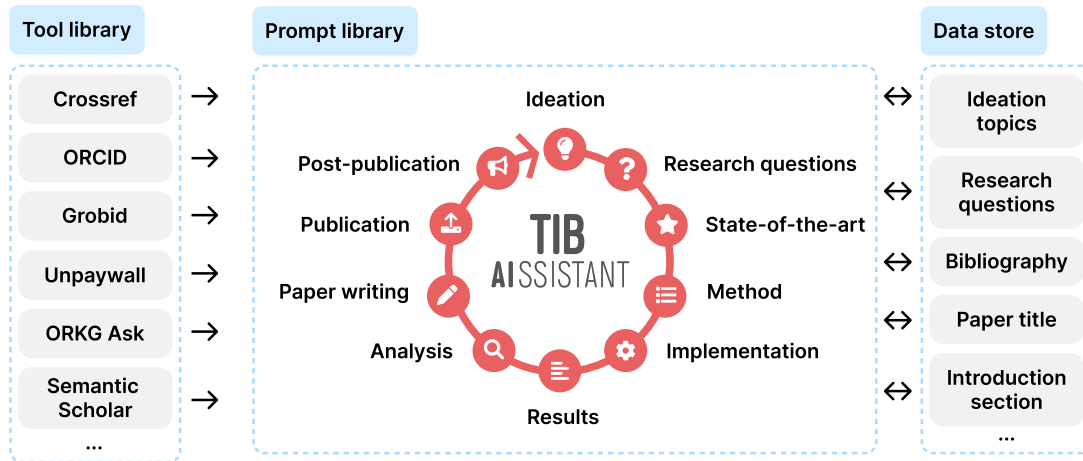
Figure 2: Example research life cycle starting from ideation until post-publication. The listed items provide concrete examples of the three core system modules, as shown in Figure 1. External tools are listed, along with items stored within the data store.

workflow within the AI assistant also fall short of our goal for a customizable and flexible platform, where users can modify prompts, select different LLMs, and integrate their tools across various disciplines.

4. Framework for AI-Assisted Research

We now describe the main components that form the foundational framework of our envisioned approach toward AI-supported research. The framework's concepts are discussed next and summarized in Figure 1. The platform provides an integrated environment for researchers, much like an IDE for software developers. It consists of a Graphical User Interface (GUI) allowing users to interact with various agents. We consider this platform as a lightweight wrapper that integrates the different components listed below. If existing approaches or tools are available, the platform should implement these services instead of attempting to replicate their functionalities.

4.1. Core System Modules

Prompt Library The Prompt Library is a collection of system prompts tailored toward specific tasks of the research life cycle. A list of prompts minimizes the need for researchers to create their own prompts, often relying on time-consuming trial-and-error. The prompt library enables researchers to share their approaches with others easily. In addition to the prompt, metadata must be assigned to the prompt to indicate what task is addressed (e.g., research question formulation, finding related literature, etc.). Multiple variants of prompts can exist for the same task, thus providing alternatives in case a prompt does not produce the expected result. Finally, users should be able to see and modify prompts when using them within the platform.

Tool Library The Tool Library integrates external services into the platform. This makes it possible to connect the platform to external tools, for example, to fetch additional publication data from Crossref and ORCID, or to fetch related work via Semantic Scholar [18], ORKG [19], or ORKG Ask [20]. Tools are called automatically, where the LLM decides, based on the description of the tool and the user's input, whether a tool should be called or not. To ensure tools can be added dynamically, a Model Context Protocol (MCP) [21] can be used. MCP provides a standardized approach to provide external access to LLMs. In this case, we are specifically interested in enabling calling tools. This enables users to integrate existing scholarly MCP servers, allowing them to access a range of scholarly tools quickly and easily. For developers, it is possible to set up an MCP server to make their tools available to the

platform. To facilitate external tool integration via MCP servers, we plan to develop an MCP GUI tool, enabling the easy setup of an MCP server for tools that utilize REST endpoints.

Data Store   The ability of different agents to communicate with each other can be accomplished via a centralized data store. Compared to keeping all generated content in the context of the LLM, this approach has several benefits: the constraint of context window size is less problematic since the data is stored in a separate store, and the agents are self-contained. They can be used in isolation, making it easier to reason about what is happening within the agent. The data store can be a relational database, storing specific data under a predefined key (e.g., research questions, bibliography, etc.). When necessary, the database can be accessed, and the respective content is added to the context. A more advanced approach could provide the LLM with a tool that allows it to access the database, enabling it to read from and write to the database automatically.

4.2. System Design Principles

Human-Machine Collaboration   In the spectrum between full automation of research and humans doing most of the required tasks, a hybrid approach where humans and machines collaborate offers the best of both worlds. In such a hybrid approach, the human researcher primarily orchestrates, directs, and reviews AI-supported processes. In this model, researchers have control at all times and review and evaluate the output created by the AI at each step. This means that user interfaces must be designed to go beyond the conventional conversational prompt-response style, allowing users to modify intermediary results and decide when to proceed to the next step. However, due to the conversational setup, processes such as iterative refinement enable users to refine AI-generated responses through a human-in-the-loop approach further. We argue that this type of control is essential for an AI-supported research assistant to be both useful and adopted by researchers. Therefore, the interface should not aim to provide an automated research pipeline, but instead offer a highly customizable environment that researchers can use to integrate AI support into their workflows.

Personalization   A sophisticated memory system is crucial for enhancing the platform's effectiveness through deep personalization. This feature enables dynamic context engineering [22], a process where only the most relevant information for a given task is selectively retrieved or fetched and then provided to the LLMs, optimizing both relevance and computational efficiency. Beyond concrete tasks, the memory system constructs a user profile that learns about domain expertise, stylistic preferences, formatting conventions, and preferred terminology over time. Moreover, the memory component maintains a comprehensive record of the user's research history. This historical knowledge would empower the assistant to guide and coordinate various (sub-)agents, ensuring their outputs are consistent and aligned with the overarching user preferences.

Customizability   The platform should be domain-agnostic and sufficiently flexible to support different workflows. Users should be able to customize the platform to support their use cases. This begins with the prompts, where users can try out different variants and edit them as needed. Additionally, the list of tools that the LLM can execute during a chat session should be modifiable to limit the scope of available tools and better direct the LLM in selecting the appropriate tool. As previously mentioned, the platform should serve as a lightweight tool connecting different services. The ability to customize the interface is therefore crucial to support a variety of use cases.

Transparency and Trustworthiness   For transparency and reproducibility reasons, for all generated artifacts, provenance data has to be recorded, capturing the creators, model name, and model version, system and user prompts, invoked tools, etc. To provide complete transparency, this provenance data should be published alongside the research paper as research data. We envision this data to be published in a machine-readable format, for example, via RO-Crates [23], to facilitate machine actionability. These aspects will also help users verify the accuracy and originality of the generated content.

Error-Tolerance   With an error-tolerant interface, we ensure that users remain in control and can modify any data generated by the AI. This means that all messages, including system and user messages, as well as the generated data, should be modifiable by the user. Additionally, the user should be able to see which data is provided as input when the LLM calls a tool. This helps determine whether the tool was called as expected. Based on our previously mentioned aspect of Human-Machine Collaboration, the user can decide which data to use and which to discard. Regarding external tool callings, since we do not control these services ourselves, we should assume that these services may respond differently than expected (e.g., because of a temporary issue, rate limiting, or changes in API specifications). In such cases, the AIssistant should be error tolerant by gracefully handling such errors.

Flexibility   We envision a centrally hosted platform. Users of the platform should have the flexibility to choose the models and LLM providers they prefer. This also serves the purpose of selecting providers based on geographical locations, legal requirements, and privacy concerns, among other factors. Smaller models can handle simple tasks, while larger models can execute more complex tasks. A limited number of free tokens can be provided to each user per day, which can be managed through an authentication system. A Bring Your Own Key (BYOK) approach can be used to allow users to use as many tokens as necessary. As an alternative, users should also be able to run the platform locally on their computers, as the source code is publicly available.

Community Driven   We aim to develop the platform in collaboration with academics from various domains. As there are virtually unlimited use cases for AI-supported research, the platform relies on community contributions to create prompts and add tools. An additional way to contribute is by adding functionalities to the platform itself.

5. Conclusion and Outlook

In this work, we laid the foundation for an AI-supported research platform. To address challenge 1, we propose a library of prompts that showcases to researchers the types of tasks that can be performed by LLMs and the tools required for these tasks. Challenge 2 is addressed by reducing the need to engineer prompts and by automatically providing the required context for a prompt through the data store. Challenge 3 is tackled by integrating external scholarly tools via tool calls and external MCP servers. Finally, challenge 4 is addressed by providing predefined research life cycle workflows, where different agents can interact with each other by means of sharing data via the data store.

A key aspect of our approach is the community effort to create and curate prompts and tools in such a manner that they are helpful for other users in the community. We aim to provide the technical means to make this possible. This includes the addition of community features, such as voting for helpful prompts and sharing custom workflows with other users. In the end, we envision that the prompt library will contain different versions of prompts aiming to accomplish the same research task. This follows the assumption that there is no one-size-fits-all approach, but that different prompt variants are helpful for different use cases.

An initial prototype is implemented, integrating the proposed framework concepts into a workable research life cycle. A demonstration of this approach is published [24]. Figure 2 depicts parts of the prototype implementation. Only domain-agnostic assistants are implemented in the prototype (i.e., ideation, research questions, state-of-the-art, paper writing). Outputs from, for example, ideation, are utilized by other assistants, such as when formulating research questions and during paper writing. Furthermore, the tools and data store items listed in the figure are also implemented. The prototype implementation shows the feasibility of our approach and demonstrates how the core system modules are integrated. The source code is available online.[1] We plan to further develop the prototype into a publicly available online service, where researchers can get started with AI-assisted research.

---

[1]https://gitlab.com/TIBHannover/orkg/tib-aissistant/web-app

Acknowledgments

Declaration on Generative AI

During the preparation of this work, the authors utilized ChatGPT, Gemini, and Grammarly to draft content, enhance content, paraphrase and reword, improve writing style, and Perform Grammar and spelling checks. After using this service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. A. Haque, S. Li, Exploring chatgpt and its impact on society, AI and Ethics 5 (2025) 791–803. doi:10.1007/s43681-024-00435-4.

[2] W. Liang, Y. Zhang, Z. Wu, H. Lepp, W. Ji, X. Zhao, H. Cao, S. Liu, S. He, Z. Huang, et al., Mapping the increasing use of llms in scientific papers, arXiv preprint arXiv:2404.01268 (2024). doi:10.48550/arXiv.2404.01268.

[3] A. J. Thirunavukarasu, D. S. J. Ting, K. Elangovan, L. Gutierrez, T. F. Tan, D. S. W. Ting, Large language models in medicine, Nature medicine 29 (2023) 1930–1940. doi:10.1038/s41591-023-02448-8.

[4] I. Grossmann, M. Feinberg, D. C. Parker, N. A. Christakis, P. E. Tetlock, W. A. Cunningham, Ai and the transformation of social science research, Science 380 (2023) 1108–1109. doi:10.1126/science.adi1778.

[5] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, Augmenting large language models with chemistry tools, Nature Machine Intelligence 6 (2024) 525–535. doi:10.1038/s42256-024-00832-8.

[6] M. Siino, M. Falco, D. Croce, P. Rosso, Exploring llms applications in law: A literature review on current legal nlp approaches, IEEE Access (2025). doi:10.1109/ACCESS.2025.3533217.

[7] M. Kazemitabaar, R. Ye, X. Wang, A. Z. Henley, P. Denny, M. Craig, T. Grossman, Codeaid: Evaluating a classroom deployment of an llm-based programming assistant that balances student and educator needs, in: Proceedings of the 2024 chi conference on human factors in computing systems, 2024, pp. 1–20. doi:10.1145/3613904.3642773.

[8] D. Long, B. Magerko, What is ai literacy? competencies and design considerations, in: Proceedings of the 2020 CHI conference on human factors in computing systems, 2020, pp. 1–16. doi:10.1145/3313831.3376727.

[9] N. Knoth, A. Tolzin, A. Janson, J. M. Leimeister, Ai literacy and its implications for prompt engineering strategies, Computers and Education: Artificial Intelligence 6 (2024) 100225. doi:10.1016/j.caeai.2024.100225.

[10] J. D. Zamfirescu-Pereira, R. Y. Wong, B. Hartmann, Q. Yang, Why johnny can't prompt: how non-ai experts try (and fail) to design llm prompts, in: Proceedings of the 2023 CHI conference on human factors in computing systems, 2023, pp. 1–21. doi:10.1145/3544548.3581388.

[11] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W.-t. Yih, T. Rocktäschel, et al., Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in neural information processing systems 33 (2020) 9459–9474.

[12] T. Schick, J. Dwivedi-Yu, R. Dessi, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, T. Scialom, Toolformer: Language models can teach themselves to use tools, in: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), Advances in Neural Information Processing Systems, volume 36, Curran Associates, Inc., 2023, pp. 68539–68551.

[13] Y. Shao, Y. Jiang, T. A. Kanell, P. Xu, O. Khattab, M. S. Lam, Assisting in writing wikipedia-like articles from scratch with large language models, 2024. doi:`https://doi.org/10.48550/arXiv.2402.14207`.

[14] J. Baek, S. K. Jauhar, S. Cucerzan, S. J. Hwang, Researchagent: Iterative research idea generation over scientific literature with large language models, 2025. doi:`10.48550/arXiv.2404.07738`.

[15] Y. Jin, Q. Zhao, Y. Wang, H. Chen, K. Zhu, Y. Xiao, J. Wang, Agentreview: Exploring peer review dynamics with llm agents, 2024. doi:`10.48550/arXiv.2406.12708`.

[16] G. Lin, T. Feng, P. Han, G. Liu, J. You, Paper copilot: A self-evolving and efficient llm system for personalized academic assistance, 2024. doi:`10.48550/arXiv.2409.04593`.

[17] C. Lu, C. Lu, R. T. Lange, J. Foerster, J. Clune, D. Ha, The ai scientist: Towards fully automated open-ended scientific discovery, 2024. doi:`10.48550/arXiv.2408.06292`.

[18] R. Kinney, C. Anastasiades, R. Authur, I. Beltagy, J. Bragg, A. Buraczynski, I. Cachola, S. Candra, Y. Chandrasekhar, A. Cohan, et al., The semantic scholar open data platform, 2025. doi:`10.48550/arXiv.2301.10140`.

[19] S. Auer, et al., Open Research Knowledge Graph: A Large-Scale Neuro-Symbolic Knowledge Organization System, in: Handbook on Neurosymbolic AI and Knowledge Graphs, IOS Press, 2025. URL: https://doi.org/10.3233/FAIA250216.

[20] A. Oelen, M. Y. Jaradeh, S. Auer, Orkg ask: A neuro-symbolic scholarly search and exploration system, arXiv preprint arXiv:2412.04977 (2024). doi:`10.48550/arXiv.2412.04977`.

[21] Introduction - Model Context Protocol — modelcontextprotocol.io, https://modelcontextprotocol.io/introduction, 2024. [Accessed 23-07-2025].

[22] L. Mei, J. Yao, Y. Ge, Y. Wang, B. Bi, Y. Cai, J. Liu, M. Li, Z.-Z. Li, D. Zhang, C. Zhou, J. Mao, T. Xia, J. Guo, S. Liu, A survey of context engineering for large language models, 2025. doi:`10.48550/arXiv.2507.13334`.

[23] S. Soiland-Reyes, P. Sefton, M. Crosas, L. J. Castro, F. Coppens, J. M. Fernández, D. Garijo, B. Grüning, M. L. Rosa, S. Leo, E. Carragáin, M. Portier, A. Trisovic, R.-C. Community, P. Groth, C. Goble, Packaging research artefacts with ro-crate, Data Science 5 (2022) 97–138. doi:`10.3233/DS-210053`.

[24] A. Oelen, S. Auer, Tib aissistant: a platform for ai-supported research across research life cycles, ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan (2025).