

A Case Study on Using Generative AI in Literature Reviews: Use Cases, Benefits, and Challenges

Eva-Maria Schön¹^a, Jessica Kollmorgen^{1,2}^b, Michael Neumann³^c and Maria Rauschenberger¹^d

¹University of Applied Sciences Emden/Leer, Emden, Germany

²University of Seville, Sevilla, Spain

³University of Applied Sciences and Arts Hannover, Hannover, Germany

Keywords: Generative AI, Literature Review, Use Cases, SLR, GenAI, AI Assistant, LLM, Large Language Model, Higher Education.

Abstract: *Context:* Literature reviews play a critical role in the research process. They are used not only to generate new insights but also to contextualize and justify one's own research within the existing body of knowledge. *Problem:* Since years, the number of scientific publications has been increasing rapidly. Therefore, conducting literature reviews can be time-consuming and error-prone. *Objective:* We investigate how integrating generative Artificial Intelligence (GenAI) tools may optimize the literature review process in terms of efficiency and methodological quality. *Method:* We conducted a single case study with 16 Master's students at a University of Applied Science in Germany. They all carried out a Systematic Literature Review (SLR) using generative AI tools. *Results:* Our study identified use cases for the application of GenAI in literature reviews, as well as benefits and challenges. *Conclusion:* The results reveal that GenAI is capable of supporting literature reviews, especially critical parts such as primary study selection. Participants can scan large volumes of literature in a short time and overcome language barriers using GenAI. At the same time, it is crucial to assess GenAI outputs and ensure adequate quality assurance throughout the research process due to technology limitations, such as hallucination.

1 INTRODUCTION

Literature reviews serve as the cornerstone of academic research and offer essential context and direction for new investigations. Their importance is increasingly recognized across disciplines, as they play a vital role in synthesizing existing knowledge and guiding future research. Literature reviews are conducted for a variety of reasons, *e.g.*, to summarize the current state of evidence on a specific treatment, technology, or topic; to identify gaps in the existing body of research; and to provide a well-informed foundation for positioning new research activities within the broader scientific landscape (Kitchenham and Charters, S., 2007; Pautasso, 2013). Literature reviews can be divided into two types: standalone work, such as review articles, or as a review needed as a background

or related work for empirical studies (Xiao and Watson, 2019).

In the past, access to academic publications was limited due to a relatively low volume of available literature and technological limitations in terms of digitization. However, in recent decades, the number of scientific publications has increased, fueled by advances in digital infrastructure and the growing global research output (Pautasso, 2013). As a result, researchers today are faced with an overwhelming volume of academic literature, making it increasingly difficult to identify and select the most relevant studies for their work. In response to the challenge of an increasing number of publications, structured and transparent methods for literature reviews, such as Systematic Literature Reviews (SLRs) (Kitchenham and Charters, S., 2007) or Rapid Reviews (Garritty et al., 2021) — have gained importance as essential tools for evidence-based research synthesis.

The emergence of generative artificial intelligence (GenAI) tools is expected to further accelerate the process of producing academic publications, thereby

^a <https://orcid.org/0000-0002-0410-9308>

^b <https://orcid.org/0000-0003-0649-3750>

^c <https://orcid.org/0000-0002-4220-9641>

^d <https://orcid.org/0000-0001-5722-576X>

amplifying the volume of available literature even further. GenAI refers to the use of generative models to produce previously unseen synthetic content in various forms, such as text, images, audio, or code, to support a wide range of tasks (García-Peñalvo and Vázquez-Ingelmo, 2023; Feuerriegel et al., 2024). Generative modeling techniques have been around for several years. However, the term “*Generative AI*” has only recently gained wide attention, following the emergence and public availability of user-friendly tools such as ChatGPT (OpenAI, 2025), which make this technology accessible to a wider audience. This trend suggests that the challenge of finding relevant studies will continue, if not increase, in the future. Consequently, the need for efficient, systematic, and goal-oriented literature review processes becomes even more critical.

This is also evident in the related work on literature reviews with GenAI. Tools such as ChatGPT, PDFgear, or Typeset have proven to be helpful for specific phases of literature reviews, such as the screening of data volumes and full texts (Castillo-Segura et al., 2024; Schryen et al., 2025). An overall increase in efficiency while maintaining a high level of accuracy has also been demonstrated (Ng and Chan, 2024; Felizardo et al., 2024). However, despite the ongoing development of the GenAI tools, there are still challenges and gaps that need to be examined and improved from a practical perspective (Ofori-Boateng et al., 2024; Schryen et al., 2025).

In this study, we investigate how the integration of GenAI tools can optimize the literature review process in terms of efficiency and methodological quality. Thus, this paper addresses two research questions to help deepen the understanding of using GenAI in literature reviews:

- **RQ 1:** Which use cases within the literature review process can be supported or automated by GenAI tools?
- **RQ 2:** What are the benefits and challenges of using GenAI tools during a literature review?

We conducted a single case study to answer our research questions. In our study, we supported master’s students in conducting a SLR with GenAI over a term at a University of Applied Science in Germany. Our paper makes three contributions to the body of knowledge:

1. We provide an overview of the use cases of how GenAI can speed up the process of an SLR, leaving more time to analyze relevant studies. In addition, we offer a selection of suitable GenAI tools for each use case as an overview.

2. We present benefits and challenges that were identified in our single case study when using GenAI tools and illustrate where human-in-the-loop is needed.
3. With this information, a lightweight approach to an SLR becomes feasible. This enables researchers to leverage our experience and conduct their next literature review in a more optimized way.

This paper is structured as follows: We provide an overview of the related work in Section 2. Next, we describe our research method in Section 3. We present our results of this single case study, including the answers to the research questions in Section 4 and discuss them in Section 5. Finally, we conclude our findings in Section 6, and give an outlook on future research.

2 RELATED WORK

With the ongoing development of GenAI tools, their use for various purposes is also being investigated. To gain an overview of the research field of (systematic) literature reviews with applied GenAI tools, we searched for literature closely related to our papers topic. The search strings are listed in the Appendix. For our search, we used four databases (SpringerLink, IEEEExplore, ACM, and ScienceDirect) with an activated year range filter since the release of ChatGPT in 2022. We argue for filtering the results since 2022, with the increased interest in both research and practice since the release of the tool by OpenAI. Since the release of ChatGPT, GenAI has become accessible to a much broader target group. Since then, many more people, not just experts, have had access to GenAI tools. To narrow down the results, we applied a thorough selection process based on pre-defined criteria. Below, we describe the five identified studies.

Ng & Chan (Ng and Chan, 2024) discussed in 2024 the potential of Artificial Intelligence (AI) in the screening phase of literature reviews. The results of the study indicated that the use of AI in SLRs reduced the amount of time and manual work required for the screening process. While the use of AI increased the efficiency and reduced the invested time from 128 human hours to less than 4 AI hours, also limitations were addressed, such as categorizing research in multidisciplinary contexts. The authors suggest a hybrid approach combining AI capabilities and human expertise to ensure accuracy in the literature screening.

Felizardo et al. (Felizardo et al., 2024) evaluated in 2024 the accuracy of ChatGPT-4.0 in the first selection stage of an SLR, meaning the screening of

title, abstract, and keywords, which includes classifying thousands of potentially relevant studies. Automating this stage could therefore save researchers time and effort. The SLRs of the study replicated with ChatGPT showed that an accuracy of about 75-86% could be achieved, which is why ChatGPT can be used as a supportive tool especially for inexperienced researchers, to reduce uncertainties. However, despite efficiency gains, critical errors (e.g., loss of relevant studies) can occur, and ChatGPT should rather act as an additional *opinion*. The authors recommended that human supplementation of the selection stage was essential to maintain the accuracy of SLRs.

Castillo-Segura et al. (Castillo-Segura et al., 2024) investigated in 2024 the use of GenAI in SLRs eligibility stage, the deep analysis of full texts. The authors compared different AI tools such as PDFgear, ChatPDF, and Typeset, which can enhance this AI full-text analysis, and analyzed the challenges of automation. Limitations such as a restriction on daily operations and varying costs were identified. The tools also had different levels of complexity, with some requiring basic programming skills to function efficiently. It also became clear in the study that further investigations into the use of GenAI in this SLR phase are necessary.

Ofori-Boateng et al. (Ofori-Boateng et al., 2024) provided in 2024 an overview of how different AI techniques, such as natural language processing, machine learning, or deep learning, can be used to automate different stages of SLRs. The study synthesized 52 studies and an online survey with systematic review practitioners to identify challenges in the automation. It was shown that the AI used for SLRs needs improvement, e.g., in the handling of diverse search queries and the eligibility stage.

Schryen et al. (Schryen et al., 2025) explored in 2025 the role of GenAI in SLRs by focusing on a human-centered approach where GenAI complements researchers rather than replacing them. The authors discussed challenges while using GenAI, such as the lack of generating novel insights and reproducing existing knowledge. They emphasized that ethical considerations must be taken into account and that the responsibility for the integrity of the work remains with the researchers.

In our article, existing research on using GenAI in literature reviews is complemented by a practice-oriented perspective. While the related work has already provided important insights into the potential, challenges, and specific phases of the SLR process with GenAI, the authors of the related work themselves emphasize the need for further practical testing and continuous development. Our study builds on the

need for further practical testing. Thus, we conducted our single case study and provide empirical insights from a real-world application context. This article contributes to the further development of a research field that is currently developing dynamically and is increasingly dependent on real-life application examples.

3 RESEARCH METHOD

The objective of this paper is to investigate how the integration of GenAI tools can optimize the literature review process in terms of efficiency and methodological quality. To this end, we applied a single case study to answer our research questions (see Section 1). Case studies are particularly suitable for investigating contemporary phenomena in their natural context, especially when the boundaries between the phenomenon and its context are not evident (Runeson and Höst, 2009; Yin, 2009).

An overview of our research method is presented in Fig. 1. We applied the conceptual structure for AI assistants in higher education (Schön et al., 2023b) to conduct our single case study in a course of the Master's program at the University of Applied Science Emden/Leer in Germany. On the one hand, we looked at the relationships between lecturers, students, GenAI tools, user behavior, and regulatory conditions such as exam formats. On the other hand, we followed the guidelines of the conceptual structure when designing the questionnaires (e.g., when asking for demographic data and previous knowledge). The single case study was conducted during the winter term 2024/2025 (September 2024 - January 2025).

3.1 Context of the Single Case Study

The public University of Applied Science Emden/Leer (Germany) has around 3,800 students and over 40 bachelors's and master's degree programs in maritime sciences, social work, and health, technology, and business. It is a founding member of the *Virtuelle Fachhochschule* (VFH), which provides different online degree programs in business administration, media informatics, and industrial engineering. The online degree programs use interactive, multimedia learning materials and state-of-the-art collaboration and communication media to implement contemporary learning scenarios on the Internet. Also, digital tools and individual learning are part of the teaching strategy.

The single case study examined involves the course Information Management (5 credit points) in

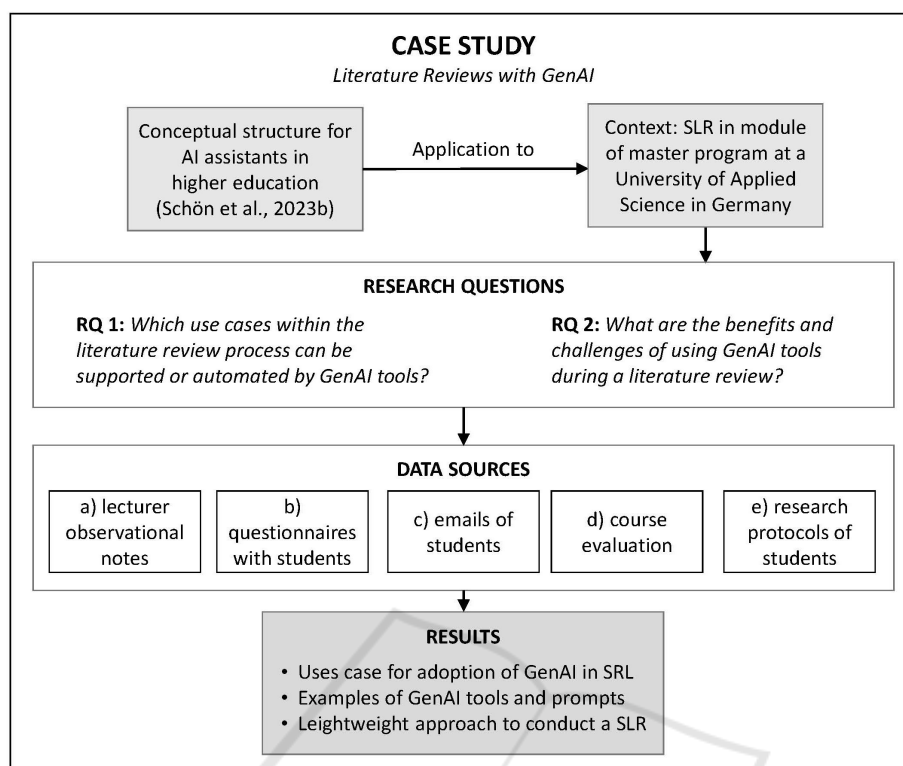


Figure 1: Research approach for the Single Case Study.

the master's degree program Business Management (M.A.) in the Faculty of Business Studies. The learning objectives of the course include classifying data, information, and knowledge, describing the different roles of information management, and evaluating models of information management. In the practical part of this course, students learn to analyze data using digital technologies and to extract and evaluate information from data by developing research questions and a research protocol in guided self-study, then carrying out their study and presenting the results in a scientific poster.

3.2 Didactic Method, Teaching Concepts, and Learning Assessment

The didactic method used for the course is a research-based learning approach. Students conduct a small-scale research project. This format enables students to acquire relevant content through guided self-study, fostering both subject expertise and methodological competence. The concept of value-based learning can thus be applied (Schön et al., 2023a). A central component of the process is familiarization with academic research methods, particularly systematic literature review techniques according to the guidelines

of (Kitchenham and Charters, S., 2007). Students began by formulating a research question, which guides the development of a structured research protocol. Subsequently, they conducted a comprehensive literature search using relevant academic databases. The retrieved literature is critically evaluated and synthesized to address the research question. The results of the analysis are then presented in a scientific format, enabling students to demonstrate their understanding of the topic, as well as their ability to conduct and communicate research in a structured and methodologically sound manner.

The assessment is designed to evaluate both the methodological rigor and the communication skills of students engaged in research-based learning. It consists of four components, each contributing to the final grade. The first component (20%) involves the development of a research protocol, in which the students define their research question, outline the methodological approach, and specify inclusion and exclusion criteria for the literature search. The second component (20%) requires students to complete a structured *Data Extraction Form*, ensuring a systematic and transparent synthesis of relevant information from the selected sources. The third component (40%) focuses on the presentation of research

findings in the form of a scientific poster. This format challenges students to communicate complex results in a concise, visually engaging, and academically sound manner. Finally, a pre-recorded presentation of the scientific poster (20%) provides an opportunity to further demonstrate understanding of the topic, the research process, and the ability to explain results clearly and professionally.

The guided self-study was supervised by the lecturer. In the weekly lectures, which were partly face-to-face and partly remote, work instructions were given with tips on how to carry out a SLR. These tips included information on how and in what way GenAI tools could be used. In addition, it was discussed in detail how GenAI may be used and what is prohibited. To this end, a traffic light system (Schön et al., 2025) was introduced for this purpose at the beginning of the term.

3.3 Sample

The single case study sample is made up of **one lecturer and 16 students**.

The **lecturer** is a professor of business informatics and holds a PhD in computer science. She has more than 7 years of teaching experience in higher education institutions and about 10 years of experience in conducting SLRs.

The group of 16 **students** can be described as follows: all students were enrolled in a master's program at the Department of Business Studies of the University of Applied Sciences Emden/Leer (14 in Master Business Management, 2 in Master Management Consulting). 7 of the students have already completed an apprenticeship. The students have Bachelor's degrees in very different disciplines (e.g., business administration, tourism management, social and health management, business psychology, or business informatics). In their own words, most of them describe their attitude towards technical innovations as open, curious, excited, and interested, but also as overwhelming. Student attitudes towards GenAI can be rated as positive with a mean value of 4.25 (*Likert scale with 5 items, (1) negative to (5) positive*). In addition, students rate their GenAI experience rather medium with a mean value of 3.06 (*Likert scale with 5 items, (1) beginner to (5) expert*). Before the single case study, they used GenAI tools like ChatGPT, Bing Chat, Copilot, Deepl, or Gemini.

The sample had prior experience with academic work through the completion of their Bachelor's thesis and therefore possessed foundational knowledge in handling scientific literature. Furthermore, participants demonstrated openness towards the use of new

technologies, making them suitable for participation in this single case study.

3.4 Data Collection and Analysis

We have used five different data sources, which are in line with Baxter (Baxter and Jack, 2008) and outlined in Figure 1. Below is a summary of the data sources. Further information can also be found in our research protocol (Schön et al., 2025).

a) Lecturer Observational Notes. At the end of every lecture, the lecturer reflected and documented on a Miro board what went well, where the students had problems, and where the GenAI tools had weaknesses or limitations. These observational notes are analyzed and incorporated into the results.

b) Questionnaires with Students. Several questionnaires were used throughout the term (Schön et al., 2025). All surveys were conducted online using Google Forms. Participation in these surveys was voluntary, leading to a very low response rate in some questionnaires.

- at the beginning of the term, we set up a questionnaire to gather information regarding the sample (e.g., socio-demographic data, attitude, and experience with GenAI)
- at the end of each lecture, we gathered a ROTI score (Return on Time Invested) and asked for the used GenAI tools
- at the end of the term, we conducted a closing survey to assess participants' attitudes and experiences with GenAI, including their evaluation of how helpful GenAI was in conducting an SLR, advantages, and disadvantages of using GenAI in SLR

c) Emails of Students. The students wrote several emails to the lecturer during the term, which included both questions and thanks regarding the proactive use of GenAI.

d) Course Evaluation. During the term, a teaching evaluation is performed centrally by an university office using a standardized questionnaire. The results are taken into account in this single case study as an additional data source.

e) Research Protocols of Students. As part of the assessment, students should submit a research protocol. An example (Schön et al., 2016) was provided to the students, which takes into account the guidelines by (Kitchenham and Charters, S., 2007). The students should add a table to the research protocol in which they document their use of GenAI and briefly describe how and for which activities the tools were used. An example of the table was made available

to the students and can be found here (Schön et al., 2025). The use of GenAI in higher education requires a change in exams (Neumann et al., 2023), and this table is an adaptation.

We triangulated qualitative and quantitative data for data analysis. The results of this single case study are presented in the next section.

4 RESULTS

This section presents the results of our single case study (see Fig. 2). First, we provide an overview of use cases that support the literature review process with GenAI. Second, we outline the benefits and challenges of adopting GenAI in literature reviews.

4.1 Overview of Use Cases

In this section, we answer our first research question *RQ1: Which use cases within the literature review process can be supported or automated by GenAI tools?* Fig. 2 depicts an overview of use cases in which the students used GenAI for their SLR.

The terms listed in the columns *Main phases of SLR* and *Stages of main phases* of Fig. 2 are based on the guidelines proposed by (Kitchenham and Charters, S., 2007), which were adopted by the students for the execution of the SLR within the context of the single case study. These guidelines provided a structured framework to guide students through the individual stages of the literature review process. The content in the columns *Use Cases* and *GenAI tools used by students* is derived from a qualitative synthesis of multiple data sources (see Fig. 1). This includes an analysis of course materials, such as task instructions for each exercise, which were developed by the lecturer in collaboration with the other senior researchers of this study. In addition, empirical data collected from students is used. The latter encompasses responses from a ROTI (Return on Time Invested) survey and elements of the research protocols submitted during the assessment of the course under investigation. Together, these data sources enabled a detailed mapping of GenAI-supported tasks across the SLR workflow, providing insights into how students utilized GenAI tools in different phases of the review process.

From the students' perspective, the use of GenAI proved to be particularly beneficial in two key areas. First, GenAI tools supported the translation of English-language academic papers, thereby significantly enhancing students' understanding of the original studies. This was especially valuable for those students with limited experience in reading scientific

texts in English. This finding is confirmed by multiple data sources, including lecturer observational notes, student questionnaires, and research protocols of students. Second, GenAI facilitated the summarization of paper contents, which accelerated the study selection process during the *main phases of SLR* while *conducting the review*. This finding is also proven by multiple of our data sources (lecturer observational notes, questionnaires with students, emails of students, research protocols of students). In particular, the otherwise time-consuming phase of *selection of primary studies*, *scanning content manually* was carried out more efficiently. As a result, students were able to process a larger number of articles within a limited timeframe, while maintaining a systematic and transparent selection strategy.

4.2 Use Cases for GenAI Adoption in Literature Reviews

In the following, we will examine the two most important steps in which GenAI helped students conduct the SLR. Our data sources (a) *lecturer observational notes*, b) *questionnaires with students*, and e) *research protocol of students* reveal that GenAI tools proved to be beneficial, especially in two steps of the SLR. This part of the process can be very time-consuming, especially depending on how much literature is available on the topic. Both steps can be found in Fig. 2 in the phase *conducting the review* and in the stage *selection of primary studies*. The use of GenAI proved to be very beneficial for this stage. On one hand, this process step could be accelerated, and on the other hand, this step could be carried out with greater precision, as the students were able to compensate for their weaknesses in terms of understanding English-language research literature.

To detail the steps, we modeled them as use cases using the Unified Modeling Language (UML) (OMG, 2017). The diagram is presented in Fig. 3. The students used the SciSpace GenAI tool for *selection of primary studies*. The tool offers the ability to extract data from several PDF files. To do this, the files must first be uploaded. The tool then creates an overview in tabular form. The columns can be customized. For the use case *scan content manually*, the students added a column in which the tool should create an evaluation of the respective primary study with regards to its relevance to their research question. Some students then also used this tool for the second use case *translation of English publications* by instructing the tool to extract the data from the papers into German. The lecturer pointed out the limitations of the GenAI tools several times during the course and

Main phases of SLR	Stages of main phases	Use Cases	GenAI tools used by students
Planning the review	Identification of the need for a review	Find a research gap on a topic, create research questions, optimizing research questions	Consensus App, ChatGPT, ChatAI, Copilot
	Specifying the research question(s)		
	Developing a review protocol	Generation of keywords, development and step-by-step optimization of a search string, find first relevant studies, optimizing text and content of the research protocol	Consensus App, ChatGPT, ChatAI, Gemini
	Evaluating the review protocol		
Conducting the review	Identification of research	Adaption of the search string to the requirements of the different digital libraries, support of the phases of the search process (step: "scan content manually" by means of a summary of the paper, translation of English publications)	ChatGPT, ChatAI, SciSpace, ASReview, DeepI
	Selection of primary studies		
	Study quality assessment	Check whether a paper is peer-reviewed, creation of the quality criteria, conducting quality assessment	ChatGPT
	Data extraction and monitoring	Support for the creation of the data extraction form, data extraction	SciSpace, ChatGPT
	Data synthesis		
Reporting the review	Formatting the main report	Spelling check, rephrasing sentences, creation of a rough structure and outline of the scientific poster	ChatGPT, ChatAI, Gemini, DeepIWrite
	Evaluating the report		

Figure 2: Overview of SLR phases, stages, use cases and GenAI tools used by the students.

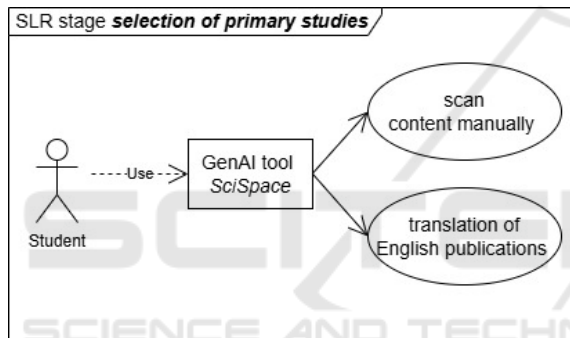


Figure 3: Use cases for the application of GenAI by the students.

asked the students to carry out a quality check of the results by comparing the translation with the original file.

4.3 Benefits and Challenges of Adopting GenAI

In this section, we answer our second research question *RQ2: What are the benefits and challenges of using GenAI tools during a literature review?* and present an overview in Table 1.

The analysis of student feedback and course materials revealed several **benefits** associated with the use of GenAI tools in the different phases of a SLR. For example, the creation of research questions was described as more efficient and precise. One student mentioned *"that happened so quickly, I have to let it sink in to understand what has happened here"*. During the selection of primary studies, tools such as *SciSpace* helped streamline the process and made it

easier for students to identify relevant sources. In the data extraction phase, GenAI supported students by generating summaries in tabular form, which simplified the organization of information. More generally, students noted that GenAI tools saved time, helped them work more efficiently, and were useful overall. They also supported students in overcoming language barriers and provided good summaries of scientific papers, which contributed to a better understanding of the content.

Despite the many benefits of GenAI tools in the context of academic research, several **challenges** were identified by students and the lecturer. When using GenAI to identify research gaps, limitations related to training data became apparent. The tools often provide only partial overviews or vague hints, making it difficult to assess the completeness and reliability of the descriptions. For example, in creating research questions, students noted that while GenAI-generated questions are often precise and well-structured, they can be formulated too narrowly, making it difficult to find relevant literature.

Another critical issue concerns the traceability and transparency of GenAI tools. Questions were raised by students and the lecturer about the reliability of sources provided by applications such as *SciSpace* or *Consensus*, especially regarding access restrictions (e.g., paywalls), unclear ranking mechanisms, and the relevance or timeliness of the results. Concerns were also voiced about whether journal rankings might be influenced by commercial interests, which could distort the research process.

In the phase of identifying primary studies, GenAI tools were found to occasionally produce hallucinations or incorrect summaries, limiting their usefulness

Table 1: Benefits and challenges of using GenAI in literature reviews

Topic	Benefits of using GenAI	Challenges of using GenAI
Research Question Creation	More efficient and precise formulation of research questions	Risk of overly narrow questions leading to no results
Finding a Research Gap	Support through summaries and hints on current topics	Limited coverage due to training data; lack of transparency and completeness
Selection of Primary Studies	Faster and more targeted identification of relevant studies (e.g., via SciSpace)	Incorrect or misleading summaries due to hallucinations of GenAI
Data Extraction	Tabular summaries generated by GenAI tools simplify the structuring of results	Require verification; potential inaccuracies in extracted data
Language and Comprehension	Overcoming language barriers and improved understanding of complex texts	Risk of relying too heavily on translations without critical reflection
Working Efficiency	Overall time-saving and more efficient workflows	Necessity of continuous quality assurance and critical review
Traceability and Transparency	-	Lack of transparency regarding source ranking, relevance, and business models of tools

in selecting the appropriate literature. More generally, the students noted that GenAI sometimes displays incorrect information or misleading references. These findings highlight the importance of critically assessing GenAI outputs and ensuring appropriate quality assurance throughout the research process. A further potential challenge identified in the student reflections was a tendency to engage less deeply with the subject matter when relying on GenAI tools. The availability of automated support appeared to reduce the degree of critical questioning and encouraged a more superficial engagement with the literature, as stated by the students themselves. This highlights the need for greater self-discipline and reflective awareness when integrating GenAI into academic work.

In summary, while GenAI tools offer significant benefits in terms of efficiency, precision, and overcoming language barriers, they also present challenges related to the quality and transparency of the results. These insights underscore the importance of balancing the use of GenAI with critical reflection and quality assurance to ensure the integrity of academic research.

5 DISCUSSION

Our single case study revealed some exciting results. We would like to revisit these topics from above and discuss the implications, as well as, outline the limitations of this study.

5.1 Implications of the Results

Literature reviews are time-consuming and often involve many humans (Pautasso, 2013; Borah et al., 2017). The use of GenAI can speed up the process. For instance, a critical point when conducting a literature review is the selection of primary studies, as it can be a challenge to filter out relevant studies from the multitude of scientific publications. The use of GenAI can support this process step. We have shown how the use cases *scan content manually* and *translation of English publications* (see Fig. 3) can be optimized with the help of GenAI tools. The students were able to use GenAI to analyze a large number of scientific papers in a short time and overcome language barriers.

However, the current GenAI tools also have disadvantages. Concerning traceability and transparency (cf. Table 1), it remains unclear according to which criteria GenAIs evaluate scientific publications, espe-

cially if they classify a publication as relevant. To ensure quality in a literature review, such an assessment of scientific literature must be reviewed by a human, as there is a risk of automation bias here. The human-in-the-loop is therefore important at this point. This goes hand in hand with the known limitations that the output generated by GenAI is not always error-free (Feuerriegel et al., 2024).

To address this, we reviewed some publications deemed relevant by GenAI tools (e.g., SciSpace) and compared them based on citation counts as well as journal and conference rankings. Our analysis revealed that established metrics such as Cite Score, Impact Factor, and conference ranking are not being used to assess relevance. Details concerning this analysis can be found in our research protocol (Schön et al., 2025). This could lead to the emergence of new business models (e.g., new ranking algorithms and paid advertisement similar to Google Ads) in the future. This development carries the risk of a stronger influence on the independence of research. Publishers have already recognized the potential of GenAI and are coming up with their own AI tools¹. Moreover, assessing the quality of primary studies remains an open question that needs to be addressed by the researcher conducting the literature review. At this point, we still need the human-in-the-loop, as the strength of the evidence supported by the quality of the selected primary studies is one of the most critical parts in literature reviews (Yang et al., 2021). This is also confirmed in the related work. Studies suggested hybrid methods (Ng and Chan, 2024) and emphasized that GenAI could not replace humans, but merely support them in terms of efficiency (Ofori-Boateng et al., 2024; Castillo-Segura et al., 2024). However, humans are still responsible for checking the work's accuracy (Felizardo et al., 2024) and integrity (Schryen et al., 2025).

Another concern highlighted by our findings relates to the knowledge acquired when conducting a literature review. We learned that there is a tendency to engage less deeply with the subject matter when relying on GenAI tools. The availability of automated support appeared to reduce the degree of critical questioning and encouraged a more superficial engagement with the literature. It is therefore questionable whether the knowledge generated is transferred to long-term memory or whether it quickly fades away. A statement by a student in a questionnaire is also thought-provoking. The student stated *“Is it still worth thinking for yourself? It's frightening that students who work 100% with GenAI get*

better grades. Are we studying more to serve?”. This statement raises the question of how the way we study today needs to change to respect the technological progress made by GenAI, and whether it makes sense to automate as much work as possible using GenAI during a study program.

When we automate steps in the process of conducting literature reviews, it is important to distinguish between different user settings: while students need to carry out the process manually in order to understand and internalize critical steps, the use of GenAI can enable experienced researchers to accelerate the literature review process.

5.2 Threats to Validity

As with any research, certain limitations arise from the methodological approach chosen. Although this study was carefully designed and conducted in accordance with established guidelines, specific limitations remain. In the following, we highlight these limitations and describe the strategies employed to minimize their influence on the study's outcomes. We used the threats to validity schema according to Wohlin et al. (Wohlin et al., 2012) and Runeson and Hoest (Runeson and Höst, 2009).

Construct Validity: One potential limitation concerning construct validity lies in the interpretation and operationalization of *“effective support”* by GenAI tools, as perceptions can vary between students. To mitigate this, we triangulated data from multiple sources (e.g., student exams, survey responses, and instructional materials) to ensure consistency and clarity in measuring the intended constructs.

Internal Validity: Given the single case study design, internal validity may be affected by uncontrolled variables, such as students' prior experience with AI tools or differences in how they engaged with the tasks. To address this, we used a conceptual structure for AI assistants in higher education (Schön et al., 2023b), applied identical task formats, and ensured guided supervision across all groups to reduce variability.

External Validity: The generalizability of our findings is limited due to the specific context of the single case study (e.g., a single higher education course and student population). However, by providing detailed contextual descriptions and linking our results to existing literature, we enable transferability and support future replication in similar educational settings.

¹see <https://www.elsevier.com/products/sciencedirect/sciencedirect-ai>

6 CONCLUSION AND FUTURE WORK

This paper presents the results of a single case study in which we investigated how integrating GenAI tools can optimize the literature review process. To this end, we observed 16 Master's students over a term as they carried out an SLR with GenAI. Based on the data analyzed, we were able to identify use cases in which the students used GenAI. Our results show that GenAI supports the selection of primary studies, in particular by supporting the process of manually analyzing the publications found and overcoming language barriers. In addition, we have identified benefits and challenges that make it possible to weigh up the use of GenAI in certain steps of the literature review process.

A limitation lies in the fact that some of the GenAI tools used suggest that relevant studies on a given topic have been selected. However, the criteria by which “*relevance*” is determined remain entirely opaque. This lack of transparency poses a challenge for inexperienced researchers and students attempting to familiarize themselves with a new topic, potentially leading to automation bias at this stage of the research process. Furthermore, our results show that students do not engage as intensively with literature due to the use of GenAI and therefore do not delve as deeply into a topic. It is precisely the examination and development of knowledge on a topic that motivates performing a literature review. And with this identified disadvantage, it is questionable whether one of the reasons for adopting literature reviews (*to provide a well-informed foundation for positioning new research activities*) can be fulfilled at all.

Our findings highlight the need for further research on how GenAI can be integrated into academic workflows without compromising the deep learning processes of students and their critical engagement with scientific content. In our future work, we aim to further explore the benefits and challenges associated with the use of GenAI in the context of literature reviews. Therefore, we are conducting further case studies to enable comparisons between different groups. In doing so, we want to motivate participants to take part in the questionnaires more actively to obtain more quantitative data.

ACKNOWLEDGEMENTS

We would like to thank the students who took part in the exciting experiment of conducting a literature review using GenAI.

GenAI tools (ChatGPT, DeepL, and Grammarly) were used for the optimization of text passages.

REFERENCES

- Baxter, P. and Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers.
- Borah, R., Brown, A. W., Capers, P. L., and Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the prospero registry. *Open*, 7:12545.
- Castillo-Segura, P., Fernández-Panadero, C., Alario-Hoyos, C., and Kloos, C. D. (2024). Enhancing research on engineering education: Empowering research skills through generative artificial intelligence for systematic literature reviews. In *2024 IEEE Global Engineering Education Conference (EDUCON)*, pages 1–10. doi: 10.1109/EDUCON60312.2024.10578328.
- Felizardo, K. R., Lima, M. S., Deizepe, A., Conte, T. U., and Steinmacher, I. (2024). Chatgpt application in systematic literature reviews in software engineering: an evaluation of its accuracy to support the selection activity. In *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, ESEM '24*, page 25–36, New York, NY, USA. Association for Computing Machinery. doi: 10.1145/3674805.3686666.
- Feuerriegel, S., Hartmann, J., Janiesch, C., and Zschech, P. (2024). Generative ai. *Business and Information Systems Engineering*, 66:111–126. doi: 10.1007/s12599-023-00834-7.
- García-Peñalvo, F. J. and Vázquez-Ingelmo, A. (2023). What do we mean by genai? a systematic mapping of the evolution, trends, and techniques involved in generative ai. *International Journal of Interactive Multimedia and Artificial Intelligence*, 8:7–16. doi: 10.9781/ijimai.2023.07.006.
- Garrity, C., Gartlehner, G., Nussbaumer-Streit, B., King, V. J., Hamel, C., Kamel, C., Affengruber, L., and Stevens, A. (2021). Cochrane rapid reviews methods group offers evidence-informed guidance to conduct rapid reviews. *Journal of Clinical Epidemiology*, 130:13–22. doi: 10.1016/j.jclinepi.2020.10.007.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
- Neumann, M., Rauschenberger, M., and Schön, E.-M. (2023). “we need to talk about chatgpt”: The future of ai and higher education. In *5th International Workshop on Software Engineering Education for the Next Generation (SEENG)*, pages 29–32. doi: 10.1109/SEENG59157.2023.00010.
- Ng, H. K. Y. and Chan, L. C. H. (2024). Revolutionizing literature search: Ai vs. traditional methods in digital divide literature screening and reviewing. In *Proceedings - 2024 6th International Workshop on Artificial Intelligence and Education, WAIE 2024*, pages 144–

148. Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/WAIE63876.2024.00033.
- Ofori-Boateng, R., Aceves-Martins, M., Wiratunga, N., and Moreno-Garcia, C. F. (2024). Towards the automation of systematic reviews using natural language processing, machine learning, and deep learning: a comprehensive review. *Artificial Intelligence Review*, 57. doi: 0.1007/s10462-024-10844-w.
- OMG, O. M. G. (2017). Unified modeling language version 2.5.1.
- OpenAI (2025). Chatgpt.
- Pautasso, M. (2013). Ten simple rules for writing a literature review. doi: 10.1371/journal.pcbi.1003149.
- Runeson, P. and Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14(2):131–164. doi: 10.1007/s10664-008-9102-8.
- Schryen, G., Marrone, M., and Yang, J. (2025). Exploring the scope of generative ai in literature review development. *Electron Markets*, 35. doi: 10.1007/s12525-025-00754-2.
- Schön, E.-M., Buchem, I., Sostak, S., and Rauschenberger, M. (2023a). *Shift Toward Value-Based Learning: Applying Agile Approaches in Higher Education*, volume 494, pages 24–41. doi: 10.1007/978-3-031-43088-6_2.
- Schön, E.-M., Kollmorgen, J., Neumann, M., and Rauschenberger, M. (2025). Research protocol for the paper entitled: A case study on using generative ai in literature reviews - use cases, benefits, and challenges. doi: <https://doi.org/10.5281/zenodo.16537498>.
- Schön, E.-M., Neumann, M., Hofmann-Stölting, C., Baeza-Yates, R., and Rauschenberger, M. (2023b). How are ai assistants changing higher education? *Frontiers in Computer Science*, 5. doi: 10.3389/fcomp.2023.1208550.
- Schön, E.-M., Thomaschewski, J., and Escalona, M. (2016). Agile requirements engineering - protocol for a systematic literature review. doi: 10.13140/RG.2.2.16615.85923.
- Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in Software Engineering*. Springer. doi: 10.1007/978-3-642-29044-2.
- Xiao, Y. and Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research*, 39:93–112. doi: 10.1177/0739456X17723971.
- Yang, L., Zhang, H., Shen, H., Huang, X., Zhou, X., Rong, G., and Shao, D. (2021). Quality assessment in systematic literature reviews: A software engineering perspective. *Information and Software Technology*, 130. doi: 10.1016/j.infsof.2020.106397.
- Yin, R. K. (2009). *Case study research: Design and methods*, volume 5 of *Applied social research methods series*. Sage, Los Angeles, 4. ed. edition. isbn: 978-1-4129-6099-1.

APPENDIX

Search Strings:

- GenAI literature review meta analysis
- Generative AI systematic literature review
- AI literature review survey meta analysis
- GenAI literature survey systematic review
- AI literature review systematic meta analysis
- ("literature review" OR "systematic review" OR "meta analysis") AND ("AI" OR "generative AI" OR "artificial intelligence")
- "systematic literature review" AND ("AI" OR "artificial intelligence")
- "systematic literature review * with" AND ("AI" OR "artificial intelligence")
- ("literature review * with" OR "systematic review * with" OR "meta analysis * with" OR "scoping * with" OR "bibliometric * with" OR "survey * with" OR "mapping * with") AND ("AI" OR "artificial intelligence" OR "generative AI" OR "generative artificial intelligence" OR "GPT")