Journal of the Association for Information Systems

Volume 25
Issue 1 *Special Issue: The Future Impact of AI on Academic Journals and the Editorial Process (pp. 37-181)*

Article 5

2024

Should We Collaborate with AI to Conduct Literature Reviews? Changing Epistemic Values in a Flattening World

Ojelanki Ngwenyama
Toronto Metropolitan University / University of Cape Town, ojelanki@torontomu.ca

Frantz Rowe
University of Nantes, frantz.rowe@univ-nantes.fr

Recommended Citation

Ngwenyama, Ojelanki and Rowe, Frantz (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.
DOI: 10.17705/1jais.00869
Available at: https://aisel.aisnet.org/jais/vol25/iss1/5

Journal of the Association for Information Systems (2024) **25**(1), 122-136
doi: 10.17705/1jais.00869

EDITORIAL

ISSN 1536-9323

Should We Collaborate with AI to Conduct Literature Reviews? Changing Epistemic Values in a Flattening World

Ojelanki Ngwenyama,[1] Frantz Rowe[2]

[1]Toronto Metropolitan University, Canada / University of Cape Town, South Africa, ojelanki@torontomu.ca
[2]Institut d'Economie et Management de Nantes, University of Nantes, France, frantz.rowe@univ-nantes.fr

Abstract

In this paper, we revisit the issue of collaboration with artificial intelligence (AI) to conduct literature reviews and discuss if this should be done and how it could be done. We also call for further reflection on the epistemic values at risk when using certain types of AI tools based on machine learning or generative AI at different stages of the review process, which often require the scope to be redefined and fundamentally follow an iterative process. Although AI tools accelerate search and screening tasks, particularly when there are vast amounts of literature involved, they may compromise quality, especially when it comes to transparency and explainability. Expert systems are less likely to have a negative impact on these tasks. In a broader context, any AI method should preserve researchers' ability to critically select, analyze, and interpret the literature.

Keywords: Artificial Intelligence (AI), Machine Learning, Expert Systems, Generative AI, Collaboration with AI, Literature Review, Control, Epistemic Values

Dov Te'eni was the accepting senior editor. This paper was submitted on July 1, 2023 and underwent two revisions. It is part of the Special Issue on The Future Impact of AI on Academic Journals and the Editorial Process.

1 Introduction

Recent rapid advances in predictive and generative AI technologies pose many new opportunities and challenges to IS scientific research. In this paper, we discuss the potentials and limitations of generative pretrained transformers (GPTs) based on large language models (LLM), and more broadly machine learning (vs. rule-based [expert] systems), for supporting core activities of literature review (LR) development. Presently, GPTs are trained using machine learning (ML) and deep learning (DL) to analyze patterns within large data sets in order to predict and generate new data based on prompts provided by users (Floridi, 2023). In this regard, GPTs can produce new images, data sets, narrative text, and so on. These capabilities introduce new and complex challenges for scientific researchers. Moreover, the challenges can differ by discipline and the specific AI technology. For example, whereas generative AI models may be legitimately used to generate data for DNA or pharmacological research (Killoran, et al., 2017; Vert, 2023), such use would be illegitimate in most types of IS research. While we do envision that data and image generation may be useful in design science IS research, that is beyond the scope of our discussion. Here, our focus is on machine learning and rule-based expert system tools for selection, analysis, and the generation of *text.* Our principal concern is: What challenges do IS scientists face when using these AI tools to support LR development?

It may seem odd to ask such questions since many of the data analysis platforms commonly used in IS research (Atlas/TI, HyperRESEARCH, NVivo, QDAMiner, etc.) already embed AI text mining and natural language processing (NLP) algorithms. Early adopters (medicine, pharmaceutical sciences, and software engineering) are using AI/ML data and text mining tools for LR and hypothesis generation (Spangler et al., 2014). However, the move to GPTs for searching and analyzing large literature databases to generate narrative text introduces new challenges for researchers. Recently, the developer of the widely used cloud-based qualitative data analysis platform Atlas/TI, collaborated with OpenAI to embed ChatGPT in the Atlas/TI platform (Atlasti, 2023). GPTs can save valuable time but they are prone to hallucination (Zheng & Zhan, 2023). Moreover, editors and publishers seem open to the possibility of legitimizing the use of AI/ML tools, GPTs included, even if they create new complications (L. Hassink, personal communication, March 2023).

In an ICIS panel eight years ago, before the recent buzz surrounding generative AI, senior scholars presented and debated opposing positions about whether AI would replace us as IS researchers. Putting forward the extreme position, M. Lynne Markus argued that there may not even be any collaboration because AI may simply replace us a lot sooner than we think (see Markus in Loebbecke et al., 2020). The exponential growth in publications is also making traditional search and analysis for LR development more difficult without any automation (Bornmann & Mutz, 2015). Whether and how this automation should be AI-supported merits further discussion. In the future, institutions governing academia may consider human IS researchers obsolete because they are governed by norms, investment, and funding strategies that prioritize efficiency. Under the pressures of technology ideology, digital giants, and the pursuit of financial interests, they may decide that we are not efficient enough or too prone to error and opt for AI instead (Ngwenyama et al., 2023), perceiving it as more disciplined, reliable, and continuously perfectible (Anders, 1956). Second, with AI power, big data has proclaimed the end of theory, because once phenomena are traceable, there are enough data to derive findings (Mayer-Schönberger & Cuckier, 2013). However, algorithmic decision-making raises tremendous ethical and explainability issues (O'Neil, 2017), which we interpret as non-epistemic and epistemic values (Grover & Rowe in Dwivedi et al., 2023). How do we as IS scientists respond to these emerging challenges? We need critical discourse about these new AI technologies to understand how they could vitalize our scientific practice and what the appropriate norms of use might be to defend against the corrosion of the legitimacy of scientific work.

Our scientific publication practices are rooted in the epistemic and ontological values that have become ingrained in our profession. Precision (of vocabulary and measures), generality, and realism are the three main values that underpin the validity of the knowledge we generate. We use any combination of these values in our empirical research strategies (Iivari, 2023). Similarly, despite claiming the qualities of interestingness, the internal consistency of ideas, rigor, and a good story (Agarwal, 2012), we are selective when conducting LRs. For example, some researchers privilege systematicity and transparency when performing certain tasks and reporting research (Paré et al., 2016) while others prioritize relevance and the interestingness of findings (Leidner, 2018). Still others give more importance to theoretical cogency and less to how the findings are obtained (Rowe & Markus 2023; Boell & Cecez-Kecmanovic, 2014). As types of reviews vary with values, what we understand by AI and the types of AI we choose might be different, even strikingly so, depending on our values. For example, would it be more appropriate to conduct a meta-analysis with AI based on machine learning that befits predictability or on an expert system that allows for greater transparency? In sum, adopting certain AI tools or approaches may have significant consequences on the evolution of our epistemic values, and this is something we should all be concerned about. In this paper, we consider the types of issues involved with some machine learning tools, possibilities for collaborating with them, implications for the defensibility of scientific discourse, and the types of LR activities that AI tools (expert systems vs. GPTs and ML/DL) can support.

2 Conceptual Framework

To frame our discussion, we use Toulmin's model of argumentation (Toulmin, 1958), and Habermas's principles of scientific discourse (Habermas, 1984; 2003), which are familiar to IS researchers.[1] In general, an LR is a substantive theoretical synthesis and discussion of a set of claims about a specific area of literature. Of course, LRs can have various theory-related goals and objectives (Rowe, 2014), such as describing the state of knowledge in a certain area, proposing a new theory based on literature analysis, or critically discussing the conceptual limitations of paradigmatic assumptions or theoretical perspectives (Rowe et al., 2023). Each type of LR has its own set of meta-requirements and is subject to specific disciplinary norms (Templier & Paré, 2018). However, it is also important to note that LRs are integral to scientific discourses that are normatively regulated by epistemic communities. We believe that the social-institutional nature of IS research (and LRs) is likely to be the Achilles' heel of ChatGPT and other ML and DL tools such as Gemini. Can they satisfy criteria for transparent scientific discourse *and* the epistemic community values of argumentation *(for defensibility* and *logical consistency)* in a way that conforms to *norms* and *paradigmatic assumptions* (Habermas, 1984; Toulmin, 1958)? Figure 1 below illustrates the model's core concepts and relationships.

[1] For some IS discussions see Ngwenyama (2023).

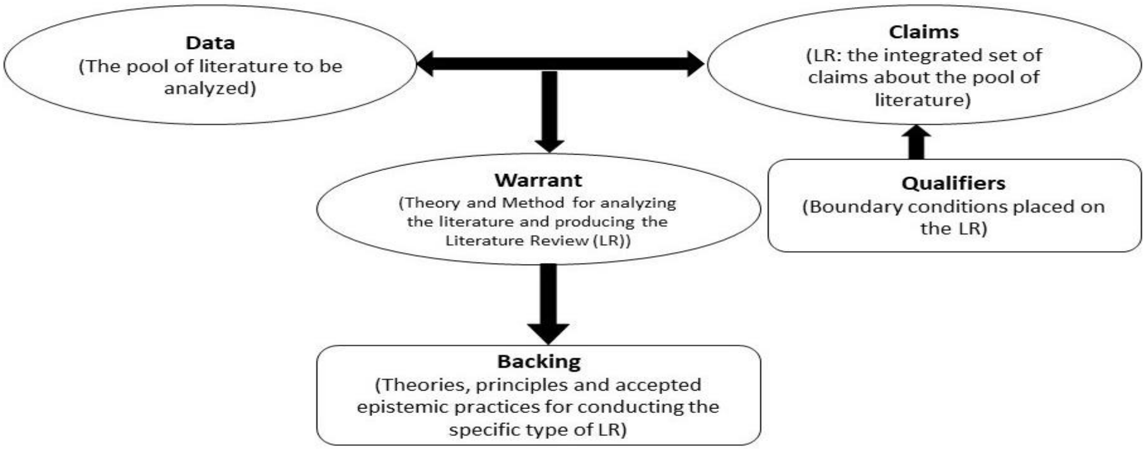

Figure 1. A Toulminian Structure of Literature Reviews (LRs) (adapted from Toulmin, 1958).

First, a lot has been written on selection and exclusion decision criteria for constituting the pool of literature (**data**) to be analyzed (vom Brocke et al., 2009). Moreover, search and selection criteria can be specified in the GPT, such that this LR task can be effectively automated. The integrated set of **claims** argued from analyzing the pool of literature (**data**) can be different depending on the type of LR the researcher is developing. For example, a more descriptive literature review would argue a very different set of claims than one proposing a new theory (Rowe, 2014). Developing literature reviews is fundamentally an interpretive process (Rowe, 2014). Furthermore, each type of literature review will use different **warrants**, or *analytical lenses* for interpreting data and assessing logical consistency of claims (Leidner, 2018), which in turn are based on different ontological and epistemological assumptions: namely, **backing**. From this perspective of scientific discourse ethics, three important issues of AI use in LR activities require critical analysis: (1) capabilities for supporting interpretation, rational argumentation, and transparency; (2) sociotechnical requirements for AI tool use; and (3) which AI tools are appropriate for which LR core activities.

3 Capabilities for Literature Analysis: Interpretation and Transparency Issues

Ram notes that there are important differences in transparency in the AI paradigms of expert systems and machine learning:[2] The expert systems approach is interpretable and transparent but the machine learning approach is not (Ram in Loebeckke et al., 2020). If we agree that literature reviews are an interpretive achievement, then we must ask two questions about **warrants** (Figure 1): What capabilities do IS scientists use when analyzing and interpreting pools of literature to develop an LR? Do AI tools, such as ChatGPT and Gemini, have any of these capabilities? Gadamer (1977) notes that interpretation is not a disembodied rule-driven activity of extracting meaning from the text. Even if we acknowledge that ChatGPT, Gemini and other GAI tools can summarize texts or produce coherent stories using statistical associations of text, this does not represent the critical thinking skills required to interpret and make meaning of a selected pool of scientific literature (Benzon, 2023). The scientist enacts meaning into the text by situating it in the emerging institutional context of meaning (Gooding, 2012) and tracing the genealogical evolution within the epistemic community (Kusch & McKenna, 2020). This process of textual interpretation involves reading, abducting, and counterfactual analysis to test and exclude possible interpretations to arrive at appropriate and defensible meanings for the epistemic context (Wenzlhuemer, 2009; Ngwenyama & Lee, 1997). Such an achievement requires the scientist to be socialized in the relevant epistemic community and deeply engaged in the emergent conversation on the subject matter (Toulmin 1985).

Situating the text in the epistemic context, counterfactual analysis, and distinguishing disciplinary and factual questions (Toulmin, 1972) are requirements that lie beyond the capabilities of ChatGPT, Gemini or other AI tools (Chomsky et al., 2023). Typically, transformers such as ChatGPT, Bard, or other GPTs will reproduce text for LRs by treating the pool of literature as data and applying a stochastic analysis to it (Bender et al., 2021). While they are capable and efficient at some functions of human cognition (Pantano & Scarpi, 2022), they lack moral and ethical reasoning (Cichocki

[2] On this see footnote 4.

& Kuleshov, 2021). They have no reflexive capabilities for reasoning with or understanding text (Chomsky et al., 2023). They do not know what they are doing (Searle, 1980), as they are not "knowing" subjects (Haack, 1979). Even though their functions are described as "machine learning," they do not learn in the classical sense (Dreyfus, 2002) and are not socialized members of an epistemic community for whom learning is a path to insight and overcoming self-deception (Habermas, 1984).

This brings us to the core value of transparency (Templier & Paré, 2018), which requires scientists to provide detailed explanations of the warrants (Figure 1) used in developing LRs. This enables the epistemic community "to scrutinize and criticize the rational merits of the arguments" (Toulmin, 1972). Reporting our methods and justifying research outputs distinguishes scientific contributions from the layperson's opinion (Rowe in Dwivedi et al., 2023). We emphasize different values, including transparency, when engaged in literature reviews or theory work (Paré et al., 2016). Typically, there is tension between: (1) interestingness, problematization, and contesting knowledge (Saalovara, 2019) and (2) systematicity and the transparency of decision choices to enable auditability and possibly replicability. Some authors may place more value on the first (Boell & Cecez-Kecmanovic, 2014; Leidner, 2018; Rowe & Markus, 2023), others on the second (Paré et al., 2016; Templier and Paré, 2018). But the consensus is that when conducting research and LRs, a minimum amount of systematicity and transparency is essential for trustworthiness (Paré et al., 2016).

4 Sociotechnical Requirements for AI Tool Support of LR Activities

Given our critique of these AI tools, what then are the sociotechnical requirements for using them effectively when developing LRs? In this respect, Kane et al. (2021), following the critical theorist Paulo Freire, proposed a set of principles for designing AI agents to support human emancipation and help overcome ethical problems, biases, and the usurpation of human agency by AI platforms.[3] Kane et al.'s principles inform the sociotechnical design of AI systems for human collaboration and offer the following recommendations for researchers:

- Avoid limiting future possibilities resulting from the backward nature of data collection.
- Avoid obscure parameters and model weights that make ML systems difficult to understand.
- Facilitate iterative dialogue and feedback to strike a *balance between freedom and authority in feedback* mechanisms for both the ML agent and the human user (Kane et al., 2021).

These principles are particularly useful for assessing the complementarity of AI tools with epistemic community values and the potential for researcher control of the tools (Sartori & Theodorou, 2022). Our AI tool choices will have highly significant implications for epistemic values (McMullin, 1983). As has been argued elsewhere, AI can also bring about various and significant social biases (O'Neil, 2017). Whatever type of AI we use, we make implicit or explicit value choices as researchers. In opting for rule-based AI, we gain control, logical consistency, explainability, and transparency. In choosing machine learning or generative AI, we give up control, transparency, logical consistency, and explainability. Knowledge of the AI tool's capabilities and expertise in using it will determine the level of control the researcher has over the AI tool and LR core activities. When planning LR projects, researchers should consider AI tools that afford them the capabilities to:

- be critical and interrogate the validity of the data (research papers and other sources) collected,
- use ML agents that codify and decodify to overcome the fundamental opacity of model weights, and
- maintain continual awareness of the amount of constraint the ML exercises over researchers, and propose adjustments in response to changing conditions (Kane et al., 2021)

5 AI for Literature Reviews: Considering Three Core Activities

Literature reviews play a significant role in scientific discourse, summarizing what we know, what we still don't know, and what we should know. They also (1) identify significant gaps or problems, (2) indicate potential future directions, and (3) sometimes theorize based on gap analysis. In our view, the LR development process comprises three core categories of activities: (1) searching, screening, and quality assessment; (2) analysis, interpretation, and problematization; and (3) writing up: explication, argumentation, and justification. In the three analysis tables below, we integrate the LR activities of Sturm and Sunyaev (2018), and Wagner et

[3] AI machine learning tools usurp human agency and are oriented to the objective world and thus have an instrumental action bias. Although they can act socially, they are not "knowing subjects" and can never claim social responsibility (an aspect of human agency).

al. (2022) and add gap analysis, problematization, argument mapping, and evidence construction. We use this framework to present our analysis and reflections on the potential of AI tools to support the different LR activities. While AI tools promise to make the literature development process more efficient (Wagner et al., 2022), they are not equally applicable to all LR development activities. Like Wagner et al. (2022), we found that AI tools offer great potential for search, screening, and data analysis but limited potential for interpretation, problem formulation, quality assessment, and data extraction (Rowe et al., 2023). Our sociotechnical analysis of ML/DL and expert systems[4] support for LR activities is indicated in the table: (+) improvement, (-) inferior compared to the best manual methods, and (?) effect is low or uncertain. We also suggest specialized training to leverage these capabilities and mitigate AI tool deficits.

5.1 Literature Search, Screening, and Quality Assessment

Problem domain specification and literature selection criteria are essential not only for quality assessment but also for disciplinary relevance and transparency. Merely providing a question and keywords to a GPT and unsupervised access to a digital archive is not good LR practice. GPTs are notorious for producing narrative text with indifference to veracity, contextual relevance, and ethical or consequential implications (Chomsky et al., 2023). However, they can be valuable tools for "rapid reviews where time is a constraint and the trade-off of high precision for lower recall is acceptable." (Wang et al., 2023, p. 1). For precision, researchers must provide strict supervision and expert prompt engineering. AI tools that enforce strict search and selection criteria can enable transparency and quality and save valuable time. However, the AI tools identified in Table 1 rarely have all these qualities. Gusenbauer and Haddaway (2020) note that ML search and retrieval algorithms can be biased (towards recent searches and publication dates) and lack transparency, consequently undermining comprehensiveness. Compared to ML tools, rule-based expert systems are highly efficient and superior for the automated search and cleaning of bibliographic data and can help avoid double counting (Walsh et al., 2022). However, bibliometric expert systems tools such as ARTIREV remain limited by the quality of the databases they search. Therefore, when using ML tools, specialized training and high-quality databases are necessary sociotechnical requirements. For instance, researcher training in screening methods and the multisourcing of databases can improve ML tool effectiveness in identifying the relevant literature. When the criteria used by the algorithm are unclear or not transparent, researchers will have difficulty satisfying traditional LR justification requirements (Wagner et al., 2022). AI platforms that allow the specification of rules for database searches, cluster analysis, and text mining can help researchers demonstrate systematicity and transparency and defend their quality standards.

5.2 Analysis, Interpretation, and Problematization

The core activities of analysis, interpretation, and problematization are all related. Analysis is the basis of interpretation, problematization, and finding gaps. Interpretation requires researchers to situate the text in the meaning context and then to infer, test, and eliminate plausible meanings before settling on the most defensible one (Ngwenyama & Lee, 1997). Problematization, on the other hand, requires researchers to unearth the underlying assumptions of specific theoretical arguments and subject them to dialectic critique in order to open up new ways of thinking (Sandberg & Alvesson, 2011). In both cases, abductive and counterfactual thinking is essential to such analysis (Wenzlhuemer, 2009). The ML/DL proxies for learning and meaning-making, which are pattern recognition and classification, do not entail capabilities of explication and justification, the standard for testing interpretation (Habermas, 2003; Toulmin, 1958). Interpretation, problematization, and gap analysis require intellectual work by researchers who are socialized into the epistemic community and deeply engaged in the emerging scientific discourse on the subject (Alvesson & Sandberg 2011). While some data analysis tasks are supportable by AI tools, such as searching for connections among concepts, most are not. Expert systems can help researchers identify and classify concepts after their characteristics have been fully described, whereas ML tools require training on thousands of examples to correctly identify and classify concepts (Kadhim, 2019; Janani & Vijayarani, 2021). Expert systems based on bibliometric techniques perform very efficiently and effectively to cluster documents and visualize them (Nakagawa et al., 2019; Walsh et al., 2022). Rowe et al. (2023) also show that expert systems that support bibliometric techniques can be iteratively combined with interpretive methods (Walsh & Rowe, 2023) to help researchers search, interpret data, or draw conclusions for various types of literature review.

[4] Regarding this third point, we acknowledge that AI uses some deductive reasoning that applies rules to fact. With this in mind, we can distinguish two distinct broad categories of AI systems (Ram in Agerfalk et al., 2022). The first, expert systems, applies rules to new situations based on the knowledge base of known facts and known rules (typically supplied by human experts). The second are data driven, deriving conclusions from big data and learning from new data. Their inductive automatic "learning" capability justifies the machine learning label, even though there are different kinds of learning (supervised or unsupervised).

Table 1. Sociotechnical Analysis of AI Support for LR Search, Screening, and Quality Assessment

LR activities	Type of AI/ML tools	AI/ML tool capabilities	Sociotechnical requirements
Search	• Machine learning (ML) based search tools, for example Convidence, LitSonar (Sturm & Sunyaev, 2018), Google Scholar, DistillerSR, litsearchr • ChatGPT (Wang et al., 2023) • Expert systems (ES) based tools such as ARTIREV (Walsh et al., 2022)	• (+) precision (as few irrelevant references as possible) (Sturm & Sunyaev, 2018 Wang et al., 2023)) • (+) recall (as comprehensive as possible) (Sturm & Sunyaev, 2018) • (-) quality assurance of inputs and outputs • (?) (1) some AI/ ML tools are subject to bias and lack transparency in their search and retrieval algorithms that can undermine the comprehensiveness of literature search (Gusenbauer & Haddaway, 2020)	• Specialized training in the specific AI/ML tools, methods, and procedures for: (1) systematic searching and filtering, (2) sourcing from literature repositories, (3) infrastructure reliability and transparency for quality
Screening	• ML screening tools—for example, ADIT, Convidence, DistillerSR, EndNote, litsearchr	• (+) recall (as comprehensive as possible) • (?) same as (1) above	• Specialized training in methods and procedures for screening: (1) multisources, (2) gray literature when relevant (Larsen et al., 2019)
Quality assessment	• Traditional tools like RevMan or SPSS available for assessing selection, attrition, and reporting biases (Wagner et al., 2022)	• (?) hard to improve by ML because even if reporting may facilitate appraisal, quality depends on human judgment (Wagner et al., 2022)	• Reporting biases for meta-analyses (Templier & Paré, 2018)

Note: (+) improvement, (-) inferior compared to the best manual methods, and (?) effect is low or uncertain.

Most qualitative data analysis platforms embed AI support for automatic coding concepts and relationships and text mining for extracting quotations. For example, QDAMiner's query-by-example facility can be trained to discover and retrieve text segments "similar in meaning" to examples researchers provide. Supervised deductive coding can also reduce the burden of exhaustive manual coding.[5] However, manual literature analysis remains the gold standard (Antons et al., 2023). While AI algorithms can code and classify concepts and relationships, they cannot interpret their meaning. For example, Gemini marks text segments it uses from the data, making them visible to the researcher. However, researchers will need to check and validate automatic codes for meaning and legitimacy within the epistemic community. These platforms also offer capabilities for identifying complex relationships and test hypotheses (see Table 2). For example, HyperRESEARCH has a goal-seeking algorithm that enables researchers to specify if-then rules for both deductive coding and hypothesis testing (Gibbs, 2018). But researchers need expertise in Boolean logic or truth tables to design effective queries. Another useful capability is topic modeling (Asmussen & Møller, 2019; Kobayashi et al., 2018; Leyersdorff et al., 2017), which can assist in probing under-researched areas. The current capabilities of AI tools for problematizing existing theoretical positions (Alvesson & Sandberg, 2011) are not convincing to domain experts (Schwitzgebel et al., 2023). Furthermore, the claims of unsupervised AI tools "generating" new theoretical insights are often nothing more than overlooked relationships in large pools of literature. Identifying relationships is not theorizing; theorizing includes assessing whether or not they are relevant to and can advance the discourse within the epistemic community. Another concern of theorizing relates to understanding ontological drifts in researchers' use of theoretical concepts (Thompson, 2011). While AI concept drift algorithms can detect shifts in user interests (Auer, 2023), they cannot detect drifts in the usage and meaning of theoretical concepts.

[5] The supervised coding approach entailed iteratively taking random samples (without replacement) from the literature pool and manually coding each new sample, then instructing Atlas/Ti to automatically code the rest of the literature in the pool based on the results of the manual coding. Repeating the process until saturation (no new codes emerge); see Ngwenyama and Nielsen (2003).

Table 2. Sociotechnical Analysis of AI Support for LR Analysis, Interpretation, and Problematization

LR activities	Type of AI/ML tools	AI/ML tool capabilities	Sociotechnical requirements
Data analysis	• Machine learning (ML) and expert systems (ES)-based data analysis tools. For example, Atlas/Ti, HyperRESEARCH, NVivo QDAMiner for supervised and unsupervised coding and analysis of large pools of literature	• (+) can improve the quality of data analysis • (+) can improve the systematic coding of large pools of literature • (+) can improve the exhaustive coding of large pools of literature • (+) can improve search and retrieval of coded text segments (as comprehensive as possible)	• Specialized training required in the specific AI/ML tool for: (1) setting up and training the ML model (Sturm et al., 2021) (2) specifying and executing supervised coding procedures for systematic and exhaustive coding (Ngwenyama & Nielsen, 2003)
Interpretation	• ML and ES-based data analysis tools: (1) supervised discovery of relationships among codes, (2) supervised hypothesis specification and testing • ML tools for text classification and topic modeling (Leyersdorff et al., 2017; Kobayashi et al., 2018).	• (-) cannot interpret text or identify socially situated meanings. Interpretation is not a disembodied rule-driven activity (Searle, 1980; Gadamer, 1977). • (+) can improve systematic and exhaustive search for theoretical conjectured relationships • (+) can improve systematic testing of researcher-specified hypotheses • (+) supervised and unsupervised search can improve topic modeling	• Researcher knowledge of the epistemic meaning context • Specialized training required in the specific AI/ML tool for: (1) Boolean logic and truth tables (2) query methods and procedures for supervised relationship discovery (3) hypothesis specification and testing (4) text classification and topic modeling.
Gap Analysis (What is still needed to know?)	• Topic modeling (unsupervised ML) such as Stanford Topic Modeling Toolbox (TMT) and MALLET, CiteSpace, CleanPoP, HisCite, Scopus	• (-) cannot identify gaps in disciplinary knowledge. Researcher expertise is essential for distinguishing what is legitimate knowledge (Toulmin, 1972; Habermas, 2003). • (?) gap analysis is dependent on researcher experience and intuition. • (+) supervised and unsupervised topic modeling can improve researchers' capabilities to assess potential gaps and determine what is missing and relevant to the epistemic community	• Researcher embeddedness in the epistemic community • Specialized training required in the specific AI/ML tool is essential: (1) methods and procedures for genealogical analysis (2) supervised and unsupervised topic modeling for probing and identifying new emerging problematics
Problematization	• Generative AI tool text generation such as ChatGPT, Bard	• (-) generative AI have difficulty problematizing (Buckingham, 2023; Schwitzgebel et al., 2023) • (?) ability to both critique current knowledge and imagine what could be promising • (?) problematization is dependent on researcher experience and intuition (Sandberg & Alvesson, 2011) • (?) theoretical concept drift may not be detectable by AI tools, researcher expertise is essential for identifying and understanding their implications (Thompson, 2011)	• Researcher embeddedness in the epistemic community • Specialized training required in the specific AI/ML tool is essential: (1) methods and procedures for genealogical analysis; (2) prompt engineering for specifying high-quality GPT outputs; (3) supervised learning methods for specializing GPTs on disciplinary literature

Note: (+) improvement, (-) inferior compared to the best manual methods, and (?) effect is low or uncertain.

5.3 Writing Up: Explication, Argumentation and Justification

As noted earlier, each type of LR has a unique set of meta-requirements (Templier & Paré, 2018), and argumentation strategies (Rowe, 2014). But each must satisfy comprehensibility criteria and justify their core claims to gain assent from the epistemic community. Some researchers argue that GPTs will soon automate scientific writing (Burger et al., 2023; Huang & Tan, 2023), but experiences so far suggest that, except for writing abstracts and literature summaries when the relevant literature has already been identified, the time saved does not outweigh the risks (Peres, et al., 2023; Zheng & Zhan, 2023). GPTs can generate "summaries" of literature but these are limited to descriptions. In order to utilize GPTs more extensively in writing up LRs, researchers would need to develop competencies in prompt engineering (Giray, 2023; Lo, 2023). Moreover, they would need to fact-check the text, as GPTs are not known for veracity (Day, 2023). While some argue that GPTs can be useful for improving the sentence structure and clarity of non-native language writers (Burger et al., 2023), highly experienced researchers developing a problematization LR would be better off doing their own writing than collaborating with a GPT. The capabilities of current GPTs are inadequate for the nuanced argumentation necessary for problematization, critically discussing limitations of a theory, paradigmatic assumptions, or explicating and justifying a new theory (Wittmann, 2023) (Table 3). Additionally, GPTs are not good at recognizing complex scientific concepts or technical terminology because they are generally not pretrained on data from the specific discipline (Buckingham, 2023). Work on developing AI algorithms for mapping and analyzing complex argumentation has progressed (Reed & Rowe, 2004; Hoffmann, 2015; Zhang, et al., 2023). But argument mapping algorithms are not yet embedded in GPTs; consequently, there is a deficit in automating the argumentation practices of epistemic communities (Hahn, & Tešić, 2023).

Confronted with increasingly large pools of literature, AI tools will be used. Our analysis suggests that ML-based techniques used with prudence for descriptive reviews can offer efficiencies when there are shared preunderstandings of the descriptive categories. Unfortunately, with increasing publishing demands, the time gained can be easily allocated to producing more low-quality publications. Researchers might also be tempted to use AI tools to reduce the burden of interdisciplinary LR development. However, this is risky, as terms (e.g., digital transformation) can have different meanings across the relevant disciplines (Markus & Rowe, 2023). Our analysis also suggests that when competent researchers use rule-based systems, time can be saved and quality attained. By competence, we mean the capability to interpret whether the outputs are relevant to the community for descriptive, explanatory, and scoping reviews (Rowe et al, 2023). Moreover, expert systems that offer capabilities to specify all the relevant rules for testing alternative explanations can be extremely helpful for theory development LRs. However, AI tools are of limited help in critical reviews that require human judgment.

Table 3. Sociotechnical Analysis of AI Support for LR Write-Up: Explication, Argumentation and Justification

Literature review activities	Type of AI/ML tools	AI/ML tool capabilities	Sociotechnical requirements
Argument mapping and analysis	• Machine learning (ML) and expert systems (ES)-based tools for argument analysis such as Araucaria, Argunet, VISAR, Rationale, ARTIREV, GPTs	• (+) useful for designing complex argumentation • (+) visualizing, summarizing and analyzing written arguments • (+) analyzing strength of claims, warrants, and backing • (-) prompt engineering interfaces for GPTs • (?) argument structures are dependent on the epistemic community • (?) GPTs are notorious for producing nonsensical arguments and fabricating "evidence" (Zhang et al. 2023; Buckingham, 2023)	• Researcher competence in epistemic community argumentation practices • Specialized training in: (1) Toulmin argumentation analysis and mapping (2) argument mining from literature (3) prompt engineering for specifying argument structure (4) prompt engineering for specifying high-quality GPT outputs
Evidence construction	• ES and supervised ML tools for text mining—for example, EndNote litsearchr	• (?) text mining for relevant quotes may help • (?) researcher expertise is required to situate quotes in the meaning context of the argument in order to avoid misquoting (Schmiedel et al. 2019).	• Researcher knowledge of the disciplinary subject and embeddedness in the epistemic community matter • Specialized training in supervised text mining and using query facilities of ES/ML tools

Note: (+) improvement, (-) inferior compared to the best manual methods, and (?) effect is low or uncertain.

6 Concluding Comments

In this opinion paper, we critically reflect on the idea that AI tools can offer support for literature search, screening, data analysis, and interpretation (Wagner et al., 2022; Rowe et al., 2023). Deeper reflection grounded on principles of scientific discourse (Habermas, 1984. 2003; Toulmin, 1958) language processing and text interpretation theory (Searle, 1980; Gadamer, 1977), confirm that, in general, AI tools can greatly help search and screen data (Van Dinter et al., 2021). However, this paper also highlights the limited potential of ML and generative AI for problem formulation, quality assessment, data extraction, and interpretation. Possibilities for collaborating with them are thus more limited than first thought and their use may be detrimental to the defensibility of scientific discourse.

While most editors currently only envisage the use of “generative AI and AI assisted technologies” as a writing assistant (L. Hassink, personal communication, March 2023), the risks often outweigh the benefits for researchers. Notably, according to Elsevier policy, authors should “only use these technologies to improve readability and language, not to replace key researcher tasks such as interpreting data or drawing scientific conclusions” (L. Hassink, personal communication, March 2023). Such editing policy has clearly been triggered by the advent of ChatGPT and generative AI that are mostly stochastic text production techniques with significant limitations (Bender et al., 2021; Rowe in Dwivedi et al., 2023). Such policy, however, may introduce the risk that papers that use what we can consider good AI for bad reasons may be rejected, while the risk of accepting papers that use problematic AI (with respect to epistemic values) for dubious reasons is not really moderated. As argued in the first part of this opinion paper, the replacement scenario is possible but not likely if we, along with publishers, fight against it.[6] Our paper fully supports the idea that “interpreting data or drawing scientific conclusions” cannot be automated because machines cannot perform these tasks correctly. Nevertheless, AI tools can assist researchers in specific analytical tasks (topic modeling, probing, clusters visualizing) necessary for interpreting data.

With the buzz surrounding machine learning AI—both predictive and generative—the world of occupations is becoming flat. By this, we mean that we live under the illusion that knowledge is immediately and easily accessible and that there are fewer and fewer differences between experts and non-experts. The launch of Gemini with accompanying advertising glitz intended to separate it from competing products, serve to deepen the public’s false consciousness about these digital technologies (Ngwenyama et al. 2023), and increase the risks that novices will accept these tools as legitimate producers of scientific knowledge. Beyond offering analysis of challenges and opportunities at the level of activities, this paper should make clear that conducting end to end literature reviews with ML techniques at each stage would be disastrous in terms of quality outputs even for the best literature review experts. In this context giving up our current epistemic values for efficiency looks like a Faustian bargain.

While it is true that we can adapt in various ways, there will be some irreversibility, and the future of education and researchers’ work and related quality outputs will depend on the aggregation of our own individual choices. Learning how to use AI *cautiously* in the common sense will not be enough to better the world. We have argued elsewhere that the cautious use of AI requires regularly *nurturing our traditional competencies* to defend ourselves and not regress, should we discover that AI does not deliver the expected results (Rowe in Dwivedi et al., 2023; Lebovitz et al, 2021). According to their preferences, experts in literature reviews may become super experts with such caution. They might avoid the risk of dependence by also cultivating their traditional expertise. But we cannot recommend AI tools for novices in literature reviews without well-designed frameworks for use. Using AI cautiously and more broadly requires reading and defending ourselves from hidden political agendas (**2**), and acting wisely (**3** and **4**). To recap, it seems important to:

1. distinguish the phenomenon (e.g., which type of AI is used for which task or activity) from the AI label and superficial categorization,
2. identify conditions for collaboration using critique (e.g., critical social theory) and how collaborations may be hampered by institutional (e.g., capitalist) logic propelling the phenomenon (Ngwenyama et al., 2023),
3. learn indirectly for a given envisioned task whether and how to use machine learning solutions to avoid disclosing more data while being able to assess whether there is more to lose than to gain (Rowe in Dwivedi et al., 2023),[7] and

[6] It is beyond the scope of this paper to review the diverse journal policies that are emerging with the advent of generative AI. Most journal policies we consulted seem to agree that generative AI use be limited to stylizing language (Shmueli et al., 2023). See the contributions from Papagiannidis et al. in Dwivedi et al. (2023). Some also suggest that when AI is used extensively, it should be listed as a co-author (Polonsky & Rotman, 2023).

[7] There is a wealth of publications, including in IS, that report examples of use that we can reflect upon regarding the use of tools and harmful unintended consequences—for example Susarla et al. (2023).

4. keep nurturing traditional competencies to maintain our ability to work as usual and revert back to it when innovative solutions remain or become too problematic as regards epistemic values, as we have argued in this paper. Researchers can adopt ML or generative AI for time efficiency gains if, and only if, their use preserves or enhances researchers' ability to critically analyze and interpret literature (Anis & French, 2023).

What will happen is not written yet. Our future will also depend a lot on how academic institutions (e.g., review processes) and publishers adapt to ML and generative AI. In domains where literature reviews guide public safety decisions (e.g., health), rigorous procedures are even more needed and new methods cannot be adopted without their adaptation to sound epistemic values. Finally, scientists may realize that non-epistemic values, such as the need to preserve a livable climate for current and existing generations, should also be taken into consideration to answer our question. We also conclude that very rational scientific behavior should do just that!

References

Agarwal, R. (2012). Editorial Notes. *Information Systems Research*, 23(4), 1087-1092.

Ågerfalk, P. J., Conboy, K., Crowston, K., Eriksson Lundström, J. S., Jarvenpaa, S., Ram, S., & Mikalef, P. (2022). Artificial Intelligence in Information Systems: State of the Art and Research Roadmap. *Communications of the Association for Information Systems*, *50*, 420-438.

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review*, *36*(2), 247-271.

Anders, G. (1956). *Die Antiquiertheit des Menschen 1: Über die Seele im Zeitalter der zweiten industriellen Revolution*. Munich: C. H. Beck.

Anis, S., & French, J. A. (2023). Efficient, explicatory, and equitable: Why qualitative researchers should embrace AI, but cautiously. *Business & Society*, *62*(6), 1139-1144.

Antons, D., Breidbach, C.F., Joshi, A.M., & Salge, T.O. (2023). Computational literature reviews: method, algorithms, and roadmap. *Organizational Research Methods*, *26*(1), 170-138.

Asmussen, C.B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, *6*(1), 1-18.

Atlasti. (2023). *The beginning of the world's first AI solution for qualitative insights.* https://atlasti.com/ai-coding-powered-by-openai

Auer, S., Barone, D. A., Bartz, C., Cortes, E. G., Jaradeh, M. Y., Karras, O., ... & Tsalapati, E. (2023). The SciQA Scientific question answering benchmark for scholarly knowledge. *Scientific Reports*, *13*(1), Article 7240.

Bender, E.M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610-623).

Benzon, William L., (2023) ChatGPT *intimates a tantalizing future; its core LLM is organized on multiple levels; and it has broken the idea of thinking* (Version 3, February 6, 2023). Available at https://ssrn.com/abstract=4336442.

Boell, S.K., & Cecez-Kecmanovic, D. (2014). A hermeneutic approach for conducting literature reviews and literature searches. *Communications of the Association for Information Systems*, *34*(12), 257-286.

Bornmann, L. & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215-2222.

Buckingham, S. (2023, May 29). *Could generative AI augment reflection, deliberation and argumentation?* Presented at the Critical Deliberative Democracy Technology Workshop, 11th International Conference on Communities & Technologies, Lahti, Finland.

Burger, B., Kanbach, D. K., Kraus, S., Breier, M., & Corvello, V. (2023). On the use of AI-based tools like ChatGPT to support management research. *European Journal of Innovation Management*, *26*(7), 233-241.

Chomsky, N., Roberts, I., & Watumull, J. (2023). The false promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html.

Cichocki, A., & Kuleshov, A.P. (2021). Future trends for human-AI collaboration: A comprehensive taxonomy of AI/AGI using multiple intelligences and learning styles. *Computational Intelligence and Neuroscience*, 2021, Article 8893795.

Day, T. (2023). A preliminary investigation of fake peer-reviewed citations and references generated by ChatGPT. *The Professional Geographer*. Advance online publication. https://doi.org/10.1080/00330124.2023.219073.

Dreyfus, H. L. (2002). Intelligence without representation-Merleau-Ponty's critique of mental representation: The relevance of phenomenology to scientific explanation. *Phenomenology and the Cognitive Sciences*, *1*(4), 367-383.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (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*, Article 102642.

Floridi, L. (2023). AI as agency without intelligence: on ChatGPT, large language models, and other generative models. *Philosophy & Technology*, *36*(1), Article 15.

Gadamer, H.-G. (1977). *Philosophical hermeneutics* (ed. D. E. Linge). University of California Press.

Gibbs, G. (2018). HyperRESEARCH. In B. B. Frey (Ed.). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE.

Giray, L. (2023). Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering, 51*, 2629-2633.

Gooding, D. C. (2012). *Experiment and the making of meaning: Human agency in scientific observation and experiment*. Springer.

Gusenbauer, M., & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, *11*(2), 181-217.

Haack, S. (1979). Epistemology with a knowing subject. *The Review of Metaphysics*, *33*(2), 309-335.

Habermas, J. (2003). *Truth and justification*. The MIT Press.

Habermas, J. (1984). *The theory of communicative action*. Beacon Press.

Hahn, U., & Tešić, M. (2023). Argument and explanation. *Philosophical Transactions of the Royal Society A*, *381*(2251), Article 20220043.

Hoffmann, M. H. (2015). Changing philosophy through technology: Complexity and computer-supported collaborative argument mapping. *Philosophy & Technology*, *28*, 167-188.

Huang, J., & Tan, M. (2023). The role of ChatGPT in scientific communication: Writing better scientific review articles. *American Journal of Cancer Research*, *13*(4), Article 1148.

Iivari, J. (2023). Inductive empiricism, theory specialization and scientific idealization in IS theory building. *Communications of the Association for Information Systems, 52*, 910-914.

Janani, R., & Vijayarani, S. (2021). Automatic text classification using machine learning and optimization algorithms. *Soft Computing, 25*, 1129-1145.

Kadhim, A. I. (2019). Survey on supervised machine learning techniques for automatic text classification. *Artificial Intelligence Review*, *52*(1), 273-292.

Kane, G. C., Young, A.G., Majchrzak, A., & Ransbotham, S. (2021). Avoiding an oppressive future of machine learning: A design theory for emancipatory assistants. *MIS Quarterly, 45*(1), 371-396.

Killoran, N., Lee, L. J., Delong, A., Duvenaud, D., & Frey, B. J. (2017). Generating and designing DNA with deep generative models. Available at https://arxiv.org/abs/1712.06148.

Kobayashi, V.B., Mol, S.T., Berkers, H.A., Kismihok, G., & DenHartog, D.N. (2017) Text mining in organizational research. *Organizational Research Methods*, *21*(3), 733-765.

Kusch, M., & McKenna, R. (2020). The genealogical method in epistemology. *Synthese*, *197*, 1057-1076.

Larsen, K., Hovorka, D., Dennis, A.R., & West, J.D. (2019). Understanding the elephant: The discourse approach to boundary identification and corpus construction for theory review articles. *Journal of the Association for Information Systems, 20*(7), 887-928.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? The dangers of training and evaluating AI tools based on experts' know-what. *MIS Quarterly, 45*(3),1501-1525.

Leidner, D. (2018). Review and theory symbiosis: An introspective retrospective. *Journal of the Association for Information Systems,* 19(6), 552-567.

Leydesdorff, L., & Nerghes, A. (2017). Co-word maps and topic modeling: a comparison using small and medium-sized corpora (n < 1,000). *Journal of the Association for Information Sciences and Technol*ogy*, 68,* 1024-1035

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), Article 102720.

Loebbecke, C., El Sawy, O. A., Kankanhalli, A., Markus, M., Te'eni, D., Wrobel, S., Rydén, P., & Obeng-Antwi, A. (2020). Artificial intelligence meets IS researchers: Can it replace us? *Communications of the Association for Information Systems, 47*, 273-283.

Markus, M. L., & Rowe, F. (2023). Guest Editorial: The Digital Transformation Conundrum: Labels, Definitions, Phenomena, and Theories. *Journal of the Association for Information Systems,* 24(2), 328-335.

Mayer-Schönberger, V., & Cuckier, K. (2013). *Big data: A revolution that will transform how we* *live, work, and think*. Houghton Mifflin Harcourt.

McMullin, E. (1983). Values in science. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Volume 2: Symposia and invited papers, pp. 3-28).

Nakagawa, S., Samarasinghe, G., Haddaway, N. R., Westgate, M. J., O'Dea, R. E., Noble, D. W. A., & Lagisz, M. (2019). Research weaving: Visualizing the future of research synthesis. *Trends in Ecology & Evolution*, *34*(3), 224-238.

Ngwenyama, O., Rowe, F., Klein, S., & Henrikssen, H. (2023) The open prison of the big data revolution: False consciousness, Faustian bargains and digital entrapment, *Information Systems Research*. Advance online publication. https://doi.org/10.1287/isre.2020.0588

Ngwenyama, O. (2023) Analysing genres of discourse in IS research: An approach to interrogating validity claims in IS scientific argumentation. In Davison, R. (Ed.), *Handbook of qualitative research methods for information systems: New Perspectives* (p. 30). Edward Elgar.

Ngwenyama, O., & Nielsen, P. A. (2003). Competing values in software process improvement: an assumption analysis of CMM from an organizational culture perspective. *IEEE Transactions on Engineering Management*, *50*(1), 100-112.

Ngwenyama, O. K., & Lee, A. S. (1997). Communication richness in electronic mail: Critical social theory and the contextuality of meaning. *MIS Quarterly*, *21*(2), 145-167.

O'Neil, C., (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Random House.

Pantano, E. & Scarpi, D. (2022). I, robot, you, consumer: measuring artificial intelligence types and their effect on consumer emotions in service. *Journal of Service Research, 25*(4), 583-600.

Paré, G., Tate, M., Johnstone, D., & Kitsiou, S. (2016). Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems*, *25*(6), 493-508.

Peres, R., Schreier, M., Schweidel, D., & Sorescu, A. (2023). On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice. *International Journal of Research in Marketing*, *40*(2), 269-275.

Polonsky, M. J., & Rotman, J. D. (2023). Should Artificial intelligent agents be your co-author? Arguments in favour, informed by ChatGPT. *Australasian Marketing Journal*, *31*(2), 91-96.

Reed, C., & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, *13*(04), 961-979.

Rowe, F. (2014). What literature review is not: Diversity, boundaries and recommendations? *European Journal of Information Systems*, *23*(3), 241-255.

Rowe, F., Kanita, N., & Walsh, I. (2023). The importance of theoretical positioning and the relevance of using bibliometrics for literature reviews, *Journal of Decision Systems*. Advance online publication. https://doi.org/10.1080/12460125.2023.2217646

Rowe, F., & Markus, M. L. (2023.) Envisioning theory and theoretical diversity: the case of digital transformation. *Journal of the Association for Information Systems*, *24*(5), 1459-1478.

Salovaara, A., Upreti, B.K., Nykänen, J. I., & Meriviki, J. (2020). Building on Shaky foundations? Lack of falsification and knowledge contestation in IS theories, methods and practices. *European Journal of Information Systems*, *29*(1), 65-83.

Sandberg, J., & Alvesson, M. (2011). Ways of constructing research questions: gap-spotting or problematization? *Organization*, *18*(1), 23-44.

Sartori, L., & Theodorou, A. (2022). A sociotechnical perspective for the future of AI: Narratives, inequalities, and human control. *Ethics and Information Technology*, *24,* Article 4.

Schmiedel, T., Müller, O., & vom Brocke, J. (2019). Topic modeling as a strategy of inquiry in organizational research: A tutorial with an application example on organizational culture. *Organizational Research Methods*, *22*(4), 941-968.

Schwitzgebel, E., Schwitzgebel, D., & Strasser, A. (2023). Creating a large language model of a philosopher. Available at https://arxiv.org/abs/2302.01339.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and brain sciences, 3*(3), 417-424.

Spangler, S., Wilkins, A. D., Bachman, B. J., Nagarajan, M., Dayaram, T., Haas, P., ... & Lichtarge, O. (2014). Automated hypothesis generation based on mining scientific literature. *Proceedings of the 20th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining (pp. 1877-1886).

Susarla, A., Gopal, R., Thatcher, J. B., & Sarker, S. (2023). The Janus effect of generative AI: Charting the path for responsible conduct of scholarly activities in information systems. *Information Systems Research, 34*(2), 399-408.

Sturm, B. & Sunyaev, A. (2018). Design principles for systematic search systems: A holistic synthesis of a rigorous multi-cycle design science research journey. *Business & Information Systems Engineering, 61*(1), 91-111.

Thompson, M. (2011). Ontological shift or ontological drift? Reality claims, epistemological frameworks, and theory generation in organization studies. *Academy of Management Review*, *36*(4), 754-773.

Templier, M., & Paré, G. (2018). Transparency in literature reviews: An assessment of reporting practices across review types and genres in top IS journals. *European Journal of Information Systems*, *27*(5), 503-550.

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.

Toulmin, S. (1972). Rationality and scientific discovery. *Proceedings of the Biennial Meeting of the Philosophy of Science Association* (pp. 387-406).

Toulmin, S. E. (1985). Pluralism and responsibility in post modern science. *Science, Technology, and Human Values, 10*(1), 28-37.

van Dinter, R., Tekinerdogan, B., & Catal, C. (2021). Automation of systematic literature reviews: A systematic literature review. *Information and Software Technology*, *136*, Article 106589.

Vert, J. P. (2023). How will generative AI disrupt data science in drug discovery? *Nature Biotechnology*, *41*, 750-751.

vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. *Proceedings of the European Conference on Information Systems.*

Wagner, G., Lukyanenko, R., & Paré, G. (2022). Artificial intelligence and the conduct of literature reviews. *Journal of Information Technology*, *37*(2), 209-226.

Walsh, I., & Rowe, F. (2023). BIBGT: combining bibliometrics and grounded theory to conduct a literature review. *European Journal of Information Systems*, *23*(4), 653-674.

Walsh, I., Renaud, A., Jeanneret Medina M., Baudet, C., & Mourmant, G. (2022). ARTIREV: An Integrated Bibliometric Tool to Efficiently Conduct Quality Literature Reviews, *Systèmes d'Information et Management*, *27*(4), 5-50.

Wang, S., Scells, H., Koopman, B., & Zuccon, G. (2023). Can chatGPT write a good boolean query for systematic review literature search? Available at https://arxiv.org/abs/2302.03495.

Wenzlhuemer, R. (2009). Counterfactual thinking as a scientific method. *Historical Social Research / Historische Sozialforschung*, *34*(2), 27-54.

Wittmann, J. (2023). Science fact vs science fiction: A ChatGPT immunological review experiment gone awry. *Immunology Letters*, *256*, 42-47.

Zhang, Z., Gao, J., Dhaliwal, R. S., & Li, T. J. J. (2023). VISAR: A human-AI argumentative writing assistant with visual programming and rapid draft prototyping. Available at https://arxiv.org/abs/2304.07810.

Zheng, H., & Zhan, H. (2023). ChatGPT in scientific writing: a cautionary tale. *The American Journal of Medicine*, *13*(8), P725-726.E6.

About the Authors

Ojelanki Ngwenyama is a professor at the Ted Rogers School of Management, Toronto Metropolitan University and a senior scholar in the Faculty of Commerce, University of Cape Town. Ojelanki is a member of the South African Academy of Science, and a Fellow of the AIS. He has a PhD from Thomas J. Watson College of Engineering, State University of New York, Binghamton, USA; and DPhil from the Faculty of Engineering, University of Pretoria, South Africa. His research focuses on critical social theory and digitalization. https://orcid.org/0000-0001-7124-430X

Frantz Rowe is a professor at Institut d'Economie et de Management de Nantes-IAE, Nantes University. Frantz is a senior member of the Institut Universitaire de France where he holds a fundamental chair on digital entrapment and digital transformation, with related projects (e.g. technical debt, generative AI). Frantz is a Fellow of the AIS. He has a PhD from the University of Paris, an ME from ENTPE, Lyons, France, and an MS from UC Berkeley, USA. https://orcid.org/0000-0001-8520-1570