

Please quote as: Tingelhoff, F., Brugger, M. & Leimeister, J. M. (2024). A guide for structured literature reviews in business research: The state-of-the-art and how to integrate generative artificial intelligence. *Journal of Information Technology (JIT)*, , 1-23. doi: 10.1177/02683962241304105

A guide for structured literature reviews in business research: The state-of-the-art and how to integrate generative artificial intelligence

Journal of Information Technology
2024, Vol. 0(0) 1–23
© Association for Information
Technology Trust 2024



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/02683962241304105
journals.sagepub.com/jinf



Fabian Tingelhoff¹ , Micha Brugger¹ and Jan Marco Leimeister^{1,2}

Abstract

Generative artificial intelligence (Gen.AI) is capable of significantly improving the breadth and depth of structured literature reviews (SLRs). However, its inclusion raises essential questions regarding the review's methodology, quality, and ethical implications. Previous research predominantly focused on the capabilities and limitations of Gen.AI to establish guidelines for research practices. However, the rapid evolution of Gen.AI often outpaces the publication of methodological papers. In response, our study adopts a criteria-centric approach, scrutinizing the scientific quality standards that Gen.AI must meet. In other words, instead of discussing *what Gen.AI can and cannot do*, we discuss *what we should allow Gen.AI to do*, irrespective of its capabilities. Our study informs researchers in the art and science of SLRs. First, we analyze the established state-of-the-art processes and associated quality standards in SLRs. From this, we synthesize a unified process and criterion set, not only underpinning a comprehensive understanding of the extant SLR methodologies but also serving as the foundational framework for integrating Gen.AI. Second, we delineate the specific scenarios conducive to incorporating Gen.AI into this fundamental framework, as well as situations where its integration may not be suitable. Our contribution is further solidified by providing a detailed, step-by-step guide—akin to a “cooking recipe”—to effectively integrate Gen.AI in SLRs, ensuring adherence to established quality criteria.

Keywords

Literature review methodology, generative artificial intelligence, research ethics, academic integrity

Introduction

Generative artificial intelligence (Gen.AI) has seen rapid developments garnering the attention of practitioners and researchers alike. Most famously, the introduction of ChatGPT in 2022 has surpassed many peoples' imagination, introducing new groundbreaking functionalities. For instance, the general public could use natural language to interact with large amounts of data, like “chatting with PDF files” (Hadi et al., 2023). These functionalities have proved similarly influential in research. Many scholars have started using Gen.AI tools for various parts of their knowledge creation processes, like identifying relevant literature (Dann et al., 2017), creating the manuscript's abstracts (Else, 2023), or even developing the entire manuscript (Dwivedi et al., 2023). This has resulted in published papers officially co-authored by Gen.AI tools (Stokel-Walker, 2023), prompting the Scientific American to attest that Gen.AI already has “thoroughly infiltrated scientific publishing” (Stokel-Walker, 2024).

However, through these unique capabilities, Gen.AI has ambivalent effects on the scientific landscape. On the one hand, it has fueled the trend of rising research publications. Even before Gen.AI, researchers struggled with the exponential growth of scientific research, increasing the risk of missing valuable contributions in their field (Cropanzano, 2009). For instance, while in 1980, approximately 650,000 academic papers (peer-reviewed papers in journals and conferences across all disciplines) were published globally, it had surged to about 3 million by 2022 (Science.org, 2023). Despite no current statistics on post-

¹University of St. Gallen, Switzerland

²University of Kassel, Germany

Corresponding author:

Fabian Tingelhoff, University of St. Gallen, Institute of Information Systems and Digital Business, Müller-Friedberg-Strasse 8, St. Gallen 9000, Switzerland.

Email: fabian.tingelhoff@unisg.ch

Gen.AI publishing, experts and scholars alike express paper submission numbers to exponentially increase because of Gen.AI usage in manuscript development (Stokel-Walker, 2023).

On the other hand, Gen.AI can support scholars in managing these large data volumes (Antons et al., 2023; Mortenson and Vidgen, 2016; Rathje et al., 2024). For instance, Gen.AI tools can assist in streamlining the identification of relevant papers, ensuring comprehensive coverage of pertinent literature (Dann et al., 2017). Furthermore, Gen.AI excels in summarizing extensive scholarly articles (Glickman and Zhang, 2024) and has been employed in automating data extraction and analysis (Rathje et al., 2024), drastically increasing the efficiency and accuracy of synthesizing and analyzing large volumes of research data, thereby enhancing the overall quality of systematic reviews (Drori & Te'eni, 2024).

In business research, both the negative and positive effects of Gen.AI are particularly pronounced for structured literature reviews (SLRs) (Tingelhoff et al., 2024). First, the negative effects of Gen.AI are especially severe for SLRs, as they are meant to consolidate past research to generate new insights, methodologies, or theories (Webster and Watson, 2002). As Gen.AI amplifies the volume of “past research,” it also significantly increases the workload and complexity for authors. Simultaneously though, SLRs are particularly apt for Gen.AI augmentation. Algorithm-based Gen.AI tools favor clear context information (like processes, task descriptions, or output formats), which perfectly maps to the structured, transparent, and reproducible nature of SLRs (Tranfield et al., 2003).

Past research on Gen.AI-assisted SLRs mainly falls into one of two categories: investigating Gen.AI’s potential by analyzing its capabilities (e.g., Schryen et al., 2024; Wagner et al., 2022); or scrutinizing how Gen.AI’s capabilities might undermine research integrity. Given the critical importance of transparency and reproducibility in the SLR process, the black-box nature of Gen.AI tools raises significant concerns (Dwivedi et al., 2023). Scholars argue that relying on Gen.AI outputs may lead to authors losing control over their SLRs, given the technologies’ issues with plagiarism, biased information, or false claims (Alavi et al., 2024; Ngwenyama and Rowe, 2024). Combining both literature streams, past research can be summarized as a discussion on the influence of Gen.AI’s capabilities on SLR standards (Jarvenpaa and Klein, 2024).

However, the speed at which Gen.AI is evolving currently outpaces publication cycles (Nguyen et al., 2022). This means the bases for Gen.AI discussions are often outdated with the paper’s publication. Consequently, we propose to flip the discussion to how research criteria shape the use of Gen.AI capabilities. In other words, instead of discussing *what Gen.AI can and cannot do*, we intend to discuss *what we should allow Gen.AI to do*, irrespective of

its capabilities. To contribute to this discussion, our investigation comprises two primary research objectives (RO). First, we aim to synthesize the current SLR process and criteria to form a basis for our further analysis. Second, we apply these integrated processes and criteria to the use of Gen.AI in SLRs to gain a new perspective on its integration. Our ROs are as follows:

RO1. To identify and analyze the established processes and guidelines for conducting state-of-the-art SLRs.

RO2. To assess how the established SLR process and guidelines influence the use of Gen.AI tools and to formulate optimal strategies for its integration.

We strive to contribute to the academic discourse in two distinct yet interconnected ways. First, our analysis of the established state-of-the-art processes and associated quality standards in SLRs culminates in the synthesis of a unified process and criterion set. This synthesis not only underpins a comprehensive understanding of the extant SLR methodologies but also serves as the foundational framework for integrating Gen.AI. The relevance of this integration extends beyond those researchers actively employing Gen.AI in their SLRs; it offers an insightful summary of established research methodologies and normative guidelines, benefiting the wider scholarly community. Second, we delineate the specific scenarios conducive to incorporating Gen.AI into this fundamental framework, as well as situations where its integration may not be suitable (Ngwenyama and Rowe, 2024). Our contribution is further solidified by providing a detailed, step-by-step guide—akin to a “cooking recipe”—to effectively integrate Gen.AI in SLRs, ensuring adherence to established quality criteria.

Using a two-by-two matrix, Figure 1 lays out the structure of this paper.

The state-of-the-art of structured literature reviews

Traditionally, SLRs are a scientific tool consolidating past research to generate new insights, methodologies, or theories, thereby serving as a general cornerstone of research (Webster and Watson, 2002). SLRs play a critical role in collating and synthesizing existing knowledge, developing new research trajectories, and contributing to the dynamic evolution of various fields (Webster and Watson, 2002). They not only frame research agendas and promote methodological transparency (Wagner et al., 2022) but also play a crucial role in acknowledging and understanding the cultural nuances within research phenomena (Kummer et al., 2012). Given their substantial influence and significance, SLRs are subject to rigorous academic scrutiny. Scholars have extensively debated the proper execution of SLRs (e.g., vom Brocke et al., 2015), their adherence to

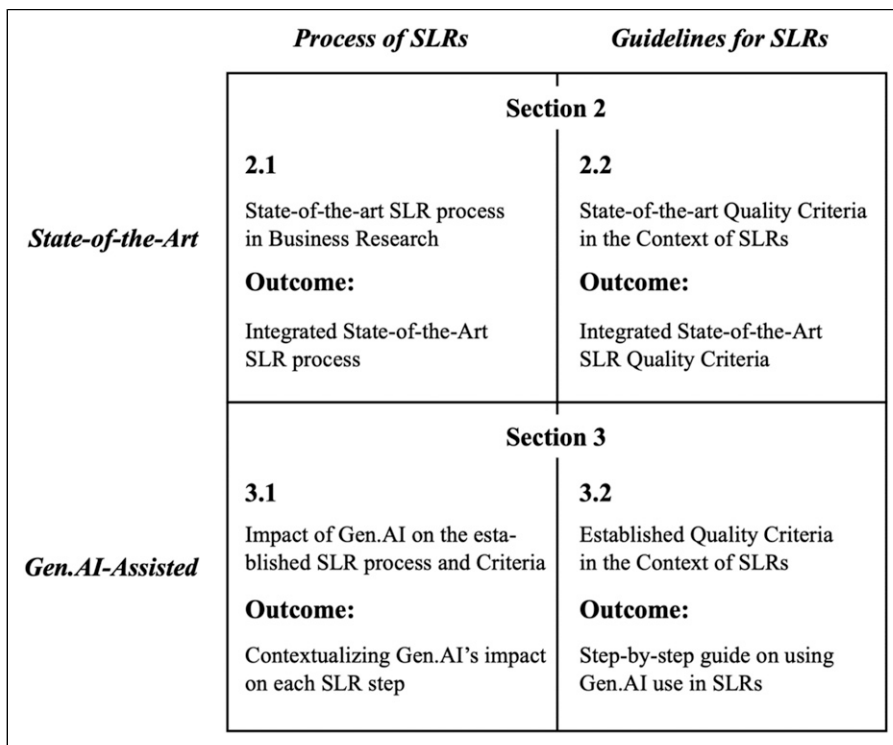


Figure 1. Structure of this paper.

quality criteria (e.g., [Tranfield et al., 2003](#)), and the effective presentation of findings (e.g., [Webster and Watson, 2002](#)). In this section, we initially discuss the state-of-the-art of the SLR process, followed by a guide on how to effectively conduct an SLR.

The state-of-the-art process of structured literature reviews

To conceptualize when and how to use Gen.AI in SLRs, it is imperative that we first understand the established process of conducting SLRs. We identified several influential method papers by examining published SLRs in business-related FT-50 journals and engaging with senior researchers across several countries and continents. We present and contrast these papers and their proposed processes in [Figure 2](#). Noticing the commonalities in the processes described in these studies (e.g., see [Snyder, 2019](#)), we categorized their tasks into four phases: (1) designing the review; (2) discovering relevant research; (3) developing the outcome of the SLR; and (4) disseminating knowledge. We discuss each phase and associated steps in detail below.

Design. The initial phase of an SLR encompasses its foundational preparation. This stage involves a thorough domain familiarization to establish a robust solid base for

the ensuing subsequent phases ([vom Brocke et al., 2015](#)). Typically, renowned scholars conduct SLRs, leveraging their extensive domain knowledge to bolster the validity of their literature selection and interpretation ([Webster and Watson, 2002](#)). Emerging scholars, especially those engaged in PhD research, may require more initial workload to achieve domain knowledge similar to senior researchers ([Ngwenyama and Rowe, 2024](#); [vom Brocke et al., 2015](#)).

Next, authors must identify the need for their SLR ([Sauer and Seuring, 2023](#)). An SLR becomes necessary, for example, when there is a need to consolidate existing knowledge on a topic ([Webster and Watson, 2002](#)), to stay abreast of developments and theoretical shifts ([Levy and Ellis, 2006](#)), or when exploring interdisciplinary connections ([Okoli and Schabram, 2015](#)).

Once the authors have developed a profound understanding of their research field and formulated the need for their SLR, they can begin its preparation. Preparing a SLR requires a systematic and clearly defined strategy ([Fisch and Block, 2018](#)). Essential steps include defining the review’s scope and objectives ([Webster and Watson, 2002](#)), choosing the right balance between breadth and depth ([Fisch and Block, 2018](#)), and documenting the process for its reproducibility in a review protocol ([Okoli, 2015](#)). A review protocol is “a plan prior to the review [that] states the criterion for including and excluding studies, the search strategy, description of the methods to be used, coding

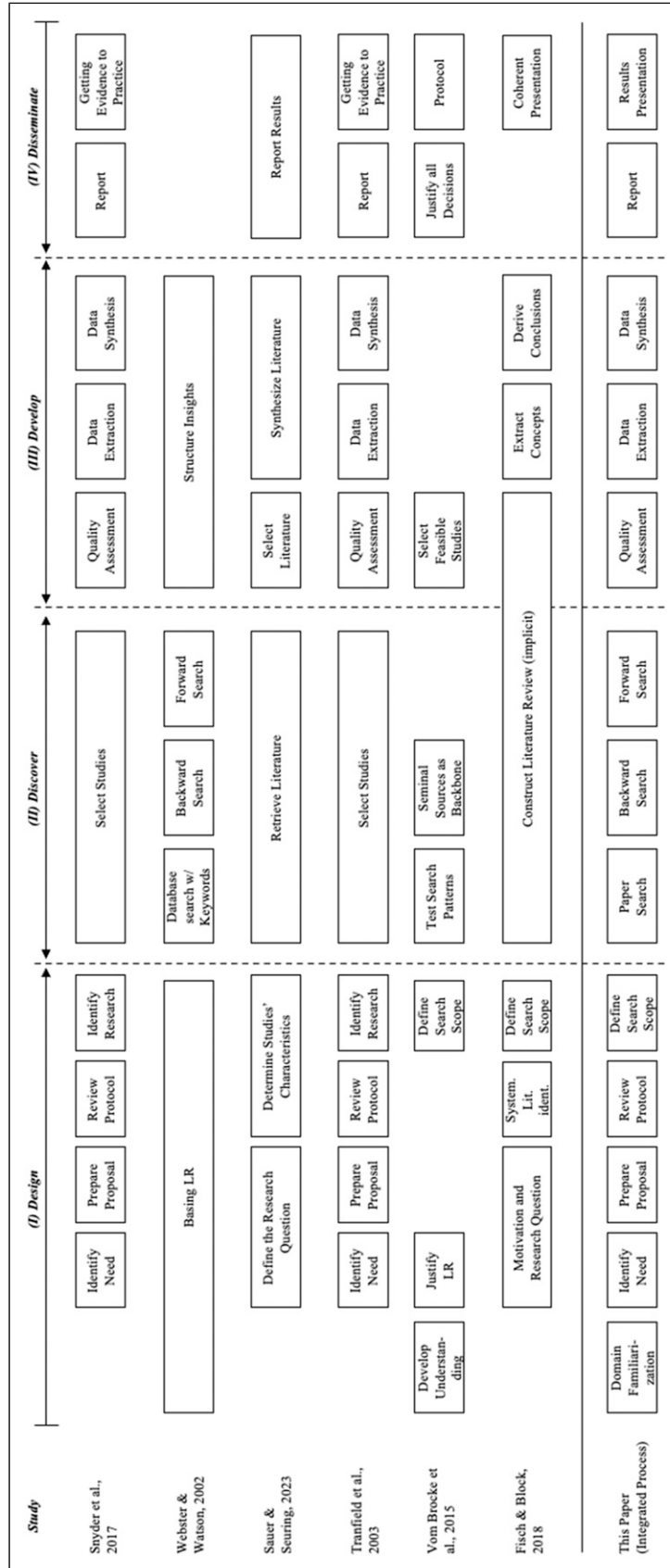


Figure 2. Consolidation of literature review processes of established method papers.

strategies and the statistical procedures to the employed” (Tranfield et al., 2003: 213).

Lastly, authors need to determine the scope of their SLR (Fisch and Block, 2018; vom Brocke et al., 2015). Usually, this refers to formulating a set of keywords and phrases that closely align with the research topic, ensuring they encompass both broad and specific aspects of the topic (Fink, 2019). Boolean (AND, OR, NOT) and logical operators (LIKE, %, *) can refine the search (Rowley and Slack, 2004). Including synonyms and related terms can broaden the scope, as Levy and Ellis (2006) recommended. Furthermore, researchers must identify criteria for the inclusion and exclusion of literature to be reviewed (Sauer and Seuring, 2023), which can regard paper characteristics (e.g., “the paper is written in English” or “the full-text is open access”) or contents (e.g., “the paper uses one theory” or “the paper assumes one perspective”).

Discover. Post-design, the discovery phase involves executing the search strategy to gather potential literature. First, researchers should conduct keyword searches within the identified databases. If unexpected or unintended results are obtained, authors can iteratively refine their decisions from the design phase (Snyder, 2019). This approach aids in retrieving a comprehensive set of articles that constitute the SLR’s foundation. A subsequent backward search allows researchers to review the references of the already identified articles to find additional sources. This method is critical for uncovering foundational works that might not surface in keyword-based searches (Webster and Watson, 2002). By examining the citations in key articles, researchers can trace the intellectual lineage of a topic, ensuring a deep and historical understanding of the research area (Okoli and Schabram, 2015). A forward search, conversely, involves looking at studies that have cited the already identified key articles. Resources such as Google Scholar and Scopus provide citation-tracking capabilities, which are essential for this task (Cronin et al., 1998). This method allows researchers to understand a topic’s evolution and current state as newer studies build upon or challenge the findings of earlier works. Forward searching is especially beneficial in rapidly evolving fields, where keeping up with the latest developments is critical (Bawden and Robinson, 2015). Together, the keyword, backward, and forward searches ensure a thorough and up-to-date SLR, capturing both the field’s historical roots and contemporary advancements.

Develop. After compiling a comprehensive list of potential sources in the development phase, each paper is rigorously evaluated for its quality, relevance, and contribution to the research question to finalize the selection (Sauer and Seuring, 2023). Initially, researchers need to evaluate the quality of each paper, focusing on the publication source’s credibility, the research methodology employed, and the

findings’ impact (Levy and Ellis, 2006; Snyder, 2019; Webster and Watson, 2002). This assessment helps to exclude studies that do not meet the scholarly standards necessary for a robust SLR. Moreover, researchers should eliminate studies that are inaccessible to the research community to ensure a transparent and replicable process (vom Brocke et al., 2015).

Data extraction and synthesis then involve categorizing key elements such as research questions, methodologies, findings, and theoretical frameworks (Rowley and Slack, 2004). Researchers should thoroughly synthesize and organize the data to uncover relevant themes and concepts (Fisch and Block, 2018), identifying patterns and gaps, and constructing a comprehensive narrative encapsulating the breadth and depth of the research field (Fink, 2019; Okoli and Schabram, 2015; Torracco, 2005). It is vital to ensure that the synthesis not only summarizes the findings of individual studies but also draws connections between them to offer a broader understanding of the field (Cronin et al., 1998).

Disseminate. The dissemination phase centers on presenting the SLR findings and associated decisions transparently and coherently. This includes justifying source selection, criteria for inclusion and exclusion, and the employed methodologies for analysis and synthesis of literature data (Levy and Ellis, 2006; Webster and Watson, 2002). Such transparency is crucial in establishing the review’s credibility and the possibility of replication, as Okoli and Schabram (2015) emphasized. In presenting results, effective techniques like visual aids and engaging narratives are vital for conveying complex information and facilitating comprehension (Snyder, 2019; Webster and Watson, 2002). Employing a concise and engaging writing style renders the SLR accessible and valuable to a broader audience, including practitioners. Consequently, authors must decide on the appropriate publication outlet and theoretically embed their SLR into its past research (Sauer and Seuring, 2023).

The state-of-the-art guidelines for structured literature reviews

As with any research paradigm, SLRs must comply with quality criteria and guidelines established by ethics committees, scientific associations, and journals. Given their crucial role in research, these SLR criteria must be carefully considered and contextualized, as detailed in established method publications (Tranfield et al., 2003; Webster and Watson, 2002). Irrespective of whether researchers decide to integrate Gen.AI or do the entire SLR manually, these criteria are the foundation of rigorous and integrity-driven SLRs (Paré et al., 2023). Consequently, they are the constant against which new methods and tools must be evaluated. Thus, to comprehend how established guidelines influence

the application (or non-application) of Gen.AI in SLRs, it is essential to first grasp the underlying criteria (Gregor, 2024).

In this study, we specifically incorporated the quality criteria from the US Institutional Review Board and the Deutsche Forschungsgemeinschaft (DFG) (as examples of national ethics committees), the Academy of Management and the Association for Information Systems' Code of Ethics (as research communities), and Nature and the Journal of Information Technology's Value Statements (as exemplary journals)—as these are found to already incorporate guidelines about the responsible use of Gen.AI for their authors and reviewers (Gregor, 2024). Through our analysis, we found eight quality criteria, emphasizing high ethical standards and scientific integrity as foundational requirements for research: beneficence, respect for persons, integrity, responsibility, rigor, impact, reproducibility, and transparency. We provide an overview of these criteria in Figure 3. While our consolidated figure refers specifically to the guidelines of the six aforementioned organizations, our extended analysis of over 30 institutions and journals showed that the consolidated criteria align closely with, and can be considered representative of, the full set. Henceforth, we proceed to explain each criterion as interpreted by these bodies, contextualized by its relevance to SLRs as discussed by scholarly research.

Beneficence. In the context of SLRs, researchers must be acutely aware of the broader impact their synthesis of existing research has on societal welfare. This requires a meticulous selection and evaluation of studies that collectively offer insights beneficial to society (Macfarlane, 2010). Researchers should aim to highlight the positive societal outcomes identified in the literature, ensuring their review underscores contributions that advance societal good. For instance, if an SLR were to review the dark sides of technological change, it should highlight mitigation strategies to positively impact how society deals with the presented problems. Additionally, it is essential to address any ethical issues reported in the studies reviewed, providing a balanced perspective that considers both benefits and potential harms (Walsham, 2012). By doing so, SLRs can guide future research and policy-making in directions that foster societal well-being and mitigate harm.

Respect for persons. In research, in general, ethical treatment of data and research subjects is crucial (Mertens and Ginsberg, 2009). While SLRs involve analyzing publicly available data rather than collecting new data, this criterion is less important to the review process. Still, when analyzing articles, authors should investigate how primary researchers obtained informed consent from their research subjects and ensured their privacy. By

thoroughly examining and reporting on these ethical practices, SLR authors can highlight the ethical rigor of the studies they review, thus maintaining public trust and upholding the ethical standards of the research community. Addressing these considerations demonstrates a commitment to ethical scholarship and reinforces the importance of ethical research practices across all stages of research.

Integrity. Research integrity is paramount in SLRs, requiring researchers to be truthful in every aspect of the review process. This principle mandates the accurate representation of the papers included in the review, strictly avoiding any form of fabrication, falsification, or misrepresentation (Resnik, 2005). For instance, suppose an experienced author is conducting an SLR in a field where they have already published many articles. Giving one's own articles more exposure in the review undermines the fair and unbiased representation of the sample, hence violating integrity. In SLRs, maintaining honesty is critical for several reasons. First, it ensures that the conclusions drawn from the review are based on objective and verifiable data, which is essential for the reliability and trustworthiness of the review. Second, accurate reporting and unbiased interpretation of the findings from the studies included in the review contribute meaningfully and authentically to the cumulative body of scientific knowledge (Comstock, 2012; Steneck, 2003). Adhering to integrity allows SLRs to provide a reliable synthesis of existing research, guiding future research directions and informing policy and practice based on solid and dependable evidence (Snyder, 2019).

Responsibility. Researchers must adhere to ethical standards and legal regulations, accepting the consequences of their decisions throughout the review process (Shamoo and Resnik, 2009). Responsibility in SLRs includes meticulously citing sources, acknowledging the work of other researchers, and attributing work fairly. For instance, authors are responsible for reading papers in their sample entirely and not just skimming over them to ensure they are not attributing wrongful information to another author (Anderson et al., 2010). Only by responsibly handling the diverse studies included in an SLR and analyzing them faithfully can researchers establish the trust and appreciation necessary for advancing a community's knowledge (Resnik et al., 2015). However, the notion of attributing work "fairly" also entails subjective interpretation; individual researchers may understand and convey the same work differently based on their perspectives and analytical frameworks. Therefore, responsibility in SLRs demands both an accurate portrayal of existing literature and an awareness of the interpretive nuances that come with assessing and citing others' work.

<i>US Institutional Review Board</i>	<i>Deutsche Forschungsgemeinschaft (AI-specific)</i>	<i>Academy of Management</i>	<i>Association of Information Systems</i>	<i>Journal of Information Technology</i>
Beneficence			Positive Change	
Respect for Persons	Respect for People's Intellectual Property Rights	Respect for People's Rights and Dignity	Respect for People's Intellectual Property Rights	Respect for People's Intellectual Property Rights
Justice	Integrity	Integrity	Integrity	Integrity
	Responsibility	Responsibility	Responsibility	Research Culture
				Rigor
				Impactful
				Reproducible
Transparency	Transparency	Transparency	Transparency	Transparency

Figure 3. Quality criteria for academic rigor and integrity.

Rigor. An SLR demands a rigorous approach to systematically identify, select, and analyze published work. This rigor involves implementing a clear, transparent methodology and encompassing specific search strategies, selection criteria, and analytical methods (Snyder, 2019; Webster and Watson, 2002). Without assessing the quality of the included studies, the review's conclusions might be based on unreliable or biased data, compromising the methodological soundness of the review. For instance, authors might use journal rankings or scientific indices as a proxy for publishing quality (Hahn, 2024). Rigor is crucial in SLRs for several reasons. Firstly, it ensures comprehensive coverage of relevant literature, which minimizes bias and enhances the review's validity (Okoli and Schabram, 2015). Secondly, a rigorous review process allows other researchers to verify and build upon the findings, which is essential for advancing scholarly research (Okoli and Schabram, 2015; Sauer and Seuring, 2023). This level of thoroughness is particularly important in dynamic and interdisciplinary fields, where evolving methodologies and paradigms require continuous reassessment and validation of findings. By maintaining rigor, SLRs contribute reliable, high-quality insights that can guide future research and inform practice, ensuring that the synthesized knowledge is both credible and valuable.

Impact. In SLRs, impact refers to the review's capacity to significantly influence future research, practice, policy, or theory (Fisch and Block, 2018). An impactful review tackles essential issues, presents new perspectives, or offers thorough frameworks that other researchers and practitioners in the field can utilize (Sauer and Seuring, 2023). It can result in developing new theories, enhancing existing systems, and advancing theoretical knowledge (Bawden and Robinson, 2015; Cronin et al., 1998). Additionally, an impactful SLR affects policy and decision-making in organizations, underscoring its importance beyond academia (Levy and Ellis, 2006; Webster and Watson, 2002). For example, authors, who set the scope of their review too narrowly, might limit the generalizability of their findings and, thus, the usefulness of their review, undermining the review's relevance and publishability.

Reproducibility. In SLRs, reproducibility means other researchers can replicate the review process and ideally reach similar conclusions using the same methods and data (Snyder, 2019). By adhering to reproducible methods, an SLR provides a transparent and systematic approach that other researchers can replicate to verify results, eliminating subjective bias and enhancing the robustness of the conclusions (Levy and Ellis, 2006; Webster and Watson, 2002). Scholars have long been advocating for systematicity and rigor in synthesizing existing knowledge, highlighting the necessity for researchers to

adhere to established guidelines and best practices when conducting literature reviews (Cram et al., 2020).

However, a reproducible SLR does not require authors to justify every paper in their sample. For instance, vom Brocke et al. (2015: 216) emphasize that "there is no reason to exclude a relevant publication from a SLR if the researcher came across it by means other than the keyword search, even by chance." This means that an SLR inherently involves some level of non-reproducibility. Still, it is crucial where reproducibility is situated in the SLR process. While paper search might have more freedom, researchers with the same paper samples and analysis framework should ideally reach the same conclusions. This is important, as reproducibility also allows for continual updates and advancements of SLRs, as future researchers can build on the established methodology to incorporate new studies and insights, ensuring the SLR remains relevant and comprehensive over time (Fink, 2019; Okoli and Schabram, 2015).

Transparency. Transparency in SLRs means clearly and explicitly documenting the review process—from the initial formulation of research questions to the selection and evaluation of sources and data analysis and synthesis methods. Transparency provides a roadmap of the researcher's intellectual journey, allowing readers to understand and evaluate the basis of the review's conclusions (Snyder, 2019). Given research's fast-paced and complex nature, clearly articulating the research scope, boundaries, and methodology is essential for the review's relevance and rigor (Levy and Ellis, 2006; Webster and Watson, 2002). Transparency also helps establish the review's and the researcher's credibility, building trust in the findings, which is essential for advancing knowledge (Okoli and Schabram, 2015; Rowley and Slack, 2004). Furthermore, transparently reporting SLR processes helps future researchers grasp the context and limits of past reviews, contributing to the field's cumulative development (Cronin et al., 1998; Torracco, 2005).

A criteria-centric perspective on Gen.AI-assisted literature reviews

In this study, we have comprehensively analyzed SLR processes and quality criteria, culminating in an integrated state-of-the-art SLR process and criteria set. Enhancing the understanding of current SLR methodologies is foundational to conducting impactful review articles (Paré et al., 2023) and establishing a baseline for integrating Gen.AI.

Gen.AI is an evolutionary step of machine learning (ML) and deep learning (DL) algorithms and has significantly transformed the landscape of academic support technologies. Initially, tools like Grammarly provided foundational

aid, primarily focusing on grammar and spelling corrections to enhance the clarity and correctness of academic writing. To this day, these tools are crucial for researchers, especially non-native English speakers, to improve their scholarly communication for clarity and understandability (Araújo et al., 2020; Mettler and Sunyaev, 2023). As computational and algorithmic capabilities advanced, the scope of these tools expanded beyond error correction. AI offers new, unique functionalities focused on analyzing, interpreting, and progressively generating natural language, but also to improve argumentative structures (Wambsganss et al., 2024). The latest stage of these developments, culminating in more sophisticated models, are referred to as Generative AI. This is possible as the underlying algorithms learn from past prompts and data, thereby refining their output to progressively become more efficient and effective (Chowdhary, 2020; Gatt and Kraemer, 2018). Particularly through the development of neural network-based models, Gen.AI tools can analyze large blocks of text, learn from extensive databases of scholarly literature, and provide contextually and academically appropriate recommendations throughout the entire research process. For instance, as early as 2017, research by Dann et al. showed that Gen.AI could outperform humans in identifying potentially relevant literature through innovative DL algorithms. Ultimately, Gen.AI is especially suitable for tasks requiring deep comprehension of language and context (Brown et al., 2020; Zhong et al., 2024), significantly extending the range of tasks that AI tools can effectively undertake (Benbya et al., 2024; Wagner et al., 2022). Hence, Gen.AI is a major advancement in computational assistance in the context of SLRs.

While the adoption of Gen.AI-assisted SLRs has grown, as evidenced by publications highlighting the potential of integrating Gen.AI into this process (Noroozi et al., 2023; Schryen et al., 2020), it is imperative for researchers to recognize that the effective use of these tools necessitates a high level of skill and domain expertise. Researchers must be capable of independently performing the tasks they delegate to Gen.AI to critically evaluate and adapt the outputs generated. If a researcher lacks the necessary skills to carry out a task manually, it becomes challenging to assess the accuracy, relevance, and reliability of the Gen.AI output.

Criteria-based implementation of Gen.AI into systematic literature reviews

Building upon our previous analyses, this section integrates the established elements to evaluate if and how Gen.AI tools might be used in the SLR process. Specifically, by mapping all quality criteria (see The state-of-the-art guidelines for structured literature reviews) against all process steps (see

The state-of-the-art process of structured literature reviews), we provide a detailed evaluation of where Gen.AI can be effectively and ethically integrated into SLRs, ensuring that the highest standards of academic rigor and integrity are maintained. A summary of our framework is presented in Table 1.

Design

Domain familiarization. Developing a thorough understanding of the relevant literature can be supported through Gen.AI tools and services trained on extensive academic literature (Dann et al., 2017; Ngwenyama and Rowe, 2024). Through their summarization and clustering capabilities, tools like Scite Assistant or Consensus can generate high-level, tabular overviews tailored to specific research questions, allowing researchers to quickly scan vast amounts of literature (Alshami et al., 2023; Benbya et al., 2024; Lim et al., 2023). However, while these tools might offer fast answers and seemingly comprehensive high-level overviews, researchers should be mindful of their inherent lack of transparency and explainability (Ngwenyama and Rowe, 2024). When using Gen.AI tools in their search process, researchers should note that these tools primarily rely on Open-Access publications and include content only up to a certain date (Lund et al., 2023). This limitation is particularly significant in fast-evolving fields like IS, and Gen.AI research in particular; it may lead to situations where authors fail to include the newest findings if they solely rely on such tools for their search. Consequently, using these tools should be followed by an extensive manual examination of the literature, ensuring one's understanding of its subtle and complex aspects to maintain the integrity and depth of the SLR (Lund et al., 2023; Ngwenyama and Rowe, 2024).

Identify need for SLR. Dependent on a comprehensive understanding of existing literature, recognizing unexplored areas and emerging trends is crucial for determining the SLR's beneficence and potential impact. Although the author should decide on the specific research focus, Gen.AI can substantially support researchers. For instance, specialized tools like Powerdrill augment general models such as GPT-4, enabling the quick inclusion and synthesis of domain-specific literature (Dann et al., 2017). These tools also critically assess the relevance of identified research gaps by incorporating lesser-known or interdisciplinary studies and serve as a feedback mechanism to direct researchers toward meaningful areas of inquiry (Dwivedi et al., 2023). However, over-relying on Gen.AI, which primarily depends on existing literature, might lead to the perpetuation of established narratives, especially in well-researched fields (Schryen et al., 2024). This highlights the necessity for researchers to critically evaluate Gen.AI outputs to avoid redundant insights and ensure the originality and innovation of their work. By employing Gen.AI,

Table I. Summary of Gen.AI assistance during the literature review process.

LR process phase and steps	Benefits to quality criteria	Risks to quality criteria	Recommendations for Gen.AI usage
(I) Design			
1. Domain familiarization	Thorough comprehension can enhance rigor	Potential biases and oversimplifications can undermine rigor and integrity	Use as starting point; complement with profound manual analysis
2. Identify need for SLR	Processing capabilities can solidify impact and comprehensiveness	Over-reliance on prevailing narratives can undermine thoroughness and originality	Critically appraise findings; complement with manual assessment
3. Prepare proposal	Alignment with established standards can enhance transparency and rigor	Over-reliance on standards can undermine originality	Use for initial suggestions; keep core synthesis researcher-led
(II) Discover			
6. Paper search	Broadened scope can benefit rigor and comprehensiveness	Potential biases can affect integrity, reliability, and transparency	Use only supplementarily; cross-reference manually
7. Backward search	Uncovered connections can enhance rigor and comprehensiveness	Potential biases and over-reliance on prevailing narratives can undermine rigor and integrity	Use for initial discovery; supplement with traditional methods
8. Forward search			
(III) Develop			
10. Extract data from papers	Streamlined analysis can enhance rigor, efficiency, responsibility	Lack of transparency can prohibit reproducibility	Verify accuracy and depth manually; ensure faithful representation of nuances
11. Synthesize and structure data	Pattern analysis can enhance rigor and efficiency	Potential biases can undermine beneficence	Use for initial pattern identification; keep core synthesis researcher-led
(IV) Disseminate			
13. Write / Present results	Clarity in writing and illustrations can increase impact	Misuse can undermine integrity and responsibility	Use for refining language, clarity, and correctness of texts and visuals; refrain from content creation

researchers can increase their efficiency in working with large amounts of literature. However, assessing if there is or is not a need for a SLR should ultimately always be done by the author. This approach ensures that the research remains impactful, beneficent, and adherent to academic standards, thus contributing significantly to the field.

Prepare proposal. When crafting research proposals, authors must align their SLRs with existing research frameworks, ensuring systematic, rigorous, and original work. Gen.AI integration can significantly enhance the structuring and conceptualization of the review. These tools are adept at aligning the review's framework with current quality standards, templates and recognized structures, thereby assisting in defining the scope and objectives of research endeavors. For example, researchers can use tools like ChatGPT-1o to generate tables or frameworks in text-based formats such as LaTeX or Mermaid. Historically, these more technical formats required specialized training. Today, Gen.AI tools allow authors to generate and adapt them using only natural language, reducing barriers and simplifying the creation of structured frameworks for their proposals. Furthermore,

authors can use Gen.AI tools to rephrase or translate their notes into a first draft for their proposal, saving time and making it easier to share with colleagues.

Prepare a review protocol. Given its instrumental role in the academic discourse (Paré et al., 2023), we advise against using Gen.AI tools in preparing a review protocol. As the process requires high precision, individual judgment, and methodological rigor, authors cannot ensure that Gen.AI judges and weighs characteristics to be reported fairly, precisely, and unbiased (Dowling and Lucey, 2023). Keeping detailed records of each step increases accuracy and facilitates critical reflection on the review process. This manual approach aligns with the quality criteria of transparency, replicability, and methodological rigor, which are fundamental to the integrity of academic research.

Define search scope. Defining the search scope involves manually selecting keywords and databases based on a deep understanding of the field and its current trends, depending strongly on the author's diligence, ethical responsibility, and transparency to ensure the reliability and trustworthiness of

the SLR findings. It is a critical task in the SLR process, as it lays the foundation for the entire review. We recommend manually selecting keywords and databases instead of relying on Gen.AI assistance, as such tools at this stage might introduce incomplete literature representations (Alshater, 2022) or subtle biases, such as the anchoring effect—where initial suggestions unduly influence subsequent decisions (Tversky and Kahneman, 1974). A manual approach ensures control and transparency, upholding the quality criteria of integrity, responsibility, and rigor in SLR findings.

Discover

Paper search. Transparency and reproducibility are essential in the SLR process, especially during the paper search step. As such, structured approaches like Boolean search queries are crucial for refining results and maintaining systematic and transparent literature searches. This ensures that all researchers can access the same data, use the same search strategies, and achieve similar results, thereby upholding the integrity of the review process.

At their core, most Gen.AI tools can be described as a natural language interface to a black-box model. As such, they neither can nor should replace academic databases as the primary resource for building the review sample. However, Gen.AI tools can improve search efficiency by assisting in the initial creation and adjustment of keyword search queries, including tailoring these queries to the syntax requirements of different databases (Wang et al., 2023). This is particularly useful since different academic databases often have slightly varying search query syntaxes and specific requirements for formatting Boolean operators, wildcards, and other search parameters, making the adaptation of the search queries both tedious and time-consuming. Furthermore, Gen.AI can assist researchers in identifying synonyms, related terms, and variations in language usage that may affect search results (Badami et al., 2023). By suggesting alternative terms and refining search queries, these tools help overcome language barriers, terminology inconsistencies, and vocabulary variations, enabling more thorough and precise literature searches (Badami et al., 2023; Min et al., 2023).

Backward and forward search. In backward and forward searches, authors need to understand and use connections between publications to find more relevant literature. Integrating visualizing services like Semantic Scholar, Research Rabbit, or Connected Papers can add a new dimension to traditional citation tracking in SLRs. Based on a set of author-selected papers, they can identify thematic and conceptual connections between papers beyond direct citation links (Wagner et al., 2022). This capability to pinpoint semantically related studies, along with providing visual maps of the literature landscape (Haddaway et al., 2022), can enable authors to gain a more comprehensive and

nuanced understanding of the subject matter, enhancing the rigor of the SLR. To illustrate, Appendix 1 shows a “Connected Papers graph.”

Moreover, Gen.AI tools can assist in automating searches, identifying relevant studies more quickly, and uncovering connections that might be missed through manual searches alone. However, ensuring the quality and relevance of the identified literature requires critical human evaluation to maintain rigor and integrity and avoid biases that might emerge from the reliance on Gen.AI’s training data (Harshvardhan et al., 2020). This approach ensures a transparent and reproducible search process, meeting critical quality criteria of academic research. Researchers can conduct a more comprehensive SLR by judiciously integrating Gen.AI’s capabilities with established, manual research techniques (Dann et al., 2017; Wagner et al., 2022).

Develop

Quality assessment of paper. Quality assessment of scientific publications involves two key steps: determining if academic works meet scholarly standards; and assessing if it contributes significantly to the research question. Whether Gen.AI can assist in achieving these goals is a much-debated question. While Wagner et al. (2022) reason that Gen.AI only possesses limited potential to assist in quality assessment, Drori and Te’eni conclude in their 2024 study that Gen.AI can indeed assist researchers in evaluating the quality of research papers by analyzing various dimensions such as contribution, soundness, and presentation. However, Gen.AI tools may inadvertently introduce biases due to limitations in their training data or algorithms, for instance, they might overemphasize certain types of research, overlook methodological nuances, or fail to capture the context-specific significance of a study. Given these limitations, we recommend not using Gen.AI tools when reviewing or assessing the quality of academic papers (Kankanhalli, 2024).

Extract data from papers. Gen.AI tools (like Humata) can support data extraction by visually connecting inferences to specific sentences and paragraphs in the sample paper. Furthermore, the structured nature of Gen.AI algorithms can ensure that data from all studies in the literature sample is extracted using an equal analysis framework, further limiting unequal representations between studies. To illustrate, Appendix 2 shows a conversation with the tool Humata about the paper by Wagner et al. (2022).

However, these tools also pose risks to these same quality criteria. Their simplicity and speed might lead to an oversimplification of complex data, potentially undermining the integrity of the analysis (Alshater, 2022). Moreover, biases in the analysis framework can lead to constitutionalized data misrepresentations (Dowling and Lucey, 2023). Additionally, the often proprietary and, thus, ultimately intransparent nature of Gen.AI’s underlying

algorithms poses a significant challenge for other researchers aiming to replicate the study (Schryen et al., 2024). Consequently, while the advantages of Gen.AI tools are compelling, researchers must exercise caution. It is crucial to rigorously check the accuracy and depth of data extracted by these tools to ensure it accurately represents the subject's detailed nuances. This careful approach balances the benefits of Gen.AI tools with the need to maintain the highest standards of transparency, reproducibility, integrity, and responsibility in academic research.

Synthesize and structure data. During the Data Synthesis phase of the SLR, researchers must critically analyze, integrate, and coherently present information from multiple sources. They are tasked with discerning key themes, interpreting data in context, and ensuring their synthesis is not only thorough but also adds original insights to the existing knowledge. Gen.AI tools with proficiency in analyzing extensive academic texts (like ResearchGPT and Consensus) provide a streamlined, efficient way to synthesize information in SLRs. Their ability to process and synthesize large data volumes far exceeds manual capacities, allowing for a more expansive and thorough analysis (Davison et al., 2023). This capability facilitates establishing extensive connections across studies, revealing patterns and insights that may be less apparent through conventional methods (Schryen et al., 2024). As a result, Gen.AI tools can significantly improve the rigor and efficiency of the analytical process. However, over-relying on Gen.AI can potentially undermine the SLR's integrity and originality, both central quality criteria. Additionally, Gen.AI's propensity to rely on existing narratives can hinder the creation of innovative contributions to the field, undermining beneficence. Therefore, we suggest using Gen.AI mainly for the initial identification of potential trends and patterns while the primary data synthesis remains under the researcher's control.

Disseminate

Report all decisions. Reporting all decisions in an SLR demands integrity, responsibility, and rigor to ensure transparency and reproducibility. Throughout the entire SLR process, we advocated for the supervised usage of Gen.AI, controlling its outcomes. Reporting Gen.AI use promotes trust in the SLR, as reviewers and readers can transparently understand the author's oversight. As a result, this reporting should be done manually, without Gen.AI assistance to maintain the SLR's integrity.

Write/present results. There is concern that authors might use Gen.AI to generate content not reflective of their own analysis or expertise (Alavi et al., 2024; Benbya et al., 2024; Else, 2023). For example, recent events at John Wiley and Sons, where over 11,300 fraudulent papers containing AI-generated

content were retracted, highlight the dangers of misusing Gen.AI (Subbaraman, 2024). This cautionary tale underscores the importance of integrity and critical evaluation when incorporating AI tools in the academic writing process. Consequently, it is crucial for authors to understand that they bear the ultimate responsibility for their work, both overall and within each sub-step. Any contributions from third parties, including co-workers or Gen.AI tools, must be critically evaluated and appropriately disclosed. Still, when used properly, tools (like Quillbot) can play a pivotal role in refining the writing style, ensuring the text is clear and concise, and, thus, promoting its impact (Dwivedi et al., 2023). Furthermore, Gen.AI tools can support authors in adhering to structural requirements in writing, such as formatting, reference, and styling (Kankanhalli, 2024). We encourage authors to use Gen.AI to enhance the language, clarity, and correctness of their texts and visual aids, yet refrain from letting Gen.AI generate content.

Criteria-based usage of Gen.AI assistance in structured literature reviews: A guide and illustration

The integration of Gen.AI in conducting SLRs is permissible only when it adheres to established academic rigor and integrity standards. To understand the current status quo of Gen.AI assistance in SLRs, we initially surveyed 20 researchers¹ of varying seniority levels about their current use of Gen.AI. Following this, we enhanced the existing process by incorporating best practices from other research disciplines. This effort resulted in the formulation of an eight-step process for employing Gen.AI assistance in SLRs.

All researchers we interviewed presented a similar status quo of how they used Gen.AI tools in SLRs. First, they identified the specific need for assistance, for instance, not finding relevant papers during their forward and backward searches. Then, all interviewed scholars would go to their preferred Gen.AI tool, which was most often ChatGPT, and execute the task (e.g., "Show me relevant papers for xy"). Most interviewees, though not all, then assessed the AI output and revised it if necessary. Subsequently, all researchers would integrate the revised output into their manuscripts.

While using Gen.AI can potentially enhance the quality of manuscripts, we identified three major challenges through interviews and our prior analysis. First, the Gen.AI tools currently available may not adequately meet researchers' specific needs. For example, ChatGPT lacks integration with scientific databases, which limits its utility in assisting with literature searches. Second, because of their black-box nature, the training data and algorithms of Gen.AI tools are usually not publicly available information. As such, researchers are often unable to understand why a tool produces specific content in response to queries about relevant publications. This intransparency makes it difficult for researchers to trust the relevance and accuracy of the output. Third, the generative nature of these tools means that they can produce

different outputs from the same prompt. For instance, using co-pilot to improve the same sentence repeatedly might yield varying results despite the identical input. This inconsistency can be problematic for researchers, undermining transparency, reproducibility, and third-party comprehensibility. To mitigate these three problems, we propose to change the status quo process in three ways: *Tool-Task fit*, *Between-Tool comparison*, and *Within-Tool comparison*.

Tool-task fit. To address the tool-task fit, we introduce a step where researchers familiarize themselves with the existing Gen.AI tools and their capabilities to make an informed decision on which tool fits their needs best.

Between-tool comparison. To mitigate the intransparency and potential biases of Gen.AI tools, we propose to use several Gen.AI tools to execute a given task and then compare the outputs. This between-tool comparison is similar to data triangulation, which is a proven scientific method to address the intransparency and potential biases of data sources (Creswell et al., 2003; Greene et al., 1989; Tashakkori and Teddlie, 2003; Venkatesh et al., 2013).

Within-tool comparison. To mitigate data quality concerns that stem from the generative nature of Gen.AI tools, we propose that Gen.AI tools should be used repeatedly with the same prompt. Using the same tool repeatedly allows for a within-tool comparison to ensure that the data extracted is all the relevant information the tool can produce (Glickman and Zhang, 2024). This approach is rooted in grounded theory (Glaser and Strauss, 1968), which introduced the concept of saturation—a point where further inquiry is unlikely to lead to the emergence of more relevant information (Guest et al., 2006).

In summary, Gen.AI assistance in SLR poses problems already seen in different research contexts (e.g., grounded theory, data triangulation). Consequently, our proposed mitigative strategies come from transferring validated and proven approaches to tackle these problems to the context of Gen.AI-assisted SLRs. This leads us to the eight-step guide for using Gen.AI for scientific task assistance, outlined in Figure 4.

In the following, we describe each step in detail. To increase the understandability and usability of our eight-step guide, we illustrate its application in an SLR example (Rathje et al., 2024). While we provide the comprehensive guide in a separate online appendix, we also want to briefly

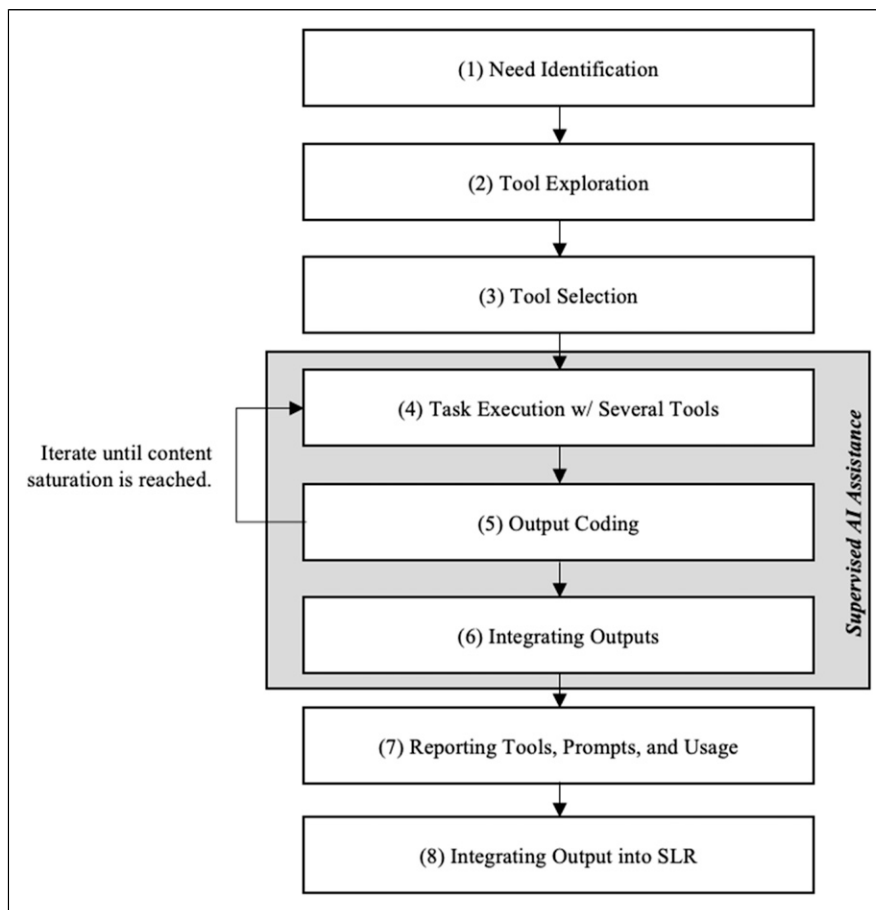


Figure 4. Eight-step guide for using Gen.AI for scientific task assistance.

illustrate the guide's application in this manuscript. Please note that we provide screenshots and further background information in the [online appendix](#) for every step.

The situation. Suppose we want to conduct a structured literature review about a novel topic that many colleagues told us about at a recent IS conference: how the metaverse reshapes B2C online retailing. Some researchers highlighted the necessity for a structured literature review on this topic (e.g., Hadi et al., 2023; Tingelhoff et al., 2024). While we have done research in similar areas already (like digital platform ecosystems), we lack the nuanced understanding of the different concepts (e.g., the retailing metaverse) necessary to conduct an SLR. Hence, to gain the most up-to-date and relevant knowledge in this area, we want to use Gen.AI assistance for the purpose of domain familiarization and deploy the eight-step guide.

Balancing AI assistance and manual work. As highlighted in this paper repeatedly, it is of utmost importance to critically assess AI-generated content. To do so, it is imperative to conduct some of the work manually and use it as a baseline to evaluate the Gen.AI outputs. So, we research FT-50 journals for recent publications on retailing in the metaverse and find two relevant papers (Hadi et al., 2023; Yoo et al., 2023). We read and analyze both papers, especially marking the publications they cite. The resulting understanding and paper database is to serve as a standard to evaluate Gen.AI tools. In other words, the outputs should not contradict the understanding of the two papers and should contain at least some of their references. With this basic understanding, we start using Gen.AI in accordance with the eight-step guide.

Need identification. Researchers need to first clearly define the specific task for which Gen.AI assistance is required and identify the necessary tool functionalities (e.g., "tool must have an integrated scientific database," specific filter criteria). Furthermore, researchers should determine tool characteristics that motivate choosing one tool over the other.

The question we seek AI assistance for is: How can organizations use the metaverse in the retail context? As a hygiene criterion, Gen.AI tools must integrate and reference a scientific database, as we need to build a reliable, peer-reviewed, scientific understanding of our domain. Additionally, as our research

question is open-ended, we can only include tools that allow for asking such questions. Moreover, to understand the context of the topic, tools' answers must summarize how the papers relate to the RQ. Furthermore, there are some motivators for selecting one capable tool over another. For instance, we know that the metaverse concept changed significantly over the past 20 years (Peukert et al., 2022). Consequently, filter criteria (i.e., publication year) would help us to generate a consistent, up-to-date conceptualization. Additionally, some tools output tabular overviews for papers they find, including columns for research methods, findings, and calls for future research (e.g., see Online Appendix Table 1). Such tables could help us to better understand the nuances between research teams' conceptualizations. In summary, our hygiene criteria are a scientific database, interfaces for open-ended questions, and summary outputs, while our motivators are filter options and tabular overviews.

Tool exploration. Next, researchers should familiarize themselves with all the Gen.AI tools. This includes experimenting with the tools occasionally to understand the tools' characteristics, like data cut-off dates, biases, or plug-ins. While researchers do not need to do this step daily, Gen.AI tools show rapid developments that should prompt researchers to update their tool knowledge from time to time. As official benchmarking reports are often not available for Gen.AI tools but only their underlying algorithmic models, we found community discussions on public forums, such as Reddit or the OpenAI user forum, particularly useful for this purpose. Not only are they usually more up-to-date than official tool websites, but they also provide more in-depth information as actual users on particular subjects and evaluations provide the data.

After understanding what we need from a tool, we next research potential Gen.AI tools. For this, we initially query "best Gen.AI tools for literature reviews" in Google. Furthermore, we go into online forums about that topic to learn from the experiences of other researchers. For example, we go to the Reddit thread titled "AI tools I found useful w/ research" in the community r/PHD. Lastly, we go to the website Future Tools, which offers an extensive list of up-to-date AI tools, which one can filter for use cases to find potentially fitting tools. This search results in eleven potential Gen.AI tools: Scite, Consensus, Elicit,

SciSpace, Semantic Scholar, IrisAI, Copy AI, ChatGPT, Research Rabbit, ChatPDF, and Scholarcy.

Tool selection. After identifying potential tools, authors should assess which of the identified tools fit their needs best. We propose comparing the tools in a table to select the most fitting tools. Authors should then move to the next steps with the top three to five tools that best fit their needs.

Next, we create a table in which each column represents one of these tools and each row one of the previous four criteria. We then investigate each individual tool and assess whether it meets each respective criterion. We provide the table in our [Online Appendix Table 1](#). Some of the eleven tools did not fulfill our hygiene factors. For instance, ChatGPT had no access to a scientific database, whereas ResearchRabbit and ChatPDF could not be asked open-ended questions, and Semantic Scholar did not summarize papers. Ultimately, we end with four Gen.AI tools that fulfilled our hygiene criteria.

Based on our previous analysis, we next decide which of the remaining four tools we want to use for our task. To decide, we input our research question into each tool and observed the outputs. While three of the tools are quite similar, Consensus gave very long answers, often rambling on about topics that are only partly linked to our research question. Consequently, we decide to use the other three tools—namely, Scite, Elicit, and SciSpace.

Task execution. Next, researchers should develop a prompt to execute the task with each chosen tool. All tools should be used with the same prompt to enable a between-tool comparison (Glickman and Zhang, 2024). It is pivotal to understand that the quality of the Gen.AI output is a function of AI literacy and prompting skills of the researcher (Knoth et al., 2024). Hence, for this step, researchers might benefit from consulting the latest prompt engineering literature to further improve the tools' output quality.

Next, we ask our open-ended question (i.e., “How can organizations use the metaverse in the retail context?”) to the three tools we selected. We report the outputs as screenshots in the [Online Appendix Figures 4-14](#) and a text-based excerpt in [Table 2](#).

Output coding. After execution, the authors should analyze the outputs from all tools. To do so, we advocate for

the grounded theory approach (Glaser and Strauss, 1968): researchers should inductively identify and code statements in the tools' outputs, grouping similar statements, resulting in an exhaustive overview of newly acquired content.

Iterations. Next, to accommodate the generative nature of Gen.AI, authors should evaluate if an additional iteration of steps 4 and 5 will likely result in new relevant information. We propose to authors to seek guidance from grounded theory as well (Guest et al., 2006). While we cannot definitively answer how many iterations researchers should run, we suggest similar thresholds to grounded theory, where, if a new interview resulted in less than, for instance, 5% of new codes, it is unlikely another interview would surface additional critical data (Braun and Clarke, 2021). As we have already suggested using coding mechanisms to make sense of outputs, using coding saturation to identify content saturation seems logical. Furthermore, the iterations enable researchers to gain a more thorough understanding of the frequency in which certain codes are mentioned across various Gen.AI tools and iterations. While a low frequency could represent an interesting and potentially under-reported research stream, it could also be a hallucination produced by the model. Thus, we strongly advise to always conduct manual checks in such cases. Ultimately, this procedure balances exploiting the Gen.AIs' knowledge with ensuring that researchers are not endlessly iterating without meaningful improvement.

Next, we begin inductive coding. To visually guide our progress, we color code: yellow for new codes and gray for code repetitions. Following the eight-step process, we repeatedly execute steps 4 and 5 with the same prompt to reach coding saturation. The first round yields, obviously, 100% new codes. We decide to finish the iterations after the fifth round, where only 5% of new codes emerge, indicating coding saturation. We provide an illustrative excerpt in [Table 2](#) alongside the coded tool outputs.

Integrating outputs. After this iterative process, authors are confronted with separate outputs from several tools (equal to the number of tools times the number of iterations). As this amount of dispersed knowledge might be overwhelming, we follow the grounded theory approach. As authors have already coded tool outputs in step 5, we suggest they should inductively combine similar codes across all outputs into higher-level, second-order themes. If the number of second-order themes is still unmanageably high, authors might want to combine them into aggregate dimensions. Ultimately, authors can then use these newly created knowledge graphs to structure the tool outputs based on content, generating one holistic, integrated overview of newly acquired knowledge.

Table 2. Example – excerpt of coded tool outputs (Yellow = New Code, Gray = Code Repetition); full coding in [online appendix](#).

Iter.	Elicit	SciSpace	SciteAssistant	Satura- tion
1	The metaverse offers a unique platform for retailers to create immersive and interactive experiences for consumers (Papagiannidis, 2010). This can be achieved by focusing on elements such as customer service, product dimension, store dimension, and 3D platform dimension (Gadalla, 2013). However, the ethical implications of data collection and usage in the metaverse must be carefully considered to ensure business sustainability (Anshari, 2022). Fashion retailing, in particular, can benefit from the metaverse by creating a more stimulating and immersive experience for consumers (Weiss, 2022).	Organizations can leverage the metaverse in the retail context by creating immersive and interactive experiences for consumers [1][2]. This involves utilizing virtual events, product demonstrations, and unique shopping experiences through digital avatars [3][4]. The metaverse offers a platform for brands to connect with consumers in novel ways, enhancing customer support, public relations, and post-purchase experiences [5]. By integrating the metaverse into their marketing strategies, businesses can reach their target audience effectively, combining the accessibility of the digital world with the immersive experiences of the physical world . Furthermore, the metaverse presents opportunities for selling digital goods and enhancing retail experiences in a 3D virtual world.	Organizations can leverage the metaverse in the retail context by focusing on creating immersive and experiential retail environments . The metaverse offers a platform where retail can transition from product-oriented to experience-oriented , enhancing customer engagement and interaction (Dwivedi et al., 2022). By embracing metaverse technologies , retailers can innovate their marketing strategies , offering consumers a "retail theater" experience that combines virtual and physical elements (Park, 2023). This shift towards experience-oriented retailing is seen as an evolution from traditional electronic retailing , emphasizing the importance of customer experience (Papagiannidis et al., 2013). Additionally, the metaverse presents opportunities for seamless integration , customer engagement, and marketing strategies, ushering in a new era for retail with boundless possibilities for those who embrace it (Singha and Mukthar, 2024).	100% new (28)
...
5	The metaverse offers a unique platform for retailers to create immersive and interactive experiences for consumers (Papagiannidis, 2010). This can be achieved by focusing on elements such as customer service, product dimension, store dimension, and 3D platform dimension (Gadalla, 2013). However, the ethical implications of data collection and usage in the metaverse must be carefully considered to ensure business sustainability (Anshari, 2022). Fashion retailing, in particular, can benefit from the metaverse by creating a more stimulating and immersive experience for consumers (Weiss, 2022).	Organisations can leverage the metaverse in the retail context by enhancing customer support, public relations, and customer experiences [1]. The metaverse provides a platform for brands to engage with consumers through immersive experiences, virtual events, and product demonstrations, fostering meaningful interactions and driving brand loyalty [2]. By creating virtual avatars and offering unique shopping experiences, brands can connect with consumers in novel ways, facilitating the sale of digital goods within the metaverse [3]. Furthermore, integrating the metaverse into marketing strategies allows businesses to reach their target audience effectively, offering a seamless blend of digital accessibility and physical-world immersive experiences, ensuring long-term brand existence and growth [4]. Additionally, as the metaverse evolves rapidly, organisations must strategize multichannel and omnichannel routes to navigate the complex platform setup and maximize marketing opportunities [5].	Organizations can leverage the potential of the metaverse in the retail context in various ways. The metaverse offers opportunities for innovation and adaptation in retailing, particularly during times of upheaval like the COVID-19 pandemic (Singha, 2024). By utilizing the metaverse, retailers can enhance the shopping experience for customers through personalized services and engaging events that captivate consumer attention (Jin, 2024). This can lead to an optimized overall shopping experience and increased customer satisfaction . Moreover, the adoption of the metaverse in retail has the potential to impact various critical areas such as marketing, supply chain management, and transportation management (Tan et al., 2023). Retailers can benefit from the metaverse by improving product verification capabilities, expanding online shopping experiences, and ensuring traceability in the supply chain (Melnichenko, 2021; Jaouhari, 2024). Additionally, the metaverse can enhance organizational efficiency and performance in the retail sector ("Exploring the unknown", 2024). Furthermore, the metaverse can be utilized to create virtual retail spaces that offer benefits beyond conventional e-commerce, providing customers with a more immersive and interactive shopping experience (Šimová et al., 2023). Retailers can explore metaverse applications that go beyond traditional retailing, considering aspects like supply chain management to enhance operational effectiveness (Queiroz et al., 2023). By embracing the metaverse, organizations in the retail sector can stay ahead of the curve and meet the evolving demands of consumers in a digitally transformed world .	5% new (3 of 66)

After the initial round and four iterations of executing and prompting tool outputs, we are left with 66 unique codes. Then, to better group similar information, we categorize these codes into 13 second-order themes and four aggregate dimensions. While we could have used Gen.AI to synthesize and integrate codes (Benbya et al., 2024), we decided against it to not further complicate this illustrative, exemplary process application. We provide an excerpt of the resulting coding framework in Figure 5, which integrates the information of the several tool outputs into one exhaustive overview of newly acquired content.

Reporting tools, prompts, and their usage. After having generated the integrated output, authors are done working directly with the Gen.AI tools and their outputs. Then, authors must thoroughly and transparently report their AI usage, including details of the tools, prompts, and iterations, along with observations on tool performance.

As we have advocated throughout this paper, it is imperative for authors to transparently and rigorously document their Gen.AI usage. As there is no standard on diligently reporting such use, we choose to report a synthesis of our six previous steps, mainly: our research question, which tools we used, our

prompt, the table showing the color-coded tool outputs of each iteration, and our coding framework. We choose these types of information as they inherently make sense to us and encompass most of our Gen.AI interactions. However, this might be different in other steps of the SLR (e.g., the paper search). The online appendix also serves as our Gen.AI report.

Integrating output into SLR. Finally, the iteratively refined output—be it a list of relevant papers, a rephrased text, or else—needs to be incorporated into the manuscript. This critical step must be carried out with utmost responsibility, ensuring the ethical and legal integrity of AI-generated content. Due to this step’s importance, authors should manually integrate this content. This method lets authors control the final content, ensuring it accurately represents their research intentions and findings.

Ultimately, we must use the integrated output to advance our SLR process. In our case, we are now much more knowledgeable in the domain of organizational adoption of retail metaverses. We can use

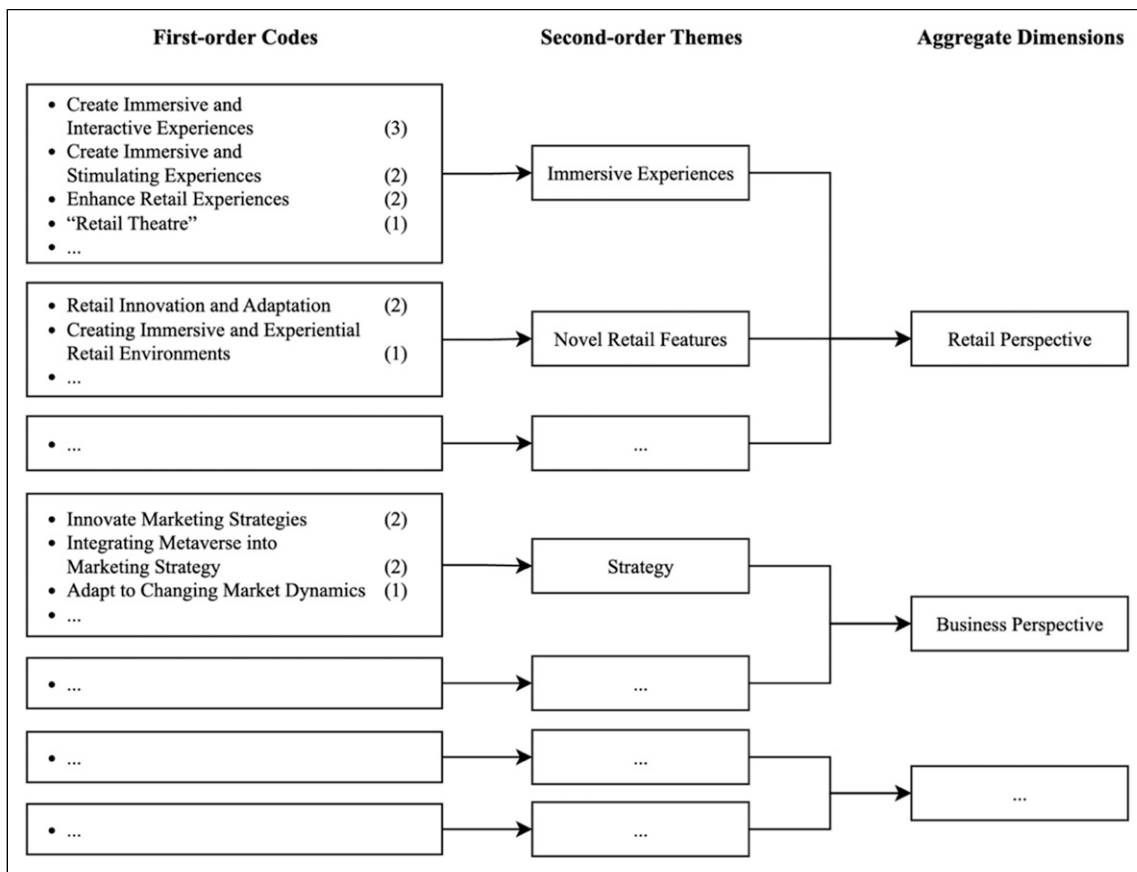


Figure 5. Example – excerpt of final coding framework from the integration of Gen.AI outputs; full Figure in [online appendix](#).

this information to better justify the need for our SLR, to better craft the keywords, and to better embed and discuss our SLR findings in the research domain. While, of course, the later full-text analysis of our sample will further deepen our understanding, using the eight-step guide helped us to rigorously and transparently seek Gen.AI assistance during the domain familiarization.

Tough questions to ask: Control versus contribution

Throughout this study, we have operated under implicit assumptions regarding the fundamental nature of SLRs. We presumed that the primary purpose of SLRs is to contribute novel insights to the academic discourse (the *why* of conducting SLRs) and that established quality criteria serve as robust, sustainable guidelines throughout the review process (the *how* of conducting SLRs). While these assumptions have historically been largely accepted, the rapid

emergence of Gen.AI necessitates a critical re-examination of these foundational beliefs (Berger, 2024).

The academic community is presently engaged in vigorous debates about integrating Gen.AI into research methodologies (e.g., Banker et al., 2024a, 2024b; Berger, 2024; Hermida Carrillo et al., 2024), including SLRs (e.g., Schryen et al., 2024; Wagner et al., 2022). Traditionally, we have placed immense value on key aspects of SLRs: structuring existing research with transparency and reproducibility; integrating information with rigor and adherence to scholarly standards; and disseminating findings effectively to maximize impact and beneficence. However, as demonstrated in our study, Gen.AI systems are increasingly capable of mimicking—and potentially surpassing—these aspects. Simultaneously, the fraction of open-access articles is continuously growing (Björk, 2017; Laakso et al., 2011; Seo, 2023) and will likely become the default publication class (Piwowar et al., 2019). This prompts pressing questions: If SLRs become partially or fully automatable, what then distinguishes an exemplary SLR? Is the hallmark of a great SLR rooted in the researcher’s unique contribution, or can it be attributed to the efficiency and comprehensiveness of AI-generated outputs?

Extending this inquiry further, we confront a pivotal dilemma: If Gen.AI could perform SLRs faster, more accurately, and on a larger scale than human researchers, should we continue to entrust the SLR process to human oversight? Current Gen.AI models often operate as opaque “black boxes,” lacking transparency and interpretability—a characteristic that may persist or evolve unpredictably in the future (Drori & Te’eni, 2024; Kankanhalli, 2024). This opacity challenges our ability to ensure accountability, reproducibility, and ethical integrity in research. Consequently, the academic community must grapple with a fundamental question: Do we prioritize the contribution of knowledge, irrespective of its source, or do we value maintaining control over the process of knowledge generation?

These profound questions extend beyond the scope and intent of our current study. Our objective has been to propose a balanced approach that navigates the spectrum between complete reliance on Gen.AI and exclusive human control. We have introduced a framework that enables researchers to leverage the transformative potential of Gen.AI in conducting SLRs while retaining essential oversight over the Gen.AI’s outputs and their integration into the scholarly narrative.

It is crucial to acknowledge that SLRs represent just one facet of research methodology. Their structured nature makes them particularly susceptible to AI integration, positioning them as an ideal initial use case within the scientific domain. However, the fundamental challenges and ethical considerations posed by all types of AI will inevitably confront scholars across all methodological paradigms. While our guide specifically pertains to SLRs in business research, we believe it can potentially be applied in other research settings as well. For instance, researchers conducting a quantitative study could potentially refer to our eight-step guide to using Gen.AI assistance when developing an experimental design. Moreover, aspects of our paper, such as the domain familiarization or paper search, can be relevant to researchers writing related-work sections. However, since we created the eight-step guide in the context of SLRs, future research is necessary to show how it needs to be modified for other application contexts. Ultimately, SLRs serve as the initial battleground—a microcosm that will shape broader paradigms of research practice in an era increasingly influenced by AI technologies.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fabian Tingelhoff  <https://orcid.org/0000-0002-7235-6001>

Supplemental Material

Supplemental material for this article is available online.

Note

1. The researchers include five professors (full, associate, and assistant), five post-docs, and ten PhD-students. Eight researchers work in Switzerland, eight in Germany, two in the Netherlands, one in the United Kingdom, and one in Australia. The interviews were conducted informally and solely concerned the participants’ current AI usage in their scientific work.

References

- Alavi M, Leidner DE and Mousavi R (2024) Knowledge management perspective of generative artificial intelligence. *Journal of the Association for Information Systems* 25(1): 1–12.
- Alshami A, Elsayed M, Ali E, et al. (2023) Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions. *Systems* 11(7): 351.
- Alshater M (2022) Exploring the role of artificial intelligence in enhancing academic performance: a case study of ChatGPT. *SSRN* (working paper): 1–22. 4312358.
- Anderson MS, Ronning EA, Devries R, et al. (2010) Extending the Mertonian norms: scientists’ subscription to norms of research. *The Journal of Higher Education* 81(3): 366–393. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995462/pdf/nihms250315.pdf>.
- Anshari M, Syafrudin M, Fitriyani NL, et al. (2022) Ethical Responsibility and Sustainability (ERS) development in a metaverse business model. *Sustainability* 14(23): 15805. DOI: [10.3390/su142315805](https://doi.org/10.3390/su142315805).
- Antons D, Breidbach CF, Joshi AM, et al. (2023) Computational literature reviews: method, algorithms, and roadmap. *Organizational Research Methods* 26(1): 107–138.
- Araújo M, Pereira A and Benevenuto F (2020) A comparative study of machine translation for multilingual sentence-level sentiment analysis. *Information Sciences* 512: 1078–1102.
- Badami M, Benattallah B and Baez M (2023) Adaptive search query generation and refinement in systematic literature review. *Information Systems* 117: 102231.
- Banker S, Chatterjee P, Mishra H, et al. (2024a) The future of large language models in social science research: Reply to Berger (2024) and Carrillo et al.(2024). *American Psychologist* 79(6): 803–804.
- Banker S, Chatterjee P, Mishra H, et al. (2024b) Machine-assisted social psychology hypothesis generation. *American Psychologist* 79(6): 789–797. DOI: [10.1037/amp0001222](https://doi.org/10.1037/amp0001222).

- Bawden D and Robinson L (2015) *Introduction to Information Science*. London: Facet Publishing.
- Benbya H, Strich F and Tamm T (2024) Navigating generative artificial intelligence promises and perils for knowledge and creative work. *Journal of the Association for Information Systems* 25(1): 23–36.
- Berger J (2024) Machines, psychology, and hypothesis generation: Commentary on Banker et al. (2024). *American Psychologist* 79(6): 798–799. DOI: [10.1037/amp0001258](https://doi.org/10.1037/amp0001258).
- Björk B-C (2017) Growth of hybrid open access, 2009–2016. *PeerJ* 5: e3878. DOI: [10.7717/peerj.3878](https://doi.org/10.7717/peerj.3878).
- Braun V and Clarke V (2021) To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative Research in Sport, Exercise and Health* 13(2): 201–216.
- Brown T, Mann B, Ryder N, et al. (2020) Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33: 1877–1901.
- Chowdhary KR (2020) Natural language processing. *Fundamentals of Artificial Intelligence*. New Delhi, India: Springer Nature India Private Limited, 603–649.
- Comstock G (2012) *Research Ethics: A Philosophical Guide to the Responsible Conduct of Research*. Cambridge: Cambridge University Press.
- Cram WA, Templier M and Paré G (2020) (Re) considering the concept of literature review reproducibility. *Journal of the Association for Information Systems* 21(5): 1103–1114.
- Creswell JW, Clark VLP, Gutmann ML, et al. (2003) Advanced mixed. In: Tashakkori A and Teddlie C (eds) *Handbook of Mixed Methods in Social & Behavioral Research*. Newcastle upon Tyne: Sage, 209.
- Cronin B, Snyder HW, Rosenbaum H, et al. (1998) Invoked on the web. *Journal of the American Society for Information Science* 49(14): 1319–1328.
- Cropanzano R (2009). Writing nonempirical articles for journal of management: general thoughts and suggestions. *Journal of Management* 35: 1304–1311.
- Dann D, Hauser M and Hanke J (2017) Reconstructing the giant: automating the categorization of scientific articles with deep learning techniques. *Proceedings der* 13: 1538–1549.
- Davison RM, Laumer S, Tarafdar M, et al. (2023) Pickled eggs: generative AI as research assistant or co-author? *Information Systems Journal* 33(5): 989–994.
- Dowling M and Lucey B (2023) ChatGPT for (finance) research: the Bananarama conjecture. *Finance Research Letters* 53: 103662.
- Drori I and Te'eni D (2024) Human-in-the-Loop AI reviewing: feasibility, opportunities, and risks. *Journal of the Association for Information Systems* 25(1): 98–109.
- Dwivedi YK, Hughes L, Wang Y, et al. (2022) Metaverse marketing: how the metaverse will shape the future of consumer research and practice. *Psychology and Marketing* 40: 750–776. DOI: [10.1002/mar.21767](https://doi.org/10.1002/mar.21767).
- Dwivedi YK, Kshetri N, Hughes L, et al. (2023) “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management* 71: 102642.
- El Jaouhari A, Arif J, Jawab F, et al. (2024) Unfolding the role of metaverse in agri-food supply chain security: current scenario and future perspectives. *International Journal of Food Science and Technology* 59(5): 3451–3460.
- Else H (2023) Abstracts written by ChatGPT fool scientists. *Nature* 613(7944): 423.
- Exploring the Unknown (2024) Exploring the unknown: Using the metaverse within the tourism industry. *Strategic direction* 40(4): 28–29. DOI: [10.1108/sd-03-2024-0034](https://doi.org/10.1108/sd-03-2024-0034).
- Fink A (2019) *Conducting Research Literature Reviews: From the Internet to Paper*. Newcastle upon Tyne: Sage Publications.
- Fisch C and Block J (2018) Six tips for your (systematic) literature review in business and management research. *Management Review Quarterly* 68: 103–106.
- Gadalla E, Keeling K and Abosag I (2013) Metaverse-retail service quality: a future framework for retail service quality in the 3D internet. *Journal of Marketing Management* 29(13-14): 1493–1517. DOI: [10.1080/0267257X.2013.835742](https://doi.org/10.1080/0267257X.2013.835742).
- Gatt A and Kraemer E (2018) Survey of the state of the art in natural language generation: core tasks, applications and evaluation. *Journal of Artificial Intelligence Research* 61: 65–170.
- Glaser BG and Strauss AL (1968) *Time for Dying*. Piscataway, NJ: AldineTransaction.
- Glickman M and Zhang Y (2024) AI and generative AI for research discovery and summarization. *Harvard Data Science Review* (Author Accepted Manuscript): 1–34.
- Greene JC, Caracelli VJ and Graham WF (1989) Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis* 11(3): 255–274.
- Gregor S (2024) Responsible artificial intelligence and journal publishing. *Journal of the Association for Information Systems* 25(1): 48–60.
- Guest G, Bunce A and Johnson L (2006) How many interviews are enough? An experiment with data saturation and variability. *Field Methods* 18(1): 59–82. DOI: [10.1177/1525822X05279903](https://doi.org/10.1177/1525822X05279903).
- Haddaway NR, Grainger MJ and Gray CT (2022) Citationchaser: a tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research Synthesis Methods* 13(4): 533–545. <https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/jrsm.1563?download=true>.
- Hadi R, Melumad S and Park ES (2023b) The metaverse: a new digital frontier for consumer behavior. *Journal of Consumer Psychology* 34(1): 142–166. DOI: [10.1002/jcpy.1356](https://doi.org/10.1002/jcpy.1356).
- Hadi MU, Qureshi R and Shah A (2023) Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Tech Rxiv* (working paper): 1–39. 23589741.
- Hahn R (2024) Garbage in, garbage out? How to avoid citing from questionable sources in systematic literature reviews.

- Literature-based Research*. <https://www.linkedin.com/pulse/garbage-out-how-avoid-citing-from-questionable-sources-sven-kunisch-q3p4e/>.
- Harshvardhan G, Gourisaria MK, Pandey M, et al. (2020) A comprehensive survey and analysis of generative models in machine learning. *Computer Science Review* 38: 100285.
- Hermida Carrillo A, Stachl C and Talaifar S (2024) A workflow for human-centered machine-assisted hypothesis generation: Commentary on Banker et al.(2024). *American Psychologist* 79(6): 800–802.
- Jarvenpaa S and Klein S (2024) New Frontiers in information systems theorizing: human-gAI collaboration. *Journal of the Association for Information Systems* 25(1): 110–121.
- Jin D (2024) Humanizing Metaverse: psychological involvement and masstige value in retail versus tourism platforms. *International Journal of Consumer Studies* 48(2): e13025.
- Kankanhalli A (2024) Peer review in the age of generative AI. *Journal of the Association for Information Systems* 25(1): 76–84.
- Knonth N, Tolzin A, Janson A, et al. (2024) AI literacy and its implications for prompt engineering strategies. *Computers & Education: Artificial Intelligence* 6: 100225.
- Kummer T-F, Leimeister JM and Bick M (2012) On the importance of national culture for the design of information systems. *Business & Information Systems Engineering* 4: 317–330.
- Laakso M, Welling P, Bukvova H, et al. (2011) The development of open access journal publishing from 1993 to 2009. *PLoS One* 6(6): e20961. DOI: [10.1371/journal.pone.0020961](https://doi.org/10.1371/journal.pone.0020961).
- Levy Y and Ellis TJ (2006) A systems approach to conduct an effective literature review in support of information systems research. *Informing Science* 9: 181–212.
- Lim WM, Gunasekara A, Pallant JL, et al. (2023) Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *International Journal of Management in Education* 21(2): 100790.
- Lund BD, Wang T, Mannuru NR, et al. (2023) ChatGPT and a new academic reality: artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing. *Journal of the Association for Information Science and Technology* 74(5): 570–581.
- Macfarlane B (2010) *Researching with Integrity: The Ethics of Academic Enquiry*. Milton Park: Routledge.
- Melnychenko O (2021) The prospects of retail payment developments in the metaverse. *Virtual Economics* 4(4): 52–60.
- Mertens DM and Ginsberg PE (2009) *The Handbook of Social Research Ethics*. Newcastle upon Tyne: Sage.
- Mettler T and Sunyaev A (2023) Are we on the right track? an update to Lyytinen et al.'s commentary on why the old world cannot publish. *European Journal of Information Systems* 32(2): 263–276.
- Min B, Ross H, Sulem E, et al. (2023) Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Computing Surveys* 56(2): 1–40.
- Mortenson MJ and Vidgen R (2016) A computational literature review of the technology acceptance model. *International Journal of Information Management* 36(6): 1248–1259.
- Nguyen QN, Sidorova A and Torres R (2022) Artificial intelligence in business: a literature review and research agenda. *Communications of the Association for Information Systems* 50(1): 7.
- Ngwenyama O and Rowe F (2024) Should we collaborate with AI to conduct literature reviews? Changing epistemic values in a flattening world. *Journal of the Association for Information Systems* 25(1): 122–136.
- Noroozi M, Moghaddam HR, Shah A, et al. (2023) An AI-assisted systematic literature review of the impact of vehicle automation on energy consumption. *IEEE Transactions on Intelligent Vehicles* 8: 3572–3592.
- Okoli C (2015) A guide to conducting a standalone systematic literature review. *Communications of the Association for Information Systems* 37: 879–910.
- Okoli C and Schabram K (2015). A guide to conducting a systematic literature review of information systems research.
- Papagiannidis S and Bourlakis MA (2010) Staging the new retail drama: at a metaverse near you. *Journal of Virtual Worlds Research* 2(5): 425–446.
- Papagiannidis S, Pantano E, See-To EW, et al. (2013) Modelling the determinants of a simulated experience in a virtual retail store and users' product purchasing intentions. *Journal of Marketing Management* 29(13-14): 1462–1492.
- Paré G, Wagner G and Prester J (2023) How to develop and frame impactful review articles: key recommendations. *Journal of Decision Systems* 33: 566–582.
- Park H and Lim RE (2023) Fashion and the metaverse: clarifying the domain and establishing a research agenda. *Journal of Retailing and Consumer Services* 74: 103413.
- Peukert C, Weinhardt C, Hinz O, et al. (2022) Metaverse: how to approach its challenges from a BISE perspective. *Business & Information Systems Engineering* 64: 401–406. DOI: [10.1007/s12599-022-00765-9](https://doi.org/10.1007/s12599-022-00765-9).
- Piowar H, Priem J and Orr R (2019) The Future of AOA: a large-scale analysis projecting open access publication and readership. *bioRxiv* (working paper): 1–36. 795310.
- Queiroz MM, Fosso Wamba S, Pereira SCF, et al. (2023) The metaverse as a breakthrough for operations and supply chain management: implications and call for action. *International Journal of Operations & Production Management* 43(10): 1539–1553.
- Rathje J, Katila R and Reineke P (2024) Making the most of AI and machine learning in organizations and strategy research: supervised machine learning, causal inference, and matching models. *Strategic Management Journal* 45(10): 1926–1953.
- Resnik DB (2005) *The Ethics of Science: An Introduction*. Milton Park: Routledge.
- Resnik DB, Rasmussen LM and Kissling GE (2015) An international study of research misconduct policies. *Accountability in Research* 22(5): 249–266. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4449617/pdf/nihms691915.pdf>.

- Rowley J and Slack F (2004) Conducting a literature review. *Management Research News* 27(6): 31–39.
- Sauer PC and Seuring S (2023) How to conduct systematic literature reviews in management research: a guide in 6 steps and 14 decisions. *Review of Managerial Science* 17: 1899–1933.
- Schryen G, Wagner G, Benlian A, et al. (2020) A knowledge development perspective on literature reviews: validation of a new typology in the IS field. *Communications of the AIS* 46: 134–186.
- Schryen G, Marrone M and Yang J (2024). Adopting generative AI for literature reviews: an epistemological perspective. In: 57th Hawaii international conference on system science, Oahu, Hawaii, USA, 3 January 2024.
- Science.org (2023) *Scienceadviser: Scientists are Publishing Too Many Papers—and That’s Bad for Science*. <https://www.science.org/content/article/scienceadviser-scientists-are-publishing-too-many-papers-and-s-bad-science>.
- Seo J-W (2023) Changes in the absolute numbers and proportions of open access articles from 2000 to 2021 based on the web of science core collection: a bibliometric study. *Science Editing* 10(1): 45–56. DOI: [10.6087/kcse.296](https://doi.org/10.6087/kcse.296).
- Shamoo AE and Resnik DB (2009) *Responsible Conduct of Research*. Oxford: Oxford University Press.
- Šimová T, Zychová K and Fejfarová M (2024) Metaverse in the virtual workplace. *Vision* 28(1): 19–34.
- Singha S and Mukthar KJ (2024) Transitioning to digital merchandise: integrating metaverse into retail offerings. In: Singla B, Shalender K and Singh N (eds) *Creator’s Economy in Metaverse Platforms: Empowering Stakeholders through Omnichannel Approach*. Hershey, PA: IGI Global, 71–96.
- Snyder H (2019) Literature review as a research methodology: an overview and guidelines. *Journal of Business Research* 104: 333–339.
- Steneck NH (2003) *ORI Introduction to the Responsible Conduct of Research*. Rockville, MD: U.S. Department of Health and Human Services, Office of Research Integrity.
- Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 613(7945): 620–621.
- Stokel-Walker C (2024) *AI Chatbots Have Thoroughly Infiltrated Scientific Publishing*. Scientific American. <https://www.scientificamerican.com/article/chatbots-have-thoroughly-infiltrated-scientific-publishing/>.
- Subbaraman N (2024) *Flood of Fake Science Forces Multiple Journal Closures*. The Wall Street Journal. <https://www.wsj.com/science/academic-studies-research-paper-mills-journals-publishing-f5a3d4bc>.
- Tan Y-C, Chandukala SR and Reddy SK (2022) Augmented reality in retail and its impact on sales. *Journal of Marketing* 86(1): 48–66.
- Tashakkori A and Teddlie C (2003) Issues and dilemmas in teaching research methods courses in social and behavioural sciences: US perspective. *International Journal of Social Research Methodology* 6(1): 61–77.
- Tingelhoff F, Klug S and Elshan E (2024) The diffusion of the metaverse: how YouTube influencers shape mass adoption. *California Management Review* 66(4): 23–50. DOI: [10.1177/00081256241246123](https://doi.org/10.1177/00081256241246123).
- Tingelhoff F, Brugger M and Leimeister JM (2024) Towards a Guide for Conventional and Gen.AI-assisted Literature Reviews. *Academy of Management Proceedings* 2024(1): 1–40. 10.5465.
- Torraco RJ (2005) Writing integrative literature reviews: guidelines and examples. *Human Resource Development Review* 4(3): 356–367.
- Tranfield D, Denyer D and Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management* 14(3): 207–222. DOI: [10.1111/1467-8551.00375](https://doi.org/10.1111/1467-8551.00375).
- Tversky A and Kahneman D (1974) Judgment under uncertainty: heuristics and biases. *Science* 185(4157): 1124–1131.
- Venkatesh V, Brown SA and Bala H (2013) Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. *MIS Quarterly* 37: 21–54.
- vom Brocke J, Simons A, Riemer K, et al. (2015) Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems* 37(1): 9. DOI: [10.17705/1CAIS.03709](https://doi.org/10.17705/1CAIS.03709).
- Wagner G, Lukyanenko R and Paré G (2022) Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology* 37(2): 209–226.
- Walsham G (2012) Are we making a better world with ICTs? Reflections on a future agenda for the IS field. *Journal of Information Technology* 27: 87–93.
- Wambsganss T, Janson A, Söllner M, et al. (2024) Improving students’ argumentation skills using dynamic machine-learning-based modeling. *Information Systems Research (Articles in Advance)*: 1–34.
- Wang S, Scells H, Koopman B, et al. (2023) Can ChatGPT write a good Boolean query for systematic review literature search? In: Proceedings of the 46th International ACM SIGIR conference on research and development in information retrieval, Taipei Taiwan, 23–27 July 2023.
- Webster J and Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. *MIS Quarterly* 26(2): xiii–xxiii. <https://www.jstor.org/stable/4132319>.
- Weiss C (2022) Fashion retailing in the metaverse. *Fashion, Style & Popular Culture* 9(4): 523–538.
- Yoo K, Welden R, Hewett K, et al. (2023) The merchants of meta: a research agenda to understand the future of retailing in the metaverse. *Journal of Retailing* 99(2): 173–192. DOI: [10.1016/j.jretai.2023.02.002](https://doi.org/10.1016/j.jretai.2023.02.002).
- Zhong Q, Wang K, Xu Z, et al. (2024) Achieving >97% on GSM8K: deeply understanding the problems makes LLMs perfect reasoners. *ArXiv Preprint arXiv* (working paper): 1–11. 2404.14963.

Author biographies

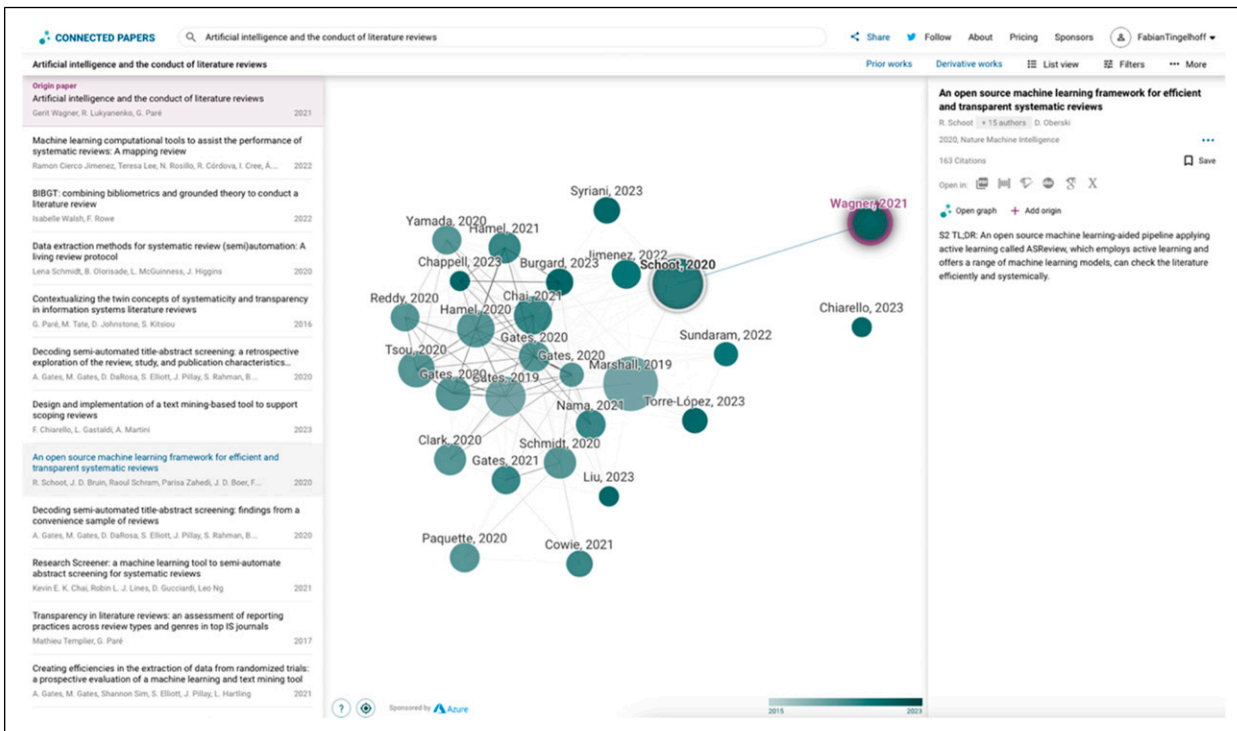
Fabian Tingelhoff is a PhD candidate and research associate at the University of St. Gallen. In his PhD, he investigates the value-generating utilization of Web3 technologies, including the metaverse and generative AI. His research has been published in various outlets, such as the California Management Review, Information Systems Frontiers, and in leading information systems conferences such as ICIS or HICSS. Throughout his PhD, he received the merit-based doctoral scholarship of the renowned German-based Konrad-Adenauer-Stiftung (KAS).

Micha Brugger is a PhD candidate and research associate at the University of St. Gallen. His research focuses on utilizing the advancements of generative artificial intelligence in teaching and educational settings. Besides his research, he utilized his data-science background to co-found an EdTech startup aiming to assist students during their final

theses by utilizing artificial intelligence to guide and structure their research endeavors.

Jan Marco Leimeister is Full Professor and Director of the Research Center for Information System Design (ITeG) at the University of Kassel, Germany. He is furthermore Full Professor and Director at the Institute of Information Systems and Digital Business, University of St. Gallen, Switzerland. His research covers Digital Business, Digital Transformation, Service Engineering and Service Management, Crowdsourcing, Digital Work, Collaboration Engineering, and IT Innovation Management. He runs research groups and projects that are funded by the European Union, German Ministries, DFG, various foundations, or industry partners. Prof. Leimeister teaches in several Executive Programs. He is a co-founder of several companies and serves as board member, consultant, coach, and speaker for numerous companies.

Appendix



Appendix I. Research graph by connected papers based on Wagner et al. (2022)

Humata
Share

Ask Humata

Summary This paper discusses the potential for artificial intelligence (AI) to transform traditional research practices, specifically in the context of literature reviews in the field of information systems research. It highlights the various steps of the literature review process and how AI-based tools can support each step. The paper also emphasizes the need for further development and improvement of these tools, as well as the importance of methodological research and best practices in using AI for literature reviews.

Example questions:

- ▶ What are the potential benefits of using AI-based tools in literature reviews, particularly in the field of information systems research?
- ▶ What are some examples of AI-based tools that can support different steps of the literature review process?
- ▶ What are the limitations of current AI-supported literature review tools and what steps can be taken to improve them in the future?

What are the potential benefits of using AI-based tools in literature reviews, particularly in the field of information systems research?

The potential benefits of using AI-based tools in literature reviews in the field of information systems research include the ability to expedite individual steps of the literature review process, automate and augment steps in different types of literature reviews, and find novel connections among distal IS phenomena. AI-based tools also offer the potential to support researchers in designing and improving different elements of search strategies, facilitate research gap verification, and aid in the development of related work sections and contribution statements [6] [4]. Additionally, AI-based tools have the potential to automate the time-consuming screening tasks, particularly the first screen based on titles and abstracts, and can complement human researchers in different ways, thereby addressing the challenge of sample sizes threatening to exceed the cognitive limits of human processing capacities [4]. Furthermore, AI-based tools can potentially point researchers to promising areas and questions, verify research gaps, and support the evaluation and validity, transparency and replicability, compatibility for recombination, and usability across the six steps of the literature review process [11] [3]. Overall, the use of AI-based tools in literature reviews has the potential to expedite review projects, address the increasing volume of reviews published in IS journals, and support tasks that fall on a spectrum between the mechanical and the creative, ultimately benefiting the research field and its stakeholders [12]. These potential benefits highlight the promising role of AI-based tools in enhancing the efficiency, effectiveness, and quality of literature reviews in the field of information systems research.

Pages: 1, 3, 9, 10, 11

Ask ...

Journal of Information Technology
Share

Artificial intelligence and the conduct of literature reviews

Gerit Wagner, Roman Lukyanenko and Guy Paré

Abstract
Artificial intelligence (AI) is beginning to transform traditional research practices in many areas. In this context, literature reviews stand out because they appear as large and rapidly growing volumes of documents, that is, partially structured (metadata), and pervade almost every type of paper published in information systems research or related social science disciplines. To facilitate researchers with some of the recent trends in this area, we outline how AI can expedite individual steps of the literature review process. Considering that the use of AI in this context is an early stage of development, we propose a comprehensive research agenda for AI-based literature reviews (AIBLRs) in our field. With this agenda, we would like to encourage design science research and a broader constructive discourse on shaping the future of AIBLRs in research.

Keywords
Artificial intelligence, machine learning, natural language processing, research data management, data infrastructure, automation, literature review

Introduction
The potential of artificial intelligence (AI) to augment and partially automate research has sparked vivid debates in many scientific disciplines, including the health sciences (Latham et al., 2017; Tachibana et al., 2018; biology (Khan et al., 2018), and management (Johnson et al., 2019). In particular, the concept of automated science is raising intriguing questions related to the future of research in disciplines that require “high-level abstract thinking, intricate knowledge of methodology and epistemology, and painstaking writing capabilities” (Johnson et al., 2019: 252). These debates resonate with scholars in Information Systems (IS), who ponder which role AI and automation can play in theory development (Tinsley et al., 2019) and in combining data-driven and theory-driven research (Mason et al., 2018). With this commentary, we join the discussion which has been initiated recently by Johnson et al. (2019) in the business discipline. The authors observe that across the multi-disciplinary discourse, two dominant narratives have emerged. The first narrative adopts a pessimistic and cautious perspective to present its audience with a choice between accepting or rejecting future research practices in which AI plays a dominant role. The second narrative acknowledges that a gradual adoption of AI-based research tools has already begun and aims at engaging its readers in a constructive debate on how to leverage AI-based tools for the benefit of the research field and its stakeholders. In this paper, our position resonates more with the latter perspective, which is focused on the mid-term instead of the long-term, and well-positioned to advance the discourse with less speculative and more actionable discussions of the specific research processes that are more amenable to applications of AI and those processes that rely more on the human ingenuity of researchers.

In this essay, we focus on the use of AI-based tools in the conduct of literature reviews. Advancing knowledge in this area is particularly promising since (1) standalone review projects require substantial efforts over months and years (Latham et al., 2017); (2) the volume of reviews published in IS journals has been rising steadily (Johnson et al., 2019); and (3) literature reviews involve tasks that fall on a spectrum between the mechanical and the creative (AI). At the same time, the process of reviewing literature is rarely conducted manually with sample sizes threatening to exceed the cognitive limits of human processing capacities. The

Department of Information Technology, HEC Montreal, Montreal, Quebec, Canada
Corresponding author: Guy Paré, Research Chair in Digital Health, HEC Montreal, 3000, Ste-Anne St, Montreal, QC H3T 1A5, Canada. Email: guy.pare@hec.ca

210 Journal of Information Technology 27(2)

has been illustrated recently by Latham et al. (2019), who estimated that in the IS field, the number of relevant papers in many research areas easily exceeds 10,000. As a consequence, some review articles, problematically, no longer aim for comprehensive coverage, often restricting their scope to few top journals. Overall, we anticipate that these trends will be reinforced in the future, further emphasizing the need to enhance fruitful collaborations between human researchers and machines, such as AI-based tools (cf. Sothor et al., 2020).

In light of these challenges, we focus on the contributions of AI which refers to the capability of performing cognitive tasks and exhibiting intelligent behavior commonly associated with human intelligence (David and Foray, 2014; Tassil and Ouy, 2019). Specifically, we are interested in approaches that are commonly referred to as “weak AI” and combine process automation (execution engines) with capabilities like machine learning (ML) or natural language processing (NLP). Machine learning refers to tools, methods, and techniques for learning and improving task performance with experience (Goodfellow et al., 2016; Mitchell, 1997), while NLP refers to computational tools, methods, and techniques for analyzing, interpreting, and increasingly generating natural language (Blanning and Schöler, 2018). Although we are particularly interested in tools powered by advanced AI, we do not discard predecessors of AI per se.

AI offers two capabilities that are particularly salient for conducting literature reviews. First, they operate on potentially large, weakly structured, and unstructured data that are provided in the form of bibliographical meta-data or full-text documents. Techniques of NLP can go beyond purely syntactic processing of text by abstracting and analyzing its semantic meaning, thereby providing in other valuable support in the searching and screening tasks. For example, news inclusion for the word “reviews” may be hard to dis-

specify. The work of Latham et al. (2019) is exemplary in this regard, developing classifiers that can automatically screen and include papers relevant to research on TAM (Technology Acceptance Model). Considering these capabilities, we expect AI to be most useful in the mechanical tasks of reviewers compared to more creative ones. At the same time, an informed discussion and methodological guidelines are necessary to identify the appropriate areas of application and to address the challenges associated with AI, such as model overfitting, biases, black box predictions, and the acceptance by the research community.

The objective of this essay is to frame the broader discourse on how AI can be applied in the individual steps of the literature review process, providing illustrative examples for prospective authors and outlining opportunities for further advancing such methods. To clarify these objectives, we coin the term AI-based literature reviews (AIBLRs), which refers to literature reviews undertaken with the aid of AI-based tools for one or multiple steps of the review process, that is, problem formulation, literature search, screening for inclusion, quality assessment, data extraction, or data analysis and interpretation. Without necessarily being driven by academic researchers, functionality for literature searches is already supported by AI, as implemented by academic literature databases and indexing services. We focus on how AI-based tools can evolve to play an even more powerful role and further automate and augment steps in different types of literature reviews. An important question for researchers is how such tools can best be leveraged in all steps of the review process and how it can be adapted to particular types of reviews. In doing so, it can be expected that different types of reviews, such as descriptive or interpretive reviews, will be more or less amenable to the use of AI. The remainder of this essay is structured as follows. In the next section we

Appendix 2. Humata references its answers to specific sections of the paper by Wagner et al. (2022)