The Application of Artificial Intelligence on Different Types of Literature Reviews - a Comparative Study

Henry Müller, Simran Pachnanda, Felix Pahl and Christopher Rosenqvist

The application of artificial intelligence on different types of literature reviews - A comparative study

Henry Müller
Department of Marketing and Strategy
Stockholm School of Economics
Stockholm, Sweden
41836@student.hhs.se

Simran Pachnanda
Department of Marketing and Strategy
Stockholm School of Economics
Stockholm, Sweden
simran.pachnanda@hhs.se

Felix Pahl
Department of Marketing and Strategy
Stockholm School of Economics
Stockholm, Sweden
41864@student.hhs.se

Christopher Rosenqvist
Department of Marketing and Strategy
Stockholm School of Economics
Stockholm, Sweden
christopher.rosenqvist@hhs.se

Abstract— **The growing number of published academic literature poses challenges to the research community which struggles to keep up with the vast number of publications through traditional research methods that are highly manual in nature. Researchers are struggling to determine the most relevant research gaps, yielding insignificant publications that constitute a waste of resources. As a consequence, AI applications are being applied increasingly to automate and facilitate the review process of these vast amounts of papers. However, scholars have so far only addressed a limited number of scientific fields and focused their efforts on one end of the spectrum in automating systematic literature reviews (SLRs). Yet, these are not sufficient to cover the full range of research questions and available data sources. This paper offers a comparative study of systematic and semi-systematic literature reviews to determine the potential of AI applications in both types of literature review processes. The analysis addresses the status quo and discusses apparent limitations of AI to automate reviews. Results are synthesized in proposing a new tool integrating various AI applications along the research process that improve the speed, quality, and cost-efficiency of the overall research process.**

Keywords—Artificial Intelligence, automation, review, literature review, systematic literature review (SLR), semi-systematic review, machine learning, natural language processing

I. INTRODUCTION

The ever-increasing volume of scholarly literature is making it difficult for researchers to stay up to date with the developments in their disciplines. While the number of published articles and journals have both been growing steadily for over two centuries now by about 3% and 3.5% per year respectively [1], scientific publications have particularly exploded in the last two decades. In 2014 alone, there were over 28,000 active peer-reviewed journals present that contributed north of 2.5 million articles [1], and that is excluding articles that were published in non-English languages. This information overload has been further exacerbated by the COVID-19 pandemic which has instigated a surge of publications in various fields, not least medical research [2], with publishers like Elsevier experiencing a rise in submissions by almost 58% between February and May 2020 compared with the same period in 2019 [3]. Keeping literature reviews up to date therefore incurs significant difficulties and inaccuracies that render the translation of knowledge into action difficult [4].

Part of this information growth can be attributed to the "Publish or Perish" culture observed in the academic world which puts a lot of weight on researchers to publish academic work to succeed in their careers. The effects of this can also be observed at the university level where the volume of publications often drive funding decisions and university rankings. Moreover, this publishing pressure has led to a rise in sub-par research with large quantities of books and articles of marginal quality being distributed as scholars are often more focused on increasing their quantity of published work, even if it comes at the cost of quality. This high volume of irrelevant articles mixed into the information pool is making it harder to screen out useful content through the traditional ways of research which is highly manual, consequently creating a demand for tools that could help fast-track this part of the research process and allow the user to direct their limited time on more abstract matters. One such solution is through the application of Artificial Intelligence (AI).

AI applications have some apparent advantages over humans in being significantly faster, accurate, and cost-efficient in performing lengthy and repetitive search activities. Another potential advantage is the increased precision and exhaustiveness of AI tools allowing researchers to better build on existing knowledge, making research much more rigorous than it currently is. Additionally, it might help reduce researcher bias such as favoring certain viewpoints and referencing highly cited papers for the sake of positioning.

While the application of AI is already prevalent in various aspects of research, a growing interest has been observed in the automation of literature reviews using AI, in particular for systematic literature reviews (SLRs) which are forming its own niche research bubble. A recent review paper [5] found 41 primary studies that focused on the application of AI tools such as machine learning (ML) and natural language processing (NLP) for automating one or more selected steps in the SLR process which were able to cut down on the workload by 30-70%. Medical researchers have suggested instituting "living" systematic reviews that are continuously updated, which could be further facilitated by automatic screening [4], [6]. However, it is important to note that while SLRs are considered the gold standard in evidence-based literature, the ultimate merit of a literature review depends heavily on the goal of the study in question. For example, for studies whose aim is to get an overview or a topic, especially from a multidisciplinary perspective, a semi-systematic literature review would be a better choice. Moreover, SLRs are by design made for handling quantitative data and might not be suited for qualitative or theoretical research which is prominent in social sciences.

While recent studies like Almasri et al. [7] have attempted to transfer the widely used PRISMA standard to social sciences and business research, systematic reviews are often inadequate for opinion-driven fields. Hence with our paper, we aim to elaborate on the differences in these two kinds of literature reviews and discuss their current ability to be automated through AI.

II. BACKGROUND

A. Research Phases

The process of conducting a research project may be divided into several phases. Whittemore & Melkus [8] defined five phases of the research process as shown in Table I. Thereby, the conceptual phase covers formulating the problem and determining the research purpose and consequently involves reviewing existing literature. It is therefore most affected by the aforementioned acceleration in research output and the related issues. At the same time, this phase is well-suited for AI applications as it consists of many repetitive, mechanical tasks. Hence, applying AI to the conceptual phase to cope with the increasing difficulty of reviewing existing literature appears a promising avenue.

TABLE I. RESEARCH PHASES

Research Phase	Definition
Conceptual phase	Formulating the research problem, reviewing the literature, determining the research purpose
Design and planning phase	Selecting a research design, developing study procedures, determining the sampling and data collection plan
Empirical phase	Collecting data and preparing data for analysis
Analytic phase	Analyzing data and interpreting the results
Dissemination phase	Communicating results to appropriate audience

Table adopted from [8]

B. Types of Literature Reviews

There are different types of literature reviews that are referred to under varying terms. Snyder [9] has classified them as *a) systematic*, *b) semi-systematic*, and *c) integrative* forms of literature reviews (Table II). These types are used to follow specific objectives and consequently apply various methods while relying on different data sets. This has implications for their aptitude for and realization of their automation via AI.

Systematic literature reviews (SLR) focus on quantitative articles to answer a specific research question, e.g. aggregating results from randomized control trials to inform evidence-based medicine. *Semi-systematic or narrative reviews* tackle broad questions and therefore rely on both quantitative and qualitative papers. *Integrative literature reviews* are a special genre commonly applied to criticize the status quo of a given field. They do not follow a systematic sampling procedure and involve considerable creative, interpretative, and conceptual work over a qualitative process that presently is little suitable for the application of AI tools. Given also its narrow scope and that such literature reviews are much less affected by the proposed complications coming with the accelerating research output, integrative literature reviews are beyond the scope of this paper.

TABLE II. TYPES OF REVIEWS

Approach	Systematic	Semi-systematic	Integrative
Typical purpose	Synthesize and compare evidence	Overview research area and track development over time	Critique and synthesize
Research question	Specific	Broad	Narrow and broad
Search strategy	Systematic	May or may not be systematic	Usually not systematic
Sample characteris -tics	Quantitative articles	Research articles	Research articles, books, other published texts
Analysis and evaluation	Quantitative	Qualitative/ quantitative	Qualitative
Examples of contribution	Evidence of effect; inform policy and practice	State of knowledge; Themes in literature; Historical overview; Research agenda; Theoretical model	Taxonomy or classification; Theoretical model or framework

Table adopted from [9]

Concerning the application of AI to the research process, especially with regards to the reviewing process, the focus in the present literature has been on SLRs [5]. However, such a narrow focus on the aggregation of quantitative, empirical studies neglects the accelerated research output throughout other areas and forms of research. With this paper, we want to draw attention to more qualitative fields and thus advocate the accessibility and design of such tools for all domains. Both systematic and semi-systematic forms are needed to provide a coherent overview and to be able to match different starting points. These may range from validating an already formulated research idea or question to coming up with relevant ideas or even fields of research, to begin with. Whereas the automation of SLRs depends merely on the identification, extraction, and aggregation of quantitative data and descriptive parameters of studies as well as more general metadata, realizing automated semi-systematic literature reviews appears more challenging. The inclusion of qualitative data and a focus on the development of concepts and theory while covering multiple, in part contrasting perspectives, complicates the aggregation of data from studies included in semi-systematic reviews.

A critical question concerns the achievability or desirability of a fully automated process for these reviews. This links to the general debate regarding AI's capacity to master tasks characterized by contextual understanding, creativity, and interpretation.

C. AI in the Literature Review process

More descriptive tasks are well suited for the use of AI tools. New technological advancements in NLP, ML, and text mining tools provide opportunities for the automation of literature reviews [5]. Especially early process steps that require little interpretation lend themselves such as the identification and selection of primary studies. This is where most tools have been developed in medical research, with recent versions using deep neural networks exhibiting good recall and precision, yet replicability across datasets is still weak for most tools causing some concerns for future development [10].

Academic research predominantly benefits from the advancement in NLP, a subfield of AI, which nowadays is

increasingly used to gain insights from vast amounts of textual data. NLP is a research area that aims to understand and process natural language text or speech [11]. It is an important tool for improving efficiency since most academics need to process large volumes of textual data and documents throughout the research process. NLP tools are mostly developed and run in conjunction with ML to train the models and identify patterns [5].

The main NLP technologies used in the automation of literature reviews are data extraction and text classification. Text classification involves models that automatically classify documents (e.g., article abstracts, full-texts, references, etc.) into categories [12]. Data extraction models attempt to identify sections of text or individual words/numbers that correspond to a particular variable of interest (e.g., extracting quantitative study results). For example, Kaur et al. [13] have used data extraction tools to label hospital discharge summaries with ICD codes, applied treatment, and doses to analyze the data more quickly. Some researchers have used text classification applications in the context of abstract screening to check whether articles meet the inclusion criteria for a particular review [5]. These models often include ML algorithms that can estimate the probability of an article being included, thereby further ranking articles according to relevance, accelerating the screening process for researchers [14]. Due to the immaturity of these solutions, ML is much rather used to expedite tasks, rather than automate them completely [4]. Along these lines, hybrid models have been proposed that enable living systematic reviews by using push and pull modes to retrieve and pre-label new articles that facilitate the integration by trained medical researchers [15].

Also, the inherent subjectivity of certain areas and the reassurance of expert humans render full automation unrealistic, resulting in almost all tools being designed as "human-in-the-loop" systems [12]. However, novel computational methods involving text mining, deep learning, and neural networks may enable new forms of synthesis that cannot be achieved by researchers, particularly in the areas of data visualization and automatic summarization of large volumes of research [12]. Recent attempts have tried to discern meaning, quality of writing, and biases from articles and create summaries [16], [17], [18], [19]. By using extrapolation, it was possible to give meaning even to poorly written texts [18] or to identify new materials or possible fields of application [19]. Although these tools have mostly been tested on specific training datasets, this constitutes a promising development, especially for viable applications in humanities and social sciences.

III. COMPARATIVE STUDY

Reviewing literature is an important part of any research project. Historically, systematic literature reviews were born out of a need to synthesize and cope with the increasing mass evidence in medicine [4], [6], [20]. SLRs and statistical meta-analyses differ from narrative reviews by using a replicable, scientific and transparent process that aims to minimize bias through exhaustive literature search and providing a detailed process description for audit [21]. To address the implicit biases in narrative reviews of management scholars, Tranfield et al. [20] conceptualized how the process of SLRs can be applied to the management

field to build a reliable body of knowledge despite using descriptive accounts and context-sensitive research.

Reviews in both fields can be structured into four phases consisting of several similar steps [9]. We perform a comparative study by revisiting the differences and challenges elaborated by Tranfield et al. [20] while addressing the potential and feasibility of using AI tools for each of the respective process steps employed in systematic and semi-systematic or narrative reviews. This is summarized in Table III.

A. Design

The specific design of each SLR is different and highly dependent on the research question(s) it is trying to answer. However, each step of the process follows a standardized procedure with strict standards that need to be met. The SLR design starts with its research question(s) that needs to be clearly formulated at the beginning and should be targeting a very specific problem. Additionally, a predefined review protocol needs to be developed before starting the review to minimize researcher bias. This covers laying out the template for each step of the SLR, from the search strategy for primary studies to the dissemination strategy. Furthermore, these protocols are critically evaluated internally or by other experts in the field before conducting the review. The conceptual and divergent nature of the design phase makes it difficult to develop standardized AI tools into this phase of the SLR process.

The initial steps of Semi-systematic reviews are similar to the aforementioned process of the SLR, as only after the formulation of the research question(s) the choice of the appropriate form of literature review is made. Unlike the SLR, a fixed review protocol is commonly not formulated for semi-systematic literature reviews due to the limited overview of the researcher regarding the issue at hand and potentially relevant field of research, including constructs and theory [22]. The exploratory nature of interpretivist research takes room for discoveries and new ideas. Consequently, little research exists concerning the application of AI in this phase, and the potential to automate the involved steps appears limited.

B. Conducting the Review

The first step in the execution stage for any review is the collection of primary studies. For SLRs, this is done through a well-defined search strategy with strict inclusion and exclusion criteria that is set during the design stage. The review needs to be transparent and replicable, requiring each step of the search process to be documented in detail. Most of the search methods involve steps that are repetitive thus giving a high potential for automation through AI. The screening step that follows the initial search is divided into multiple stages starting from an initial screening of all possible material that might be relevant to the study where keeping a high level of recall is essential. This first screen, which usually focuses on the title and abstracts, has high potential for automation as it requires many repetitive decisions. The subsequent screening is considerably more involved and required evaluating the full texts of the articles, in which case AI tools like NLPs might be applicable but only to a certain extent as the final selection might require some expert human intervention. Recent tools such as Dextr show that semi-automated methods only yield

marginal losses in recall and precision, while drastically reducing the extraction time compared to manual workflows [23]. The final step in this phase is the quality assessment of the collected material, which is a highly scrutinized step for SLRs and has only shown a low-level potential for automation given the current state of AI. Further, the accuracy of automated literature screening tools varies widely across different algorithms and review topics [24]. As reviewers attempt to consider each publication in relation to the review topic, literature screening systems can achieve up to 95% sensitivity, which comes at the expense of specificity leading to additional human assessments [24]. However, certain NLP and ML-powered software like RobotReviewer [25], is showing promising potential for expansion. Researchers agree that this is one of the most time-consuming steps in the process of writing an SLR, thus highly beneficial in terms of time and cost when paired with AI tools.

As for semi-systematic reviews, relevant (sub-)fields of research need to be identified. This process is commonly undertaken in a subjective, non-systematic manner by researchers talking to their peers or building upon personal experience [20]. Due to the broader scope of the pursued research questions, this often results in interdisciplinary setups, definitions, and conceptualizations that may not be clearly delineated in the beginning, meaning related constructs and synonyms may have to be unveiled to adequately capture the relevant fields [22]. This may cause a need for scoping studies to assess the relevance and size of literature, which constitutes an additional time-intensive and complex step due to the need for cross-disciplinary perspectives. Although not addressed in present literature, AI tools may be introduced in this step to help translate the current non-systematic and potentially biased process into a more systematic encounter of fields of research suited to approach a given research question. However, the automation potential of the identification step is limited as determining the right fields involves significant contextual understanding and at times creative new combinations or applications of distant research yield significant scientific progress. Automated content analysis (ACA) can assist in classifying (related) concepts often by measuring the frequency of certain words. Researchers can then cluster concepts and frameworks to classify literature or even parts of it yielding different outputs such as trend analyses, concepts maps, and co-occurrence records [14]. In addition, ML applications are capable of calculating the probability of a paper being selected therefore ranking higher and saving the researcher valuable time.

The selection process involves two major steps, including the identification of related research with help of the prior tools and the assessment of the research quality to decide about its inclusion in the sample. Concerning the former step and as mentioned in the context of SLRs, the application of AI can yield significant benefits in aggregating primary research, as search and text classification algorithms may process vast amounts of data and papers from various databases.

TABLE III. COMPARISON - SYSTEMATIC VS SEMI-SYSTEMATIC LITERATURE REVIEW

Phase	Steps included	Systematic	AI Tools	Semi-systematic	(Potential) AI Tools
Design	- Commissioning a review - Specifying the research question - Developing a review protocol - Evaluating the review protocol	- Often commissioned to researchers (review panel) - Well defined research question(s) - Protocol defined prior to undertaking review, often cross-verified with experts - Largely free of researcher bias	No widely used AI tools for this phase - difficult to develop due to the subjective nature of the phase	- Low level of standardization - Low consensus over research questions - Context-specific and many extraneous factors - Usually informal/ad hoc process involving peers and supervisor - Strict planning often inappropriate	- Difficult to develop an AI tool for this phase, similar reasoning as SLRs
Conduct	- Identification of research - Selection of primary studies - Study quality assessment	- Detailed search strategy - Rather broad field (scoping pre-study) - Well defined inclusion/ exclusion criteria - Review must be transparent and replicable - Protocol incl. coding strategy and statistical procedures	- Search: RobotSearch, TheoryOn, Litbaskets - Screening: Colandr, ASReview, RobotAnalyst, Abstrackr - Unsupervised concept seeding	- Subjective, non-systematic identification - Semi-independent sub-fields - Selection based on interest, seldom critical appraisal - Raw data often not available for qualitative - No recording of inclusion/exclusion criteria	- NLP: Concept Seeding (Hand-Seeding) - NLP: Text classification - NLP: Data Extraction - ML: Relevance scores - ML: Quality assessment
Analysis	- Data extraction and monitoring - Data synthesis	- Ideally draw upon raw data to create study in its own right - Quantitative synthesis (meta-analysis) - Predominantly quantitative but can have narrative elements	- Extraction: Nvivo, ExaCT, SRDR+, Robot Reviewer, WebPlotDigitizer - Synthesis: RevMan	- Qualitative and quantitative data - Generally, more narrative, with higher levels of subjectivity - Qualitative tools/methods to counteract biases - Sensemaking and contextual understanding necessary - Might include statistical aggregation, qualitative aggregation, narrative summary	- Automated Content Analysis (ACA) to analyze qualitative data - Concept classification, relationship of concepts, syntax - Descriptive analysis of research streams

Phase	Steps included	Systematic	AI Tools	Semi-systematic	(Potential) AI Tools
Writing	- Specifying dissemination mechanisms - Formatting the main report - Evaluating the report	- Standardized reporting structure	- No widely used AI tools for this phase	- No standardized reporting structure - Explanatory power added through analogy, metaphor, and homology	- Visualize research streams

Yet, one major constraint is the accessibility of research. Often, relevant contributions to ongoing debates are only present in unpublished papers, conference papers or manuscripts. Consequently, researchers must make a trade-off between volume and actuality. Here, unpublished sources might be added manually by the researchers after preliminary screening and selection steps to complement the relevant published papers.

The selection of retrieved articles is commonly based on quality assessment [9], [20]. Thereby, research paradigms differ regarding agreed-upon quality criteria as research questions, methodology and instruments vary between them. Unlike in predominantly quantitative research such as medicine, where double-blind randomized control trials represent the highest methodological standard and other statistical methods may be evaluated based on sample sizes and other error metrics, things are more equivocal in semi-systematic literature reviews [20]. The decisive question is whether the chosen methodology matches the formulated research question and whether the methodology was performed thoroughly [22]. Presently, the educated eye of the researcher and journal rankings are needed to make this judgment. As of today, it remains open if supervised ML algorithms will be able to sufficiently automate or merely support this process. In combining NLP and ML, quality ratings based on defined criteria of the respective research community or related rankings based on overall quality and relevance could be involved.

C. Analysis

In this phase, all the relevant quantitative and qualitative data from the selected literature needs to be extracted and converted into structured data files. The process of data extraction is highly manual, error-prone and time-consuming, which makes it a valuable step for automation. The quality of NLP and text mining techniques have reached significant maturity in recent years which has created a wide variety of tools for data extraction that can be easily applied to both systematic and semi-systematic literature reviews. Literature reviews differ in the required data inputs, i.e. the part of the paper and the type of data contained in the various parts, varying between qualitative and quantitative. For SLRs, the data extraction part ranges from simple title and abstract analysis in the initial phases to reviewing materials from the full length of the shortlisted articles. Extracting the former is significantly faster than the latter which is often affected by non-standardized formats of articles across journals. Hence the degree of automation achieved for SLRs at this step ranges depending on the phase of the review. In the case of semi-systematic reviews, gathering sufficient data to broad and complex research questions would require text mining algorithms to process full-text documents. Hence, papers with limited access, e.g. to title, abstract, and bibliography may need to be excluded. In addition, the qualitative and more conceptual nature of the papers investigated here can entail relevant information (i.e. text fragments) presented as part of visualizations that most tools cannot access. Tools transforming images to text will be necessary to extract this information.

In the analysis phase, researchers cluster the derived research into clusters along varying dimensions depending on the research fields, research question, and other contextual factors. This process involves considerable analysis and insight of large amounts of papers making it a time intensive and mechanical task. Here, ML applications relying on supervised learning could help classify research based on predefined categories. Previously mentioned tools like concept maps and co-occurrence records can help researchers gain a deeper understanding of the literature [14]. Moreover, there is a potential in unsupervised learning algorithms to yield relevant clusters of papers and articles that researchers would have not identified given their constraints in time and attention. A considerable drawback with qualitative studies is that raw data (e.g. transcripts) are commonly not shared as part of the study and only anecdotal evidence is provided with the papers. Hence, no aggregation across papers may take place employing text mining technology. However, some qualitative synthesis methods could be explored in more detail to develop AI tools such as realist synthesis or meta-synthesis. While these two methods are fundamentally different from systematic reviews, they both aim to improve upon traditional narrative reviews by adopting explicit and rigorous processes for synthesis to achieve a greater level of understanding and reach a level of conceptual or theoretical development that goes beyond what has been achieved in a single empirical study [20], [28]. A further downside of semi-systematic reviews often exists in their lack of documentation regarding decisions made concerning the inclusion or exclusion of papers in the original sample and concerning data included in the analysis. Here, technical solutions may facilitate the documentation of decisions made, i.e. sources or papers included/excluded, filters applied, search strings, retrieval dates, or else.

As mentioned, the extracted data may involve both qualitative and quantitative sources. Hence, the analysis needs to be adjusted accordingly. In addition, the ontological and epistemological perspective of semi-systematic reviews is predominantly interpretivist, meaning that the multiplicity of research perspectives on a given issue is acknowledged and investigated. Consequently, the analysis may involve two stages if the review represents a meta-analysis of several research traditions' perspectives. In the analysis stage, each given perspective is analyzed separately, before the findings are synthesized in a second step [22]. The analysis works out the key characteristics of the narrative of a given perspective including core theories, preferred study designs, and key findings. Here, the analysis of quantitative data per tradition follows the same pattern as described for SLRs in that metrics and parameters are aggregated. The aggregation of qualitative data states a more challenging encounter concerning the application of

AI in the process as the analysis involves sensemaking and contextual understanding. Given the limitations of AI concerning interpretative tasks, this step can not be automated but supported. Data can be extracted as codes from the text using ACA tools that can understand syntax and the relationship of certain concept seeds [17], [18].

In the synthesis stage, the narratives of the perspectives stemming from the selected research fields are compared and contrasted. The focus lies on how groups have conceptualized the topic, how it has been theorized, the methodological approaches, and the derived findings. The latter are commonly seen as "higher-order" data that should be analyzed interpretatively by the researcher. The synthesis may occur on higher (i.e. commonalities, differences, tensions, patterns) or lower levels of abstraction. At lower levels, statistical aggregation and qualitative aggregation across perspectives bear potential for automation. Lastly, narrative summaries do presently not show potential for automation.

D. Structuring and Writing

At this stage, the SLR needs to be curated to fit the targeted audience which can include both academics and practitioners. Furthermore, the publishing medium needs to be considered as such types of reviews are suited for both journal and conference papers, which often have a page limit, as well as Ph.D. theses that require significant amounts of detail.

Very few AI-based tools have been developed so far for this step in either review type which can be explained through the creative nature of writing and the researchers' weaving in of professional opinions and interpretations. Furthermore, the purpose of our study is to help researchers apply these rigorous but time-consuming review processes into their studies through the assistance of AI-based tools to improve the overall quality of their research. We do not aim to describe these literature reviews as a study topic on their own. Hence, we do not go into too much detail for this phase.

IV. CONCLUSION & RECOMMENDATION

To make sense of the masses of published research, automation of literature reviews will be essential. For this, however, the entire spectrum of reviews needs to be covered. Helping both evidence-based and context-specific research in finding relevant research questions and validating conceived ideas, will effectively direct research efforts to relevant and promising areas and thereby improve the cost-efficiency of research spending. AI tools have to be designed according to the different objectives of reviews, in particular, positivist vs. interpretivist; quantitative vs. qualitative vs. mixed, and broad vs. narrow scoping. The comparison of systematic and semi-systematic reviews revealed similarities in process steps, yet differences concerning the processing of data corresponding to the underlying ontological assumptions. These lead to additional complications regarding automatability as creative and interpretative tasks need to be performed by researchers. We believe that end-to-end automation may only be possible for a subset of the needed reviews. Thus, we propose that systems are designed in a sequential manner involving man and machine, i.e. human-in-the-loop systems where steps are intermediated by human assessment and decision-making. With this paper we aim to draw attention to somewhat neglected disciplines that can greatly benefit from many of the technologies and processes in this growing field.

Due to the enormous potential of semi-automation, we intend to develop a tool that integrates available AI applications in a way that allows for a seamless process experience for researchers within the conceptual phase of any given research project. This will improve the rigor of research overall and bring more transparency to the review process in predominantly interpretivist sciences. The user journey should be customizable to the research's objectives and methodology. There will be a graphical user interface (GUI) that allows for use even without programming skills. In the long run, the aim is to further integrate along the research process to also support researchers in the planning, execution, and dissemination of the research (i.e. beyond the literature review). Additional features are conceivable such as collaborative tools, automatic summarization, highlighting of open problem spaces and new research ideas, as well as the evaluation of quality. Thinking about the application of AI in an inverse manner, these tools can also help evaluate the quality and novelty of work-in-progress papers, suggest potential enhancing concepts and frameworks, and propose suitable journals for publishing. The combination of human's creativity and criticality and the accuracy and thoroughness of AI can potentially reorient academic research and accelerate knowledge creation.

REFERENCES

[1]   M. Ware and M. Mabe, "The STM report: An overview of scientific and scholarly journal publishing," 2015.

[2]   S. Aviv-Reuven and A. Rosenfeld, "Publication patterns' changes due to the COVID-19 pandemic: A longitudinal and short-term scientometric analysis," Scientometrics, pp. 1–24, 2021.

[3]   F. Squazzoni, G. Bravo, F. Grimaldo, D. Garcıa-Costa, M. Farjam and B. Mehmani, "Only Second-Class Tickets for Women in the COVID-19 Race. A Study on Manuscript Submissions and Reviews in 2329" Elsevier Journals, 2020.

[4]   J. H. Elliott et al., "Living Systematic Reviews: An Emerging Opportunity to Narrow the Evidence-Practice Gap," *PLoS Medicine*, vol. 11, no. 2, p. e1001603, 2014.

[5]   R. van Dinter, B. Tekinerdogan, and C. Catal, "Automation of systematic literature reviews: A systematic literature review," Information and Software Technology, p. 106589, 2021.

[6]   J. Thomas et al., "Living systematic reviews: 2. Combining human and machine effort." Journal of clinical epidemiology, vol. 91, pp. 31-37, 2017.

[7]   H. Almasri, N. Zakuan, M. S. Amer and M. R. Majid, "A developed systematic literature review procedure with application in the field of digital transformation," Studies of Applied Economics, vol. 39, no. 4, 2021.

[8]   R. Whittemore and G. D. Melkus, "Designing a research study," The Diabetes Educator, vol. 34, no. 2, pp. 201–216, 2008.

[9]   H. Snyder, "Literature review as a research methodology: An overview and guidelines," Journal of business research, vol. 104, pp. 333–339, 2019.

[10]   W. Kusa, A. Hanbury and P. Knot, "Automation of Citation Screening for Systematic Literature Reviews using Neural Networks: A Replicability Study," arXiv preprint, 2201.07534, 2022.

[11]   G.G Chowdhury,"Natural language processing". Ann. Rev. Info. Sci. Tech., 37: 51-89, 2003.

[12]   I. Marshall & B. Wallace, "Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. Systematic Reviews.", 2019.

[13]   R. Kaur, J. A. Ginige and O. Obst, "A Systematic Literature Review of Automated ICD Coding and Classification Systems using Discharge Summaries," arXiv preprint, 2107.10652, 2021.

[14]   G. C. Nunez-Mir, B. V. Iannone III, B. C. Pijanowski, N. Kong, and S. Fei, "Automated content analysis: addressing the big literature challenge in ecology and evolution," Methods in Ecology and Evolution, vol. 7, no. 11, pp. 1262–1272, 2016.

[15]   H. He et al., "A Hybrid Approach to Semi-Automate the Screening Process for Living Systematic Reviews and Meta-Analysis," https://hehuan2112.github.io/static/file/amia21-all/screener-paper.pdf

[16]   L. Cheng, D. A. Katz-Rogozhnikov, K. R. Varshney and I. Baldini, "Automated Meta-Analysis: A Causal Learning Perspective," arXiv preprint, 2104.04633, 2021

[17]   A. Gupta, D. Chugh and R. Katarya, "Automated news summarization using transformers," In Sustainable Advanced Computing, pp. 249-259. Springer, Singapore, 2022.

[18]   C. M. Ormerod, A. Malhotra and A. Jafari, "Automated essay scoring using efficient transformer-based language models," arXiv preprint, 2102.13136, 2021.

[19]   P. Shetty and R. Ramprasad, "Automated knowledge extraction from polymer literature using natural language processing," Iscience, 24(1), 101922, 2021

[20]   D. Tranfield, D. Denyer, and P. Smart, "Towards a methodology for developing evidence-informed management knowledge by means of systematic review," British journal of management, vol. 14, no. 3, pp. 207–222, 2003.

[21]   D.J Cook, C.D Mulrow, R.B Haynes, "Systematic reviews: synthesis of best evidence for clinical decisions", Ann Intern Med, 1997.

[22]   G. Wong, T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson, "RAMESES publication standards: Meta-narrative reviews," Journal of Advanced Nursing, vol. 69, no. 5, pp. 987–1004, 2013.

[23]   V. R. Walker et al., "Evaluation of a semi-automated data extraction tool for public health literature-based reviews: Dextr," Environment international, vol.159, 107025, 2022.

[24]   Y. Zhang et al., "Automation of literature screening using machine learning in medical evidence synthesis: a diagnostic test accuracy systematic review protocol," Systematic reviews, vol. 11(1), pp. 1-7, 2022.

[25]   J. Marshall, J. Kuiper, B.C Wallace. "RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials". J Am Med Inform Assoc, 2016

[26]   B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[27]   G. Wagner, R. Lukyanenko and G. Pare, "Artificial intelligence and the conduct of literature reviews." Journal of Information Technology, 2021.

[28]   R. Campbell et al., "Evaluating meta-ethnography: a synthesis of qualitative research on lay experiences of diabetes and diabetes care," Social science & medicine, vol. 56, no. 4, pp. 671–684, 2003.