Beyond Automation: Decoding the Alignment of Artificial Intelligence Applications with Systematic Review Guidelines

Além da Automação: Decifrando o Alinhamento de Aplicações de Inteligência Artificial Com os Manuais de Revisões Sistemáticas

Más Allá de la Automatización: Descifrando la Alineación de las Aplicaciones de Inteligencia Artificial con las Guías de Revisiones Sistemáticas

RESUMO
Revisões Sistemáticas (RS) representam uma metodologia consolidada para a síntese de evidências científicas na área da saúde, sua condução exige rigor metodológico, preconizado pelos manuais JBI e Cochrane. Avanços tecnológicos, como a Inteligência Artificial (IA) foram integrados às RS, automatizando etapas e otimizando recursos. Este estudo identificou como as aplicações baseadas em IA utilizadas na elaboração de RS da área da saúde se alinham a estes manuais, avaliando 29 estudos que empregaram IA em diferentes etapas da RS. A análise revelou que 51,7% (15 estudos) atenderam aos manuais, enquanto os 48,3% (14 estudos) não atenderam. A etapa de Seleção (primeira triagem), representou 89,7% dos estudos (26 de 29). Enquanto etapas como formulação de estratégia de busca, avaliação de risco de viés e síntese de resultados não foram abordadas. Conclui-se que, para garantir a confiabilidade das RS apoiadas por IA, é necessário alinhar essas ferramentas às diretrizes metodológicas dos manuais, bem como de um esforço conjunto entre desenvolvedores de softwares e a comunidade científica.
DESCRITORES: Revisões sistemáticas como assunto, Metodologia – revisão sistemática, Armazenamento e recuperação da informação - saúde, Inteligência artificial, Aprendizado de máquina.

ABSTRACT
Systematic Reviews (SR) represent a well-established methodology for synthesizing scientific evidence in the healthcare field, and their conduct requires methodological rigor as outlined in the JBI and Cochrane manuals. Technological advances, such as Artificial Intelligence (AI), have been integrated into SRs, automating stages and optimizing resources. This study identified how AI-based applications used in the development of healthcare SRs align with these manuals, evaluating 29 studies that employed AI in different stages of SRs. The analysis revealed that 51.7% (15 studies) adhered to the manuals, while 48.3% (14 studies) did not. The Selection stage (first screening) represented 89.7% of the studies (26 out of 29). Stages such as search strategy formulation, risk of bias assessment, and results synthesis were not addressed. It is concluded that to ensure the reliability of AI-supported SRs, it is necessary to align these tools with the methodological guidelines of the manuals, as well as foster collaboration between software developers and the scientific community.
KEYWORDS: Systematic Reviews as Topic, Methodology – systematic review, Information Storage and Retrieval – health, Artificial Intelligence, Machine Learning.

RESUMEN
Las Revisiones Sistemáticas (RS) representan una metodología consolidada para la síntesis de evidencia científica en el área de la salud, y su realización requiere un rigor metodológico, según lo estipulado por los manuales JBI y Cochrane. Avances tecnológicos, como la Inteligencia Artificial (IA), han sido integrados a las RS, automatizando etapas y optimizando recursos. Este estudio identificó cómo las aplicaciones basadas en IA utilizadas en la elaboración de RS en el área de la salud se alinean con estos manuales, evaluando 29 estudios que emplearon IA en diferentes etapas de las RS. El análisis reveló que el 51,7% (15 estudios) cumplió con los manuales, mientras que el 48,3% (14 estudios) no lo hizo. La etapa de Selección (primer filtro) representó el 89,7% de los estudios (26 de 29). Mientras que etapas como la formulación de estrategias de búsqueda, la evaluación del riesgo de sesgo y la síntesis de resultados no fueron abordadas. Se concluye que, para garantizar la fiabilidad de las RS apoyadas por IA, es necesario alinear estas herramientas con las directrices metodológicas de los manuales, así como fomentar la colaboración entre los desarrolladores de software y la comunidad científica.
PALABRAS CLAVE: Revisiones sistemáticas como tema, Metodología – revisión sistemática, Almacenamiento y recuperación de información – salud, Inteligencia artificial, Aprendizaje automático.

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

RECEIVED: 01/27/2025 **APPROVED:** 02/05/2025

How to cite this article: Zanela M, Carvalho DR, Junior RM. Beyond Automation: Decoding the Alignment of Artificial Intelligence Applications with Systematic Review Guidelines. Saúde Coletiva (Edição Brasileira) [Internet]. 2025 [acesso ano mês dia];15(93):14761-14770. Disponível em: DOI: 10.36489/saudecoletiva.2025v15i93p14761-14770

Mariluci Zanela
Doctoral student of the Postgraduate Program in Information Management at the Federal University of Paraná.
ORCID: https://orcid.org/0009-0001-5436-5609

Deborah Ribeiro Carvalho
Professor of the Postgraduate Program in Information Management at the Federal University of Paraná.
ORCID: https://orcid.org/0000-0002-9735-650X

Ricardo Mendes Junior
Professor of the Postgraduate Program in Information Management at the Federal University of Paraná
ORCID: https://orcid.org/0000-0003-4947-0364

INTRODUCTION

Systematic Reviews (SR) in the health field are studies that use a transparent and impartial methodology to identify, evaluate and synthesize scientific evidence. [1,2] Defined and reproducible methods are followed, allowing the consolidation of previously evaluated studies, [3]

To ensure the quality and standardization of SR, methodological guidelines are used, such as the Joanna Briggs Institute Manual [4] and the Cochrane Handbook. [5] These manuals provide guidance for conducting evidence-based SRs, covering criteria such as feasibility, suitability, relevance and effectiveness of health interventions. [6]

However, performing an SR is an intensive and challenging process, which requires screening a large volume of studies, managing heterogeneous data and rigorous control to minimize biases, i.e., methodological errors in the selection and analysis of studies. [7]

AI-based tools are already being applied to SR development in stages such as initial screening, where pattern recognition algorithms help automatically classify large volumes of literature, speeding up the process and minimizing the risk of bias. This makes it possible to optimize critical steps in the process, reducing the workload of researchers, improving accuracy in study screening, and increasing consistency in data extraction. [8]

AI applications used in the development of RS include methods such as Machine Learning (supervised, unsupervised and reinforcement learning), deep neural networks (Deep Learning), algorithms such as Support Vector Machine (SVM) and Naïve Bayes, as well as advanced Natural Language Processing (NLP) technologies, such as Bidirectional Encoder Representations from Transformers (BERT) and generative models, such as Generative Pre-trained Transformer (GPT). [9]

Given this context, this study aims to identify how AI-based applications used in the preparation of systematic reviews in the health area align with the JBI and Cochrane manuals.

The justification for the research is the growing adoption of AI-based tools to support the conduct of SR. These tools, although capable of improving stages of the process, require compliance with criteria of transparency, systematicity, and reproducibility. Despite this, to date, no studies have been found that systematically evaluate the alignment of these applications with the guidelines of the JBI and Cochrane manuals, which highlights a critical gap in the scientific literature. Thus, this article seeks to fill this gap, contributing to an understanding of these technologies in the field of health SR.

METHOD

This research was conducted through an exploratory bibliographic review, aiming to identify, from published SRs, information to respond to the study objective. [10] The methodological process followed the steps of: defining a search strategy, selecting data sources, screening and extracting information and, finally, synthesizing and interpreting the data extracted from the selected studies.

The strategy used in the search was: ('artificial intelligence' OR 'machine intelligence' OR 'machine learning software' OR 'machine learning program' OR 'data processing' OR 'natural language processing' OR 'semi supervised machine learning' OR 'artificial intelligence chatbot' OR 'artificial intelligence software') AND ('systematic review (topic)' OR 'systematic reviews' OR 'study selection'

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

OR 'screening' OR 'screening method' OR 'citation analysis' OR 'methodological studies' OR 'methodological problems' OR 'methodological quality' OR 'method detection limit' OR 'qualitative research' OR 'qualitative study' OR 'search algorithm' OR 'search strategy').

The selected databases were: MEDLINE/PubMed (National Library of Medicine), Embase (Elsevier), BVS/LILACS (BIREME), CINAHL (EBSCOhost), Cochrane Library, Scopus (Elsevier), Web of Science (Clarivate Analytics), Preprints and OpenGrey.

SR articles that reported the use of AI applications in the preparation of SR in the health area were selected. In this study, the tools were defined as software, platforms, frameworks, APIs, Chatbots and virtual assistants, methods, heuristics or pre-trained AI models, automatic and/or semi-automatic. Studies that used AI-based software or methods only for data management or to assist in the diagnosis and/or treatment of diseases, as well as other types of medical documents, such as electronic medical records, were excluded.

The articles collected from the databases were stored in the Rayyan software. This software was used to remove duplicate references, select the studies included in this review, perform the initial screening, and record exclusions.

A spreadsheet structured in Microsoft Excel© was used to extract, discuss, and present the results. In this spreadsheet, aspects such as author, year of publication, application of AI, and SR stages were recorded, in accordance with the JBI and Cochrane Manuals.

The categorization of the SR stages also followed the methodological guidelines of the JBI and Cochrane manuals, which present similar guidelines, as shown in Table 1.

Table 1 - Systematic Review Steps

Step	Description
PLANNING	
Formulation of the SR Question	Define a clear and objective question using research acronyms (PICO, PICOT, PECOS)
Eligibility Criteria	Define inclusion and exclusion criteria to determine which texts will be included in the review. These criteria should be based on the research question and should consider factors such as type of study, population, intervention, outcome and publication period.
Search strategy	The search strategy should be planned using standardized health descriptors such as MeSH and Boolean operators to maximize the sensitivity and specificity of the search.
Databases	The choice of databases should be related to the research question, and can be general health databases such as MedLine, Embase, Lilacs, as well as specific databases such as Cinahl, Cochrane Library, PEDro, PsycInfo, as well as multidisciplinary databases such as Scopus, Web of Science and gray literature: Google Scholar, Open Grey
Protocol	Develop and register a detailed and transparent protocol, describing the methods and strategies to be used, register the protocol in one of the following platforms: PROSPERO, Cochrane Database of Systematics Reviews, JBI Evidence Synthesis
EXECUTION	
Location of Studies	Conduct searches in the databases systematically, simultaneously, and export the results to reference managers
Removal of Duplicate Records	Removal of duplicate references through reference managers
Selection of Studies	Study selection should occur in two screenings: 1st screening: reading of titles and abstracts; by 2 blind reviewers. 2nd screening: reading of the full text: by 2 blind reviewers. Both are conducted by independent reviewers and, in case of disagreement, a third reviewer is called, ensuring less influence of bias. Eligible works after the initial screening are analyzed in depth, and it is recommended to justify the exclusion of each study and keep a record of these decisions through a flowchart (PRISMA, 2020).
Quality Assessment	Establish methodological quality, reliability of studies, and assessment of risk of bias using instruments (Rob2, Robin-s; JBI Sumari).

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

Data Extraction	Define in advance the information that will be extracted from the articles, such as: type of study, location and period in which it was conducted, inclusion and exclusion criteria, follow-up time, number of participants, intervention, outcome presented (extraction table). This step must be carried out in pairs independently.
Statistical Analysis	It is recommended when data from more than one study are homogeneous, appropriate in relation to the methodology used and with their respective similar clinical outcomes.
Synthesis of Results	Present a qualitative or quantitative analysis (synthesis can be narrative, meta-analysis, graphical or tabular)
Quality of Evidence Assessment	Assess the certainty of the evidence for each outcome analyzed in the review using the Grading of Recommendations Assessment, Development and Evaluation - GRADE system
REPORT	
Report	Disclose the findings of the RS in a transparent manner, following the guidelines of the PRISMA 2020 Reporting Guide

Source: The authors, adapted from JBI[4], Cochrane[5].

To assess alignment with the recommendations of the JBI and Cochrane manuals in systematic reviews that used AI applications, the following criteria were established:

Meets: the AI application performs SR steps in alignment with the JBI and Cochrane manuals (Table 1);

Does not meet: the AI application performs SR steps in disagreement with the JBI and Cochrane manuals (Table 1).

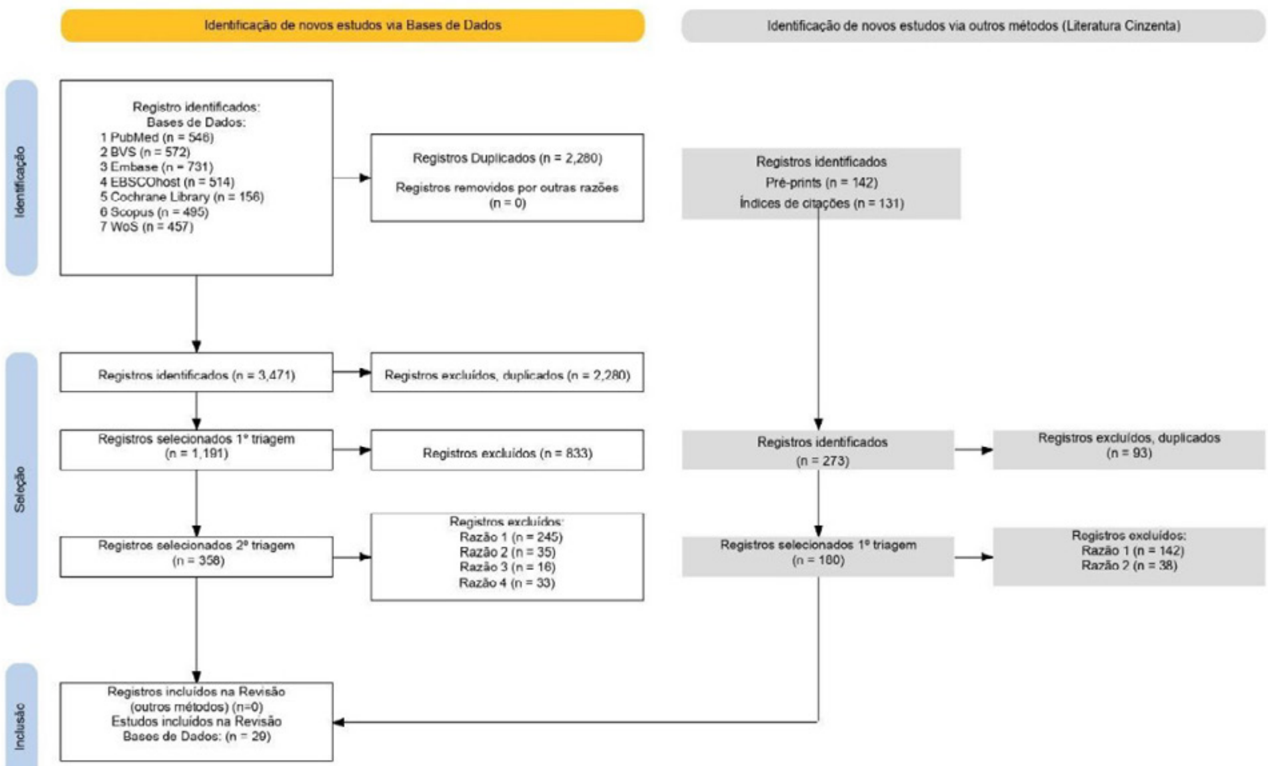
RESULTS

The searches were carried out simultaneously on June 22, 2024, in the databases and resulted in 3,471 studies. Of these, 2,280 were excluded because they were duplicate references, leaving 1,191 for the 1st screening. At this stage, 833 studies that were unrelated to the objective of the review were excluded. This left 358 articles for the 2nd screening. At this time, 309 studies were excluded for the following reasons: 245 studies in which AI was not used to prepare reviews (reason 1); 35 studies excluded because they did not use AI tools or methods (reason 2); 16 studies excluded because they were review studies or secondary studies (reason 3); 33 studies excluded from Randomized Clinical Trials, or intervention studies that were not part of this review (reason 4). Thus, a total of 29 studies were selected for data extraction.

Furthermore, searches in the gray literature resulted in 273 studies. After removing 93 duplicates, 180 studies remained for the first screening. Of these, 180 were excluded for the following reasons: 142 preprints (reason 1) and 38 did not meet the eligibility criteria of this review (reason 2). No studies from the gray literature were included in this article (Figure 1).

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

Figure 1 - PRISMA flowchart - study identification



SOURCE: The authors, adapted from Page et al.[1]

Twenty-nine studies published between 2012 and 2023 were analyzed. These studies evaluated AI applications at different stages of SR development in the health area. To identify the alignment of AI applications with the JBI and Cochrane manuals, the results were organized according to Table 1 – SR Stages. Each application was categorized as "Meets" or "Does not meet", as shown in Table 2:

Table 2 - Studies included in the review

Authors / Year	AI Applications	SR Step	Alignment with JBI Cochrane Manuals
Bekhuis, Demner-Fushman, 2012[11]	SVM; K-NN; NB; CNB; EvoSVM	Selection1st screening	Doesn't meet
Jonnalagadda, Petitti, 2013[12]	NB; SVM; FCNB	Selection1st screening	Doesn't meet
Kim, Choi, 2014[13]	SVM	Selection1st screening	Doesn't meet
Blake, Lucic, 2015[14]	GLM; SVM	Selection1st screening	Doesn't meet
Rathbone; Hoffmann; Glasziou, 2015[15]	Abstrackr	Selection1st screening	Meets
Hashimoto, Kontonatsios, Miwa, Ananiadou, 2016[16]	*Rede Neural*	Selection1st screening	Doesn't meet
Przybyła, Brockmeier, Kontonatsios, et al. 2018[17]	Robot Analyst	Seleção2ª triagem	Meets

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

Tsafnat, Glasziou, Karystianis, Coiera, 2018[18]	GATE	Selection1st screening	Meets
Bucheli Guerrero, 2019[19]	SASR	Selection1st screening	Doesn't meet
Gartlehner, Wagner, Lux, et al., 2019[20]	DistillerAI	Selection1st screening	Meets
Gates, Guitard, Pillay, et al. 2019[21]	Abstrackr; DistillerAI RobotAnalyst	Selection1st screening	Meets
Gates, Gates, Sebastianski, et al. 2020[22]	Abstrackr	Selection1st screening	Meets
Howard, Phillips, Tandon, et al. 2020[23]	SWIFT-Active Screener	Selection1st screening	Meets
Orgeolet, Foulquier, Misery, et al. 2020[24]	BIbliography BOT - BIBOT	Location of studies in databases	Doesn't meet
Popoff, Besada, Jansen, et al. 2020[25]	SVM; NB; CART	Selection1st screening	Meets
Burns, Etherington, Cheng-Boivin, Boet, 2021[26]	DistillerAI	Selection1st screening	Meets
Chai, Lines, Gucciardi, Ng, 2021[27]	Research Screener	Selection1st screening	Meets
Pham, Jovanovic, Bagheri,Antony, et al. 2021[28]	Mineração de texto; LDA; SVD; PLN	Selection1st screening	Meets
Qin, Liu, Wang, et al. 2021[29]	PNL; BERT; LightGBM	Selection1st screening	Meets
Borissov, Haas, Minder, et al. 2022[30]	Deduklick	Duplicate removal	Meets
Facchinetti, Benetti, Giuffrida, Nocera, 2022[31]	Slr-kit	Selection1st screening	Doesn't meet
Muller, Ames, Jardim, Rose, 2022[32]	Lingo3G	Selection1st screening	Meets
Reis, Oliveira, Fritsch, et al. 2023[33]	Rayyan; Abstrackr; Colandr	Selection1st screening	Meets
Kebede, Cornet, Fortner, 2023[34]	NB; SVM; SVD	Selection1st screening	Doesn't meet
Li, Kabouji, Bouhadoun, et al. 2023[35]	Rayyan; Abstrackr; SWIFT-Review	Selection1st screening	Doesn't meet
Natukunda, Muchene, 2023[36]	Latent Dirichlet Allocation	Selection1st screening	Doesn't meet
Oude Wolcherink, Pouwels, van Dijk, et al. 2023[37]	ASReview	Selection1st screening	Doesn't meet
Qureshi, Shaughnessy, Gill, et al. 2023[38]	ChatGPT	Question and Search Strategy	Doesn't meet
van Dijk Brusse-Keizer, Bucsán, et al. 2023[39]	ASReview	Selection1st screening	Meets

SOURCE: The authors, 2025.

The analysis of the 29 articles revealed that 51.7% (15 studies) complied with the JBI and Cochrane manuals in at least one stage of the SR, while 48.3% (14 studies) did not. The suitability of the applications varied considerably according to the stage of the SR in which they were used, highlighting differences in implementation and alignment with the manuals.

The majority of the studies analyzed applied AI tools in the Study Selection phase (1st Screening), representing 26 of the 29 studies (89.7%). Of these 26, half (13 applications) demonstrated alignment with the manuals, while the other half failed to meet the established methodological criteria. Among the studies that stood out positively are: Burns et al. [26], with the application of DistillerAI; and Rathbone et al. [15], with Abstrackr. Both studies met the criteria in the initial screening. In contrast, the studies by Bucheli Guerrero [19], with SASR; and Oude Wolcherink et al.[37], with ASReview, did not meet the manuals due to deficiencies in meeting methodological criteria, such as the lack of traceable justifications for study exclusions.

In the Duplicate Removal stage,

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

only one study was evaluated: Borissov et al.[30], with Deduklick. This demonstrated alignment with the manuals. For the Research Question Formulation and Search Strategy phase, the study by Qureshi et al. [38], with ChatGPT, was evaluated but did not meet the recommendations of the manuals.

The chronological analysis revealed a significant evolution in AI capabilities over the years. In the period from 2012 to 2016, the initial technologies highlighted in these studies, such as SVM (Support Vector Machine), a supervised learning algorithm designed to classify data, such as neural networks – computational structures inspired by the functioning of the human brain and capable of modeling complex patterns through interconnected processing layers – did not show alignment with the established methodological criteria. Between 2018 and 2021, an increase in the proportion of software aligned with the manuals was observed. Studies such as those by Przybyła et al., [17], with Robot Analyst, and Burns et al. [26], with DistillerAI stand out. Both met the methodological criteria in the stages, such as the 1st screening and the 2nd screening. In the most recent period, between 2022 and 2023, the study by Van Dijk et al. [39] stood out for the application of ASReview, which demonstrated compliance with the manuals. However, when the same software was used by Oude Wolcherink et al. [37], the lack of traceable justifications for study exclusions compromised their transparency and, consequently, their compliance with the manuals.

In general, the studies that stood out the most were those that used AI tools in operational steps, such as initial screening and removal of duplicates, showing alignment with the manuals. On the other hand, steps related to planning and formulating search strategies proved to be problematic, reflecting limitations both in the tools used and in the way they were implemented by the researchers.

DISCUSSION

The results presented in this study show that, although the use of AI-based applications in SR is advancing, only half of the studies evaluated meet the JBI and Cochrane methodological guidelines. This scenario reveals a contrast between the operational stages, in which the application of AI has proven effective, and the critical stages of planning and analysis, which still depend on human supervision.

The selection phase, specifically the first screening, was covered in 23 of the 29 studies evaluated, Rathbone et al.[15], with the application of the Abstrackr Software, Gartlehner et al. [20] with DistillerAI and Li et al.[35] with SWIFT-Review have demonstrated effectiveness in significantly reducing manual workload.

On the other hand, the study by Przybyła et al. [17], with the Robot Analyst software, it complied with the manuals, however, in the automated scenario, it excluded up to 70% of the relevant records, which may compromise the reliability of the SR results. The study by Gartlehner et al. [20], with the DistillerAI software, although it was efficient in reducing the workload (99%), presented a high loss of records (97%) in the automated scenario. These data reinforce that human supervision remains essential to ensure the quality and integrity of the selection process.

Programs and platforms that use advanced algorithms, such as ASReview, cited in the study by Van Dijk et al. [39], have shown efficiency in selecting the first screening by combining models such as Naïve Bayes and TF-IDF. In the study by Oude Wolcherink et al. [37] no clear criteria were presented for excluding studies, which makes it difficult to replicate and validate the results, compromising their application.

Deduplication was one of the most successful steps in the use of AI. Borissov et al. [30] presented Deduklick and Reis et al.[33] presented Rayyan, both of which ensured traceability and standardization, automatically and effectively identifying and eliminating duplicates. Deduklick [30] stood out for using similarity measures and generating reports in PRISMA format.

Despite the advances, the AI applications evaluated for automated database searches, such as BiBot studied by Orgeolet et al. [24],did not comply with the manuals. Limitations such as the absence of authorized health descriptors, such as MeSH, DeCS, and Boolean operators, compromised the sensitivity and specificity of the search strategies.

Steps such as search strategy formulation, bias risk assessment, and result synthesis were not addressed in the articles analyzed in this study. This reflects a significant gap in the application of AI in SR, since these steps require critical and contextual judgment, characteristics that current technologies still do not adequately replicate.

CONCLUSION

This article sought to identify how SRs that use AI-based tools in the health area align with the JBI and Cochrane manuals, which establish methodological guidelines to ensure transparency, reproducibility, and quality of reviews. The results indicated that, despite advances in AI applications, especially in operational steps such as initial screening and record deduplication, only 15 of the 29 studies met the manuals.

The applications that met the methodological guidelines of the manuals stood out for their traceability, standardization and transparency. An example of this was the work of Borissov et al. [30], with application of the Deduklick and by Pham et al. tool

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

[28], with text mining, which exemplify how AI can be successfully applied in operational stages of SR. On the other hand, studies such as Oude Wolcherink [37], with the ASReview, and Orgeolet [24], with BIBOT, presented flaws due to the failure to document the criteria adopted in the decisions made.

These findings show that the limitations observed are not intrinsically linked to AI applications, but to the way they are used by researchers. The lack of adherence to the guidelines of the JBI and Cochrane manuals compromises the reliability of SRs, even when advanced technological tools are used. The absence of standardized strategies, the limited use of the tools' functionalities, and the reliance on non-transparent decisions compromise the validity and reproducibility of the results. For AI-supported SRs to be consistent and useful in evidence-based practice, it is essential that the tools are compatible with the guidelines established by the manuals.

Therefore, the objective of this study was achieved. The results allowed us to identify that AI-based applications comply with the JBI and Cochrane manuals, but this alignment is still limited. This is due to both the limitations of the tools and their inadequate use by researchers.

As a contribution, this review highlights the need to raise awareness among researchers about the methodological guidelines of the manuals, as well as a joint effort between software developers and the scientific community. This would aim to promote robust and integrated solutions, improve the functionalities of AI applications, and train users for their methodologically correct application. These measures are considered essential for these tools to fulfill their role in strengthening SR, ensuring greater efficiency without compromising scientific quality.

This study was conducted based on published SRs, which restricts the findings to the AI-based applications described in these sources. The heterogeneity in objectives, methodologies, and data reported by the evaluated studies makes it difficult to generalize the results, since not all applications were tested under uniform conditions, which may have limited the detailed understanding of the technologies. Despite these limitations, this work provides an important basis for discussing the advances, challenges, and potential of the use of AI in SRs in the health area.

REFERENCES

1. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JMMJ, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;29(372). Disponível em: https://doi.org/10.1136/bmj.n71

2. Clarke M, Chalmers I. Reflections on the history of systematic reviews. BMJ Evid Based Med. 2018;23(4):121-122. Disponível em: https://doi.org/10.1136/bmjebm-2018-110968

3. Ahn HS, Kim HJ. An introduction to systematic review. J Korean Med Assoc. 2014;57(1):49-59. Disponível em: https://www.researchgate.net/publication/272396557_An_introduction_to_systematic_review

4. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, Editors. JBI Manual for Evidence Synthesis. JBI; 2024. Disponível em: https://synthesismanual.jbi.global

5. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., Editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.5, 2024. Disponível em; https://training.cochrane.org/handbook/current

6. Gusenbauer M, Haddaway NR. What every researcher should know about searching – Clarifying the fundamentals of literature searching. Res Synth Methods. 2020;11(3):136-147. Disponível em: https://doi.org/10.1002/jrsm.1427

7. Ioannidis JPA. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Quarterly. 2016;94(3):485-514. Disponível em: https://doi.org/10.1111/1468-0009

8. Blaizot A, Veettil SK, Saidoung P, Moreno-Garcia CF, Wiratunga N, Aceves-Martins M, Lai NM, Chaiyakunapruk N. Using artificial intelligence methods for systematic review in health sciences: a systematic review. Res Synth Methods. 2022;13(3):353-362. Disponível em: https://doi.org/10.1002/jrsm.1553

9. Tóth B, Berek I, Gulácsi I, et al. automation of

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

systematic reviews of biomedical literature: a scoping review of studies indexed in Pubmed. Syst Rev. 2024;13:174. Disponível em: https://doi.org/10.1186/s13643-024-02592-3.

10. Gil AC. Como elaborar projetos de pesquisa. 6. ed. São Paulo: Atlas; 2017.

11. Bekhuis T, Demner-Fushman D. Screening nonrandomized studies for medical systematic reviews: a comparative study of classifiers. Artif Intell Med. 2012;55(3):197-207. Disponível em: https://doi.org/10.1016/j.artmed.2012.05.002

12. Jonnalagadda S, Petitti D. A new iterative method to reduce workload in systematic review process. Int J Comput Biol Drug Des. 2013;6(1-2):5-17. Disponível em; https://doi.org/10.1504/ijcbdd.2013.052198

13. Kim S, Choi J. Improving the performance of text categorization models used for the selection of high quality articles. Healthc Inform Res. 2012;18(1):18-28. Disponível em: https://doi.org/10.4258/hir.2012.18.1.18

14. Blake C, Lucic A. Automatic endpoint detection to support the systematic review process. J Biomed Inform. 2015 Aug;56:42-56. Disponível em: https://doi.org/10.1016/j.jbi.2015.05.004

15. Rathbone J, Hoffmann T, Glasziou P. Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. Syst Rev. 2015;4:80. Disponível em: https://doi.org/10.1186/s13643-015-0067-6

16. Hashimoto K, Kontonatsios G, Miwa M, Ananiadou S. Topic detection using paragraph vectors to support active learning in systematic reviews. J Biomed Inform. 2016;62:59-65. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC4981645/

17. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam MA, McNaught J, von Elm E, Nolan K, Ananiadou S. Prioritising references for systematic reviews with RobotAnalyst: a user study. Res Synth Methods. 2018;9(3):470-488. Disponível em: https://doi.org/10.1002/jrsm.1311

18. Tsafnat G, Glasziou P, Karystianis G, Coiera E. Automated screening of research studies for systematic reviews using study characteristics. Syst Rev. 2018 Apr 25;7(1):64. Disponível em: https://doi.org/10.1186/s13643-018-0724-7

19. Bucheli Guerrero VA. Desarrollo del estado del arte en investigación: una herramienta basada en inteligencia artificial. Rev Politécnica. 2019;15(30):70-81. Disponível em; http://dx.doi.org/10.33571/rpolitec.v15n30a7

20. Gartlehner G, Wagner G, Lux L, Affengruber L, Dobrescu A, Kaminski-Hartenthaler A, Viswanathan M. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: a user study. Syst Rev. 2019;8(1):277. Disponível em: https://doi.org/10.1186/s13643-019-1221-3

21. Gates A, Guitard S, Pillay J, Elliott SA, Dyson MP, Newton AS, Hartling L. Performance and usability of machine learning for screening in systematic reviews: a comparative evaluation of three tools. Syst Rev. 2019;8(1):278. Disponível em; https://doi.org/10.1186/s13643-019-1222-2

22. Gates A, Gates M, Sebastianski M, Guitard S, Elliott SA, Hartling L. The semi-automation of title and abstract screening: a retrospective exploration of ways to leverage Abstrackr's relevance predictions in systematic and rapid reviews. BMC Med Res Methodol. 2020;20(1):139. Disponível em: https://doi.org/10.1186/s12874-020-01031-w

23. Howard BE, Phillips J, Tandon A, Maharana A, Elmore R, Mav D, Sedykh A, Thayer K, et al. SWIFT-Active Screener: accelerated document screening through active learning and integrated recall estimation, Environ Int. 2020;138:105623. Disponível em: https://doi.org/10.1016/j.envint.2020.105623

24. Orgeolet L, Foulquier N, Misery L, Redou P, Pers JO, Devauchelle-Pensec V, Saraux A. Can artificial intelligence replace manual search for systematic literature? Review on cutaneous manifestations in primary Sjögren's syndrome. Rheumatology (Oxford). 2020;59(4):811-819. Disponível em; https://doi.org/10.1093/rheumatology/kez370

25. Popoff E, Besada M, Jansen JP, Cope S, Kanters S. Aligning text mining and machine learning algorithms with best practices for study selection in systematic literature reviews. Syst Rev.

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770

2020 Dec 13;9(1):293. Disponível em: https://doi.org/10.1186/s13643-020-01520-5

26. Burns JK, Etherington C, Cheng-Boivin O, Boet S. Using an artificial intelligence tool can be as accurate as human assessors in level one screening for a systematic review. Health Info Libr J. 2024;41(2):136-148. Disponível em: https://doi.org/10.1111/hir.12413 Epub 2021. PMID: 34792285.

27. Chai KEK, Lines RLJ, Gucciardi DF, Ng L. Research Screener: a machine learning tool to semi-automate abstract screening for systematic reviews. Syst Rev. 2021;10(1):93. Disponível em: https://doi.org/10.1186/s13643-021-01635-3

28. Pham B, Jovanovic J, Bagheri E, Antony J, Ashoor H, Nguyen TT, Rios P, Robson R, Thomas SM, Watt J, Straus SE, Tricco AC. Text mining to support abstract screening for knowledge syntheses: a semi-automated workflow. Syst Rev. 2021;10(1):156. Disponível em: https://doi.org/10.1186/s13643-021-01700-x .

29. Qin X, Liu J, Wang Y, Liu Y, Deng K, Ma Y, Zou K, Li L, Sun X. Natural language processing was effective in assisting rapid title and abstract screening when updating systematic reviews. J Clin Epidemiol. 2021;133:121-129. Disponível em: https://doi.org/10.1016/j.jclinepi.2021.01.010

30. Borissov N, Haas Q, Minder B, Kopp-Heim D, von Gernler M, Janka H, Teodoro D, Amini P. Reducing systematic review burden using Deduklick: a novel, automated, reliable, and explainable deduplication algorithm to foster medical research. Syst Rev. 2022;11(1):172. Disponível em: https://doi.org/10.1186/s13643-022-02045-9 .

31. Facchinetti T, Benetti G, Giuffrida D, Nocera A. Slr-kit: a semi-supervised machine learning framework for systematic literature reviews. Knowledge-Based Systems. 2022;251:109266. Disponível em: https://doi.org/10.1016/j.knosys.2022.109266.

32. Muller AE, Ames HMR, Jardim PSJ, Rose CJ. Machine learning in systematic reviews: Comparing automated text clustering with Lingo3G and human researcher categorization in a rapid review. Res Synth Methods. 2022 Mar;13(2):229-241. Disponível em: https://doi.org/10.1002/jrsm.1541 .

33. Reis AHS, de Oliveira ALM, Fritsch C, Zouch J, Ferreira P, Polese JC. Usefulness of machine learning softwares to screen titles of systematic reviews: a methodological study. Syst Rev. 2023;12(1):68. Disponível em: https://doi.org/10.1186/s13643-023-02231-3 .

34. Kebede MM, Le Cornet C, Fortner RT. In-depth evaluation of machine learning methods for semi-automating article screening in a systematic review of mechanistic literature. Res Synth Methods. 2023 Mar;14(2):156-172. Dispon[ivel em: https://doi.org/10.1002/jrsm.1589 .

35. Li J, Kabouji J, Bouhadoun S, Tanveer S, Filion KB, Gore G, Josephson CB, Kwon CS, et al. Sensitivity and specificity of alternative screening methods for systematic reviews using text mining tools. J Clin Epidemiol. 2023;162:72-80. Disponível em: https://doi.org/10.1016/j.jclinepi.2023.07.010 .

36. Natukunda A, Muchene LK. Unsupervised title and abstract screening for systematic review: a retrospective case-study using topic modelling methodology. Syst Rev. 2023 Jan 3;12(1):1. Disponível em: https://doi.org/10.1186/s13643-022-02163-4 .

37. Oude Wolcherink MJ, Pouwels XGLV, van Dijk SHB, Doggen CJM, Koffijberg H. Can artificial intelligence separate the wheat from the chaff in systematic reviews of health economic articles? Expert Rev Pharmacoecon Outcomes Res. 2023;23(9):1049-1056. Disponível em: https://doi.org/10.1080/14737167.2023.2234639 .

38. Qureshi R, Shaughnessy D, Gill KAR, Robinson KA, Li T, Agai E. Are ChatGPT and large language models "the answer" to bringing us closer to systematic review automation? Syst Rev. 2023 Apr 29;12(1):72. Disponível em: https://doi.org/10.1186/s13643-023-02243-z .

39. Van Dijk SHB, Brusse-Keizer MGJ, Bucsán CC, Van der Palen J, Doggen CJM, Lenferink A. Artificial intelligence in systematic reviews: promising when appropriately used. BMJ Open. 2023;13:e072254. Disponível em: https://doi.org/10.1136/bmjopen-2023-072254 .

DOI: https://doi.org/10.36489/saudecoletiva.2025v15i93p14761-14770