Debates and Perspectives Paper

JIT

Journal of Information Technology
2022, Vol. 37(2) 209–226

DOI: 10.1177/02683962211048201
Journals.sagepub.com/jinf

Artificial intelligence and the conduct of literature reviews

Gerit Wagner, Roman Lukyanenko and Guy Paré

Abstract

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

Keywords

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

Introduction

The potential of artificial intelligence (AI) to augment and partially automate research has sparked vivid debates in many scientific disciplines, including the health sciences (Adams et al., 2013; Tsafnat et al., 2014), biology (King et al., 2009), and management (Johnson et al., 2019). In particular, the concept of automated science is raising intriguing questions related to the future of research in disciplines that require "high-level abstract thinking, intricate knowledge of methodologies and epistemology, and persuasive writing capabilities" (Johnson et al., 2019: 292). These debates resonate with scholars in Information Systems (IS), who ponder which role AI and automation can play in theory development (Tremblay et al., 2018) and in combining data-driven and theory-driven research (Maass et al., 2018). With this commentary, we join the discussion which has been resumed recently by Johnson et al. (2019) in the business disciplines. The authors observe that across this multi-disciplinary discourse, two dominant narratives have emerged. The first narrative adopts a provocative and visionary perspective to present its audience with a choice between accepting or rejecting future research practices in which AI plays a dominant role. The second narrative acknowledges that a gradual adoption of AI-based research tools has already begun and aims at engaging its readers in a constructive debate on how to leverage AI-based tools for the benefit of the research field and its stakeholders. In this paper, our position resonates more with the latter perspective, which is focused on the mid-term instead of the long-term, and well-positioned to advance the discourse with less speculative and more actionable discussions of the specific research processes that are more amenable applications of AI and those processes that rely more on the human ingenuity of researchers.

In this essay, we focus on the use of AI-based tools in the conduct of literature reviews. Advancing knowledge in this area is particularly promising since (1) standalone review projects require substantial efforts over months and years (Larsen et al., 2019), (2) the volume of reviews published in IS journals has been rising steadily (Schryen et al., 2020), and (3) literature reviews involve tasks that fall on a spectrum between the mechanical and the creative . At the same time, the process of reviewing literature is mostly conducted manually with sample sizes threatening to exceed the cognitive limits of human processing capacities. This

Department of Information Technologies, HEC Montréal, Montréal, Québec, Canada

Corresponding author:
Guy Paré, Research Chair in Digital Health, HEC Montréal, 3000, chemin de la Côte-Sainte-Catherin Montréal, Québec H3T 2A7, Canada.
Email: guy.pare@hec.ca

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

In light of these challenges, we focus on the contributions of AI which refers to the capability of performing cognitive tasks and exhibiting intelligent behavior commonly associated with human intelligence (Russell and Norvig, 2016; Taulli and Oni, 2019). Specifically, we are interested in approaches that are commonly referred to as "weak AI" and combine process automation (execution engines) with capabilities like machine learning (ML) or natural language processing (NLP). Machine learning refers to tools, methods, and techniques for learning and improving task performance with experience (Goodfellow et al., 2016; Mitchell, 1997), while NLP refers to computational tools, methods, and techniques for analyzing, interpreting, and increasingly generating natural language (Manning and Schütze, 1999). Although we are particularly interested in tools powered by advanced AI, we do not discard predecessors of AI per se.

AI offers two capabilities that are particularly salient for conducting literature reviews. First, they operate on potentially fuzzy, weakly structured, and unstructured data that are provided in the form of bibliographical meta-data or full-text documents. Techniques of NLP can go beyond purely syntactic processing of text by abstracting and analyzing its semantic meaning, thereby promising to offer valuable support in the searching and screening tasks. For example, papers including the word "review" may be hard to distinguish on a syntactic level, but using semantic techniques, NLP performs much better in dissociating whether "review" refers to a literature review or a customer review. An example applying such techniques to IS research is offered by Sidorova et al. (2008), who illustrate the topics prevalent in top-tier IS journals based on latent Dirichlet allocation (LDA) models. This paper clearly shows the advantages of LDA models, which allow unobserved (latent) topics to emerge from the analysis of bags of words. The application of NLP techniques has further been considered useful for generating semantic topics from samples of papers and thereby allowing researchers to explore the literature from a more abstract perspective (Mortenson and Vidgen, 2016). Second, advanced supervised ML techniques, such as deep learning, can be trained to replicate the decisions of researchers. This relieves researchers of the task of explicating and codifying myriads of rules, and even more significantly, it can automate decisions for which exact rules are hard to specify. The work of Larsen et al. (2019) is exemplary in this regard, developing classifiers that can automatically screen and include papers relevant to research on TAM (Technology Acceptance Model). Considering these capabilities, we expect AI to be most useful in the mechanical tasks of reviews compared to more creative ones. At the same time, an informed discourse and methodological guidelines are necessary to identify the appropriate areas of application and to address the challenges associated with AI, such as model overfitting, biases, black box predictions, and the acceptance by the research community.

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

Artificial intelligence–based support for the literature review process

The literature review process involves both creative and mechanical tasks, which creates exciting opportunities for advanced AI-based tools[1] to reduce prospective authors' efforts for time-consuming and repetitive tasks and to dedicate more time to the creative tasks that require human interpretation, intuition, and expertise (Tsafnat et al., 2014). To familiarize the reader with state-of-the-art knowledge in this area, we consider each step of the review process in turn, outlining current AI-based tools as well as the potential for AI-based tool support. Corresponding opportunities for

further tool development and improvement are outlined in the following agenda. The overview is in line with the steps of the review process that Templier and Paré (2018) have synthesized from the methodological literature. Table 1 provides a brief summary, explaining whether the step is amenable to AI-support and pointing the reader to corresponding tools.

We collected the evidence by surveying previous literature reviews of AI-based tools (e.g., Al-Zubidy et al., 2017; Harrison et al., 2020; Jonnalagadda et al., 2015; Kohl et al., 2018; Marshall and Wallace, 2019; Tsafnat et al., 2014; Van Dinter et al., 2021) and online registries (i.e., www.systematicreviewtools.com). Since AI-based tools are constantly evolving, with some not applicable to IS research and some no longer maintained, we briefly tested those deemed relevant for our main purpose. This overview is by no means comprehensive and aims at illustrating promising examples for IS researchers. In the following paragraphs,

Table 1. AI-based tools for steps of the review process.

Step	AI-based tools	Potential for AI-support
1. Problem formulation	• Programming libraries supporting thematic analyses based on LDA models (example paper: Antons and Breidbach, 2017) • GUI applications and programming libraries supporting scientometric analyses (Swanson and Smalheiser, 1997)	• Moderate potential with AI potentially pointing researchers to promising areas and questions or verifying research gaps
3. Search	• TheoryOn (Li et al., 2020) enables ontology-based searches for constructs and construct relationships in behavioral theories • Litbaskets (Boell and Wang, 2019) supports researchers in setting a manageable scope in terms of journals covered • LitSonar (Sturm and Sunyaev, 2018) offers syntactic translation of search queries in for different databases; it also provides (journal) coverage reports	• Very high potential since the most important search methods consist of steps that are repetitive and time-consuming, that is, amenable to automation
4. Screen	• ASReview (Van de Schoot et al., 2021) offers screening prioritization • ADIT approach of Larsen et al. (2019) for researchers capable of designing and programming ML classifiers	• High potential for semi-automated support in the first screen, which requires many repetitive decisions • Moderate potential for the second screen, which requires considerable expert judgment (especially for borderline cases)
5. Quality assessment[a]	• Statistical software packages (e.g., RevMan) • RobotReviewer (Marshall et al., 2015) for experimental research	• Low-to-moderate potential for semi-automated quality assessment
5. Data extraction	• Software for data extraction and qualitative content analysis (e.g., Nvivo and ATLAS.ti) offers AI-based functionality for qualitative coding, named entity recognition, and sentiment analysis • WebPlotDigitizer and Graph2Data for extracting data from statistical plots	• Moderate potential for reviews requiring a formal data extraction (descriptive reviews, scoping reviews, meta-analyses and qualitative systematic reviews) • High for objective and atomic data items (e.g., sample sizes), low for complex data which has ambiguities and lends itself to different interpretations (e.g., theoretical arguments and main conclusions)
6. Data analysis and interpretation	• Descriptive synthesis: Tools for text-mining (Kobayashi et al., 2017), scientometric techniques and topic models (Nakagawa et al., 2019; Schmiedel et al., 2019), and computational reviews aimed at stimulating conceptual contributions (Antons et al., 2021) • Theory building: Examples of inductive (computationally intensive) theory development (e.g., Berente et al., 2019; Lindberg, 2020; Nelson, 2020) • Theory testing: Tools for meta-analyses (e.g., RevMan and dmetar)	• Very high potential for descriptive syntheses • Moderate potential for (inductive) theory development and theory testing • Low non-existent potential for reviews adopting traditional and interpretive approaches

[a]Applicable to meta-analyses and qualitative systematic reviews.

we adopt a granular perspective, highlighting AI-support for individual tasks that authors can ultimately orchestrate in an overarching data processing and tool chain.[2]

Step 1: Problem formulation

The first step of a literature review requires authors to identify and clarify the research questions and central concepts or theories (Templier and Paré, 2018). In addition, authors can be advised to complete an initial verification of the research gap (Müller-Bloch and Kranz, 2015), which may involve assessing whether the gap has already been addressed, whether the research question allows for a substantial contribution that exceeds previous work, and whether it is indeed important to address the gap (Rivard, 2014).

We expect moderate potential for AI-support in the problem formulation step, in which we focus on the identification and verification of research gaps. Given that the natural sciences have witnessed some exciting advances with regard to detecting new research gaps and promising starting points for hypotheses, it can be hoped that some of this work will eventually be applied and adapted to the social sciences. For instance, there are path-breaking advances with regard to automated generation of hypotheses (and experimental testing in automated laboratories) in biochemistry (King et al., 2009) and machine-learning approaches for scientometric, literature-based discovery in computer science (Thilakaratne et al., 2019). These areas are certainly more predestined for the initial application of AI because they do not necessarily raise complex ethical decisions of research involving human participants, or the fuzziness of behavioral theories and constructs prevalent in IS research (Li et al., 2020). Nevertheless, research in the social sciences may eventually be inspired by these approaches. Overall, they could stimulate researchers in identifying areas in need of review papers or further research more broadly. Still, we expect that human judgment will remain a necessary ingredient for this step in the near future, especially for research generating questions through problematization (Alvesson and Sandberg, 2011). Beyond the discovery of research gaps and areas in need of a review paper, there is some potential for supporting researchers in verifying whether the gaps are still open by identifying identical or similar knowledge contributions and previous review papers. In the discovery and verification, the use of AI is likely to involve uncertainty, requiring researchers to make final decisions regarding the treatment of research gaps.

In sum, current tool support for this initial step is still in the early stages of development with published approaches implementing their code on an individual basis as opposed to drawing on mature GUI-based tools. Researchers with programming skills can find inspiration in previous works exploring and detecting research gaps and opportunities for inter-disciplinary research at the intersection of literatures or research topics. For instance, this can be achieved by applying and advancing scientometric methods (e.g., Evans and Foster, 2011; Swanson and Smalheiser, 1997) and LDA topic models (e.g., Antons and Breidbach, 2017).

Step 2: Literature search

In this step, authors construct the literature sample by applying different search methods, including database searches, table-of-content scans, citation searches, and complementary searches (cf. Templier and Paré, 2018). Depending on the goal of the review, authors can aim at a coverage that is comprehensive, representative, or selective (Cooper, 1988). Corresponding search methods can be assembled in complex search strategies, involving several iterations, collaborative work of research teams, and AI-based tools. Due to the heterogeneity of the information retrieval process, the variety of data sources (e.g., journals, conference proceedings, books, and different forms of grey literature), and the plethora of data quality problems (e.g., incomplete or incorrect meta-data), appropriate data management strategies are vital. Ultimately, they should enable transparent reporting (Paré et al., 2016; Templier and Paré, 2018), as well as repeatability and reproducibility (Cram et al., 2020).

Recent work in IS provides compelling advances, especially with regard to the prevalent database searches. We highlight three tools that have been published recently. First, the TheoryOn search engine (Li et al., 2020) offers a complementary option to traditional databases by allowing researchers to execute ontology-based searches for individual constructs and construct relationships across behavioral theories. Second, Litbaskets (Boell and Wang, 2019) can inform the design of search strategies. This web-based tool allows researchers to assess the potential volume of database search results based on pre-specified keywords and different sets of journals which can be adjusted in a flexible way. Third, LitSonar (Sturm and Sunyaev, 2018) greatly facilitates the execution of database searches by automatically translating search queries for a range of literature databases relevant for IS reviews (e.g., EBSCO, AIS eLibrary, and ProQuest). This tool is particularly promising because it provides a coverage report, potentially identifying possible database embargos (periods in which journals are not indexed), and thereby alleviating insufficiencies of academic databases.

Overall, this step tends to be time-consuming with many mechanical tasks potentially lending themselves to automation (Carver et al., 2013; Johnson et al., 2019). The need for automation and AI-support is particularly salient when considering the rapid growth of research output (Larsen et al., 2019) and the inefficiency of investing valuable time of academic experts to complete repetitive and mechanical tasks.

Step 3: Screening for inclusion

In the screening step, authors work with the search results to dissociate the relevant papers from those that must be excluded from the review. This step is typically divided into a first (more inclusive) screening based on titles and abstracts and a second (more restrictive) screening based on full-texts (Templier and Paré, 2018). In manual screening processes, researchers execute the tedious task of checking hundreds or thousands of papers. In this process, they are likely to experience fatigue, which may interfere with their ability to accurately dissociate cognitively demanding borderline cases. This explains why methodologists have recommended to conduct a first screen in which researchers exclude papers that are clearly irrelevant (based on titles and abstracts) and to deliberately retain challenging papers for a second round of screening. To ensure screening performance in the second round, researchers typically work with smaller samples (after excluding a bulk of papers in the first screen), consult the full-text documents, apply more specific (pre-defined) exclusion criteria, and execute parallel independent assessment with team decisions on the final borderline cases. The most rigorous screening procedures have been suggested for reviews aimed at theory testing (Templier and Paré, 2018), in which erroneous inclusion decisions may have more significant and measurable effects on the conclusions of the review.

AI-based tool support for screening has been evolving over the years (Harrison et al., 2020). While many tools suffer from severe restrictions for screening IS research (e.g., operating primarily on health sciences databases and requiring PubMed IDs), ASReview (Van de Schoot et al., 2021), which has been published recently, offers an option for researchers in IS. This tool is a particularly promising exemplar since it combines inspectability of the code (published under the Apache-2.0 License), extensibility (availability of code and documentation, implementation using popular Python ML-libraries), ongoing validation efforts, and interoperability (offering import and export of common Research Information System Format and Comma-separated values files). Implementing a range of ML classifiers (including Naive Bayes, Support Vector Machines, Logistic Regression, and Random forest classifiers), it learns from initial inclusion decisions and leverages these insights to present researchers with a prioritized list of papers (i.e., the titles and abstracts), proceeding from those most likely to be included to those least likely. This allows researchers to efficiently work through a prioritized list of papers in the first inclusion screen and even rely on automated exclusion by stopping to screen after decisions exceed a certain number of excluded papers in a row (e.g., n = 100). Borderline cases can be retained for a second screen in which decisions are based on full-text documents.

Furthermore, researchers with programming skills can adapt the discourse approach proposed by Larsen et al. (2019). The authors point out that reviews of popular theories may be infeasible, especially when thousands of relevant papers have been published. In such cases, they propose the thought-provoking possibility to consider sampling papers randomly from the set of relevant (called theory-contributing) papers, as identified by machine-learning algorithms. In line with random sampling in empirical studies, Larsen et al. (2019) thereby suggest that random selection may be useful in literature reviews to obtain a view of the literature that is representative for the scientific discourse. As a result, the Automated detection of implicit theory approach is particularly promising for developing reviews that use a theory as the unit of analysis (Theory and Review Articles) and encounter excessive amounts of research following up on the selected theory.

We expect the potential for AI-support to be high for the first screen and moderate for the second screen. The first screen seems particularly amenable to partial automation and AI-support because it is more inclusive and does not require final exclusion decisions for borderline cases. This arguably requires machines to have adequate capabilities of reading and "understanding" abstracts and titles. In contrast, the second screen is dedicated to disentangling the remaining cases, which can be particularly challenging since IS research is not standardized as strictly as other disciplines. In contrast to the health sciences and biology, for instance, the lack of widely used taxonomies for IS constructs, standard vocabulary for keywords (e.g., MeSH terms), and descriptive paper titles makes it difficult to achieve required classification performance in the second screen (cf. O'Mara-Eves et al., 2015). This challenge applies to humans and machines alike. Taken together, the screen and search (considered as an information retrieval task) should primarily be evaluated in terms of the recall, that is, the proportion of papers that are successfully retrieved (i.e., relevant). Authors of literature reviews traditionally target a high recall by executing comprehensive searches and thereby accept very low precision and the corresponding screening burden (Li et al., 2020). One implication of Li et al.'s (2020) work is that AI-supported, ontology-based searches may effectively prevent some of the screening burden through better precision. Overall, the two screening steps are therefore among the most time-consuming activities of the literature review process (Carver et al., 2013). When considering potential AI-support of this step, the reliability of manual screening processes should not be overestimated, even if the screen is conducted by academic experts. In fact, recent evidence in the health sciences suggests a base rate of 10% disagreements between inclusion screens conducted independently (Wang et al., 2020). This indicates that it may even be possible to augment and improve screening activities of researchers by having AI-based tools identify inconsistent and potentially erroneous screening decisions.

Step 4: Quality assessment

The quality assessment involves checking primary empirical studies for methodological flaws and other sources of bias (Higgins and Green, 2008; Kitchenham and Charters, 2007; Templier and Paré, 2018). This step is intended to assess the degree to which the conclusions of reviews aimed at theory testing may be affected by different types of biases (e.g., selection, attrition, and reporting bias). It is recommended to conduct these procedures in a parallel and independent way to ensure high reliability (Templier and Paré, 2018).

We believe the potential for AI-based tools supporting these procedures is low to moderate for two reasons. First, assessing (methodological) quality is a challenging task which requires expert judgment, making it difficult to achieve high inter-coder agreement (Hartling et al., 2009). Second, sample sizes in IS reviews (meta-analyses and qualitative systematic reviews) are not excessively large, that is, manual assessments are still manageable. Following methodological guidelines for quality appraisal and risk of bias assessment, IS researchers conducting meta-analyses and systematic literature reviews can leverage traditional tools like RevMan (cf. Bax et al., 2007) or corresponding packages of statistical software environments like R and SPSS. Further AI-based tools like RobotReviewer (Marshall et al., 2015) can also be applicable for meta-analyses in IS. While focusing on risk of bias assessment of randomized controlled trials in the life sciences, RobotReviewer is an excellent exemplar for explainable AI, allowing researchers to interactively trace ratings in each domain of bias to its origin in the full-text document.

Step 5: Data extraction

Data extraction requires researchers to identify relevant fragments of qualitative and quantitative data and to transfer them to a (semi) structured coding sheet (Templier and Paré, 2018). It is more salient in descriptive reviews, scoping reviews, and reviews aimed at theory testing compared to reviews that are more selective and interpretive such as narrative reviews and theory development reviews.

Current tool support in the IS field reflects the moderate potential of AI in this area, with authors relying on general tools for qualitative data analysis, such as ATLAS.ti and NVivo (which are starting to implement NLP and machine learning algorithms for tasks such as automated qualitative coding, named entity recognition and sentiment analysis), or specialized tools for extracting data from tables or statistical plots, such as WebPlotDigitizer or Graph2Data.

We expect the potential for supporting this step with AI to be moderate. Learning from ongoing data extraction decisions prospective tools could progressively improve, highlight the more promising fragments in a given paper, and facilitate the transfer and organization of data into corresponding repositories. We do not expect full automation of more significant data items in the near future. Even in the health sciences, which have established relatively consistent reporting practices, corresponding tools designed to extract study characteristics like the PICO (population, intervention, context, and outcome) elements are still in the early stages of development (Jonnalagadda et al., 2015).

Step 6: Data analysis and interpretation

The final step of the review process can take various forms, depending on the type of review (Templier and Paré, 2018). Some reviews put more emphasis on elegant narratives which convey insightful and deeply hermeneutic interpretations while others are designed to eliminate any subjectivity which may interfere with the accuracy of aggregated evidence or descriptive overviews.

Depending on the main knowledge building activities (Schryen et al., 2020), IS researchers can use different tools. For descriptive syntheses, there is a range of established tools for text-mining (Kobayashi et al., 2017), as well as tools for analyzing and visualizing topics, theories, and research communities based on scientometric techniques, computational techniques, or LDA models (Balducci and Marinova, 2018; Nakagawa et al., 2019; Thilakaratne et al., 2019), for instance. Further promising tools originally applied to unstructured data in contexts such as technology adoption (Laurell et al., 2019), corporate communication (van Zoonen and van der Meer, 2016), or corporate social responsibility (Tate et al., 2010) could be adapted to the needs of IS review papers. For inductive work, there is an increasing amount of research on computationally intensive techniques that leverage data for theory generation (e.g., Berente et al., 2019; Lindberg, 2020; Nelson, 2020). Finally, for theory testing, there is a range of applications and libraries for meta-analyses, such as RevMan (cf. Bax et al., 2007) or the R package dmetar.

In assessing the potential for future AI-based tools to support data analysis, we need to take into account that this step can take various forms. In pre-theoretical reviews, AI-based tools offer capabilities to generate descriptive insights, for example, based on topic modeling (Kunc et al., 2018; Mortenson and Vidgen, 2016; Schmiedel et al., 2019). As a promising bridge toward conceptual contributions, Antons et al. (2021) advance the notion of computational reviews. Specifically, they suggest to leverage descriptive and scientometric analyses to stimulate researchers in pursuing conceptual goals of explicating, envisioning, relating, and debating. Theory development in IS appears to be most amenable to AI-support when it follows an inductive approach (Berente et al., 2019). We have yet to witness exemplars of AI-driven theory development in a

behavioral research domain which comes close to the ingenuity and creativity displayed in some of the strongest theory and review papers. After all, it is important to remember that new assemblages of constructs and relationships are not a sufficient condition for a strong theoretical contribution. As pointed out by Johnson et al. (2019), it is the "why" associated with the relationships, the underlying theoretical rationale (Whetten, 1989), which is critical and one of the open challenges for AI-based theory development in the near future. In contrast to the creative and unstructured endeavor of theory development, the process of aggregating evidence from primary studies in order to test hypotheses and theories (most notably in a meta-analysis) can largely be supported by AI-based tools using the extant methodological literature as guidance.

A research agenda

In this section, we outline an agenda suggesting how IS researchers can focus and coordinate their efforts in advancing AILRs. Nurturing a vibrant AILR tradition is a task for the entire scholarly community, including design science researchers, behavioral scientists, methodologists, reviewers, and journal editors as well as authors of primary research papers. The AILR-centric agenda for research, design, and action, as displayed in Figure 1, covers three levels: (I) supporting infrastructure, (II) methods and tools, and (III) research practice.

The agenda proceeds from technical questions of how research is stored and made accessible (Level I) to specific questions of how methods and tools can support the process of conducting AILRs (Level II), to overarching community-centric questions of how IS research could facilitate the conduct of AILRs (Level III). For each level, we suggest fruitful opportunities, focusing on research and design (primarily on Levels I and II) as well as actionable advice targeting individuals and the IS community as a whole (primarily on Level III). With this broad agenda, we emphasize that AI and AILRs raise interesting questions on how we conduct and synthesize research.

Level I: Supporting infrastructure

Technical infrastructure can greatly facilitate or constrain AILRs. The diversity of infrastructure needed to support a vibrant AILR tradition within IS points to a wide range of related opportunities for research and design. We cover quality assurance, smart search technologies, and enhanced databases.

Quality assurance of AILR inputs. One of the prominent characteristics of AILRs is scalability. The expanding computational resources allow ever greater numbers of research papers to be extracted and analyzed with some projects reporting tens (Larsen et al., 2019) and others reporting hundreds of thousands (Dang et al., 2009) of papers processed. In such cases, the technical quality of paper documents would vary (e.g., regarding Optical character recognition and the inclusion of additional, non–content-related text), and the scale of work necessitates novel methods and metrics for preprocessing and establishing the quality of inputs into AILRs (Antons et al., 2021). Considering the scale, an important consideration in undertaking this work is the need to establish and, if necessary and possible, improve quality of inputs automatically, with little human intervention. Furthermore, methods and tools should attempt to cover a diverse range of AILR inputs, including traditional research papers, grey literature, and objects of research, such as IT artifacts. These challenges create fertile opportunities for collaboration with researchers working on information

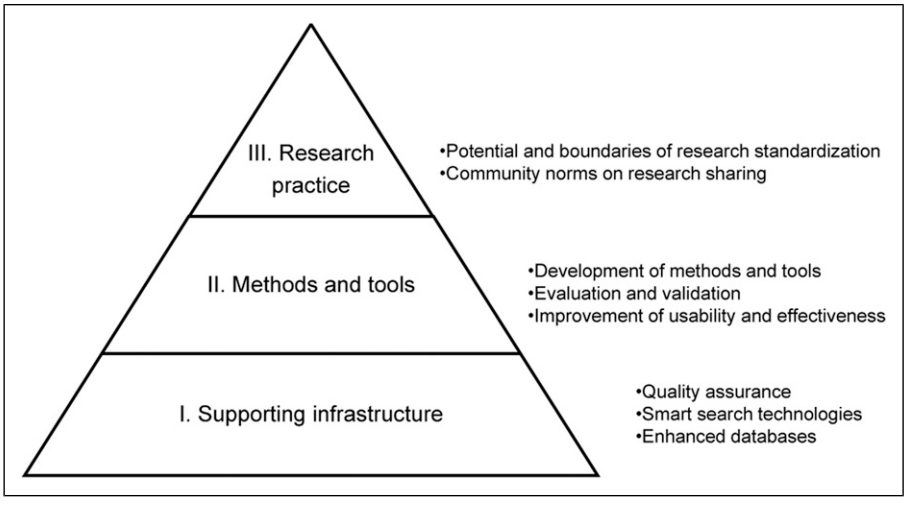

Figure 1. A research agenda for AILR-centric research, design, and action.

quality of big heterogenous data within IS, computer science and beyond, many of whom also utilize AI methods (Batini et al., 2015; Kenett and Shmueli, 2016; Lukyanenko et al., 2019; Wahyudi et al., 2018).

Smart search technologies. To facilitate information retrieval from databases and subsequent analyses, more research is needed on smart search technologies. This involves going beyond a mere word matching and seeking to understand the intent and meaning behind a search query. First, work is needed on better syntactic interpretation of search words, such as research on query parsing and validation (Russell-Rose and Shokraneh, 2019). Indeed, machine learning has become a major driver in NLP improvements. One notable opportunity here is the development of IS-specific NLP query parsing algorithms, as parsing text is partially context-dependent (Eisenstein, 2019).

Second, to go from syntactic parsing to understating the meaning and intent of a search query, AILR tools must possess the ability to infer and reason beyond the information provided. Here, domain-specific ontologies can greatly facilitate deeper semantic interpretation of search queries. Unlike simple tags or labels, ontologies capture nuanced semantics of the domain, including concept definitions and relationships among concepts (e.g., "ERP" is a *kind of* "enterprise IT," "design principles" are *synonymous with* "design guidelines," "UTAUT" is an *extension of* "TAM"). Research opportunities pertain both to IS-specific and foundational ontologies. Work on IS-specific ontologies has already been undertaken in the past (Alter, 2005; Lukyanenko, 2020), and the need to support AILRs could provide a new impetus to this line of work. Indeed, in fast-paced disciplines like IS, for such ontologies to be useful, they need to be current and up to date, necessitating continuous research attention.

We call for further work on ontology-based indexing to improve discoverability of scholarly content (cf. Li et al., 2020). This could involve the use of domain ontologies to assign specific labels to papers and their content (e.g., constructs, construct relationships, theories, themes, and methodologies), potentially alleviating the problems of "concept drift" or buzzwords in IS (O'Mara-Eves et al., 2015; vom Brocke et al., 2015).

Further design science and empirical research is needed on foundational ontologies, which commonly provide the basis for domain ontologies, describing primitive constructs and rules for using them in domain ontologies (March and Allen, 2014; Wand and Weber, 1988). With the development of new ontologies constituting a research opportunity, surveys and evaluations of different ontologies will also be needed, especially for the objective of creating an IS-specific domain ontology to guide AILRs.

Enhanced databases. We envision enhancements of IS databases and complementary repositories which can greatly facilitate AILRs. First, recognizing that there are many areas in which scholarly databases could improve, we call for research advancing coverage reports and improving *interoperability of academic databases*. A prevalent challenge for literature reviews in the social sciences, in general, and IS, in particular, is the lack of databases comprehensively curating research published in the main outlets, including journals and conferences (vom Brocke et al., 2015), accompanied by increasing volatility of database indices and search algorithms (Cram et al., 2020). This requires researchers to search multiple sources and apply multiple search techniques (Papaioannou et al., 2009; Templier and Paré, 2018). The ground-breaking paper of Sturm and Sunyaev (2018) illustrates how journal coverage reports could enable substantially more targeted and efficient literature searches. We further emphasize that limited interoperability (accessibility via APIs) is still a major obstacle breaking the data processing pipeline between the database and local repositories of the research team, introducing manual database queries and duplicate checking as potential sources of errors.

Second, data curation initiatives could benefit from the interplay of supervised ML and crowdsourcing platforms, targeting annotation, quality control, and synthesis. For example, the *Cochrane crowd* is a section of the larger online medical database which actively solicits volunteers "to help categorise and summarise healthcare evidence."[3] Beyond the IS theory wiki,[4] which is curated by online editors in a manner similar to Wikipedia, we are not aware of any such efforts in the IS discipline. Considering the potential value such repositories may bring, we call for more work on crowdsourcing for research, including on how to motivate volunteer researchers to categorize and analyze IS literature and how to evaluate and improve quality of the contributions.

Third, we highlight opportunities for advancing complementary database repositories for models and artefacts. To ensure long-term viability and cross-fertilization of AILRs, there are opportunities to design reusable repositories for NLP and ML models, whitelists and lexicons which can then be shared with the community (Dalgali and Crowston, 2019) and incorporated into AILR routines. This can be especially powerful for capturing recurring patterns in IS literature, such as AI models capable to understanding common dependent variables (e.g., intention to adopt), which could be reused as module components for specific literature reviews (e.g., adoption of electronic health record technologies).

In addition, tools and algorithms are needed to extract and analyze non-textual subject matter of research papers,

especially the IT artifacts. IS researchers have begun to conduct reviews of the features of IT artifacts, known as "design archaeology" (Gleasure, 2014; Chandra Kruse et al. 2019). Corresponding algorithms for automated reviews of digital artifacts can be based on recent advances in image mining, process mining, or computer vision. This is a unique opportunity native to the IS discipline, which stands to enrich AILRs broadly.

Overall, advancing quality assurance and domain-ontologies, as well as information curated in enhanced databases and local repositories will enable prospective authors to execute AILRs more efficiently, to expedite ML training, and to enhance ML models by infusing domain knowledge into NLP algorithms. These advances can facilitate research progress through corresponding tools and methods supporting distinct steps of the literature review process.

Level II: Methods and tools

There are vast opportunities for methodological and tool-centric research on AILRs. We outline promising avenues for research and design in each individual step of the review process and offer complementary recommendations on cross-cutting concerns, covering the need for advancing evaluation studies, conceptions of validity, and the notion of transparency in AILRs.

Step 1: Problem formulation. Future design-oriented research targeting problem formulation could facilitate posing innovative questions, discovering unanticipated patterns in the literature, and promoting novel research perspectives on the phenomena under investigation. This problem of discovering novel insights from big data repositories has already been discussed in IS (Rai, 2016) and investigated in various contexts (e.g., Germonprez et al., 2007; Kallinikos and Tempini, 2014; Lukyanenko et al., 2019). The lessons from these studies could be transferred to the retrieval of information from knowledge repositories and subsequent discovery of novel problems. A second stream of research opportunities relates to facilitating the verification of research gaps (Müller-Bloch and Kranz, 2015). For instance, this could be accomplished based on ML classifiers identifying previous review papers (cf. Tsafnat et al., 2014). While health scientists have dedicated repositories of review papers at their disposal (e.g., the Cochrane library), IS researchers do not have easy access to an overview of published reviews. A corresponding artifact could enable prospective authors to verify whether their research questions have already been addressed and to make more informed statements on related review papers, especially when confronted with a rapidly growing volume of papers and reviews published outside the top-tier journals. Such tools could provide a prioritized list of reviews on related topics (e.g., based on sample overlap), facilitating research gap verification, the development of related work sections, and contribution statements.

Step 2: Literature search. We highlight two opportunities to leverage AI in the search step. First, backward citation searches (wherein the literature is drawn from the references in a focal article) are rarely reported in IS review papers despite initial evidence for their effectiveness (Jalali and Wohlin, 2012; Papaioannou et al., 2009). The tedious task of scanning multiple (in all likelihood overlapping) reference sections would easily lend itself to AI-supported extraction, consolidation, and merging of reference data. Consideration of additional cues (e.g., frequency and contexts of citations) could even provide a basis for prioritizing or filtering de-duplicated search results. Second, we expect further progress regarding tools aimed at documenting, analyzing, and justifying individual search strategies (cf. Templier and Paré, 2018). IS researchers could use initial work on syntactic search query validation (Russell-Rose and Shokraneh, 2019) to build tools supporting researchers in designing and improving different elements of search strategies. Examples of promising starting points include the analysis and justification of the scope in terms of publication outlets covered and the selection of search terms in database searches.

Step 3: Inclusion screen. The screening tasks provide the most significant opportunities for advancing AI-based tools. Most importantly, this pertains to supporting the time-consuming first screen (based on titles and abstracts), in which AI-based tools and humans can complement each other in different ways (see Larsen et al., 2019; O'Mara-Eves et al., 2015, van de Schoot et al., 2021). Many promising tools designed for reviews of health research (e.g., Wallace et al., 2012) could serve as an inspiration for design-oriented work in IS. Since IS research does not follow comparable standards regarding research reporting and regulated vocabulary, such as the MeSH terms in the health sciences (cf. O'Mara-Eves et al., 2015), modifications and careful evaluation are necessary. Beyond supporting the screen, further AI-based tools could target two mechanical, tedious tasks intertwined with the screen, namely the acquisition of full-texts between the first and second screen (Thomas et al., 2017; Tsafnat et al., 2014) and the identification of studies reporting results from the same dataset (Templier and Paré, 2018).

Step 4: Quality assessment. There are several opportunities for advancing AI-based tool support for the quality appraisal step. This is underlined by the fact that current research practices regarding the reporting of quality assessment in meta-analyses of IS research are insufficient, even for meta-analyses published in top-tier journals

(Templier and Paré, 2018). While methods papers in IS are beginning to recognize the critical role of quality assessment, corresponding procedures are a more established element in general meta-analysis methods papers (e.g., Higgins and Green, 2008; Hunter and Schmidt, 2014).

Corresponding design science research could draw inspiration from risk of bias assessment tools like *RobotReviewer* (Marshall et al., 2015) and advance AI-supported quality appraisal of non-experimental, observational, and cross-sectional research designs. In a first step, this would be greatly facilitated by classifiers dedicated to detecting the research designs and methods of primary papers. In a second step, design-oriented work could aim at (partial) automation of quality assessment based on checklists and criteria for observational studies (Shamliyan et al., 2010), surveys (Pinsonneault and Kraemer, 1993), positivist case studies (Dubé and Paré, 2003), or Delphi studies (Paré et al., 2013).

Step 5: Data extraction. Future AI-based tools supporting the data extraction step may adopt two perspectives on extant research. First, there is the view that papers contain data and evidence that have a singular interpretation and should lend themselves to efficient extraction and analysis. Consistent with the positivist paradigm (which holds that there is single ground truth accessible through empirical studies), corresponding tools may spot relatively isolated fragments of a paper to provide researchers with target categories like methodological characteristics or details of the research design. There are several opportunities of transferring and adapting corresponding tools aimed at extracting study characteristics (Jonnalagadda et al., 2015). Second, consistent with philosophical paradigms recognizing that multiple interpretations may exist simultaneously (e.g., interpretivism, critical social theory, or critical realism), research can be viewed as a discourse in which authors of review papers examine different arguments and possibly diverging interpretations of the same observations (Avison and Malaurent, 2014; Klein and Myers, 1999) in the process of forming a synthesis. The discourse evolving around particular theoretical models, that is, the focus of Larsen et al. (2019), is one very promising way to operationalize this perspective. Future work in this area is unlikely to succeed when considering isolated fragments. Instead, identifying main arguments and influential ideas will require consideration of a broader context. This pertains to an argument's position in the overall structure of the paper (Prester et al., 2020), overarching ontologies (Li et al., 2020), or its relation to previous work and the meaning associated with cited papers (Hassan et al., 2020; Small, 1978). It will require researchers to go beyond basic NLP models and improve NLP algorithms for context-sensitive data extraction, analysis, and synthesis which are capable of capturing and representing higher order linguistic structures and deeper meaning. This research could contribute to work on deep linguistic processing, context extraction, word-sense disambiguation, and other open ML and NLP problems (Dörpinghaus and Stefan, 2019; Raganato et al., 2017; Stanovsky et al., 2017; Wang et al., 2020). We believe this second, discourse-oriented perspective illustrates how tools in the social sciences may differ from those in the natural sciences.

Step 6: Data analysis and interpretation. Regarding the final step, we focus on knowledge integration and inductive theory development. With knowledge integration still posing challenges for prospective authors, we call for the development of discipline-specific algorithms to address issues especially prevalent in IS, such as similarity assessment of IS measurement items or IS constructs, and IS construct integration. Here, important progress can be made by anchoring solutions in IS domain knowledge. For example, similarity judgment for reflective vs formative constructs rely on different assumptions about the relationships between measurement items, and knowledge of whether a given IS construct is generally reflective or formative can be useful. Hence, different algorithmic similarity metrics can be used depending on the type of constructs involved. For formative constructs, measurement items are expected to be quite dissimilar from one another, as the items represent different dimensions of the higher order construct (e.g., measurement items of information quality's dimensions of perceived accuracy are quite different from those which measure perceived information completeness, see, e.g., Xu et al., 2013). In this case, to ensure that these measurement items can be automatically related, a domain-specific ontology can be helpful. In contrast, for reflective constructs, the algorithms can expect to detect high synonymity among the measurement items (e.g., as in the case of items which measure perceived ease of use), which could be a signal that these measurement items belong to the same reflective construct.

Work on integrating constructs is already being pursued, including such artefacts as *CID1* (Larsen and Bong, 2016), *ADIT* (Larsen et al., 2019), ideational impact classifiers (Prester et al., 2020), *RefMod-Miner* (Hake et al., 2017), or *TheoryOn* (Li et al., 2020). Developing theories lies at the core of IS research (Leidner, 2018; Rivard, 2014; Webster and Watson, 2002). AILRs, which can operate on larger volume of literature, provide an ideal basis for large-scale inductive theory development (Berente et al., 2019; Choudhury et al., 2018; Nelson, 2020), resulting in substantive opportunities for future studies. Here, many open questions remain, for instance, ascertaining which steps of the theory-building process based on literature are most ripe for automation. Another key challenge is assurance of validity and rigor in inductive generation of components needed for theory construction (i.e., constructs, relationships,

and boundary conditions) from heterogenous literature sources which may vary in quality and contain tacit assumptions. Automated discovery of causal chains from literature is a booming practice in medicine and bioinformatics (Hossain et al., 2012). However, it has yet to be widely utilized in IS, which offers a lucrative prospect of finding novel connections among distal IS phenomena (e.g., organizational IT, social media, and Internet of things). Although, as we stated earlier, we do not envision tools to replace humans; here, we are excited to see how a theory development could benefit from scalable automated pattern discoveries in IS.

All steps. We conclude this section by outlining four aspects relevant to AI-based tools across the six steps of the literature review process: (1) evaluation and validity, (2) transparency and replicability, (3) compatibility for recombination, and (4) usability. These aspects point to the need for methodological research as well as an accompanying discourse in the IS community, both critical pillars for the development of best practices (further discussed in level 3 of this agenda).

Evaluation and validity. Evaluation methods and studies of the feasibility, effectiveness, and utility of AILR methods and tools are needed, considering that they operate at the intersection of data, theory, and complex computational software systems. It is unlikely that a single evaluation type (e.g., behavioral experiment) could lend sufficient and comprehensive support for the various considerations in developing and using AILR tools. Presently, little guidance exists on what constitutes an effective evaluation of these tools, and researchers have started to draw attention to the lack of methodological guidance on performing such evaluations (Li et al., 2020). Drawing inspiration from Li et al.'s (2020) multi-stage evaluation approach, we encourage future work on the evaluation of AILR methods and tools, potentially combining ML experiments, behavioral field and laboratory studies, and applicability checks. As part of reporting AILRs, as well as designing and evaluating AILR tools, researchers invariably make assertions and claims about properties, behavior, and value of these tools, raising questions of validity in this context. Validity deals with justification of claims (such as inferences and conclusions) in research studies (Lindzey et al., 1998), including literature reviews (Paré et al., 2016). The area of AI developed a set of validation procedures to establish the performance of automated classifiers. The most common measures are precision, recall, their harmonic mean (F-measure), or the area under the receiver operating characteristic curve (AUC), which captures the diagnostic ability of a classifier based on varying discrimination thresholds (O'Mara-Eves et al., 2015). These measures may be used to assess the validity of AILR findings (Larsen and Bong, 2016; Prester et al., 2020). However, a recent review of design validities, including AI-based validities (Larsen et al., 2020), concluded that we continue to lack specialized design science validities needed to establish common principles of rigor when designing and applying such artifacts as ML and NLP. In the context of evaluation and validity, many open questions remain: How can bias in AILRs arising from using unrepresentative literature sources or specific local design choices (e.g., feature engineering decisions) be detected and mitigated? Do we need specialized validity categories (Larsen et al., 2020) to capture the nuances of automated literature reviews? For AILRs such validities could be concerned with the quality of inputs (e.g., research papers and other variables, such as hyperparameters), whether the internal model's characteristics could interfere with the outcomes and conclusions, and whether language interpretation is valid and appropriate to the domain of IS and specific contexts of the study.

Transparency and replicability. There has been concern and growing effort to make IS research more transparent to support replication of findings and application of IS knowledge in practice (Burton-Jones et al., 2021). Accordingly, the traditional literature review process seeks to be transparent and replicable in IS and beyond (Paré et al., 2016; Templier and Paré, 2018). Typically, researchers query databases (e.g., *Web of Science*) with explicit keywords and then follow pre-defined steps for screening, extracting, and analyzing the results. Yet, decisions can sometimes be idiosyncratic, leading to reproducibility concerns (Cram et al., 2020). AILRs may make literature extraction and coding more reproducible, as the same ML or NLP logic can be applied to new datasets, further allowing reviews to incorporate recently published papers or add new journals. At the same time, AILRs raise new transparency concerns. The problem is two-fold. First, AI frequently relies on black box models for search, data extraction, and classification tasks. The very power of such approaches (e.g., deep learning neural networks) lies in their ability to form thousands of extremely nuanced and complex rules resulting from millions or even billions of iterations over training data (Castelvecchi, 2016; Holzinger, 2016). Indeed, the complexity of the resulting models is so high that the scientists themselves may not fully understand how the algorithms work exactly (Hutson, 2018a). Second, even if AI models themselves are simple and relatively interpretable (e.g., regression models or decision trees), the process of generating these models may be opaque, lacking rigor and systematization. For example, in training ML models, many local choices need to be made (e.g., input normalization, dimensionality reduction, missing value imputation, feature engineering, and hyperparameter selection). In the absence of standardized procedures, these are commonly performed in an ad hoc manner, with great reliance on the

experience and intuition of researchers as well as trial and error approaches (Anderson et al., 2013; Duboue, 2020). These choices are then seldomly explained or shared. This means that it could be unclear how to produce these exact models given the inputs (Castellanos et al., 2021). This makes it challenging to follow the logic and replicate AILRs using traditional human coders. In short, the lack of *model preparation transparency* is a key culprit for the current *reproducibility crisis* in ML (Hutson, 2018b; Jones, 2018), which could also affect AILRs. This creates an opportunity for research on the most appropriate and effective strategies in the AILR context. To improve replicability of literature reviews (Cram et al., 2020), research on transparency and explainability of ML and NLP models is needed. Such efforts can build on the extant research in the computer science (Castelvecchi, 2016; Gunning and Aha, 2019; Knight, 2017), and we invite IS scholars to contribute to these efforts in the context of AILRs. More broadly, we hope future researchers can begin addressing such questions as: How can the AI-based literature identification and search be made more transparent? How can automated data extraction and analysis become more explainable? Also, how can transparency be improved when AI is used at multiple stages of the literature review process? Until these questions can be answered to a degree of satisfaction of the research community, transparency is likely to remain a persisting constraint on AILRs.

Of special importance is development of methods for undertaking local decisions appropriate for the AILR context. Using ML and NLP algorithms involves making a large number of specific, local decisions. For example, before running an ML algorithm, a researcher has to specify parameters (hyperparameters). Furthermore, to improve performance, the data (here, contents of research articles) are typically transformed using a variety of techniques (e.g., normalization, dimensionality reduction, and missing value imputation).

Compatibility for recombination. It is also important to design AILR tools with future recombination in mind (Beller et al., 2018). Presently, a major limitation of many tools, especially those aiming at automating multiple steps of the review process, is that they do not effectively integrate with preexisting components. Thus, while such tools are accessible, they may not be as powerful, as they restrict researchers in incorporating the most effective tools for the task. Consistent with previous calls (Al-Zubidy et al., 2017; Germonprez et al., 2007; O'Connor et al., 2018), we encourage more research on making AILR tools more flexible, modular, and tailorable (which should also contribute to a greater AILR transparency). Promising packages and ongoing projects in this area can be found in the Evidence Synthesis Hackathon Series, which has been initiated recently.[5]

Usability. Developing AILR tools is not only a technical problem, but is also a usability challenge, resulting in ample opportunities for research on human factors in tool use. Poor usability has been a persistent concern in this area (Marshall and Wallace, 2019) with current tools often created on an ad hoc basis, implementing idiosyncratic interfaces, which are difficult to understand. The tools also face the challenging tension between offering simple interfaces that are accessible to users without AI knowledge while at the same time allowing advanced users to adapt, modify, and combine algorithms. We note that *ASReview* is a promising exemplar in this regard since it offers a graphical user interface as well as command-line access to the underlying code (Van de Schoot et al., 2021). These issues create an opportunity for researchers in human–computer interaction and usability research to study and improve upon the interface design and process flow of AILR tools. As the portfolio of methods and tools supporting AILRs continuously expands, we encourage IS researchers to join forces in review teams, bringing together researchers with expertise in state-of-the-art tools and the broader spectrum of NLP and ML techniques with others who have experience in theory development, for example. There are many interesting and exciting possibilities of integrating the unique strength of AI and human–computer interaction and usability researchers in the collaborative process of conducting AILRs (cf. Raisch and Krakowski, 2020; Seeber et al., 2020).

Level III: Research practice

AILRs require broader considerations, which we believe, should involve the entire IS community, including authors of papers surveyed by AILRs, their reviewers, community thought leaders, and innovators interested in improving the way we conduct research. We highlight two broad streams of discussion pertaining to standardization and sharing.

Standardization debate within IS. To support AILR practice within IS, a discussion is needed on the potential and boundaries of greater standardization within the discipline. Clearly, our discipline being so diverse would not be well-served if we begin to straitjacket ideas through umbrella standardizations. However, some local cases of standardization (e.g., use common evaluation metrics, such as F-scores or AUCs in ML studies) may become beneficial. Indeed, many AILRs rely on limited information, such as paper abstracts only (cf., Sidorova et al., 2008), and difficulties in separating relevant elements and sections of the paper may introduce noise in the analysis. Identifying similar entities within papers remains a challenge due to lack of agreed upon conventions for describing and presenting, for example, theoretical constructs or measurement items (Endicott et al., 2017). To continue building a cumulative research body of knowledge, the IS discipline

could consider adopting common naming conventions and domain ontologies. For instance, this could mean avoiding giving the same construct different names (see Larsen and Bong, 2016) or consistent naming of standard sections of research papers (e.g., "methods"and "results"), as long as it does not detract from the ability to present the results in a unique manner. A. The rationale is that even technically perfect tools (like researchers) would struggle to extract and interpret information from sources which use ambiguous, confusing language, and presentation. Standardization, however, brings its own challenges and invites additional considerations, especially in disciplines of great diversity, such as IS. Rather than endorsing the need to standardize, we call on the community to engage in the debates about its merits and limitations. In particular, we suggest considering which areas of IS and type of papers are most amenable to standardization, where standardization may bring benefits, while constantly remaining cognizant of the negative effects of standardization, some of which may not be easy to anticipate at the onset. It is important to ensure that in pursuing the goal of integrated science, we do not alter the spirit of those contributions, which purposefully strive for multiple, nuanced, and at times, contradictory perspectives and interpretations (Avison and Malaurent, 2014; Klein and Myers, 1999).

Debate on sharing complementary research outputs. For AILRs to go beyond text of papers and analyze other research outputs (e.g., IT artifacts, empirical data, and ML models), we need to develop stronger sharing tradition in the IS discipline. Corresponding calls to improve data and IT artifact sharing practices within the IS discipline are mounting (Lukyanenko and Parsons, 2020; Maass et al., 2018), culminating with the recent MISQ Editorial on transparency, where sharing of research components is a major recommendation (Burton-Jones et al., 2021). Sharing components of research is not without challenges, such as protection of privacy or intellectual property rights of software code or adding to an already long list of things to do for scientists. Furthermore, as the IS discipline is investigating properties of IT artifacts (in addition to human behavior), it is not unreasonable to foresee research which uses advanced computational techniques (such as computer vision) to mine properties of IT artifacts and use those as units of LR analysis. Motivated by its potential benefits to AILRs, we call on the community to investigate technical approaches, requisite infrastructure (e.g., at the conference and journal levels), and community practices (e.g., during review stage) for data, model, and artifact sharing. This includes making the data open and publicly accessible, complete with appropriate meta-data to facilitate identification of the data semantics. Our call joins a chorus of suggestions made by other researchers. Thus, Maass et al. (2018) impel researchers to "play a *proactive* role in which they prepare data for future problems, needs, or changes. As part of this role, IS researchers will need to anticipate concerns that go beyond a single research project to generate approaches and infrastructures that build capacity for, and facilitate work across, multiple research settings and projects (e.g., to address problems such as data sharing)." (p. 1266). While community-level norms for sharing may take time to emerge (Burton-Jones et al., 2021), authors could lead this effort by voluntarily sharing those components of their own work they deem appropriate. In doing so, authors would make their papers more accessible to the AILR tools of the future and hence increase the exposure and potential impact of their work.

Concluding remarks

Scholars in many scientific disciplines share excitement about the opportunities of leveraging AI in support of various research tasks. In this essay, we explored how literature reviews can benefit from AI-support, summarizing the current state-of-research and sketching opportunities for future research, design, and action. While some trends target (partial) automation of repetitive tasks, others are more ambitious, advancing the use of AI in the analysis and interpretation steps. Not unexpectedly, such visionary approaches are met with excitement from some and reservations from others. In this opinionated discourse, we emphasize that AILRs are not an end in itself but a means to the end of making a strong contribution to knowledge and theory development. We expect top IS journals to continue their tradition of championing papers that thoughtfully integrate previous research streams, develop new theories, or elaborate on existing ones (Leidner, 2018; Rivard, 2014; Webster and Watson, 2002). While AI can certainly automate repetitive tasks and support others, there is no doubt that these contributions require human interpretation and insightful syntheses, as well as novel explanation and theory building. Having surveyed a range of promising examples of AI-based tools for literature reviews, we recognize that much remains to be done to support the more repetitive tasks and to facilitate insightful contributions. We therefore propose a multi-level agenda for AILR-centric research, design, and action. Our main ambition is to foster a vibrant and constructive AILR tradition in IS, which offers exciting opportunities for the entire research community, including authors and reviewers, as well as external stakeholders from other disciplines and the industry. Especially for design science researchers, there is significant potential for advancing AI-based tools and methods beyond the IS discipline. We hope that our vision encourages scholars to engage in debates and reflections on how AI can be leveraged for the progress of research in IS and its neighboring disciplines.

Declaration of conflicting interests

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

Funding

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

ORCID iD

Guy Paré https://orcid.org/0000-0001-7425-1994

Notes

1. We use the term "tool" in a broad sense, covering applications offering graphical user interfaces (GUI), statistical packages, as well as programming libraries, for example.
2. For the sake of completeness, we recognize that some tools support multiple steps of the review process (e.g., Covidence, Rayyan QCRI, Parsifal, SRDB.PRO, and SESRA). These tools tend to focus on data, workflow, and collaboration management functionality without necessarily drawing on AI capabilities. In this commentary, we focus on tools supporting individual steps because they tend to be more amenable to code inspection and extension (i.e., published under open source, non-commercial licenses), as well as independent validation.
3. https://crowd.cochrane.org/
4. https://is.theorizeit.org/wiki/Main_Page
5. https://www.eshackathon.org/

References

Adams CE, Polzmacher S, and Wolff A (2013) Systematic reviews: work that needs to be done and not to be done. *Journal of Evidence-Based Medicine* 6(4): 232–235.

Alter S (2005) Architecture of Sysperanto: a model-based ontology of the is field. *Communications of the Association for Information Systems* 15(1): 1–40.

Alvesson M and Sandberg J (2011) Generating research questions through problematization. *Academy of Management Review* 36(2): 247–271.

Al-Zubidy A, Carver JC, Hale DP, et al. (2017) Vision for SLR tooling infrastructure: prioritizing value-added requirements. *Information and Software Technology* 91: 72–81.

Anderson MR, Antenucci D, Bittorf V, et al. (2013) Brainwash: a data system for feature engineering. In: Proceedings of the biennial conference on innovative data systems research, 2013, Asilomar, CA, 2013.

Antons D and Breidbach CF (2017) Big data, big insights? Advancing service innovation and design with machine learning. *Journal of Service Research* 21(1): 17–39.

Antons D, Breidbach CF, Joshi AM, et al. (2021) Computational literature reviews: method, algorithms, and roadmap. *Organizational Research Methods*: 1094428121991230.

Avison D and Malaurent J (2014) Is theory king?: questioning the theory fetish in Information Systems. *Journal of Information Technology* 29(4): 327–336.

Balducci B and Marinova D (2018) Unstructured data in marketing. *Journal of the Academy of Marketing Science* 46(4): 557–590.

Bandara W, Furtmueller E, Gorbacheva E, et al. (2015) Achieving rigour in literature reviews: insights from qualitative data analysis and tool-support. *Communications of the Association for Information Systems* 34(8): 154–204.

Batini C, Rula A, Scannapieco M, et al. (2015) From data quality to big data quality. *Journal of Database Management* 26(1): 60–82.

Bax L, Yu L-M, Ikeda N, et al. (2007) A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC Medical Research Methodology* 7(1): 1–9.

Beller E, Clark J, Tsafnat G, et al. (2018) Making progress with the automation of systematic reviews: principles of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews* 7(1): 77. BioMed Central: 1–7.

Berente N, Seidel S and Safadi H (2019) Research commentary-data-driven computationally intensive theory development. *Information Systems Research* 30(1): 50–64.

Boell SK and Wang B (2019) wwwlitbaskets.io, an IT artifact supporting exploratory literature searches for Information Systems research. In: Proceedings of the Pacific Asia conference on information systems (eds KK Wei, WW Huang, JK Lee, et al.), X'ian, China, 8–12 July 2019.

Burton-Jones A, Boh WF, Oborn E, et al. (2021) Editor's Comments: advancing research transparency at MIS quarterly: a pluralistic approach. *MIS Quarterly* 45(2): iii–xviii.

Carver JC, Hassler E, Hernandes E, et al. (2013) Identifying barriers to the systematic literature review process. In: International symposium on empirical software engineering and measurement, Baltimore, MD, USA, 10–11 October 2013.

Castellanos A, Castillo A, Tremblay MC, et al. (2021) Improving machine learning performance using conceptual modeling. In: AAAI spring symposium: Combining machine learning with knowledge engineering, 2021.

Castelvecchi D (2016) Can we open the black box of AI? *Nature* 538(7623): 20–23.

Choudhury P, Allen R and Endres M (2018) Developing theory using machine learning methods. Available at: http://ssrn.com/abstract=3251077 http://ssrn.com/abstract=3251077.

Cooper HM (1988) Organizing knowledge syntheses: a taxonomy of literature reviews. *Knowledge in Society* 1(1): 104–126.

Cram WA, Templier M, Templier M, et al. (2020) (Re)considering the concept of literature review reproducibility. *Journal of the Association for Information Systems* 21(5): 1103–1114.

Dalgali A and Crowston K (2019) Sharing open deep learning models. In: Proceedings of the hawaii international conference on system sciences, Grand Wailea, Hawaii, USA, January 8–11, 2019.

Dang Y, Zhang Y, Chen H, et al. (2009) Arizona literature mapper: an integrated approach to monitor and analyze global bioterrorism research literature. *Journal of the American Society for Information Science and Technology* 60(7): 1466–1485.

Dörpinghaus J and Stefan A (2019) Knowledge extraction and applications utilizing context data in knowledge graphs. In: Proceedings of the federated conference on computer science and information systems, Leipzig, Germany. September 1-4, 2019, pp. 265–272.

Dubé L and Paré G (2003) Rigor in information systems positivist case research: current practices, trends, and recommendations. *MIS Quarterly* 27(4): 597–636.

Duboue P (2020) *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge, UK: Cambridge University Press.

Eisenstein J (2019) *Introduction to Natural Language Processing*. Cambridge, MA: MIT press.

Endicott J, Larsen K, Lukyanenko R, et al. (2017) Integrating scientific research: Theory and design of discovering similar constructs. In: AIS SIGSAND Symposium, Cincinnati, Ohio, 2017, pp. 1–7.

Evans JA and Foster JG (2011) Metaknowledge. *Science* 331(11): 721–725.

Germonprez M, Hovorka D, Hovorka D, et al. (2007) A theory of tailorable technology design. *Journal of the Association for Information Systems* 8(6): 351–367.

Gleasure R. Conceptual design science research? How and why untested meta-artifacts have a place in IS. In: Proceedings of the international conference on design science research in information systems and technology. Miami, FL, USA, 2014, pp. 99–114.

Goodfellow I, Bengio Y and Courville A (2016) *Deep Learning, Adaptive Computation and Machine Learning Series*. Cambridge, MA: MIT Press.

Gunning D and Aha D (2019) DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine* 40(2): 44–58.

Hake P, Fettke P, Neumann G, et al. Extracting business objects and activities from labels of German process models. In: Proceedings of the international conference on design science research in information system and technology, Karlsruhe, Germany, May 30 - June 1, 2017, pp. 21–38.

Harrison H, Griffin SJ, Kuhn I, et al. (2020) Software tools to support title and abstract screening for systematic reviews in healthcare: an evaluation. *BMC Medical Research Methodology* 20(7): 1–12.

Hartling L, Ospina M, Liang Y, et al. (2009) Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *British Medical Journal* 339(1): 1–6.

Hassan NR, Prester J and Wagner G (2020) Seeking out clear and unique Information Systems Concepts: a natural language processing approach. In: Proceedings of the European conference on information systems (eds MLF Rowe and R El Amrani), Marrakech, Morocco, 15–17 June 2020.

Higgins JPT and Green S (2008) *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons, Ltd.

Holzinger A (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop?. *Brain Informatics* 3(2): 119–131.

Hossain MS, Gresock J, Edmonds Y, et al. (2012) Connecting the dots between PubMed abstracts. *PLoS One* 7(1): 1–23.

Hunter JE and Schmidt FL (2014) *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. 2nd edition. Thousand Oaks, CA: Sage.

Hutson M (2018a) Has artificial intelligence become alchemy? *Science* 360(6388): 478–479.

Hutson M (2018b) Artificial intelligence faces reproducibility crisis. *Science* 359(6377): 725–726.

Jalali S and Wohlin C (2012) Systematic literature studies: database searches vs. backward snowballing. In: Proceedings of the ACM-IEEE international symposium on empirical software engineering and measurement, Lund, Sweden, September 20-21, 2012, pp. 29–38.

Johnson CD, Bauer BC and Niederman F (2019) *The Automation of Management and Business Science*. Academy of Management Perspectives 35(2): 292–309.

Jones M (2018) How do we address the reproducibility crisis in artificial intelligence? *Forbes*. Available at: https://www.forbes.com/sites/forbestechcouncil/2018/10/26/how-do-we-address-the-reproducibility-crisis-in-artificial-intelligence/#54236c5b7688.

Jonnalagadda SR, Goyal P, and Huffman MD (2015) Automating data extraction in systematic reviews: a systematic review. *Systematic Reviews* 4(1): 78.

Kallinikos J and Tempini N (2014) Patient data as medical facts: social media practices as a foundation for medical knowledge creation. *Information Systems Research* 25(4): 817–833.

Kenett RS and Shmueli G (2016) *Information Quality: The Potential of Data and Analytics to Generate Knowledge*. Hoboken, NJ: John Wiley & Sons.

King RD, Rowland J, Oliver SG, et al. (2009) The automation of science. *Science* 324(5923): 85–89.

Kitchenham BA and Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. EBSE Technical Report.

Klein HK and Myers MD (1999) A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly* 23(1): 67–94.

Knight W (2017) *Darpa Is Funding Projects that Will Try to Open up AI's Black Boxes*. MIT Technology Review. Available at: https://www.technologyreview.com/2017/04/13/152590/the-financial-world-wants-to-open-ais-black-boxes/. Accessed on Sept 25, 2020.

Kobayashi VB, Mol ST, Berkers HA, et al. (2017) Text mining in organizational research. *Organizational Research Methods* 21(3): 733–765.

Kohl C, Mcintosh EJ, Unger S, et al. (2018) Online tools supporting the conduct and reporting of systematic reviews and systematic

maps: a case study on cadima and review of existing tools. *Environmental Evidence* 7(8): 1–17.

Chandra Kruse L, Seidel S and vom Brocke J (2019) Design archaeology: Generating design knowledge from real-world artifact design. In: Proceedings of the International Conference on Design Science Research in Information Systems and Technology, Worcester, MA, USA, June 4-6, 2019, pp. 32–45.

Kunc M, Mortenson MJ and Vidgen R (2018) A computational literature review of the field of system dynamics from 1974 to 2017. *Journal of Simulation* 12(2): 115–127.

Larsen KR, Bong CH and Bong CH (2016) A tool for addressing construct identity in literature reviews and meta-analyses. *MIS Quarterly* 40(3): 529–551.

Larsen K, Hovorka D, Dennis AR, et al. (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.

Larsen K, Lukyanenko R, Muller R, et al. (2020) Validity in design science research. In: Proceedings of the International Conference on Design Science Research in Information Systems and Technology, Kristiansand, Norway, December 2-4, 2020.

Laurell C, Sandström C, Berthold A, et al. (2019) Exploring barriers to adoption of virtual reality through social media analytics and machine learning–an assessment of technology, network, price and trialability. *Journal of Business Research* 100: 469–474.

Leidner DE (2018) Review and theory symbiosis: an introspective retrospective. *Journal of the Association for Information Systems* 19(6): 552–567.

Li J, Larsen K, Abbasi A, et al. (2020) TheoryOn: a design framework and system for unlocking behavioral knowledge through ontology learning. *MIS Quarterly* 44(4): 1733–1772.

Lindberg A (2020) Developing theory through integrating human and machine pattern recognition. *Journal of the Association for Information Systems* 21(1): 90–116.

Lindzey G, Gilbert D, and Fiske ST (1998) *The Handbook of Social Psychology*. New York: Oxford University Press.

Lukyanenko R (2020) A journey to BSO: evaluating earlier and more recent ideas of Mario Bunge as a foundation for information and software development. In: Exploring Modeling Methods for Systems Analysis and Development, Grenoble, France, 2020, pp. 1–15.

Lukyanenko R, Parsons J and Parsons J (2020) Research perspectives: design theory indeterminacy: what Is it, how can it be reduced, and why did the polar bear drown? *Journal of the Association for Information Systems* 21(5): 1343–1369.

Lukyanenko R, Parsons J, Wiersma YF, et al. (2019) Expecting the unexpected: effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly* 43(2): 623–648.

Maass W, Parsons J, Purao S, et al. (2018) Data-driven meets theory-driven research in the era of big data: opportunities and challenges for Information Systems research. *Journal of the Association for Information Systems* 19(12): 1253–1273.

Manning CD and Schütze H (1999) *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

March ST and Allen GN (2014) Toward a social ontology for conceptual modeling. *Communications of the Association for Information Systems* 34(70): 1347–1359.

Marshall IJ and Wallace BC (2019) Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic Reviews* 8(1): 163.

Marshall IJ, Kuiper J and Wallace BC (2015) RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *Journal of the American Medical Informatics Association* 23(1): 193–201.

Mitchell TM (1997) *Machine learning*. Burr Ridge, IL: McGraw-Hill, 45, pp. 870–877.

Mortenson MJ and Vidgen R (2016) A computational literature review of the technology acceptance model. *International Journal of Information Management* 36(6): 1248–1259.

Müller-Bloch C and Kranz J (2015) A framework for rigorously identifying research gaps in qualitative literature reviews. In: Proceedings of the international conference on information systems (eds T Carte, A Heinzl and C Urquhart), Fort Worth, Texas, USA, 16 December 2015.

Nakagawa S, Samarasinghe G, Haddaway NR, et al. (2019) Research weaving: visualizing the future of research synthesis. *Trends in Ecology & Evolution* 34(3): 224–238.

Nelson LK (2020) Computational grounded theory: a methodological framework. *Sociological Methods & Research* 49(1): 3–42.

O'Connor AM, Tsafnat G, Gilbert SB, et al. (2018) Moving toward the automation of the systematic review process: a summary of discussions at the second meeting of the International Collaboration for the Automation of Systematic Reviews (ICASR). *Systematic Reviews* 7(3): 1–5.

O'Mara-Eves A, Thomas J, McNaught J, et al. (2015) Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews* 4.

Papaioannou D, Sutton A, Carroll C, et al. (2009) Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Information & Libraries Journal* 27(2): 114–122.

Paré G, Cameron A-F, Poba-Nzaou P, et al. (2013) A systematic assessment of rigor in information systems ranking-type Delphi studies. *Information & Management* 50(5): 207–217.

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

Pinsonneault A and Kraemer K (1993) Survey research methodology in management information systems: an assessment. *Journal of Management Information Systems* 10(2): 75–105.

Prester J, Wagner G, Schryen G, et al. (2020) Classifying the ideational impact of Information Systems review articles: a

content-enriched deep learning approach. *Decision Support Systems* 140: 113432.

Raganato A, Bovi CD and Navigli R (2017) Neural sequence learning models for word sense disambiguation. In: Proceedings of the conference on empirical methods in natural language processing, Copenhagen, Denmark, September 7-11, 2017, pp. 1156–1167.

Rai A (2016) Editor's comments: synergies between big data and theory. *MIS Quarterly* 40(2): iii–ix.

Raisch S and Krakowski S (2020) Artificial intelligence and management: the automation-augmentation paradox. *Academy of Management Review* 46(1): 192–210.

Rivard S (2014) Editor's comments: the ions of theory construction. *MIS Quarterly* 38(2): iii–xiv.

Russell SJ and Norvig P (2016) *Artificial Intelligence: A Modern Approach*. Malaysia: Pearson Education Limited.

Russell-Rose T and Shokraneh F (2019) 2Dsearch: Facilitating reproducible and valid searching in evidence synthesis. *BMJ Evidence-Based Medicine* 24(Suppl 1): A36.

Schmiedel T, Müller O and 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.

Schryen G, Wagner G, Benlian A, et al. (2020) A knowledge development perspective on literature reviews: validation of a new typology in the IS field. *Communications of the Association for Information Systems* 46: 134–168.

Seeber I, Bittner E, Briggs RO, et al. (2020) Machines as teammates: a research agenda on AI in team collaboration. *Information & Management* 57(2): 1–22.

Shamliyan T, Kane RL, Dickinson S, et al. (2010) A systematic review of tools used to assess the quality of observational studies that examine incidence or prevalence and risk factors for diseases. *Journal of Clinical Epidemiology* 63(10): 1061–1070.

Sidorova A, Evangelopoulos N, Valacich JS, et al. (2008) Uncovering the intellectual core of the information systems discipline. *MIS Quarterly* 32(3): 467–482.

Small HG (1978) Cited documents as concept symbols. *Social Studies of Science* 8(3): 327–340.

Stanovsky G, Eckle-Kohler J, Puzikov Y, et al. (2017) Integrating deep linguistic features in factuality prediction over unified datasets. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, July 30 - August 4, 2017, pp. 352–357.

Sturm B and 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.

Swanson DR and Smalheiser NR (1997) An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence* 91(2): 183–203.

Tate WL, Ellram LM and Kirchoff JF (2010) Corporate social responsibility reports: a thematic analysis related to supply chain management. *Journal of Supply Chain Management* 46(1): 19–44.

Taulli T and Oni M (2019) *Artificial Intelligence Basics*. 1st edition. Berkeley, CA: Apress.

Templier M and 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.

Thilakaratne M, Falkner K and Atapattu T (2019) A systematic review on literature-based discovery: general overview, methodology, & statistical analysis. *ACM Computing Surveys* 52(6): 1–34.

Thomas J, Noel-Storr A, Marshall I, et al. (2017) Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology* 91: 31–37.

Tkaczyk D, Collins A, Sheridan P, et al. Machine learning vs. Rules and out-of-the-box vs. retrained. In: Proceedings of the ACM/IEEE on Joint Conference on Digital Libraries, Fort Worth, Texas, USA, June 3-7, 2018.

Tremblay MC, van der Meer D and Beck R (2018) The effects of the quantification of faculty productivity: perspectives from the design science research community. *Communications of the Association for Information Systems* 43: 625–661.

Tsafnat G, Glasziou P, Choong MK, et al. (2014) Systematic review automation technologies. *Systematic Reviews* 3: 1–15

van de Schoot R, de Bruin J, Schram R, et al. (2021) An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence* 3, 125–133.

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

vom Brocke J, Simons A, Riemer K, et al. (2015) Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. *Communications of the Association for Information Systems* 37(9): 205–224.

Wahyudi A, Kuk G and Janssen M (2018) A process pattern model for tackling and improving big data quality. *Information Systems Frontiers* 20(3): 457–469.

Wallace BC, Small K, Brodley CE, et al. Deploying an interactive machine learning system in an evidence-based practice center: Abstrackr. In: Proceedings of the ACM SIGHIT International Health Informatics Symposium, Miami, Florida, USA, January 28-30, 2012, pp. 819–824.

Wand Y and Weber R (1988) An ontological analysis of some fundamental information systems concepts. *Proceedings of the International Conference on Information Systems*: 213–226.

Wang Z, Nayfeh T, Tetzlaff J, et al. (2020) Error rates of human reviewers during abstract screening in systematic reviews. *PLoS One* 15(1): e0227742.

van Zoonen W and van der Meer TGLA (2016) Social media research: the application of supervised machine learning in organizational communication research. *Computers in Human Behavior* 63: 132–141.

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

Whetten DA (1989) What constitutes a theoretical contribution? *Academy of Management Review* 14(4): 490–495.

Xu J, Benbasat I, Benbasat I, et al. (2013) Integrating service quality with system and information quality: an empirical test in the e-service context. *MIS Quarterly* 37(3): 777–794.

Author Biographies

Gerit Wagner is a postdoctoral fellow at HEC Montréal. His research focuses on literature reviews, the impact of research methods, digital health, and digital platforms for knowledge work. His research has been published in international journals, including the Journal of Strategic Information Systems, Journal of Medical Internet Research, Information & Management, and Decision Support Systems. He is a member of the Association for Information Systems and he regularly serves as a reviewer for Information Systems journals and conferences.

Roman Lukyanenko is an associate professor of information systems at HEC Montreal, Canada. He obtained his PhD from Memorial University of Newfoundland, Canada. In his research, Roman investigates and develops innovative information technology solutions to support management of natural resources, development of smart cities and decision making in healthcare systems. Roman published over 120 scientific conference and journal articles, including in leading scientific journals, such as Nature, MIS Quarterly, Information Systems Research, and Conservation Biology. Roman's recent book, "A Message Almost Lost", offers a new perspective on humanity's existential issues. Roman is the current president of AIS SIGSAND and a co-developer of https://sigsand.com/.

Guy Paré is Professor of Information Technology and holds the Research Chair in Digital Health at HEC Montréal. His current research interests involve the barriers to adoption, effective use, and impacts of e-health technologies as well as literature review approaches and methods. His publications have appeared in top-ranked journals including MIS Quarterly, Journal of Information Technology, European Journal of Information Systems, Journal of the Association for Information Systems, Information & Management, Journal of the American Medical Informatics Association, and Journal of Medical Internet Research.