Management Review Quarterly
https://doi.org/10.1007/s11301-025-00522-8

A hybrid framework for creating artificial intelligence-augmented systematic literature reviews

Faisal Saeed Malik[1] · **Orestis Terzidis**[1]

Received: 9 January 2025 / Accepted: 17 April 2025

Abstract

The integration of artificial intelligence (AI), particularly generative AI (GenAI) and large language models (LLMs), into systematic literature reviews (SLRs) represents a transformative advancement in research methodologies. This paper proposes a hybrid framework combining AI's computational power with the epistemological rigor of human expertise, anchored in transparency, validity, reliability, comprehensiveness, and reflective agency. Through three interconnected phases—design, study collection, and interpretation—the framework employs AI model selection, knowledge base curation, and iterative prompt engineering to enhance scalability, uncover interdisciplinary connections, and ensure methodological integrity through robust human oversight. It addresses key SLR challenges, including handling vast datasets, ensuring reproducibility, and maintaining epistemic rigor while leveraging advanced AI capabilities. Key innovations include cyclical validation, inter-model comparisons, and sensitivity testing to enhance trustworthiness and mitigate biases. The framework aligns AI processes with ethical standards and research objectives by emphasizing domain-specific LLMs, reliability metrics, and standardized reporting protocols. It establishes SLRs as a foundation for advancing knowledge in complex, interdisciplinary research landscapes, harmonizing AI efficiency with human expertise.

JEL M00 · M1

✉ Faisal Saeed Malik
Faisal.malik@kit.edu

Orestis Terzidis
Orestis.terzidis@kit.edu

1 The Institute of Entrepreneurship, Technology-Management, and Innovation (EnTechnon), Karlsruhe Institute of Technology (KIT), Fritz Erler Str. 1-3, Gebäude 01.85, 76131 Karlsruhe, Germany

Published online: 29 April 2025

Keywords Systematic literature reviews (SLRs) · Artificial intelligence (AI) · Human factor · Generative AI (GenAI) in research methods · Epistemological principles · Management research methodologies

1 Introduction

Systematic Literature Reviews (SLRs) have long served as a cornerstone of academic research, offering a structured approach to synthesizing existing knowledge, identifying research gaps, and advancing theoretical and practical insights. Foundational contributions by Webster and Watson (2002) established principles of rigor, transparency, and reproducibility, which continue to form the methodological backbone of SLRs (Aguinis et al. 2023; Block and Fisch 2020; Fisch and Block 2018; Grant and Booth 2009; Hiebl 2023; Sauer and Seuring 2023). Subsequent advancements by Tranfield et al. (2003) and Denyer and Tranfield (2009) addressed the growing complexities of interdisciplinary research landscapes, while frameworks like PRISMA codified best practices for systematic and transparent reviews (Alvesson and Sandberg 2020; Else 2023; Grant and Booth 2009; Kraus et al. 2020; Liberati et al. 2009a, b; Okoli and Schabram 2010; Snyder 2019). However, despite these methodological advancements, the exponential growth of academic publications, the rise of interdisciplinary knowledge production, and the increasing demand for scalability present new challenges to traditional SLR methodologies.

However, the emergence of artificial intelligence (AI), particularly generative AI (GenAI) and large language models (LLMs), represents a paradigm shift in SLR methodologies (Etsenake and Nagappan 2024; Gao et al. 2024; Lorenz et al. 2024; Wu et al. 2024). These technologies automate labor-intensive tasks such as literature retrieval, thematic synthesis, and citation mapping, significantly enhancing efficiency and precision (Alshami et al. 2023; Bolaños et al. 2024; Dwivedi et al. 2023; Vom Brocke et al. 2009; von Krogh et al. 2023). AI also enables researchers to uncover interdisciplinary connections and synthesize diverse perspectives that traditional review methods may overlook (Anthony et al. 2023; Gatrell et al. 2024; Grimes et al. 2023; Pan et al. 2023). However, integrating AI into SLRs raises critical epistemological and methodological concerns, particularly regarding reliability, transparency, and contextual relevance (Ngwenyama and Rowe 2024; Schlagwein and Willcocks, 2023; Wach et al. 2023). AI-generated hallucinations—fabricated or misleading outputs—pose risks to research integrity and have led to skepticism about the use of AI in scholarly inquiry (Schryen et al. 2024; Shaheen et al. 2023; Tingelhoff et al. 2024; Wu et al. 2023). As AI adoption increases, recent studies highlight the urgency of developing robust validation frameworks to ensure AI-driven methodologies align with established academic standards (Aguinis et al. 2018, 2023; Aytug et al. 2012; Cronin and George 2023; Glaser et al. 2024; Hiebl 2023; Kunisch et al. 2018; Mees-Buss et al. 2022; Schryen et al. 2020).

Therefore, this study introduces a hybrid methodological framework that integrates AI's computational capabilities with the epistemological rigor of human expertise. This framework is grounded in five core epistemic principles—transparency, validity, reliability, comprehensiveness, and reflective agency—ensuring that

AI enhances, rather than compromises, the quality and reliability of SLRs (Schryen et al. 2015; Kraus et al. 2024; Sharma and Bansal 2023). A central contribution of this study is its emphasis on validation mechanisms with humans-in-the-loop, which counteract AI hallucinations and reinforce trust in AI-assisted research. Unlike conventional AI-based SLR approaches that focus solely on automation, this framework incorporates cyclical validation, inter-model comparisons, and human verification at each phase of the review process, ensuring that AI-generated insights remain methodologically sound and contextually relevant (Alshami et al. 2023; Benbya et al. 2024; Dwivedi et al. 2023).

The framework is operationalized through three interconnected phases—design, study collection and selection, and interpretation—which incorporate iterative prompt engineering, reliability checks, and sensitivity testing to ensure alignment with established research standards (Korzynski et al. 2023; Schlagwein and Willcocks, 2023). The framework enhances scalability and interdisciplinary integration by automating literature retrieval and thematic synthesis while mitigating biases inherent in AI-driven methodologies. Furthermore, this study provides concrete recommendations for researchers to implement AI-augmented SLRs responsibly, bridging the gap between automation and epistemological rigor.

This research positions AI as an augmentation tool rather than a substitute for human expertise, reinforcing the importance of critical engagement and ethical considerations in AI-assisted reviews (Bouschery et al. 2023; Ngwenyama and Rowe 2024). Ethical concerns such as algorithmic opacity, bias, and the credibility of AI-generated outputs are explicitly addressed within the framework, promoting responsible AI use in academic research (Hiebl 2023; Kulkarni et al. 2024; Schryen et al. 2024; Shaheen et al. 2023). By contributing to the ongoing scholarly discourse on AI in research methodologies, this study offers a structured and empirically grounded approach to integrating AI into systematic reviews while preserving methodological rigor and fostering interdisciplinary collaboration (Anthony et al., 2023; Gatrell et al. 2024; Grimes et al. 2023; Lorenz et al. 2024; Sauer and Seuring 2023). Ultimately, this framework enables researchers to navigate the complexities of contemporary research environments, ushering in an era of efficient, ethical, and epistemologically sound interdisciplinary studies.

2 Epistemological background for SLRs

A sound understanding of the epistemological background is crucial for successfully integrating AI within the SLR process. This understanding will serve as a guiding framework, ensuring that AI is employed to enhance, rather than compromise, the core scientific objectives of SLRs. A sound understanding of epistemology is fundamental to the rigor and credibility of SLRs (Berghofer 2022; Sharma and Bansal 2023). Epistemology, the study of knowledge and its acquisition, provides a foundation for assessing how evidence is collected, analyzed, and interpreted in academic research (Aguinis and Solarino 2019; Kunisch et al. 2023). SLRs are a cornerstone of knowledge synthesis, ensuring that research findings are comprehensive, replicable, and theoretically and methodologically robust (Gibbert and Ruigrok 2010; Köhler et al. 2022).

Three dominant epistemological traditions in management and organizational research—empiricism, rationalism, and constructivism—have shaped literature reviews (Audi and Goldman 1989; Aytug et al. 2012; Harley and Cornelissen 2022; Pratt et al. 2022). Empiricism emphasizes objective observation and systematic evidence collection (Aguinis et al. 2011), rationalism highlights logical reasoning and coherence in knowledge synthesis (Mees-Buss et al. 2022), and constructivism underscores the role of interpretation in shaping research insights (Antons et al. 2023). These traditions provide a philosophical grounding for how SLRs should be structured, validated, and interpreted (Villiger et al. 2022).

To ensure methodological rigor and epistemic consistency in SLRs, we propose a framework based on five key principles: transparency, validity, reliability, comprehensiveness, and reflective agency (Aguinis et al. 2023; Köhler et al. 2022). These principles are not arbitrarily chosen but are grounded in established methodological scholarship (Aguinis et al. 2018, 2023; Choi 2018; Gond et al. 2023; Hofer 2017; Williams et al. 2021) and serve as theoretical safeguards for producing credible and reproducible reviews (Breslin and Gatrell 2023; Harley and Cornelissen 2022). The following sections elaborate on these principles, justify them, and explore their implications for AI-augmented SLRs.

2.1 Transparency

Transparency is a foundational principle of SLRs, ensuring that each stage of the review process is explicitly documented, reproducible, and subject to scrutiny (Aguinis et al. 2011, 2023; Block and Fisch 2020; Hiebl 2023; Kunisch et al. 2023). This principle mandates transparent reporting on search criteria, study selection, inclusion and exclusion decisions, and data extraction methods to allow other researchers to replicate and validate findings (Almotahari and Glick 2010; Duan et al. 2019; Köhler et al. 2022; Usberti 2019; Zupic and Čater 2015). A transparent review process requires explicit methodological documentation of how literature is searched, classified, and synthesized (Harley and Cornelissen 2022; Pratt et al. 2022). Without transparency, reproducibility is compromised, making it difficult for future researchers to build upon or verify the findings presented (Schryen et al. 2015).

Frameworks such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and RAMESES (Realist and Meta-narrative Evidence Syntheses: Evolving Standards) provide standardized reporting guidelines that reinforce the role of transparency in ensuring scientific rigor (Aguinis et al. 2018; Köhler et al. 2022; Liberati et al. 2009a, b). Additionally, transparency extends beyond procedural documentation to theoretical positioning, where scholars must explicitly state their assumptions, conceptual definitions, and methodological choices (Aguinis et al. 2023; Lorenz et al. 2024; Schryen et al. 2024). This level of clarity strengthens the credibility of synthesized knowledge and supports an open and verifiable research process (Aytug et al. 2012; Kraus et al. 2020; Villiger et al. 2022).

2.2 Validity

Validity in SLRs ensures that the review process and findings accurately reflect the research question and theoretical objectives (Aguinis and Solarino 2019; Kunisch et al. 2023). In methodological scholarship, validity is commonly divided into internal and external, crucial for literature reviews (Köhler et al. 2022; Mees-Buss et al. 2022). Internal validity concerns the credibility of included studies and the robustness of their theoretical and empirical claims (Breslin and Gatrell 2023; Pratt et al. 2022). Ensuring that the studies reviewed are methodologically rigorous is vital, as flawed research design or weak empirical evidence can compromise the theoretical conclusions drawn (Harley and Cornelissen 2022).

Conversely, external validity addresses the generalizability of findings, determining whether the synthesized literature applies beyond the specific context of the review (Aguinis et al. 2023; Hiebl 2023; Shaheen et al. 2023). This requires assessing whether the selected studies' scope, diversity, and representativeness support broad theoretical claims (Bobko et al. 2007; Köhler et al. 2022; Kunisch et al. 2018, 2023). To maintain validity, SLRs often employ methodological triangulation, cross-validating sources and findings to ensure coherence between theoretical perspectives and empirical insights (Breslin and Gatrell 2023; Schryen et al. 2015, 2020).

2.3 Reliability

Reliability refers to the consistency of the review process, ensuring that independent researchers conducting the same SLR would reach comparable findings (Gibbert and Ruigrok 2010; Köhler et al. 2022; Pollock 1984). Achieving reliability requires the application of structured methodologies, such as systematic search strategies, predefined coding schemes, and intercoder agreement procedures (Aguinis et al. 2023; Locke et al. 2022; Pratt 2008). By codifying literature selection processes, scholars can minimize subjectivity and establish a consistent basis for including or excluding studies (Schryen et al. 2015, 2024).

Reliability also depends on replicability tests, where multiple independent searches using identical criteria validate the robustness of findings (Mees-Buss et al. 2022; Villiger et al. 2022). When combined with transparency, reliability enhances cumulative knowledge building, ensuring that future research can build upon a stable and replicable body of literature (Aguinis et al. 2023; Bonett and Wright 2014; Madill et al. 2000; Pratt et al. 2022). Reliability gains importance in the context of non-deterministic methods like using LLM and GenAI. We will expand on this during the paper.

2.4 Comprehensiveness

Comprehensiveness ensures that an SLR captures the full breadth of relevant literature, distinguishing it from narrative reviews prone to selective biases (Clark et al. 2021; Köhler et al. 2022; Williams et al. 2021). A thorough literature search must incorporate diverse sources, including peer-reviewed journal articles, theoretical advancements, and empirical contributions (Antons et al. 2023; Berghofer 2022;

Sauer and Seuring 2023; Zupic and Čater 2015). Theoretical inclusivity is also essential, as integrating perspectives from multiple paradigms within a research domain strengthens the depth and breadth of insights generated (Adams et al. 2017; Hiebl 2023; Mees-Buss et al. 2022).

Comprehensiveness also reinforces an SLR's reliability and reproducibility (Sharma and Bansal 2023). A transparent and detailed account of study selection and synthesis criteria ensures that future researchers can replicate the review process and validate its conclusions (Aguinis et al. 2023; Cronin and George 2023; Gond et al. 2023; Köhler et al. 2022; Williams et al. 2021).

2.5 Reflective agency

Reflective agency highlights the active role of researchers in shaping the review process, interpreting findings, and refining theoretical contributions (Kunisch et al. 2023; Madill et al. 2000; Pratt 2008; Zagzebski 2020). Scholars must critically evaluate methodological trade-offs, ensuring that inclusion criteria align with the overarching research goals (Aguinis et al. 2023; Duan et al. 2019; Köhler et al. 2022). This process requires epistemic reflexivity, where researchers remain aware of their biases, theoretical assumptions, and interpretative choices (Harley and Cornelissen 2022; Mees-Buss et al. 2022).

Reflective agency is vital in theory-building through SLRs, where different scholars might derive distinct theoretical insights from the same body of literature (Alavi et al. 2024; Hoon 2013; Schryen et al. 2015; Simsek et al. 2023). Ensuring self-awareness and methodological integrity strengthens the epistemological soundness of SLRs and enhances their contribution to management research (Aguinis and Solarino 2019; Aytug et al. 2012; Benbya et al. 2024; Choi 2018; Köhler et al. 2022).

Critical thinking, interpretation, ethical considerations, and contribution to discourse are essential for an SLR project. Reflective agency thus complements the other epistemological principles of transparency, validity, reliability, and comprehensiveness. While these principles ensure that the review process is rigorous, unbiased, and thorough, the reflective agency highlights the indispensable role of the researchers' critical thinking and judgment (Hofer 2017). By incorporating a reflective agency, SLRs can achieve higher intellectual creativity and science (Ameen et al. 2022). It ensures that the review is not just a mechanical aggregation of existing studies but a thoughtful and meaningful contribution to the body of knowledge (Lorenz et al. 2024; Ngwenyama and Rowes, 2024).

3 Traditional methodological framework for SLRs

SLRs are integral to academic research, enabling the synthesis of existing knowledge, identifying research gaps, and generating theoretical and practical insights. This methodological framework traditionally comprises three phases: **(i)** planning and formulating research questions, **(ii)** searching and screening relevant literature, and **(iii)** analyzing, synthesizing, and reporting findings (Hiebl 2023; Sauer and Seuring 2023; Snyder 2019). These phases ensure rigor and relevance, providing a robust foundation for scholarly inquiries and addressing practical challenges. This section explores traditional SLR methodologies while setting the stage for their evolution through the integration of generative AI (Alavi et al. 2024; Benbya et al. 2024; Endrigo Sordan et al. 2024; Farrelly and Baker 2023; Najafi et al., 2023).

3.1 Core phases in SLRs

3.1.1 Planning and formulating research questions

The planning phase establishes the scope and objectives of an SLR, serving as the foundation for methodological rigor. Well-formulated research questions align the review process with theoretical priorities and practical relevance, ensuring transparency, validity, and comprehensiveness (Tranfield et al. 2003). This phase involves developing theoretical frameworks to provide structure and context for the inquiry, as Denyer and Tranfield (2009) and Eisenhardt (1989) highlighted. Grounding research in empirical observations enhances its practical applicability (Baškarada 2014). Additionally, engaging in the academic community ensures alignment with current trends and increases the review's impact (Kumar et al. 2024). PICO (Population, Intervention, Comparison, Outcome) and SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research Type) are primarily used to structure research questions in evidence-based research, particularly in medicine and the social sciences. However, these frameworks also serve as a **tool for operationalizing research questions** by **guiding the development of search strategies** in systematic reviews. Systematically defining key terms and inclusion criteria helps ensure conceptual clarity and comprehensive coverage in database searches (Grant and Booth 2009). The iterative nature of this process, underscored by Fisch and Block (2018), requires clarity and adaptability, while advancements in AI tools like ChatGPT and Llama enhance efficiency by synthesizing literature and identifying research gaps, reducing time-intensive efforts (Alshami et al. 2023; Anderson et al. 2023; Deepa et al. 2024).

3.1.2 Searching and screening literature

The second phase systematically identifies and selects relevant literature, ensuring comprehensiveness and reliability in the review process. PRISMA guidelines play a critical role in promoting transparency and reproducibility by requiring rigorous selection process documentation (Liberati et al. 2009a, b). Automation tools, including databases like Scopus and Web of Science, combined with AI-driven search algorithms, improve efficiency and ensure inclusiveness (Chintalapati and Pandey 2022). AI advancements streamline this phase by automating keyword searches, categorizing data, and uncovering interdisciplinary connections, addressing the challenges of comprehensiveness in traditional reviews (Alshami et al. 2023; Benbya et al. 2024; Bolaños et al. 2024; Schryen et al. 2024). These technologies reduce bias and the labor intensity of manual screening (Pan et al. 2023). Block and Kuckertz (2024) emphasize that AI-driven methodologies transform traditional practices, enabling researchers to manage large datasets while maintaining methodological rigor.

3.1.3 Analyzing, synthesizing, and reporting

The final phase focuses on extracting insights, synthesizing findings, and presenting them in a structured and reflective manner. Thematic synthesis is commonly used to organize findings into coherent themes, providing clarity and comprehensiveness (Webster and Watson 2002). Reflective synthesis further contextualizes findings within broader theoretical and practical frameworks, ensuring validity and relevance (Denyer and Tranfield 2009). AI tools enhance this phase by automating thematic clustering and generating predictive insights, facilitating nuanced analysis and the identification of novel research directions (Deepa et al. 2024). Hybrid models integrating AI with traditional methodologies can enrich the reporting process (Block and Kuckertz 2024; Tingelhoff et al. 2024). These innovations align with the principle of reflective agency by enabling dynamic and impactful reporting, supported by AI visualization tools that engage diverse audiences effectively (Alavi et al. 2024; Benbya et al., 2024; Bilgram and Laarmann 2023; Locke et al. 2022; Reyes-Menendez et al. 2023).

3.2 Bridging to AI-augmented systematic reviews

Generative AI tools address critical gaps in traditional SLR methodologies by enhancing comprehensiveness, reducing bias, and enabling reflective synthesis (Farrelly and Baker 2023). These tools facilitate identifying interdisciplinary connections and improve objectivity through algorithmic processes (Grimes et al. 2023; Pan et al. 2023). However, the role of human cognition remains pivotal, ensuring that AI outputs align with nuanced research objectives and retain contextual relevance (Bouschery et al. 2023; Dwivedi et al. 2023; Lu 2019; Zhang and Lu 2021).

Researchers act as "prompters" in the AI-augmented review process, guiding the application of these tools to maintain methodological integrity. They define goals that align with theoretical and practical priorities (Tranfield et al. 2003). Validation of outputs requires researchers to apply critical scrutiny and cross-reference findings to safeguard scholarly rigor (Block and Kuckertz 2024; Noy and Zhang 2023). Moreover, researchers synthesize AI-driven patterns into robust theoretical and practical frameworks, emphasizing the collaborative nature of AI-human integration (Anthony et al. 2023; Simsek et al. 2023). AI-augmented methodologies enhance the traditional contributions of literature reviews by introducing scalability and interdisciplinarity while maintaining critical oversight (Alavi., et al., 2024; Bolaños et al. 2024; Sharma and Bansal 2023). Schryen et al. (2015) outlined six key pathways of contribution. These pathways can be advanced significantly through AI integration.

First, AI-driven thematic analyses and clustering reveal patterns often overlooked in manual reviews, with researchers contextualizing these findings within broader epistemological frameworks to ensure their relevance (Bolaños et al. 2024). Second, generative AI identifies relationships across disparate fields, fostering reinterpretations of existing knowledge, while researchers critically evaluate and integrate these insights to maintain epistemic validity (Gatrell et al. 2024; Grimes et al. 2023). Third, AI facilitates the synthesis of meta-knowledge, expanding theoretical boundaries under the guidance of researchers to ensure coherence and originality in newly emerging frameworks (Dwivedi et al. 2023). Fourth, AI-powered meta-analyses provide tools for validating and refining existing theories, with researchers safeguarding the empirical and theoretical rigor of the findings (Aguinis and Solarino 2019; Deepa et al. 2024; Kunisch et al. 2023). Fifth, AI algorithms highlight research gaps by identifying inconsistencies and unexplored areas, enabling actionable inquiries, while researchers provide oversight to ensure these gaps align with scholarly priorities (Pan et al. 2023). Lastly, AI supports the development of forward-looking research agendas, with researchers ensuring these pathways are significant and actionable within the academic discourse (Aguinis et al. 2023; Kumar et al. 2024; Maaravi et al. 2020; Reyes-Menendez et al. 2023). Integrating AI's computational strengths with researchers' intellectual oversight enhances SLRs' rigor, interdisciplinarity, and impact, marking a substantial advancement in academic research methodologies.

3.2.1 Comparison of literature review approaches

Traditional methodologies excel in rigor and transparency but are limited in scalability and interdisciplinarity (Alvesson and Sandberg 2020; Rammal 2023; Redondo-Rodríguez et al. 2024; Tranfield et al. 2003). AI-augmented approaches address these limitations by automating labor-intensive processes and uncovering hidden patterns (Dwivedi et al. 2023; Haenlein and Kaplan 2019; Maaravi et al. 2020; Tingelhoff et al. 2024). Hybrid methodologies combine these strengths, advancing epistemological rigor and practical relevance (Bolaños et al. 2024; Block and Kuckertz 2024; Schryen et al. 2024).

The hybridization of methodologies allows the integration of diverse approaches (see Table 1).

The hybrid SLR methodology combines traditional systematic review practices with generative AI innovations to address methodological gaps while adhering to transparency, reproducibility, validity, and community engagement principles. Integrating established rigor with AI-driven capabilities enhances efficiency, scalability, and adaptability without compromising methodological integrity. This approach integrates the structured rigor of systematic reviews (Denyer and Tranfield 2009; Kitchenham 2004) with AI's rapid data processing and synthesis capabilities (Dwivedi et al. 2023), enabling comprehensive analyses with incredible speed and precision. It utilizes AI to identify contradictions and gaps in theoretical frameworks, advancing reflexive and innovative insights (Lindebaum and Fleming 2024). Additionally, it builds on Kraus et al. (2020) by synthesizing qualitative and quantitative evidence through AI-driven cross-domain analysis, ensuring a balanced and holistic understanding of research phenomena.

By combining the flexibility of narrative reviews (Snyder 2019) with AI-driven clustering and thematic analysis (Webster and Watson 2002), the hybrid SLR allows for real-time adaptation to evolving themes and interdisciplinary insights. Furthermore, it enhances the reflective synthesis process of traditional reviews by leveraging AI to uncover interdisciplinary connections and emerging themes (Bolaños et al. 2024; Locke et al. 2022; Maaravi et al., 2021; Schryen et al., 2024).

Table 1 Hybridization of literature review methodologies

Approach	Key Features	Advantages	Limitations	Innovations in Hybrid SLR
Concept-Centric Review	Focuses on themes, concepts, and theoretical frameworks (Webster and Watson 2002).	Provides clarity by categorizing research into distinct themes and concepts.	Can overlook interconnections and emergent relationships between concepts.	AI-driven clustering and thematic analysis uncover latent patterns and interdisciplinary connections (Bolaños et al. 2024; Dwivedi et al. 2023).
Problem-Centric Review	Challenges theoretical assumptions through problematization (Alvesson and Sandberg 2020).	Encourages critical thinking, reflexivity, and innovative theory development.	Relies heavily on researcher expertise and interpretation; time-intensive process.	AI tools identify contradictions and tensions in theoretical frameworks to support targeted problematization and hypothesis generation (Tingelhoff et al. 2024).
Systematic Literature Review	Structured, replicable process with predefined inclusion criteria (Denyer and Tranfield 2009; Liberati et al. 2009a, b).	Ensures rigor, transparency, and reproducibility through detailed documentation and replicable methodologies.	Labor-intensive and less adaptive to rapidly emerging or interdisciplinary themes.	AI automates large-scale data retrieval, synthesis, and iterative updates, improving efficiency while maintaining transparency (Block and Kuckertz 2024; Dwivedi et al. 2023).
Narrative Review	Provides a broad overview of a field or topic (Snyder 2019).	Flexible and incorporates diverse perspectives, enabling comprehensive overviews.	Subjective interpretations can result in inconsistencies and lack methodological rigor.	AI-generated narrative summaries enhance structure, reduce bias, and streamline narrative synthesis (Mariani et al. 2023; Tingelhoff et al. 2024).
Mixed Methods Review	Combines qualitative and quantitative evidence for holistic insights (Kraus et al. 2020).	Integrates numerical trends with in-depth thematic insights, balancing breadth and depth of analysis.	Requires expertise in both methodologies; synthesis can be challenging without extensive training.	AI-augmented synthesis enables scalable integration of mixed-method data, ensuring methodological coherence and actionable insights (Amoozad Mahdiraji et al. 2023; Lee 2021).
Generative AI-Augmented Review	Leverages AI to synthesize and summarize large datasets, identifying patterns (Dwivedi et al. 2023; Bolaños et al. 2024).	Enhances efficiency and scalability, uncovering patterns across large and complex datasets.	Risk of overreliance on automation; potential biases in AI-driven selection and interpretation.	Extends AI capabilities for abductive reasoning and iterative theory-building processes, integrating human validation for rigor (Prikshat et al. 2023; Tingelhoff et al. 2024).
Practical Review	Focused on actionable recommendations bridging theory and practice (Fisch and Block 2018).	Effectively translates research into practice-relevant insights, addressing real-world challenges.	Risk of oversimplification, potentially compromising depth of theoretical insights.	AI automates generation of actionable recommendations while linking findings to broader theoretical frameworks (Block and Kuckertz 2024).

Table 1 (continued)

Approach	Key Features	Advantages	Limitations	Innovations in Hybrid SLR
Technology-Driven Review	Explores the role of technology in advancing methodologies and identifying research opportunities (Tingelhoff et al. 2024).	Facilitates rapid identification of patterns, trends, and gaps, especially in technology-intensive domains.	Requires expertise in AI tools and additional oversight to ensure accurate and comprehensive synthesis.	Interactive researcher-AI collaboration fosters iterative problem-solving and contextual adaptation, enhancing methodological flexibility and interdisciplinary research (Anthony et al. 2023; Grimes et al. 2023; Mariani et al. 2023).

This hybrid approach advances literature review methodologies (Block and Kuckertz 2024), offering a robust framework to tackle complex, interdisciplinary research challenges while maintaining academic rigor and relevance (see Table 1).

4 Key levers for authors using LLMs in SLRs

The various forms of AI have significant potential in SLRs, offering potent enhancements in efficiency, accuracy, and scalability (Dwivedi et al. 2023; Tingelhoff et al. 2024). In this context, let us first explore a fundamental distinction between different types of AI and highlight the critical role of LLMs in SLRs. AI systems can be broadly categorized into rule-based and probabilistic models. Rule-based AI operates on predefined rules and logic, making decisions based on explicit instructions. These systems are highly transparent, explainable, and predictable but can be limited in handling complex, unstructured data (Adadi and Berrada, 2018). In contrast, probabilistic AI relies on statistical methods to make predictions and decisions (Nguyen et al. 2019). These models, including LLMs and many other machine learning (ML) algorithms, learn from data and can adapt to new information (Cai et al. 2023; Kasneci et al. 2023; Tapeh and Naser 2022). While they are more flexible and powerful in dealing with complex datasets, their decision-making processes are less transparent and explainable (Ge et al., 2024; Lê and Schmid 2022; Sharma and Bansal 2023; Wach et al. 2023).

LLMs have revolutionized the field of natural language processing (NLP) and have significant implications for SLRs (Kasneci et al. 2023). LLMs can process and generate human-like text, making them invaluable for various stages of the SLR process, including literature search, data extraction, and summarization and synthesis (Li and Li 2024; Porsdam Mann et al. 2023; Sumbal et al. 2024). Despite their advantages, integrating LLMs into SLRs must be approached with caution. Ensuring the accuracy and reliability of LLM-generated outputs, maintaining transparency, and addressing ethical considerations are paramount (Davison et al. 2024; Lorenz et al. 2024; Lund et al. 2023; Wu et al. 2024).

Against this background, it becomes clear that using LLMs in SLRs is promising and challenging. To unlock the potential and overcome the challenges, authors must first know how to interact with an LLM (Thorp 2023). Researchers have three primary 'levers' within their sphere of influence: the choice of the AI model and application, the knowledge base used to fine-tune the model, and the prompt given to the

LLM (Korzynski et al. 2023). These elements significantly impact the effectiveness and reliability of AI in the SLR process (Hiebl 2023; von Krogh et al. 2023).

The first lever is the selection of the AI model and its application. Different AI models offer varying capabilities and strengths, making the choice of model crucial for the specific needs of an SLR (Anthony et al. 2023; Glaser et al. 2024). Most researchers are aware of the mainstream models like GPT, Gemini, LlaMA, Claude. or MISTRAL, but tens of thousands of models are available on the platform 'Hugging Face' (Hugging Face, n.d.). Choosing one or multiple AI models therefore is an important step (Nazir and Wang 2023). LLMs have shown particular promise in SLRs due to their ability to process and generate text. These models can automate literature searches, screen abstracts, and even assist in data extraction and synthesis (Alavi et al. 2024; Tingelhoff et al. 2024; Wittenborg et al. 2024). The choice of model and the application environment through which it is accessed should align with the specific requirements of the review, considering factors such as the complexity of the data, the need for transparency, and the desired level of automation.

The second lever is the knowledge base used to 'fine-tune' the AI model. Fine-tuning involves training the foundational model on a specific dataset to enhance performance on particular tasks. The quality and relevance of the knowledge base are critical, as they directly influence the model's accuracy and reliability (Hiebl., 2023; Villiger et al. 2022; Williams et al. 2021). For example, fine-tuning an LLM on a dataset of journal articles selected with a specific search term can make it possible to screen and extract relevant content (Hayden et al. 2024). Selecting a comprehensive and high-quality knowledge base ensures that the AI model is well-equipped to handle the specific nuances and requirements of the SLR (Anthony et al. 2023; Grimes et al. 2023; Simsek et al. 2023). This process can involve curating datasets from reputable sources, ensuring diversity in the data to cover various aspects of the research topic (Aguinis et al. 2023; Glaser et al. 2024; Hiebl 2023; Kulkarni et al. 2024).

The third lever is the prompt given to the AI model. Prompt engineering involves crafting specific instructions or queries that guide AI in generating the desired output. The effectiveness of AI in SLRs can be significantly enhanced by developing precise and contextually relevant prompts (Korzyński et al., 2023; Robledo et al. 2021). For instance, prompts can be designed to focus the AI on identifying key themes, extracting specific data points, or coherently summarizing findings (Aguinis et al. 2023; Kraus et al. 2024; Redondo-Rodríguez et al. 2024; Schryen et al. 2024; Wang et al. 2023).

By strategically leveraging these three elements – choice of AI model and application, knowledge base for fine-tuning, and prompt engineering – authors can enhance the efficiency, accuracy, and reliability of AI in systematic literature reviews (Korzyński et al., 2023; Kunisch et al. 2023; Sharma and Bansal 2023). These levers provide framework elements for effectively integrating AI into the SLR process, ensuring that the technology is a valuable tool in advancing research.

Figure 1 visualizes the AI-augmented research cycle. The cycle begins with the researcher, who leverages their expertise to guide the overall direction of the process. The researcher provides essential inputs, including a curated set of papers as the knowledge base and carefully crafted prompts to direct the AI models. One or multiple AI models process these inputs, which utilize the knowledge base and respond

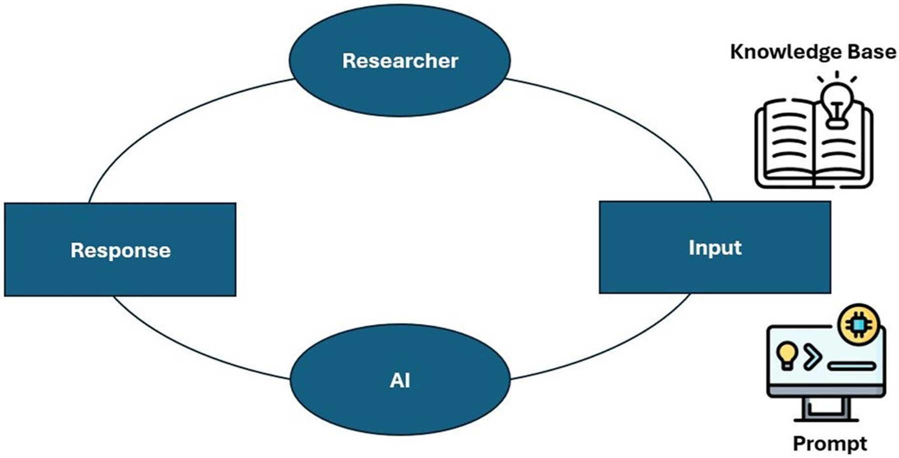

Fig. 1 Augmenting Cognition with AI.
Source: *Authors (2025)*

to the prompts with relevant information and insights (Marchionini 2008). The AI models generate responses based on the prompts, which the researcher then reviews and analyzes. This dynamic interaction between the researcher and AI enhances the efficiency and depth of various steps in systematic literature reviews, creating a continuous loop of input, processing, and response (see Fig. 1). The following section will show how this translates into a consistent framework for AI-augmented SLRs.

5 Framework for AI-augmented SLRs

This section presents a comprehensive and methodologically rigorous framework for incorporating AI into the SLR process. It sets it apart by addressing established methodologies and emerging challenges. The framework integrates cutting-edge advancements in AI while embedding critical epistemological principles, ensuring that the scholarly rigor of traditional SLRs is not compromised but enhanced.

By combining technological sophistication with methodological robustness, this framework contributes uniquely to the evolving landscape of SLR methodologies (Dell'Acqua et al., 2023; Mollick 2024). The framework comprises five interrelated stages: choice of AI model, integration of a domain-specific knowledge base, iterative prompt engineering, validation of outputs, and employment to the research question at hand (see Fig. 2). Each stage addresses critical methodological and epistemological challenges, balancing technological capabilities and scholarly rigor.

The first stage, the choice of an AI model, involves systematically landscaping available tools, followed by their evaluation based on predefined criteria (see Fig. 2). These criteria include domain specificity, adaptability, performance metrics such as sensitivity and precision, and compatibility with the intended research objectives. For instance, while general-purpose models like GPT-4 offer robust capabilities, specialized models such as PubMedBERT or ClinicalBERT demonstrate superior perfor-

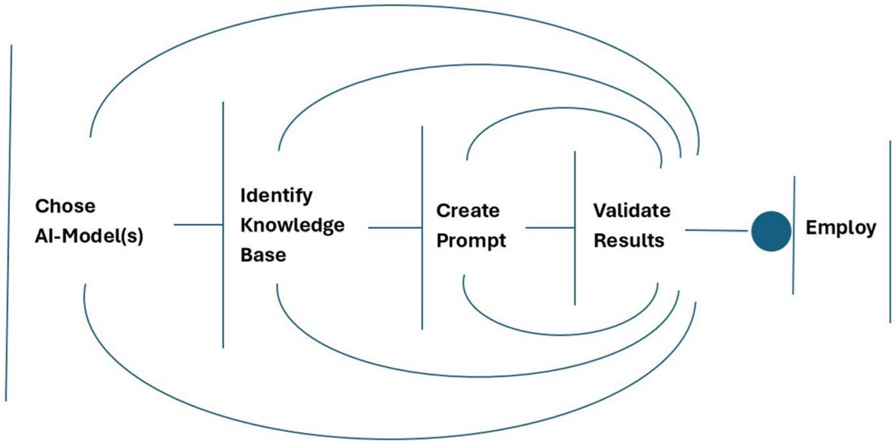

Fig. 2 Three AI Validation Iterations.
Source: *Authors (2025)*

mance in niche areas like medical literature reviews (Alavi et al. 2024; Alshami et al. 2023; Benbya et al. 2024; Dwivedi et al. 2023; Nazir and Wang 2023; O'Connor 2022). This stage ensures the chosen model's alignment with the study's contextual and methodological requirements, enhancing relevance and rigor. LLMs like GPT-4, along with specialized variants such as PubMedBERT and ClinicalBERT, have demonstrated significant efficacy in tasks such as title and abstract screening, as well as thematic clustering, particularly within domain-specific contexts (Alshami et al. 2023; Dehouche 2021; Dergaa et al. 2023; Li and Li 2024; Lindebaum and Fleming 2024; Shen et al. 2023). This evaluation emphasizes domain relevance, ensuring that the selected model aligns with the contextual requirements of the research. Additionally, fine-tuning pre-trained models with domain-specific datasets enhances their relevance and applicability (Anthony et al. 2023; Bolaños et al. 2024; Choudhary et al. 2023; Gao et al. 2024; Gatrell et al. 2024).

The second stage involves integrating a curated knowledge base, the foundation for model optimization. The quality and scope of this knowledge base are determined through stringent selection criteria, including the relevance of included datasets to the research objectives, their methodological rigor, and their alignment with the epistemological principles underlying the SLR. High-quality datasets are carefully chosen to refine the AI model's performance, ensuring that outputs are reliable and contextually relevant (Benbya et al. 2024; Tingelhoff et al. 2024).

This meticulous curation process not only enhances the alignment of AI-generated insights with research goals but also ensures that the methodological integrity of the SLR is upheld. Integrating well-defined and high-quality datasets is critical in refining AI's performance, ensuring that outputs align with specified research objectives. This process requires careful selection of datasets based on relevance, methodological rigor, and domain-specific applicability (Dwivedi et al. 2019; Schryen et al. 2024). For instance, integrating datasets curated for medical research has significantly enhanced AI outputs' precision and contextual reliability, making them more effective for systematic literature reviews. Prompt engineering plays a pivotal role here, enabling researchers to craft precise and contextually relevant instructions that guide AI processes. Advanced tools such as LangChain facilitate the seamless integration of prompts with AI models, optimizing their output quality (Alshami et al. 2023).

The third and fourth stages—iterative prompt engineering and validation—are structured as cyclical processes to maximize flexibility and ensure the continual refinement of AI outputs. This cyclical design is preferred over sequential steps due to its inherent adaptability and responsiveness to real-time feedback. Unlike linear methodologies, a cyclical approach enables iterative refinements that dynamically address emerging inconsistencies or gaps, thereby enhancing AI-generated outputs' precision and contextual relevance (Haenlein and Kaplan 2019; Tingelhoff et al. 2024). For example, in a non-deterministic generative AI environment, iterative cycles allow researchers to fine-tune prompts and validate the outputs continuously, ensuring alignment with evolving research objectives (Anthony et al. 2023; Davison et al. 2024; Farrelly and Baker 2023). This adaptability is particularly advantageous in interdisciplinary research contexts, where diverse datasets and methodologies demand frequent adjustments to maintain methodological rigor and epistemological integrity. By revisiting and fine-tuning prompts iteratively, researchers can enhance the alignment of AI-generated outputs with the research objectives, promptly addressing emerging inconsistencies or gaps.

This iterative process is particularly effective for tackling the non-deterministic nature of generative AI models, ensuring that output remains consistent, reliable, and contextually relevant (Farrelly and Baker 2023). The cyclical approach also integrates human oversight at every loop, blending computational efficiency with the critical judgment necessary for maintaining methodological rigor and epistemological integrity. This design enhances the accuracy of AI-augmented SLRs and fosters a deeper engagement with the complexities of interdisciplinary research landscapes. Prompt engineering involves iterative testing and refinement, enabling researchers to optimize model outputs through continuous adjustment. Validation ensures the reliability, reproducibility, and methodological rigor of AI-augmented processes. Inter-model comparisons and sensitivity testing ensure methodological rigor in AI-augmented SLRs. These processes minimize false negatives and enhance accuracy by leveraging cross-validation techniques across AI models.

For instance, inter-model validation allows researchers to identify discrepancies by comparing outputs from various models, thereby addressing biases inherent in individual algorithms. Sensitivity testing further examines the robustness of AI outputs by systematically varying input parameters to assess their impact on results. These methods are particularly effective in highlighting inconsistencies and ensuring that the findings remain reliable and contextually relevant. Empirical studies have demonstrated the efficacy and reliability of this approach, with sensitivity metrics ranging between 89% and 98% in diverse applications (Alshami et al. 2023; Dwivedi et al. 2023; Gao et al. 2024; Kharawala et al. 2024; Mariani et al. 2023; Tingelhoff et al. 2024). Researchers can maintain the review process's epistemological integrity by integrating inter-model comparisons with human oversight while leveraging AI tools' computational efficiency (Block and Kuckertz 2024; Lorenz et al. 2024; Schlagwein and Willcocks, 2023). Moreover, human oversight remains indispensable, integrating computational insights with contextual expertise to maintain epistemological integrity (Aguinis et al. 2023; Block and Kuckertz 2024; Glaser et al. 2024; Grimes et al. 2023; Hiebl 2023; Lee, 2021; Ngwenyama and Rowe 2024; Schlagwein and Willcocks, 2023; Simsek et al. 2023; Stokel-Walker, 2023).

A key innovation in this framework is the iterations validation system. This approach includes inter-model validation—where outputs from different AI models are cross-validated for consistency—and intra-model validation—where iterative prompt adjustments refine a single model's output. This dual approach addresses the challenges posed by the non-deterministic nature of generative AI, providing a robust mechanism for ensuring reliability and reproducibility (Pan et al. 2023).

The final stage involves deploying the refined AI-augmented workflow for study collection, selection, synthesis, and communication. AI tools significantly reduce manual effort while maintaining high methodological standards by automating labor-intensive tasks like citation mapping, thematic clustering, and narrative synthesis (Hiebl 2023; Snyder 2019). Empirical evidence suggests that these tools can reduce workloads by over 90%, facilitating rapid yet thorough analyses (Alshami et al. 2023; Kharawala et al. 2024). Thematic clustering capabilities further enhance the synthesis of actionable insights, particularly in interdisciplinary research contexts (Gatrell et al. 2024; Grimes et al. 2023; Kraus et al. 2020).

This framework represents a transformative approach to SLRs by offering a robust methodology that bridges AI's computational strengths with human judgment's intellectual rigor and contextual sensitivity (Choudhary et al. 2023). It prioritizes scalability, interdisciplinarity, and the synthesis of diverse knowledge streams, making it particularly relevant for complex research domains like management and organizational studies (Mariani et al. 2023; Mariani and Wirtz 2023; Reyes-Menendez et al. 2023). The practical implications of this framework include enhanced accuracy, reduced workload, and improved methodological rigor, which enable researchers to address the challenges of navigating extensive and diverse research landscapes.

6 Reflection on epistemic practices in AI-augmented SLRs

SLRs are underpinned by foundational epistemological principles that ensure methodological rigor, credibility, and alignment with scientific standards. These principles—transparency, validity, reliability, comprehensiveness, and reflective agency—are as crucial for traditional methodologies as AI-augmented approaches. Incorporating these principles within AI-driven SLRs ensures the integrity of reviews in the rapidly evolving fields of management and organizational studies (Aguinis et al. 2018, 2023; Alavi et al. 2024; Anthony et al. 2023; Berghofer 2022; Choi 2018; Hiebl 2023; Gond et al. 2023; Hofer 2017; Williams et al. 2021).

Transparency remains the cornerstone of SLR methodologies, guaranteeing that every stage of the review process—from study selection to data extraction—is systematically documented and accessible for replication and critical assessment (Aguinis et al. 2023; Harley and Cornelissen 2022; Pratt et al. 2022). While PRISMA guidelines provide a structured framework for ensuring transparency, AI tools further enhance this by maintaining detailed decision logs and systematically tracing automated processes (Aytug et al. 2012; Hiebl 2023; Liberati et al. 2009a, b). This dual approach facilitates reproducibility and strengthens the credibility of reviews (Bolaños et al. 2024; Schryen et al. 2024; Usberti 2019).

Validity ensures that SLR findings accurately address the research question and are derived from rigorous empirical evidence (Bobko et al. 2007; Kunisch et al. 2018). AI contributes to internal and external validity by triangulating data across multiple sources, identifying inconsistencies, and cross-referencing evidence (Aguinis and Solarino 2019; Kunisch et al. 2023). Such capabilities are particularly impactful in management and organizational studies, where diverse contexts and data streams often complicate the validation process (Alshami et al. 2023; Dwivedi et al. 2023; Grant and Booth 2009; Schryen et al. 2020).

Reliability underscores the consistency and reproducibility of the SLR process. Traditional workflows rely on strict adherence to predefined methodologies, but AI tools bring an additional layer of standardization by automating workflows, minimizing human error, and ensuring consistency across iterative analyses (Aguinis et al. 2023; Locke et al. 2022; Pollock 1984; Pratt et al. 2022). This is especially relevant in management studies, where reliability augments confidence in the scalability of findings and their application in various organizational contexts (Aguinis et al. 2023; Vom Brocke et al. 2009).

In the context of AI-augmented SLRs, ensuring reliability is paramount due to the non-deterministic behavior of LLMs. The inherent variability in LLM outputs can challenge the epistemic principle of reliability. Our framework employs an iterative approach incorporating multiple reliability check layers to address this (Bonett and Wright 2014; Hiebl 2023). Firstly, inter-model reliability is assessed by comparing outputs from different AI models to identify consistent patterns and discrepancies. Secondly, intra-model reliability is evaluated by examining the consistency of responses to variations in prompts within the same model. Finally, the reliability of AI-generated results is validated through human inspection of manageable datasets, ensuring that the findings are robust and credible. This triangulated approach enhances the reliability of AI-augmented SLRs, providing a comprehensive and trustworthy synthesis of the literature.

Comprehensiveness is pivotal for capturing a holistic view of the research landscape, ensuring the inclusion of all relevant studies and perspectives (Mees-Buss et al. 2022; Sauer and Seuring 2023; Zupic and Čater 2015). Traditional methods emphasize systematic and thorough coverage, while AI-powered technologies elevate this principle by automating large-scale searches, uncovering interdisciplinary linkages, and mitigating the risk of omitting critical studies. This is particularly significant for organizational research, which often draws on diverse and cross-disciplinary insights (Schryen et al. 2015; Williams et al. 2021).

Reflective Agency highlights the essential role of researchers in shaping the review process through critical thinking, methodological oversight, and ethical decision-making (Duan et al. 2012; Else 2023; Zagzebski 2020). While AI tools enable efficiency and scalability, human judgment ensures that the review process aligns with epistemological values and contributes to meaningful scientific discourse (Hoon 2013; Pratt 2008; Simsek et al. 2023). In management and organizational studies, the reflective agency ensures that SLRs aggregate knowledge and foster nuanced synthesis and innovative theoretical contributions (Ngwenyama and Rowe 2024; Lorenz et al. 2024; Reyes-Menendez et al. 2023; Vatankhah et al. 2024).

In AI-augmented SLRs, the reflective agency is crucial to ensure the responsible and well-managed use of AI technologies (Harley and Cornelissen 2022; Mees-Buss et al. 2022). Our framework incorporates a 'researcher-in-the-loop' approach, emphasizing human oversight and decision-making at every stage. This approach includes three key iterations: the reflective choice of AI models, the reflective selection of a knowledge base for fine-tuning foundational models, and the conscious and well-documented practice of prompt engineering. By embedding reflective agency within each validation loop, we ensure that researchers actively engage with the AI processes, making informed and deliberate choices. This enhances the reliability and validity of the AI-augmented SLRs and fosters a responsible and ethical integration of AI, aligning with the broader epistemic principles of transparency and reliability (Lorenz et al. 2024).

By synthesizing these principles with AI's computational capabilities, hybrid methodologies provide a robust framework for SLRs in management research. This integration fosters a dynamic approach that bridges the gap between traditional rigor and the demands of contemporary research, enabling scholars to address complex organizational challenges with methodological robustness and interdisciplinary adaptability (Russell and Norvig 2020; Schryen et al. 2024).

6.1 Application of AI into SLRs

The integration of AI into SLRs necessitates a deliberate alignment between epistemological principles and computational advancements. This alignment is central to the hybrid framework proposed in this study, which embeds iterative validation loops across three critical elements: AI models, knowledge bases, and prompt engineering (Alshami et al. 2023; Dwivedi et al. 2023; Tingelhoff et al. 2024). This section demonstrates how the principles of transparency, validity, reliability, comprehensiveness, and reflective agency are operationalized through AI-augmented SLR methodologies while also addressing broader implications for research practices (Kraus et al. 2024; Vom Brocke et al. 2009).

6.2 Illustrative case: AI-assisted application in an SLR

Consider a study focused on digital transformation strategies in multinational corporations to provide a concrete example of how researchers might apply the AI validation framework. A research team conducting an SLR in this domain integrates AI tools to enhance literature retrieval, thematic clustering, and citation mapping. The first step in their methodology involves selecting an LLM-based AI tool, such as OpenAI's GPT-4 or a domain-specific LLM, which is fine-tuned using relevant corpora from management research (Dwivedi et al. 2019; Jatobá et al. 2023; Mariani et al. 2023). Before applying AI to the full dataset, iterative testing ensures that the model accurately recognizes domain-specific concepts and relationships.

Once the AI model is selected, the next step is curating and expanding the knowledge base. The research team constructs the initial dataset using structured inclusion criteria based on established frameworks such as PRISMA (Grant and Booth 2009; Shahzadi et al. 2024). A small, manually validated sample is used to assess AI-generated classifications and to ensure alignment with existing management theories and empirical studies. As the dataset expands, human verification plays a critical role in maintaining methodological consistency and thematic integrity.

The final stage involves refining prompt engineering techniques to optimize AI performance. The researchers experiment with different prompt structures to refine the AI-generated thematic syntheses, iteratively modifying these prompts to enhance specificity, reduce biases, and improve contextual accuracy (Bilgram and Laarmann, 2023; Bolaños et al. 2024; Pan et al. 2023; Schryen et al. 2020). This structured validation process enhances the reliability of AI-generated outputs, reduces errors, and increases transparency in the review process. By systematically applying human oversight at every stage, the research team effectively mitigates AI hallucinations and aligns AI-driven insights with existing scholarly frameworks.

6.3 Practical recommendations for researchers

To enhance the applicability of this framework, the study expands on practical recommendations for researchers and research teams adopting AI-augmented SLRs. Selecting and fine-tuning AI models requires researchers to choose domain-specific LLMs that are pre-trained on relevant academic corpora, ensuring enhanced conceptual precision in literature synthesis (Anthony et al. 2023; Grimes et al. 2023; Gatrell et al. 2024; Gond et al. 2023). Before scaling up the literature synthesis process, preliminary model testing should be conducted on a small sample dataset to verify the accuracy and relevance of AI-generated insights (Dwivedi et al. 2023; Kharawala et al. 2024).

In developing validation protocols for AI-generated insights, cyclical validation techniques should be employed to compare AI-generated summaries with human-extracted themes, ensuring accuracy and consistency (Shaheen et al. 2023; Schryen et al. 2024). Inter-model comparisons can further enhance this process by evaluating variability in AI-generated outputs, reducing the risk of single-model biases (Bouschery et al. 2023; Ngwenyama and Rowe 2024). Transparency and reproducibility are critical to AI-augmented SLRs, and researchers should maintain detailed documentation of AI-generated results, including search parameters, dataset modifications, and thematic clustering processes (Liberati et al. 2009a, b; Webster and Watson 2002). Decision logs should be clearly outlined to facilitate replicability and enhance research credibility (Usberti 2019).

To mitigate biases in AI-augmented reviews, iterative human-AI interaction should be incorporated to identify and correct biases in AI-generated literature syntheses (Hiebl 2023). Sensitivity testing should be conducted to ensure that findings remain stable across different AI model configurations, thereby safeguarding the integrity of AI-assisted research (Alshami et al. 2023; Dwivedi et al. 2023; Grant and Booth 2009; Kumar et al. 2024; Tingelhoff et al. 2024). These practical recommendations provide a structured, repeatable approach to AI validation in SLRs, ensuring methodological rigor while embracing the potential of AI-driven research.

6.4 Future research directions in management and organizational studies

Building on these insights, future research should focus on advancing domain-specific LLMs, refining prompt engineering strategies, and developing standardized validation frameworks to enhance the reliability and impact of AI-augmented SLRs in management and organizational studies. The development of tailored LLMs trained specifically for management and organizational research can improve the precision of literature reviews by incorporating field-specific terminologies and conceptual frameworks (Aguinis et al. 2023; Mariani et al. 2023; Reyes-Menendez et al. 2023). Enhancing prompt engineering strategies through adaptive algorithms can mitigate inconsistencies in AI-generated outputs by responding to context-specific demands in management and organizational research (Anthony et al. 2023; Grimes et al. 2023; Gatrell et al. 2024; Gond et al. 2023; Kharawala et al. 2024; Kumar et al. 2024).

Further, establishing standardized validation frameworks will ensure the reproducibility and robustness of AI-assisted SLRs. Benchmarks for reliability must be developed to facilitate the adoption of AI tools across diverse organizational research contexts, promoting consistency and methodological integrity (Alavi et al. 2024; Benbya et al. 2024; Dwivedi et al. 2023; Tingelhoff et al. 2024). Addressing these research gaps will enable future studies to foster scalability, enhance methodological robustness, and support interdisciplinary research while maintaining epistemic rigor.

7 Conclusion

This study highlights the transformative potential of AI in advancing SLRs. By introducing a hybrid framework rooted in epistemological principles—transparency, validity, reliability, comprehensiveness, and reflective agency—the paper establishes a robust methodological foundation for AI-augmented SLRs (Aguinis et al. 2023; Mees-Buss et al. 2022). The framework integrates AI's computational capabilities, such as landscaping, fine-tuning, iterative prompt engineering, validation, and deployment, with human cognition to ensure methodological rigor, scalability, and transparency. This synergy positions AI-augmented SLRs as a crucial tool for addressing complex and interdisciplinary research challenges (Dwivedi et al. 2023; Kraus et al. 2024; Schryen et al. 2024).

The proposed framework makes significant contributions by embedding epistemological principles into the SLR process and emphasizing a cyclical approach to validation. Using multiple AI models or combining AI outputs with human insights can help cross-verify the findings. This triangulation increases the validity, reduces the risk of AI hallucinations, and ensures that the results are consistent with human interpretations and reflective agency. This iterative process, including inter-model comparisons and sensitivity testing, enhances AI-generated insights' reliability and contextual relevance (Lê and Schmid 2022; Ngwenyama and Rowes, 2024; Sharma and Bansal 2023; Tingelhoff et al. 2024). These contributions address the challenges posed by probabilistic AI models, ensuring trustworthiness and ethical integrity in research practices. Furthermore, the framework provides actionable strategies for researchers, such as tailoring AI models, curating domain-specific knowledge bases, and employing iterative prompt engineering, thereby aligning AI outputs with scholarly standards (Alvesson and Sandberg 2020; Bolaños et al. 2024; Benbya et al. 2024; Grimes et al. 2023; Lorenz et al. 2024).

The practical implications of this framework extend to the broader research community by advocating standardized protocols and metrics for evaluating AI-augmented methodologies. For example, integrating LLMs with comprehensive datasets facilitates thematic clustering and nuanced analyses, enhancing the depth and breadth of literature reviews (Grant and Booth 2009; Liberati et al. 2009a, b). The inclusion of sensitivity testing ensures robustness and reliability in findings, enabling researchers to address potential biases in AI-generated outputs preemptively (Alavi et al. 2024; Li and Li 2024; Locke et al. 2022; Vom Brocke et al. 2009; von Krogh et al. 2023).

Future research should focus on refining domain-specific applications of this framework, particularly the development of LLMs tailored to the unique needs of various academic disciplines. Additionally, efforts to quantify the reliability of AI-generated insights and establish standardized reporting guidelines will further enhance transparency and reproducibility (Grant and Booth 2009; Köhler et al., 2022; Kulkarni et al. 2024; Tingelhoff et al. 2024). Ethical considerations, such as algorithmic bias and data privacy, must be systematically addressed to align AI applications with broader academic and societal values (Aguinis., 2023; Dwivedi et al. 2023; Lorenz et al. 2024; Lund et al., 2023; Madill et al. 2000; Mees-Buss et al., 2022).

In conclusion, this study underscores the transformative potential of AI in reimagining SLR methodologies while maintaining their epistemological underpinnings. The proposed framework serves as a guide and calls to action for researchers to embrace AI as a collaborative partner in pursuing knowledge. By fostering a balanced synergy between human expertise, scientific community input, and computational efficiency, this approach offers a pathway to more comprehensive, reliable, and reflective syntheses of academic knowledge. This vision aligns with the broader goal of leveraging technology to extend the boundaries of human understanding while upholding scientific rigor and ethical integrity (Antons et al. 2023; Gond et al. 2023; Hofer 2017; Lorenz et al. 2024; Saura et al. 2024).

Funding Open Access funding enabled and organized by Projekt DEAL.
Project DEAL will provide it.

Data availability We do not analyze or generate any datasets because our work proceeds with a conceptual literature review approach.

Declarations

Competing interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access 6:52138–52160.

Adams RJ, Smart P, Huff AS (2017) Shades of Grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. Int J Manage Reviews 19(4):432–454

Aguinis H, Pierce CA, Bosco FA, Dalton DR, Dalton CM (2011) Debunking Myths and urban legends about meta-analysis. Organizational Res Methods 14(2):306–331

Aguinis H, Ramani RS, Alabduljader N (2018) What you see is what you get? Enhancing methodological transparency in management research. Acad Manag Ann 12(1):83–110

Aguinis H, Ramani RS, Alabduljader N (2023) Best-Practice recommendations for producers, evaluators, and users of methodological literature reviews. Organizational Res Methods 26(1):46–76

Aguinis H, Solarino AM (2019) Transparency and replicability in qualitative research: the case of interviews with elite informants. Strateg Manag J 40(8):1291–1315

Alavi M, Leidner DE, Mousavi R (2024) Knowledge management perspective of generative artificial intelligence. J Association Inform Syst 25(1):1–12

Almotahari M, Glick E (2010) Context, content, and epistemic transparency. Mind 119:1067–1086

Alshami A, Elsayed M, Ali E, Eltoukhy AE, Zayed T (2023) Harnessing the power of ChatGPT for automating systematic review process: methodology, case study, limitations, and future directions. Systems 11(7):351

Alvesson M, Sandberg J (2020) The problematizing review: A counterpoint to Elsbach and Van Knippenberg's argument for integrative reviews. J Manage Stud 57(6):1290–1304

Ameen N, Sharma GD, Tarba S, Rao A, Chopra R (2022) Toward advancing theory on creativity in marketing and artificial intelligence. Psychol Mark 39(9):1802–1825

Amoozad Mahdiraji H, Arabi S, Beheshti H, M., and, Vrontis D (2023) A mixed-method analysis of Industry 4.0 technologies in value generation for collaborative consumption companies. *Management Decision*, Vol. ahead-of-print No. ahead-of-print

Anderson N, Belavy DL, Perle SM, Hendricks S, Hespanhol L, Verhagen E, Memon AR (2023) AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports and exercise medicine manuscript generation. BMJ Open Sport Exerc Med, 9(1), e001568

Anthony C, Bechky BA, Fayard AL (2023) Collaborating with AI: taking a system view to explore the future of work. Organ Sci 34(5):1672–1694

Antons D, Breidbach CF, Joshi AM, Salge TO (2023) Computational literature reviews: method, algorithms, and roadmap. Organizational Res Methods 26(1):107–138

Audi R, Goldman A (1989) Epistemology and cognition. Philos Phenomenol Res 49:733

Aytug ZG, Rothstein HR, Zhou W, Kern MC (2012) Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. Organizational Res Methods 15(1):103–133

Baškarada S (2014) Qualitative case study guidelines. Qualitative Rep 19(40):1–25

Benbya H, Strich F, Tamm T (2024) Navigating generative artificial intelligence promises and perils for knowledge and creative work. J Association Inform Syst 25(1):23–36

Berghofer P (2022) How to supplement mentalist evidentialism: what are the fundamental epistemological principles?? Theoria 88:679–700

Bilgram V, Laarmann F (2023) Accelerating innovation with generative AI: AI-augmented digital prototyping and innovation methods. IEEE Eng Manage Rev 51(2):18–25

Block JH, Fisch C (2020) Eight tips and questions for your bibliographic study in business and management research. Manage Rev Q 70:307–312

Block J, Kuckertz A (2024) What is the future of human-generated systematic literature reviews in an age of artificial intelligence? Manage Rev Q 74:1975–1980

Bobko P, Roth PL, Buster MA (2007) The usefulness of unit weights in creating composite scores: A literature review, application to content validity, and Meta-Analysis. Organizational Res Methods 10(4):689–709

Bolaños F, Salatino A, Osborne F, Motta E (2024) Artificial intelligence for literature reviews: Opportunities and challenges. *Artificial Intelligence Review, 57(259).*

Bonett DG, Wright TA (2014) Cronbach's alpha reliability: interval estimation, hypothesis testing, and sample size planning. J Organizational Behav 36(1):3–15

Bouschery SG, Blazevic V, Piller FT (2023) Augmenting human innovation teams with artificial intelligence: exploring transformer-based Language models. J Prod Innov Manage 40(2):139–153

Breslin D, Gatrell C (2023) Theorizing through literature reviews: the miner-prospector continuum. Organizational Res Methods 26(1):139–167

Cai Y, Mao S, Wu W, Wang Z, Liang Y, Ge T,… and, Wei F (2023) Low-code LLM: Graphical User Interface over Large Language Models. *arXiv preprint arXiv:2304.08103*

Chintalapati S, Pandey SK (2022) Artificial intelligence in marketing: A systematic literature review. Int J Market Res 64(1):38–68

Choi P (2018) Reliabilism, scepticism, and evidentia in Ockham. Br J History Philos 27:23–45

Choudhary V, Marchetti A, Shrestha YR, Puranam P (2023) Human-AI ensembles: when can they work? J Manag, 01492063231194968

Clark WR, Clark LA, Raffo DM, Williams RI (2021) Extending fisch and block's (2018) tips for a systematic review in management and business literature. Manage Rev Q 71:215–231

Cronin MA, George E (2023) The why and how of the integrative review. Organizational Res Methods 26(1):168–192

Davison RM, Chughtai H, Nielsen P, Marabelli M, Iannacci F, van Offenbeek M et al (2024) The ethics of using generative AI for qualitative data analysis. Inform Syst J 34(5):1433–1439

Deepa R, Sekar S, Malik A, Kumar J, Attri R (2024) Impact of AI-focussed technologies on social and technical competencies for HR managers–A systematic review and research agenda. Technol Forecast Soc Chang 202:123301

Dehouche N (2021) Plagiarism in the age of massive generative Pre-trained Transformers (GPT-3). Ethics Sci Environ Politics 21:17–23

Dell'Acqua F, McFowland III, Mollick E, Lifshitz-Assaf ER, Kellogg H, Rajendran K, S.,… and, Lakhani KR (2023) Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Technology and Operations Mgt. Unit Working Paper*, (24–013)

Denyer D, Tranfield D (2009) Producing a systematic review. In: Buchanan DA, Bryman A (eds) The Sage handbook of organizational research methods. Sage Publications Ltd., pp 671–689

Dergaa I, Chamari K, Zmijewski P, Ben Saad H (2023) From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. Biology Sport 40(2):615–622

Duan Y, Edwards JS, Dwivedi YK (2019) Artificial intelligence for decision making in the era of big Data–evolution, challenges and research agenda. Int J Inf Manag 48:63–71

Duan Y, Ong VK, Xu M, Mathews BP (2012) Supporting decision making process with ideal software agents—What do business executives want? Expert Syst Appl 39(10):8724–8738

Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, Williams MD (2019) Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manag 48:63–71

Dwivedi YK, Kshetri N, Hughes L, Slade E, Jeyaraj A, Kar AK, Wirtz J (2023) So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. Int J Inf Manag 71:102642

Eisenhardt KM (1989) Building theories from case study research. Acad Manage Rev 14(4):532–550

Else H (2023) Abstracts written by ChatGPT fool scientists. Nature 613(7945):423

Endrigo Sordan J, Andersson R, Antony J, Lopes Pimenta M, Oprime PC (2024) How industry 4.0, artificial intelligence and augmented reality can boost digital lean six Sigma. Total Qual Manage Bus Excellence 35(13–14):1542–1566

Etsenake D, Nagappan M (2024) Understanding the Human-LLM Dynamic: A Literature Survey of LLM Use in Programming Tasks. *arXiv preprint arXiv:2410.01026*

Farrelly T, Baker N (2023) Generative artificial intelligence: implications and considerations for higher education practice. Educ Sci 13(11):1109

Fisch C, Block J (2018) Six tips for your (systematic) literature review in business and management research. Manage Rev Q 68(2):103–106

Gao J, Gebreegziabher SA, Choo KTW, Li TJJ, Perrault ST, Malone TW (2024), May A Taxonomy for Human-LLM Interaction Modes: An Initial Exploration. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1–11)

Gatrell C, Muzio D, Post C, Wickert C (2024) Here, there and everywhere: on the responsible use of artificial intelligence (AI) in management research and the peer-review process. J Manage Stud 61(3):739–751

Ge Z, Huang H, Zhou M, Li J, Wang G, Tang S, Zhuang Y (2024, October) Worldgpt: empowering llm as multimodal world model. In: Proceedings of the 32nd ACM internationalconference on multimedia (pp 7346–7355).

Gibbert M, Ruigrok W (2010) The "what"and "how"of case study rigor: three strategies based on published work. Organizational Res Methods 13(4):710–737

Glaser VL, Sloan J, Gehman J (2024) Organizations as algorithms: A new metaphor for advancing management theory. J Manage Stud 61(6):2748–2769

Gond J-P, Mena S, Mosonyi S (2023) The performativity of literature reviewing: constituting the corporate social responsibility literature through Re-presentation and intervention. Organizational Res Methods 26(2):195–228

Grant MJ, Booth A (2009) A typology of reviews: an analysis of 14 review types and associated methodologies. Health Inf Libr J 26(2):91–108

Grimes M, von Krogh G, Feuerriegel S, Rink F, Gruber M (2023) From scarcity to abundance: scholars and scholarship in an age of generative artificial intelligence. Acad Manag J 66(6):1617–1624

Haenlein M, Kaplan AM (2019) A brief history of artificial intelligence: on the past, present, and future of artificial intelligence. Calif Manag Rev 61(4):5–14

Harley B, Cornelissen J (2022) Rigor with or without templates? The pursuit of methodological rigor in qualitative research. Organizational Res Methods 25(2):239–261

Hayden N, Gilbert S, Poisson LM, Griffith B, Klochko C, Moy L, Wolfe S (2024) Performance of GPT-4 with vision on text- and image-based ACR diagnostic radiology in-training examination questions. Radiology, 312(3), e240153

Hiebl MR (2023) Sample selection in systematic literature reviews of management research. Organizational Res Methods 26(2):229–261

Hofer B (2017) Shaping the epistemology of teacher practice through reflection and reflexivity. Educational Psychol 52:299–306

Hoon C (2013) Meta-synthesis of qualitative case studies: an approach to theory Building. Organizational Res Methods 16(4):522–556

Hugging Face (n.d.). Hugging Face – The AI community building the future. Retrieved January 1, 2025, from https://huggingface.co

Jatobá MN, Ferreira JJ, Fernandes PO, Teixeira JP (2023) Intelligent human resources for the adoption of artificial intelligence: a systematic literature review. J Organizational Change Manage, (ahead-of-print).

Kasneci E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, Kasneci G (2023) ChatGPT for good? On opportunities and challenges of large Language models for education. Learn Cult Social Interact, *38*, Article 100671.

Kharawala S, Isaacs S, Jindal D, Gandhi P (2024) Artificial intelligence in systematic literature reviews: Part 1 – AI-aided title/abstract screening. *Bridge Medical White Paper*. Retrieved from https://bridgemedical.org. Retreived on January 2, 2025

Köhler T, Smith A, Bhakoo V (2022) Templates in qualitative research methods: origins, limitations, and new directions. Organizational Res Methods 25(2):183–210

Kitchenham B (2004) Procedures for performing systematic reviews. *Keele University Technical Report TR/SE-0401.*

Korzynski P, Mazurek G, Krzypkowska P, Kurasinski A (2023) Artificial intelligence prompt engineering as a new digital competence: analysis of generative AI technologies such as ChatGPT. Entrepreneurial Bus Econ Rev 11(3):25–37

Kraus S, Bouncken RB, Yela Aránega A (2024) The burgeoning role of literature review articles in management research: an introduction and outlook. RMS 18(2):299–314

Kraus S, Breier M, Dasí-Rodríguez S (2020) The Art of crafting a systematic literature review in entrepreneurship research. Int Entrepreneurship Manage J 16:1023–1042

Kulkarni M, Mantere S, Vaara E, van den Broek E, Pachidi S, Glaser VL et al (2024) The future of research in an artificial intelligence-driven world. J Manage Inq 33(3):207–229

Kumar S, Rao P, Singhania S, Verma S, Kheterpal M (2024) Will artificial intelligence drive the advancements in higher education? A tri-phased exploration. Technol Forecast Soc Chang 201:123258

Kunisch S, Denyer D, Bartunek JM, Menz M, Cardinal LB (2023) Review research as scientific inquiry. Organizational Res Methods 26(1):3–45

Kunisch S, Menz M, Bartunek JM, Cardinal LB, Denyer D (2018) Feature topic at organizational research methods: how to conduct rigorous and impactful literature reviews?? Organizational Res Methods 21(3):519–523

Lee B (2021) A coherentist justification of epistemic principles and its merits. Acta Anal 36:533–551

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Moher D (2009b) The PRISMA statement for reporting systematic reviews and meta-analyses. PLoS Med, 6(7), e1000097

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Moher D (2009a) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *Bmj*, *339*

Li J, Li J (2024) A map of exploring human interaction patterns with LLM: insights into collaboration and creativity. ArXiv, abs/2404.04570.

Lindebaum D, Fleming P (2024) ChatGPT undermines human reflexivity, scientific responsibility and responsible management research. Br J Manag 35(2):566–575

Lê JK, Schmid T (2022) The practice of innovating research methods. Organizational Res Methods 25(2):308–336

Locke K, Feldman M, Golden-Biddle K (2022) Coding practices and iterativity: beyond templates for analyzing qualitative data. Organizational Res Methods 25(2):262–284

Lorenz F, Lorenzen S, Franco M, Velz J, Clauß T (2024) Generative artificial intelligence in management research: a practical guide on mistakes to avoid. Manage Rev Q, 1–21

Lund BD, Wang T, Mannuru NR, Nie B, Shimray S, Wang Z (2023) ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the largelanguage models in scholarly publishing. J Assoc Inf Sci Technol 74(5):570–581.

Lu Y (2019) Artificial intelligence: A survey on evolution, models, applications and future trends. J Manage Analytics 6(1):1–29

Maaravi Y, Heller B, Shoham Y, Mohar S, Deutsch B (2020) Ideation in the digital age: literature review and integrative model for electronic brainstorming. RMS 14(3):561–591

Maaravi Y, Heller B, Shoham Y, Mohar S, Deutsch B (2021) Ideation in the digital age: literature review and integrative model for electronic brainstorming. Rev Manag Sci 15:1431–1464.

Madill A, Jordan A, Shirley C (2000) Objectivity and reliability in qualitative analysis: realist, contextualist and radical constructionist epistemologies. Br J Psychol 91(Pt 1):1–20

Marchionini G (2008) Human–information interaction research and development. Libr Inform Sci Res 30(3):165–174

Mariani MM, Machado I, Magrelli V, Dwivedi YK (2023) Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. Technovation 122:102623

Mariani M, Wirtz J (2023) A critical reflection on analytics and artificial intelligence based analytics in hospitality and tourism management research. Int J Contemp Hospitality Manage 35(8):2929–2943

Mees-Buss J, Welch C, Piekkari R (2022) From templates to heuristics: how and why to move beyond the Gioia methodology. Organizational Res Methods 25(2):405–429

Mollick E (2024) Co-intelligence: living and working with AI. Penguin Portfolio

Najafi B, Najafi A, Farahmandian A (2023) The impact of artificial intelligence and blockchain on six sigma: a systematic literature review of the evidence and implications. IEEETrans Eng Manag 71:10261–10294.

Nazir A, Wang Z (2023) A comprehensive survey of ChatGPT: advancements, applications, prospects, and challenges. Meta-Radiology 1(2):100022

Nguyen N-T, Nguyen TV, Kashani A, Ngo T, Bordas SPA (2019) Deep neural network with high-order neuron for the prediction of foamed concrete strength. Computer-Aided Civ Infrastruct Eng 34(4):316–332

Ngwenyama O, Rowe F (2024) Should we collaborate with AI to conduct literature reviews? Changing epistemic values in a flattening world. J Association Inform Syst 25(1):122–136

Noy S, Zhang W (2023) Experimental evidence on the productivity effects of generative artificial intelligence. Science 381(6654):187–192

O'Connor S (2022) Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? Nurse Educ Pract 66:103537
Okoli C, Schabram K (2010) A guide to conducting a systematic literature review of information systems research. Sprouts: Working Papers Inform Syst 10(26):1–49
Pan SL, Nishant R, Tuunanen T, Nah FF-H (2023) Literature review in the generative AI era—How to make a compelling contribution. J Strategic Inform Syst 32(3):4
Pollock J (1984) Reliability and justified belief. Can J Philos 14:103–114
Porsdam Mann S, Earp BD, Nyholm S, Danaher J, Møller N, Bowman-Smart H et al (2023) Generative AI entails a credit–blame asymmetry. Nat Mach Intell 5(5):472–475
Pratt MG (2008) Fitting oval pegs into round holes: tensions in evaluating and publishing qualitative research in top-tier North American journals. Organizational Res Methods 11(3):481–509
Pratt MG, Sonenshein S, Feldman MS (2022) Moving beyond templates: A bricolage approach to conducting trustworthy qualitative research. Organizational Res Methods 25(2):211–238
Prikshat V, Islam M, Patel P, Malik A, Budhwar P, Gupta S (2023) AI-Augmented HRM: literature review and a proposed multilevel framework for future research. Technol Forecast Soc Chang 193:122645
Najafi B, Najafi A, Farahmandian A (2023) The impact of artificial intelligence and blockchain on six Sigma: A systematic literature review of the evidence 71:10261–10294.
Rammal HG (2023) Systematic literature reviews: steps and practical tips. Advancing methodologies of conducting literature review in management research. Emerald Publishing Limited, pp 27–35
Redondo-Rodríguez MD, Díaz-Garrido E, Pérez-Bustamante Yábar DC, Ramón-Jerónimo M (2024) Á. Entrepreneurship and artificial intelligence: a bibliometric analysis. *The Journal of Technology Transfer*, 1–33
Reyes-Menendez A, Clemente-Mediavilla J, Villagra N (2023) Understanding STI and SDG with artificial intelligence: A review and research agenda for entrepreneurial action. Technol Forecast Soc Chang 196:122785
Robledo S, Grisales Aguirre AM, Hughes M, Eggers F (2021) Hasta La Vista, baby – will machine learning terminate human literature reviews in entrepreneurship? J Small Bus Manage 61(3):1314–1343
Russell SJ, Norvig P (2020) *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson
Sauer PC, Seuring S (2023) How to conduct systematic literature reviews in management research: A guide in 6 steps and 14 decisions. RMS 17(1):1–35
Saura JR, Škare V, Dosen DO (2024) Is AI-based digital marketing ethical? Assessing a new data privacy paradox. J Innov Knowl 9(4):100597
Schlagwein D, Willcocks L (2023) ChatGPT et al.': the ethics of using (generative) artificial intelligence in research and science. J Inform Technol 38(3):232–238
Schryen G, Marrone M, Yang J (2024, January) Adopting generative AI for literature reviews: an epistemological perspective. In *HICSS* (pp. 4020–4029)
Schryen G, Wagner G, Benlian A (2015) Theory of knowledge for literature reviews: an epistemological model, taxonomy and empirical analysis of IS literature. In *ICIS 2015 Proceedings*, 2015
Schryen G, Wagner G, Benlian A, Paré G (2020) A knowledge development perspective on literature reviews: Validation of a new typology in the IS field. *Communications of the AIS*, *46*
Shaheen N, Shaheen A, Ramadan A, Hefnawy MT, Ramadan A, Ibrahim IA, Hassanein ME, Ashour ME, Flouty O (2023) Appraising systematic reviews: A comprehensive guide to ensuring validity and reliability. Front Res Metrics Analytics 8:1268045
Shahzadi G, Jia F, Chen L, John A (2024) AI adoption in supply chain management: a systematic literature review. J Manuf Technol Manage 35(6):1125–1150
Sharma G, Bansal P (2023) Partnering up: including managers as research partners in systematic reviews. Organizational Res Methods 26(2):262–291
Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, Moy L (2023) ChatGPT and other large Language models are double-edged swords. Radiology, 307(2), e230163
Simsek Z, Fox B, Heavey C (2023) Systematicity in organizational research literature reviews: A framework and assessment. Organizational Res Methods 26(2):292–321
Snyder H (2019) Literature review as a research methodology: an overview and guidelines. J Bus Res 104:333–339
Stokel-Walker C (2023) ChatGPT listed as author on research papers: many scientists disapprove. Nature 613(7945):620–621
Sumbal A, Sumbal R, Amir A (2024) Can ChatGPT-3.5 pass a medical exam? A systematic review of ChatGPT's performance in academic testing. J Med Educ Curriculum Dev 11:23821205241238640

Tapeh TG, A., and, Naser MZ (2022) Artificial intelligence, machine learning, and deep learning in structural engineering: A scientometrics review of trends and best practices. Arch Comput Methods Eng 29(6):3751–3796

Thorp HH (2023) ChatGPT is fun, but not an author. Science 379(6630):313

Tingelhoff F, Brugger M, Leimeister JM (2024) A guide for structured literature reviews in business research: the state-of-the-art and how to integrate generative artificial intelligence. J Inform Technol 39(4):489–505

Tranfield D, Denyer D, Smart P (2003) Toward a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 14(3):207–222

Usberti G (2019) Inference and Epistemic Transparency. *Topoi*, 1–14

Vatankhah S, Bamshad V, Arici HE, Duan Y (2024) Ethical implementation of artificial intelligence in the service industries. Serv Ind J, 1–25

Villiger J, Schweiger SA, Baldauf A (2022) Making the invisible visible: guidelines for the coding process in meta-analyses. Organizational Res Methods 25(4):716–740

Vom Brocke J, Simons A, Riemer K, Niehaves B, Plattfaut R, Cleven A (2009) Reconstructing the giant: on the importance of rigor in documenting the literature search process. ECIS Proc 161:2206–2217

von Krogh G, Roberson Q, Gruber M (2023) Recognizing and utilizing novel research opportunities with artificial intelligence. Acad Manag J 66(2):367–373

Wach K, Duong CD, Ejdys J, Kazlauskaitė R, Korzynski P, Mazurek G et al (2023) The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Bus Econ Rev 11(2):7–30

Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z et al (2023) Scientific discovery in the age of artificial intelligence. Nature 620(7972):47–60

Webster J, Watson RT (2002) Analyzing the past to prepare for the future: writing a literature review. MIS Q 26(2):xiii–xxiii

Williams RI Jr, Clark LA, Clark WR, Raffo DM (2021) Re-examining systematic literature review in management research: additional benefits and execution protocols. Eur Manag J 39(4):521–533

Wittenborg T, Karras O, Auer S (2024) SWARM-SLR: Streamlined workflow automation for machine-actionable systematic literature reviews. *arXiv preprint arXiv:2407.18657.*

Wu S, Fung M, Qian C, Kim J, Hakkani-Tur D, Ji H (2024) Aligning LLMs with Individual Preferences via Interaction. *arXiv preprint arXiv:2410.03642*

Wu T, He S, Liu J, Sun S, Liu K, Han QL, Tang Y (2023) A brief overview of ChatGPT: the history, status quo and potential future development. IEEE/CAA J Automatica Sinica 10(5):1122–1136

Zagzebski LT (2020) Must knowers be agents? In *Epistemic values: Collected papers in epistemology* (online ed.). Oxford Academic

Zhang C, Lu Y (2021) Study on Artificial intelligence: the state of the Art and future prospects. J Industrial Inform Integr 23:100224

Zupic I, Čater T (2015) Bibliometric methods in management and organization. Organizational Res Methods 18(3):429–472