ORIGINAL ARTICLE

OBJN
Online Brazilian Journal of Nursing
ISSN: 1676 - 4285 | www.objnursing.uff.br

Submission: 28-May-2026
Approved: 26-Aug-2026

STUDY SELECTION ASSISTED BY CLAUDE AI ARTIFICIAL INTELLIGENCE: A METHODOLOGICAL GUIDE

SELEÇÃO DE ESTUDOS ASSISTIDA PELA INTELIGÊNCIA ARTIFICIAL *CLAUDE AI*: UM GUIA METODOLÓGICO

SELECCIÓN DE ESTUDIOS ASISTIDA POR LA INTELIGENCIA ARTIFICIAL CLAUDE AI: UNA GUÍA METODOLÓGICA

Rômulo Severo Sampaio[1] ORCID: 0000-0002-0657-2646
Yasmim Ribeiro Fracaroli[2] ORCID: 0009-0004-2400-2323
Anna Paula Mendes Marques de Lima Franco[2] ORCID: 0009-0009-8847-7914
Marianne Alves Rodrigues[2] ORCID: 0009-0001-6504-9089
Francini Castilha do Nascimento[2] ORCID: 0000-0001-7860-0493
Cecília Santos de Lima[2] ORCID: 0009-0004-3706-2368
Isabelle Cristinne Pinto Sampaio Costa[2] ORCID: 0000-0002-2611-8643

[1] Universidade Professor Edson Antônio Velano. Alfenas, Minas Gerais, Brazil.
[2] Universidade Federal de Alfenas (UNIFAL-MG). Alfenas, Minas Gerais, Brazil.

How to cite: Sampaio RS, Fracaroli YR, Franco APMML, Rodrigues MA, Nascimento FC, Lima CS, et al. Study selection assisted by Claude AI artificial intelligence: a methodological guide. Online Braz J Nurs. 2026;25(Suppl1):e20267039. https://doi.org/10.17665/1676-4285.20267039

RESUMO

Objetivo: Apresentar um guia metodológico para a seleção de estudos em revisões de literatura com o auxílio do *Claude AI*. **Método:** Estudo de reflexão teórico-metodológica, fundamentado em revisão narrativa da literatura e na experiência empírica dos autores na condução de sínteses de evidências. **Resultados:** Propõe-se um fluxo metodológico estruturado em cinco etapas: (1) exportação e deduplicação inicial no *EndNote*; (2) deduplicação adicional e fragmentação do corpus em blocos anuais no *Rayyan*; (3) criação de um projeto no *Claude AI*, incluindo os critérios de elegibilidade; (4) triagem de títulos e resumos assistida por inteligência artificial (IA), com geração automatizada de planilha contendo a decisão e a justificativa para cada registro; e (5) retorno ao *Rayyan* para revisão humana independente e posterior integração ao fluxograma PRISMA. Esse processo configura um modelo teórico-operacional que fornece recomendações práticas para a definição dos critérios de elegibilidade, elaboração estruturada de prompts e registro transparente das decisões. **Conclusão:** O guia estabelece uma estrutura inovadora para a incorporação da IA na etapa de seleção de estudos em revisões de literatura. Essa abordagem acelera as fases iniciais da síntese de evidências e orienta o uso responsável da tecnologia. Como limitação, aponta-se a ausência de validação empírica multicêntrica do fluxo proposto, o que demanda sua aplicação e avaliação em pesquisas futuras.

Descritores: Inteligência Artificial; Inteligência Artificial Generativa; Prática Clínica Baseada em Evidências; Metodologia como Assunto; Pesquisa; Revisão Sistemática.

ABSTRACT

Objective: To present a methodological guide for study selection in literature reviews assisted by Claude AI. **Method:** A theoretical-methodological reflection study grounded in a narrative literature review and the authors' empirical experience in conducting evidence syntheses. **Results:** A five-stage structured methodological workflow is proposed: (1) export and initial deduplication in EndNote; (2) additional deduplication and corpus segmentation into annual blocks in Rayyan; (3) creation of a project in Claude AI incorporating the eligibility criteria; (4) AI-assisted title and abstract screening with automated generation of a spreadsheet containing the decision and justification for each record; and (5) return to Rayyan for independent human review and subsequent integration into the PRISMA flow diagram. This process establishes a theoretical-operational model that provides practical recommendations for defining eligibility criteria, developing structured prompts, and transparently documenting decisions. **Conclusion:** The guide establishes an innovative framework for incorporating AI into the study selection phase of literature reviews. This approach accelerates the initial stages of evidence synthesis and promotes the responsible use of technology. A noted limitation is the lack of multicenter empirical validation of the proposed workflow, which warrants its application and evaluation in future research.

Descriptors: Artificial Intelligence; Generative Artificial Intelligence; Evidence-Based Practice; Methodology as a Subject; Research; Systematic Review.

RESUMEN

Objetivo: Presentar una guía metodológica para la selección de estudios en revisiones de literatura con el auxilio de Claude AI. **Método:** Estudio de reflexión teórico-metodológica, fundamentado en una revisión narrativa de la literatura y en la experiencia empírica de los autores en la conducción de síntesis de evidencias. **Resultados:** Se propone un flujo metodológico estructurado en cinco etapas: (1) exportación y deduplicación inicial en EndNote; (2) deduplicación adicional y fragmentación del corpus en bloques anuales en Rayyan; (3) creación de un proyecto en Claude AI, incluyendo los criterios de elegibilidad; (4) tamizaje de títulos y resúmenes asistido por inteligencia artificial (IA), con generación automatizada de una hoja de cálculo que contiene la decisión y la justificación para cada registro; y (5) retorno a Rayyan para la revisión humana independiente y su posterior integración al diagrama de flujo PRISMA. Este proceso configura un modelo teórico-operacional que proporciona recomendaciones prácticas para la definición de los criterios de elegibilidad, la elaboración estructurada de prompts y el registro transparente de las decisiones. **Conclusión:** La guía establece una estructura innovadora para la incorporación de la IA en la etapa de selección de estudios en revisiones de literatura. Este enfoque acelera las fases iniciales de la síntesis de evidencias y orienta el uso responsable de la tecnología. Como limitación, se señala la ausencia de validación empírica multicéntrica del flujo propuesto, lo que demanda su aplicación y evaluación en futuras investigaciones.

Descriptores: Inteligencia Artificial; Inteligencia Artificial Generativa; Práctica Clínica Basada en la Evidencia; Metodología como un Tema; Investigación; Revisión Sistemática.

Editors:
Rosimere Ferreira Santana (ORCID: 0000-0002-4593-3715)
Geilsa Soraia Cavalcanti Valente (ORCID: 0000-0003-4488-4912)
Emerson Willian Santos de Almeida (ORCID: 0000-0002-6846-021X)

Publisher:
Escola de Enfermagem Aurora de Afonso Costa – UFF
Rua Dr. Celestino, 74 – Centro, CEP: 24020-091 – Niterói, RJ, Brazil
Journal email: objn.cme@id.uff.br

Corresponding author:
Yasmim Ribeiro Fracaroli
E-mail: yasmim.fracaroli@sou.unifal-mg.edu.br

What is already known:

- Traditional screening methods in literature reviews require substantial time investment and present operational limitations in the evidence synthesis process.
- Generative AI has the potential to accelerate title and abstract screening in health research.
- Methodological gaps remain regarding the standardization, transparency, and validation of AI use in systematic reviews.

What this article adds:

- The use of Claude AI can reduce the time spent on initial study screening for health literature reviews.
- The proposed methodological workflow enhances the transparency and standardization of decisions during AI-assisted study selection.
- Integrating Claude AI, Rayyan, EndNote, and PRISMA promotes greater traceability and organization in the evidence synthesis process.

INTRODUCTION

Review studies are central instruments in Evidence-Based Practice (EBP), as they enable the compilation and critical appraisal of available knowledge on a given topic. Consequently, they provide qualified support for clinical decision-making and health policy formulation. Such studies reliably synthesize knowledge, facilitate the identification of gaps within the investigated field, and guide the formulation of new research questions[1-3].

Screening articles by reading titles and abstracts represents one of the most exhausting and time-consuming stages of a literature review, directly impacting the validity and reproducibility of findings. Estimates indicate that a conventional systematic review requires, on average, six to 18 months of work. The study selection phase accounts for a substantial portion of this timeframe, rendering it particularly vulnerable to human fatigue and interpretive drift among reviewers[4,5].

Reference managers, such as EndNote, and platforms dedicated to collaborative screening, such as Rayyan, have become established as indispensable resources in review workflows. Although these technologies automate the deduplication and organization of the corpus of manuscripts retrieved from searches, eligibility decisions remain centered on human reading, interpretation, and cognition, requiring the individual assessment of each record[6,7].

Artificial Intelligence (AI) has transformed review studies by automating stages such as literature searching, screening, and data synthesis, thereby conferring greater speed and precision to the methodological process[8]. More recently, Large Language Models (LLMs)—systems capable of comprehending, generating, and structuring complex text—have stood out, significantly expanding technological support options for study selection in these investigations[9].

Models such as ChatGPT (OpenAI), Gemini (Google), and Claude AI (Anthropic) have demonstrated promising performance in preliminary evaluations of title and abstract screening. Within specific domains, these tools have achieved sensitivity and specificity comparable to those of human reviewers[10,11]. Nevertheless, human validation remains imperative to ensure methodological quality, mitigate bias, and guarantee contextualized interpretations[8,12].

Among available technologies, Claude AI exhibits features that make it uniquely attractive for health-related reviews. The tool offers an extensive context window (up to 200,000 tokens), supports the creation of projects with persistent attached files, and processes large batches of bibliographic records within a single interaction. Furthermore, it generates auditable, natural-language justifications for each article inclusion or exclusion decision[13]. These capabilities enable the maintenance of consistent eligibility criteria throughout screening and confer traceability to the process. Such attributes are fundamental to the transparency advocated by the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines, as well as the Cochrane Handbook and JBI manuals[14-16].

Despite recent advances, a gap persists in the literature regarding the lack of systematized methodological guidance for integrating LLMs into the study selection stage in a structured, critical, and ethical manner. Current scientific output tends to focus on comparing performance between AI and human reviewers or on theoretical discussions regarding the risks of automation. Consequently, there is a lack of a practical guide linking intermediate steps—such as reference management, deduplication, corpus segmentation, and tool configuration—to the final study eligibility decision[16-18].

Against this background, this study aims to present a methodological guide for study selection in literature reviews assisted by generative AI—specifically the Claude AI model—with a focus on optimizing evidence synthesis in healthcare.

METHOD

This study proposes a methodological guide grounded in theoretical-methodological reflection. The process was conducted through a narrative literature review on the use of generative AI in evidence synthesis, combined with the authors' experience in conducting different types of reviews (systematic, integrative, and scoping). The development of the guide was informed by the international PRISMA guidelines[14], the Cochrane Handbook[15], and the JBI Manual for Evidence Synthesis[16], while also adopting recognized precepts for developing guidance manuals in healthcare[19].

The reflection took place in a structured manner across five stages: (1) retrieval and critical analysis of methodological publications and empirical studies on the application of LLMs in study screening[8,12,17,18]; (2) mapping the functionalities of these tools against PRISMA guideline requirements for title and abstract screening; (3) comparative analysis between the methodological recommendations of international reference organizations (such as Cochrane and JBI) and the operational capabilities and limitations of Claude AI; (4) systematization of the authors' practical experience in the iterative application of the proposed workflow to identify opportunities, methodological risks, and requirements for responsible tool use; and (5) consolidation

of the identified stages into a methodological workflow, culminating in the proposed guide and operational flowchart presented in the Results section.

The literature search was conducted across the following information sources: Medical Literature Analysis and Retrieval System Online (PubMed/MEDLINE), Scopus, Web of Science Core Collection (WoS), and Google Scholar. A timeframe from 2020 to 2025 was established, justified by the recent commercial release of LLMs. The descriptors "artificial intelligence", "large language model", "systematic review", "screening", "title and abstract", and "Claude AI", along with their Portuguese equivalents, were combined using the Boolean operators AND and OR. Studies published in English, Portuguese, or Spanish, with no restrictions on methodological design, that addressed the use of AI in evidence synthesis were included. Articles lacking a clear methodological description or falling outside the thematic scope were excluded.

Because this is a reflective, non-systematic narrative review, prior protocol registration on platforms such as the International Prospective Register of Systematic Reviews (PROSPERO) or the Open Science Framework (OSF) was waived, in accordance with the nature of the study design[17].

Literature screening occurred in two phases: reading of titles and abstracts, followed by full-text assessment of potentially relevant records. Data analysis was conducted interpretively, focusing on identifying strategies, limitations, and recommendations for using AI in article selection. Priority was given to experiences applicable to the healthcare context, with an emphasis on methodological proposals and empirical validation studies.

The authors' empirical experience was systematically incorporated through practical application and critical reflection on the use of Claude AI while conducting a systematic review focused on taxane desensitization protocols in women with breast cancer. This research is currently ongoing and duly registered in PROSPERO (CRD420261325732). Selecting this study as an application scenario stemmed from its synergy with the research group's line of inquiry, enabling the testing and refinement of the guide within a real scientific production setting. During testing, annual blocks of bibliographic records—previously deduplicated in EndNote and Rayyan—were processed by uploading the eligibility criteria as persistent context files within a dedicated Claude AI project. The research team critically evaluated the decisions and justifications generated by the model, facilitating the iterative refinement of the method documented herein.

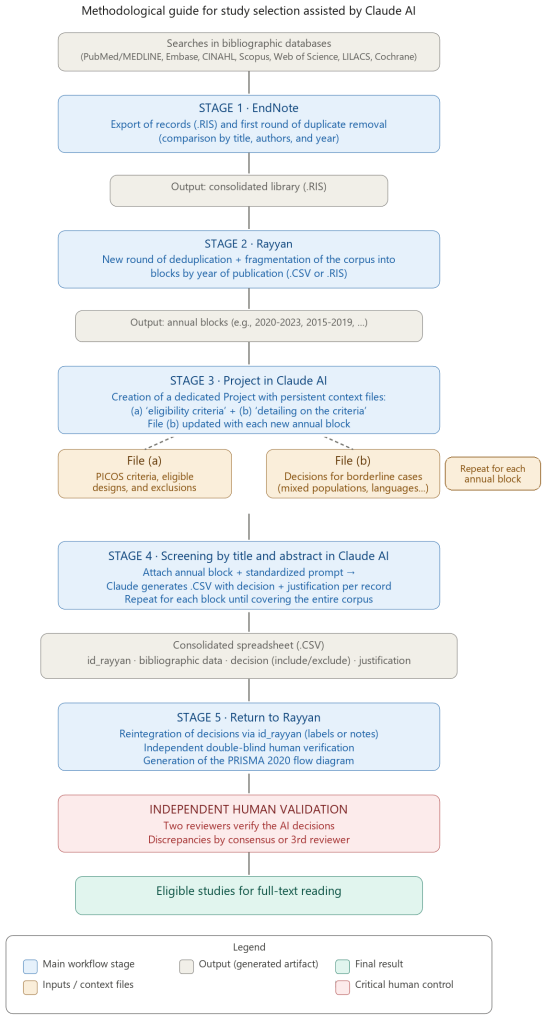
Regarding data security and privacy, the workflow was executed using the Claude Pro version (claude.ai interface). The "Help Improve Claude" privacy setting was disabled to prevent prompts and attached files from being used to train future models by Anthropic[20]. Information supplied to the tool was restricted to the bibliographic metadata of the records (identifier, title, authors, year, journal, DOI, abstract, and keywords). No full-text articles, unpublished data, or other sensitive materials were submitted.

The resulting methodological proposal was structured into five sequential stages, described in the Results section and illustrated in an operational flowchart. Given its theoretical-methodological nature without direct involvement of human subjects, ethical review by a Research Ethics Committee was waived, in full compliance with Law No. 14.874/2024 and Resolutions No. 466/2012 and No. 510/2016 of the Brazilian National Health Council.

RESULTS

The proposed methodological guide is structured into five sequential stages, integrating traditional reference management and collaborative screening tools (EndNote and Rayyan) with Claude AI as an assistive resource for title and abstract eligibility decisions. Figure 1 presents the schematic representation of the workflow.

Figure 1 – Operational flowchart of the methodological guide for study selection in literature reviews assisted by Claude AI. Alfenas, MG, Brazil, 2026



Source: developed by the authors, 2026.

Stage 1 – Export from information sources and initial deduplication in EndNote

Following the execution of search strategies across the selected information sources (PubMed/MEDLINE, Embase, LILACS, Scopus, Web of Science, and the Cochrane Library), all retrieved records must be exported in a format compatible with the EndNote reference manager. Prioritizing files with the .RIS extension or the native format of

each database is recommended[5].

In EndNote, records retrieved from the different sources are consolidated into a single library, where the first round of duplicate removal is performed using the "Find Duplicates" function. Configuring the software to compare title, authors, and publication year is suggested. Furthermore, manual review of the dataset prior to permanent deletion is imperative, as typographical and formatting variations—such as the presence or absence of the Digital Object Identifier (DOI) and differing author citation formats—can lead to false negatives in the automated algorithm[5,15]. The output of this stage consists of a consolidated and partially deduplicated library, which should be re-exported in .RIS format for the subsequent phase.

Stage 2 – Further deduplication and segmentation in Rayyan

Subsequently, the library is imported into Rayyan, an intelligent platform widely used for article screening in reviews. The tool offers a free version with restricted features and paid plans with advanced capabilities[6]. Rayyan features a proprietary algorithm for detecting and removing duplicates that demonstrates higher sensitivity than EndNote across various scenarios, justifying a second round of deduplication at this stage[6,15].

Once deduplication is complete, the consolidated corpus is segmented into smaller blocks organized by publication year. Annual segmentation provides relevant operational advantages: (i) it facilitates temporal tracking of publications and auditing of the screening process; (ii) it generates file sizes compatible with the Claude AI context window, preventing data truncation; and (iii) it enables simultaneous work among reviewers or partial resumption in the event of interruptions. Each block is exported in .CSV or .RIS format and must contain, at a minimum, the following fields: unique identifier, title, authors, year, journal, DOI, abstract, and keywords.

Stage 3 – Project creation in Claude AI and configuration of eligibility criteria

The "Projects" feature in Claude AI enables the creation of a dedicated workspace in which context files can be persistently attached and remain accessible across all interactions linked to that environment[11]. This functionality forms the core of the proposed strategy, as it transforms the eligibility criteria into a fixed decision-making reference, ensuring consistency throughout the evaluation of different document blocks.

Creating a specific Project for each review and instructing it with two complementary context documents is recommended: a) Eligibility criteria document (narrative version): containing the elements typically described in the review protocol, such as population, intervention, comparison, outcomes, eligible study designs, and exclusion criteria. b) Operational detailing document: explicitly stating the interpretive decisions for handling borderline cases. This may include, for example, the approach to studies with mixed populations, the precise definition of the intervention, the handling of systematic reviews as secondary sources, the policy regarding publication languages, and other conventions agreed upon by the team.

Separating the eligibility criteria document from the operational detailing document—a strategy successfully adopted in the authors' experience—offers two central advantages. First, it preserves the integrity of the original protocol (usually registered on platforms such as PROSPERO or OSF). Second, it allows the refinement and versioning of operational rules throughout screening without altering the formally registered document. This confers flexibility to the process while safeguarding traceability. Table 1 summarizes the minimum elements recommended for compiling these two documents.

Table 1 – Minimum recommended elements for configuring eligibility criteria in the Claude AI Project. Alfenas, MG, Brazil, 2026

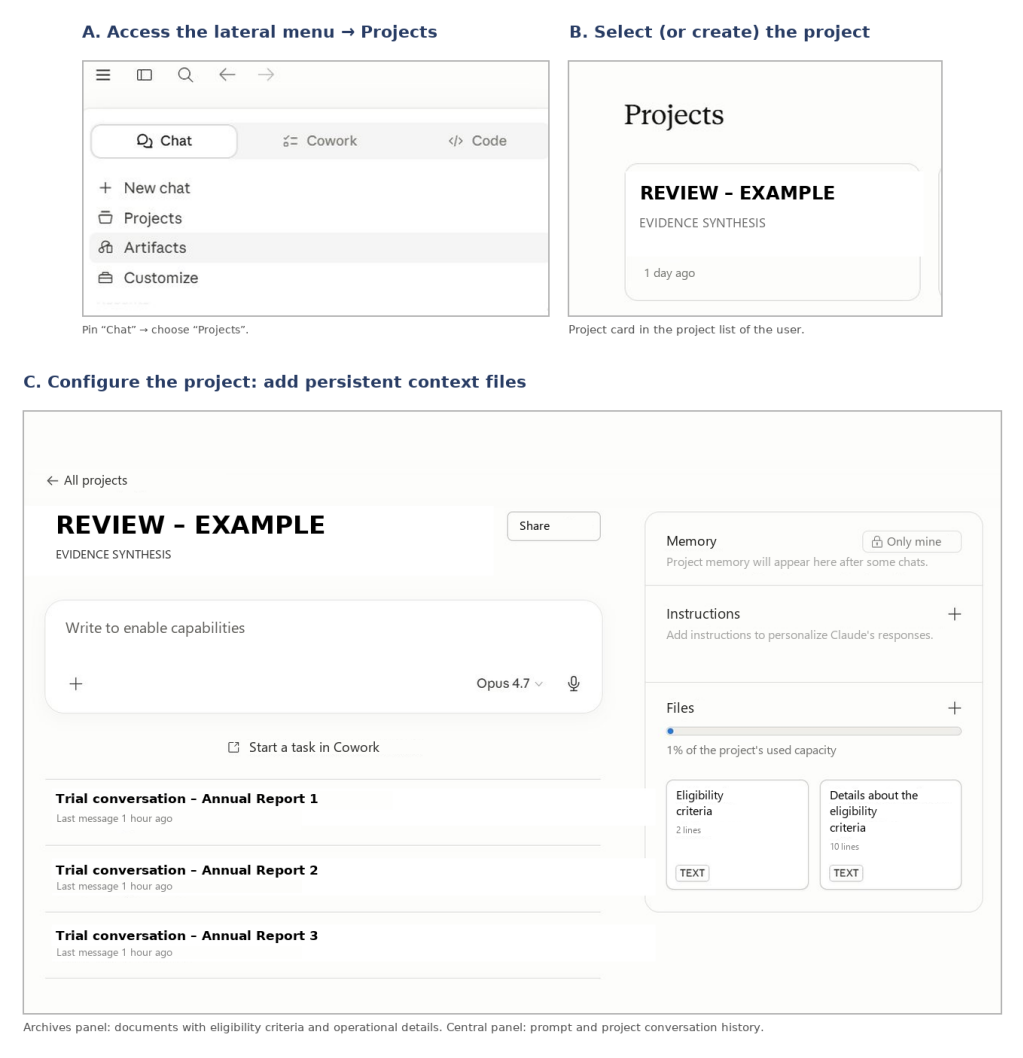
Element	Expected content
Population	Clear definition of participants, specifying age, sex, clinical condition, and required diagnostic criteria.
Intervention	Precise description of the intervention, including its operational definition (e.g., minimum number of stages, doses, routes of administration).
Comparison	Comparison group or reference standard adopted (when applicable).
Outcomes	List of primary and secondary outcomes considered eligible.
Eligible designs	Accepted study types (RCT, NRCT, cohort, case-control, etc.) and explicitly excluded designs.
Exclusion criteria	Situations warranting immediate exclusion, such as out-of-scope age ranges, presence of other pathologies, and editorial formats.
Languages and period	Policy on accepted languages and search timeframe.
Borderline cases	Decision rules for mixed populations, separately extractable data, prior reviews, studies without abstracts, and other ambiguous scenarios.

Source: developed by the authors, 2026.

Step-by-step guide for creating the Project in Claude AI

To support the operationalization of this stage, the sequence of procedures for creating and configuring the Project in Claude AI is described below and illustrated in Figure 2.

Figure 2 – Step-by-step guide for creating and configuring a Project in Claude AI aimed at assisted screening in literature reviews. Alfenas, MG, Brazil, 2026



Source: developed by the authors, 2026.

Step 1 - Access the side menu and select "Projects": After registering or logging into the interface (https://claude.ai/new), the user should open the left-hand navigation menu and select the "Projects" option (Figure 2 - Panel A).

Step 2 - Create and name a new Project: On the listing screen, select the option to create a new project. Using an easily identifiable naming convention (e.g., the review's acronym or short title) and adding a brief description specifying the type of synthesis (systematic, integrative, scoping review, etc.) is recommended. This standardization optimizes searchability and teamwork organization (Figure 2 - Panel B).

Step 3 - Attach persistent context files: In the right-hand panel of the interface, under the "Files" section, the two guiding screening documents must be attached: the eligibility criteria and the operational detailing. These files remain available across all interactions, ensuring the consistent application of the same parameters to every annual block evaluated (Figure 2 - Panel C). Additionally, the interface allows entering general guidelines in the "Instructions" field to direct the model's behavior. The "Memory" panel automatically records relevant information accumulated across interactions, strengthening the traceability of the method.

Stage 4 – AI-assisted title and abstract screening in Claude AI

Once the Project is configured, screening execution begins. Each annual block of records is attached individually to a new conversation within the Project. The entered prompt must explicitly request: (i) reading the titles and abstracts of all items in the file; (ii) applying the previously attached eligibility criteria and operational detailing; and (iii) generating a spreadsheet (.CSV) containing the record identifier, essential bibliographic metadata, the verdict (include for full-text reading or exclude), and the corresponding textual justification.

The individualized justification per record is the central element of this method, as it acts as an auditing and accountability mechanism. Unlike systems that provide only binary classifications, Claude AI can articulate in natural language the exact criterion underlying the decision (e.g.,

"narrative review on taxanes without description of a formal desensitization protocol" or "preclinical study conducted in cell culture"). This transparency enables subsequent human verification and mitigates disagreements.

Downloading and archiving the spreadsheets generated for each block is recommended. These serve as artifacts of the selection process and become part of the research's auditable repository. Table 2 presents a prompt template based on the authors' experience and iteratively refined throughout testing.

Table 2 – Prompt template used for study screening in Claude AI. Alfenas, MG, Brazil, 2026

Main request:
"Screen the studies available in the attached file named '[block_name].csv' based on reading the titles and abstracts, in accordance with the eligibility criteria established and available in the project files named 'eligibility criteria' and 'operational detailing of eligibility criteria'.
Expected output:
Produce a new .csv file containing, for each analyzed record, the following columns: (1) screening_id — sequential identifier assigned during screening; (2) rayyan_id (or key) — original identifier of the record in Rayyan; (3) study title; (4) publication year; (5) journal; (6) volume, issue, and pages; (7) authors; (8) DOI; (9) language; (10) abstract; (11) keywords; (12) PubMed ID (when available); (13) screening_decision — 'Include for full text' or 'Exclude'; (14) justification — textual explanation grounded in the specific eligibility or exclusion criterion that motivated the decision.
Additional instructions:
Adopt a conservative approach: in case of doubt regarding whether the criteria are met, mark the record as 'Include for full text' and justify the uncertainty. For records without an abstract, base the decision on the title and keywords, indicating this limitation in the justification. Maintain 8-bit Unicode Transformation Format (UTF-8) (with BOM) encoding and a comma separator in the generated .csv file, enclosing all fields in quotation marks to preserve content containing commas and semicolons".

Source: developed by the authors, 2026.

Stage 5 – Return to Rayyan for verification and integration into PRISMA

Following the completion of screening by Claude AI, reintegrating the decisions into the Rayyan environment is imperative. This procedure allows for: (i) the execution of independent human verification; (ii) the consolidation of all data within a single auditable environment; and (iii) the automated generation of the PRISMA flow diagram, encompassing precise counts of inclusions and exclusions at each methodological phase.

Operationally, the annual spreadsheets generated by the AI must be compiled into a single master spreadsheet. The rayyan_id column must be retained as the primary key, as it enables the exact localization of each originally imported record. Rigorous verification of these identifiers prior to re-importing into the platform is recommended to resolve any potential duplicates or encoding errors.

In Rayyan, AI decisions can be applied by assigning custom labels (e.g., 'AI-include' and 'AI-exclude'). This approach preserves the tool's verdict in an informational layer distinct from that generated by the researchers. Another plausible strategy is inserting the model-generated justifications into the "notes" field of each record, optimizing comparative analysis. The choice of method should weigh the volume of the corpus, the size of the review team, and the preferences of the research group.

Independent human validation—conducted by two reviewers in a double-blind format within Rayyan—must evaluate the entirety of the decisions issued by the AI. Disagreements between the model and a researcher, or among the researchers themselves, must be resolved by consensus or through the assessment of a third reviewer, aligning with the precepts of the Cochrane Handbook and JBI[15,16]. This methodological rigor establishes the AI in the role of an automated first reviewer while maintaining indispensable human oversight, which guarantees the quality of the synthesized evidence.

Once validation is finalized, Rayyan enables the quantitative export of identified, deduplicated, included, and excluded studies, providing the inputs required to construct the PRISMA flow diagram. Finally, the final review report is recommended to explicitly document the use of generative AI in the title and abstract screening phase. A methodological note detailing the adopted validation procedures must be included, in strict compliance with reporting guidelines for the use of artificial intelligence in review studies[21].

Transparency and registration of the application experience

The systematic review used as a practical scenario for testing the proposed workflow is registered in the PROSPERO database (CRD420261325732) and is pub-

licly available. The complete quantitative results of this screening—including the number of records identified, deduplicated, screened by Claude AI, and confirmed during independent human verification, along with the corresponding PRISMA flow diagram—will be detailed in the study's primary publication. This review is currently underway by the same research group. The decision not to present the quantitative data at this time aims to prevent duplicate data publication and preserve the originality of the evidence synthesis for its main article.

It should be emphasized that the central objective of this guide is to describe and substantiate the developed operational workflow, rather than to report the findings of a specific systematic review. Such results will be published in due course in a dedicated manuscript, with appropriate cross-citation to this methodological study.

Given the growing adoption of AI tools to support study selection and screening, Table 3 presents a comparative analysis of manual screening, the use of Claude AI, and the application of other technologies. The table highlights the main differences regarding operational, contextual, and functional aspects.

Table 3 – Comparison among manual screening, Claude AI, and other AI tools in supporting study selection. Alfenas, MG, Brazil, 2026

Aspect	Manual screening	Claude AI	Other AIs (e.g., ChatGPT, Gemini, ASReview)
Average time per block (~400 records)	8 to 15 hours per reviewer.	Minutes to a few hours, depending on file size.	Variable. Tools with active learning (such as ASReview) progressively reduce effort.
Consistency across batches	Susceptible to fatigue and interpretive drift over time.	High, provided context files remain unchanged in the Project.	Depends on session memory. ChatGPT and Gemini require reconfiguration in each new conversation.
Justification per record	When recorded, it is often brief.	Automatically generated in natural language and linked to eligibility criteria.	ASReview outputs a probability score; ChatGPT and Gemini generate free text.
Context window	Not applicable.	Extensive; supports large batches of records without truncation.	ChatGPT and Gemini feature competitive windows; ASReview does not use LLMs.
Persistent criteria	The reviewer consults the protocol document.	Context files remain pinned to the Project environment.	Requires manually pasting or re-attaching criteria in each new interaction.
Cost	Researcher working time.	Monthly subscription; direct use via the Claude.ai interface.	Variable. ASReview is free; other commercial solutions charge per usage or subscription.
Auditability	High, provided reviewers record detailed decisions.	High (generates a standardized spreadsheet containing the verdict and justification for each record).	Variable, depending on the architecture and features of the tool used.

Source: developed by the authors, 2026.

DISCUSSION

The present study did not aim to develop a new AI technology, nor to configure an autonomous agent capable of operating without human oversight. Claude AI was used in its native form through prompts manually directed by the researchers within a dedicated Project, acting as an assisted language model rather than an independent execution agent. The innovation proposed in this study therefore lies in the systematized articulation of a replicable operational workflow. This workflow integrates established tools (EndNote and Rayyan) with the structured use of commercial LLMs, grounded in an original set of practical recommendations. These recommendations include: separating formal eligibility criteria from operational detailing for borderline cases; standardizing prompts with an auditable output structure; using unique identifiers ('rayyan_id') as a traceability mechanism across virtual environments; and applying an independent human verification protocol following AI screening.

Unlike previous studies—which have focused on comparing performance between AI and human reviewers or on theoretical discussions regarding the risks of automation[8,12,18]—this guide offers a theoretical-operational model empirically tested in an ongoing systematic review. Consequently, it transforms the ad hoc use of language models into a transparent, auditable process aligned with international evidence synthesis guidelines. Furthermore, the workflow fulfills specific recommendations for the transparent reporting of AI use as a methodological tool in literature reviews[21].

In this regard, the proposed method should be analyzed through three interdependent lenses: as an operational efficiency strategy, as a methodological intervention with inherent risks, and, fundamentally, as a driver of transformative impact on scientific practice in healthcare. Incorporating AI into the study selection stage reconfigures the methodological work of reviews by shifting operational effort to computational systems. This frees up greater cognitive investment from reviewers for inference, interpretation, and critical appraisal of analytical decisions. However, this transformation introduces new technical dependencies—such as the chosen model, its version, and its parameters—as well as new sources of variation that require rigorous governance to maintain traceability and scientific accountability[9].

Potential of Claude AI in optimizing study selection

In terms of efficiency, Claude AI demonstrates high potential by shifting the operational burden of title and abstract screening to computational systems. Recent studies quantify this advantage by showing that LLM-based tools can achieve precision and recall comparable to those of human reviewers, while significantly reducing the time required for manuscript selection. Comparative analyses demonstrate F1-scores above 90% in screening tasks, reducing evaluation time from days to hours when analyzing hundreds of bibliographic records[10,22].

This optimization allows human resources to be reallocated to stages of the research requiring higher cognitive demand, such as critical appraisal of findings and qualitative synthesis. Additionally, Claude AI's extensive context window and its Projects feature with persistent files enable the processing of large batches of records without artificial fragmentation of the workflow, representing a relevant operational advantage over other AI tools[13].

Given the growing adoption of these technologies to support article selection and screening, Table 3 (presented in the Results section) offers a comparative analysis between Claude AI and other systems applicable to the same purpose, highlighting differences in operational, contextual, and functional aspects. The synthesized data demonstrate that, although platforms such as ASReview offer specific advantages in cost and progressive active learning, Claude AI stands out due to the combination of three factors: an extensive context window, persistence of criteria throughout the entire screening process, and generation of natural-language justifications linked to each decision. Together, these attributes substantially enhance the auditability of the methodological process.

Ethical and methodological challenges

The methodological impact of AI on the validity and reliability of study selection warrants caution. Although automation reduces human errors stemming from fatigue or inconsistency, it introduces systemic risks that require rigorous management. Algorithmic biases—notably selection, measurement, and confounding biases—stand out, as they can shift the center of gravity of the synthesis toward scenarios more compatible with the model's training corpus. Such deviation can generate asymmetries and conceptually fragile classifications[23-25].

In the context of article selection, selection bias may manifest when models trained on databases predominantly featuring certain languages, study designs, or geographic regions underperform when evaluating research from underrepresented contexts, thereby compromising the equity of the synthesis[23]. Furthermore, the phenomenon of "hallucinations" (the generation of plausible yet factually incorrect information) represents a unique epistemological risk. For records lacking abstracts or containing minimally informative abstracts, the AI may produce seemingly well-substantiated decisions based on inferences that cannot be verified in the source text[22,26].

The occurrence of false positives (undue inclusion of ineligible studies) and false negatives (undue exclusion of eligible studies) constitutes a central concern in any AI-assisted screening strategy, as it directly impacts the completeness and validity of the final synthesis. In the practical experience underpinning this guide, this risk was mitigated through two approaches: explicitly instructing the model to adopt a conservative stance when facing uncertainty (prioritizing inclusion for full-text reading over premature exclusion) and conducting independent human verification of all generated decisions. This approach shifts the predominant risk toward false positives, which are considered methodologically less harmful at this stage since they will be rectified during full-text reading. False negatives, conversely, would irreversibly compromise the scope of the synthesis.

The reliability of decisions, in turn, depends intrinsically on the quality and stability of the prompts used. Generic or inconsistent prompts across different runs can introduce undue variability into screening, even when eligibility criteria remain unchanged. Standardizing the prompt and establishing it as an auditable component of the process, as proposed in this guide, represent not merely good operational practices, but essential methodological requirements to ensure the reproducibility of AI-assisted screening[21].

To mitigate these risks within the established workflow, incorporating containment mechanisms is recommended. Key measures include explicitly instructing the model to restrict its inferences to data contained within the title and abstract; requiring a textual justification tied to a specific eligibility criterion; and performing cross-verification by independent human reviewers. Transparency and reproducibility are non-negotiable pillars. Therefore, the scientific community must adopt open science practices by meticulously documenting the AI model, its version, the prompt, and the context files utilized[23,27].

Security, privacy, and transparency in AI use

The use of generative AI tools in research processes raises pressing considerations regarding data security and confidentiality, particularly when handling metadata from unpublished scientific output or ongoing review protocols. During the practical application of this guide, disabling the "Help Improve Claude" setting in the Claude Pro version served as an essential safeguard. The objective was to prevent bibliographic records and eligibility criteria from being incorporated into the training data for future versions of the tool[20].

It should be noted that disabling this setting only prevents conversations from being used for algorithm training; it does not eliminate data storage on Anthropic's servers during the retention period stipulated in its privacy policy. Moreover, the consumer version does not offer the guarantees provided by enterprise plans (such as Claude for Work), which feature strict data protection agreements, or by Application Programming Interface (API) implementations configured with Zero Data Retention. This distinction is crucial in reviews involving sensitive data, vulnerable populations, or unpublished information. In such scenarios, adopting environments with robust contractual confidentiality safeguards is substantially more appropriate.

In the methodological workflow proposed herein, restricting data input to non-confidential bibliographic metadata (title, abstract, authors, year, DOI, and keywords) reduces the exposure of confidential information. However, it does not entirely eliminate the risk, as abstracts of ongoing reviews may occasionally reveal strategic directions of yet-unpublished research.

Researchers are therefore advised to rigorously assess the nature and sensitivity of their data prior to submission to the tool. It is imperative to observe the current privacy policy—which is subject to updates—and evaluate the suitability of the subscribed plan to the nature of the study, adopting a level of contractual protection commensurate with the processed information[20].

Study limitations

Despite its relevance, this study has methodological limitations. The primary limitation lies in the absence of multicenter quantitative validation of the proposed workflow. Practical application focused on a single thematic domain, precluding accurate estimation of the method's performance under relevant contextual variations, such as different fields of knowledge, specialized terminologies, distinct publication languages, and varied synthesis typologies. Furthermore, the effectiveness of Claude AI is inherently dependent on the specific model and version employed, whose future updates may alter outcomes. This dynamic nature imposes an imperative need for granular technical documentation to ensure methodological reproducibility[28].

Implications for healthcare research and practice

For healthcare research—a field characterized by an exponential volume of publications and an urgent need for rapid responses to guide clinical practice and public policy—adopting automated workflows represents a strategic advancement. However, researchers must assume the role of critical curators of information. Claude AI acts as a catalyst for the operational stage, whereas the professional's clinical and methodological expertise is preserved for interpreting complex contexts, capturing the nuances of care, and evaluating the applicability of evidence within health services (with particular emphasis on the Brazilian Unified Health System in the local context).

Within this new technological landscape, human validation transcends mere quality control to become an organic part of the methodological process. The reviewer's role is reconfigured: rather than performing exhaustive duplicate screening of the entire corpus, their attention converges on reconciling discrepancies between AI decisions and initial human assessments. Cognitive effort is thus directed toward critical appraisal and contextualized interpretation of the evidence[22].

Additionally, the process of drafting and configuring eligibility criteria within the Project fosters deep reflection on the review protocol. Frequently, this step uncovers operational ambiguities that would otherwise emerge only during the course of screening. This attribute endows the method with a substantial pedagogical dimension, proving particularly valuable in the continuing education of researchers in evidence synthesis.

Future directions

Future investigations should focus on developing specific methodological guidelines for using AI in study selection, establishing minimum reporting standards (specification of the model, version, prompt, parameters, and audit trails). Conducting multicenter external validation studies is also imperative. Such research should benchmark AI performance against human reviewers across heterogeneous thematic domains, utilizing precise sensitivity and specificity metrics while also evaluating the tool's concrete impact on final synthesis conclusions[28,29]. Solidifying these guidelines will establish the necessary foundations for the responsible, ethical, and auditable use of generative AI in healthcare evidence production.

CONCLUSION

Incorporating Claude AI—a LLM developed by Anthropic—into the study selection stage of literature reviews modernizes the evidence synthesis process by automating one of the most time-consuming phases of scientific production: title and abstract screening. This theoretical-methodological study presents a five-stage structured workflow proposal to guide the responsible use of this technology. The approach integrates scientific rigor with Natural Language Processing (NLP) capabilities, demonstrating that automation can be safely adopted by establishing clear protocols, indispensable human oversight, and continuous validation.

Despite Claude AI's high capacity to interpret scientific language and apply eligibility criteria—when configured as persistent context within a dedicated Project—its use demands caution due to inherent risks of algorithmic biases and "hallucinations". This caveat is particularly critical in reviews intended to inform clinical decisions or shape health policies. Therefore, ethical and responsible use of the tool requires ensuring transparency, traceability, and reproducibility across all methodological stages.

In terms of practical implications, researchers are recommended to: (1) configure the Project in Claude AI with persistent context files encompassing the eligibility criteria and their corresponding operational detailing to ensure analytical consistency across screening blocks; (2) utilize structured, iteratively validated prompts, explicitly requiring a textual justification for each evaluated record; (3) perform prior deduplication in EndNote and Rayyan, segmenting the bibliographic corpus into annual blocks to optimize processing and enable auditing; and (4) maintain independent duplicate human validation of the decisions generated by Artificial Intelligence, resolving any discrepancies by consensus or through the assessment of a third reviewer.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

USE OF ARTIFICIAL INTELLIGENCE

Claude AI was used as methodological and operational support during the preparation of the article, including assistance in organizing the content and generating the images presented in the manuscript. ChatGPT was used as an auxiliary tool for spell checking, improving text fluency, and refining scientific writing. Furthermore, all methodological decisions, analyses, interpretations, data organization, and scientific content of the study were fully carried out and validated by the authors.

REFERENCES

1. Farhat N, Tsaioun K, Saunders-Hastings P, Morgan RL, Ramoju S, Hartung T, et al. Systematic review in evidence-based risk assessment. ALTEX. 2022;39(3):463-79. https://doi.org/10.14573/altex.2004111. PMID: 34585732.
2. Wu Q, Nabi G. Systematic reviews and data analyses: guide to clinical practice and research. Scott Med J. 2023;68(3):69-71. https://doi.org/10.1177/00369330231193888. PMID: 37605529.
3. Canto GDL. Revisões sistemáticas da literatura: guia prático. Curitiba: Brazil Publishing; 2020.
4. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? PLoS Med. 2010;7(9):e1000326. https://doi.org/10.1371/journal.pmed.1000326. PMID: 20877712.
5. Borah R, Brown AW, Capers PL, Kaiser KA. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. BMJ Open. 2017;7(2):e012545. https://doi.org/10.1136/bmjopen-2016-012545. PMID: 28242767.
6. Bramer WM, Giustini D, Kramer BMR, Anderson PF. The comparative recall of Google Scholar versus PubMed in identical searches for biomedical systematic reviews: a review of searches used in systematic reviews. Syst Rev. 2013;2:115. https://doi.org/10.1186/2046-4053-2-115. PMID: 24360284.
7. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan - a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4. PMID: 27919275.
8. Ge L, Agrawal R, Singer M, Kannapiran P, Molina JADC, Teow KL, et al. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. Syst Rev. 2024;13(1):269. https://doi.org/10.1186/s13643-024-02682-2. PMID: 39456077.
9. Maity S, Saikia MJ. Large language models in healthcare and medical applications: a review. Bioengineering (Basel). 2025;12(6):631. https://doi.org/10.3390/bioengineering12060631. PMID: 40564447.
10. Guo E, Gupta M, Deng J, Park YJ, Paget M, Naugler C. Automated paper screening for clinical reviews using large language models: data analysis study. J Med Internet Res. 2024;26:e48996. https://doi.org/10.2196/48996. PMID: 38214966.
11. Syriani E, David I, Kumar G. Screening articles for systematic reviews with ChatGPT. J Comput Lang. 2024;80:101287. https://doi.org/10.1016/j.cola.2024.101287
12. 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(7):e072254. https://doi.org/10.1136/bmjopen-2023-072254. PMID: 37419641.
13. Claude [Internet]. San Francisco: Anthropic; 2026 [cited 2026 May 18]. Available from: https://claude.ai/
14. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. https://doi.org/10.1136/bmj.n71. PMID: 33782057.
15. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 6.5 (updated August 2024). London: Cochrane; 2024 [cited 2026 May 18]. Available from: https://www.cochrane.org/handbook
16. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, editors. JBI Manual for Evidence Synthesis [Internet]. Adelaide: JBI; 2024 [cited 2026 May 18]. Available from: https://synthesismanual.jbi.global/. https://doi.org/10.46658/JBIMES-24-01
17. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs. 2005;52(5):546-53. https://doi.org/10.1111/j.1365-2648.2005.03621.x. PMID: 16268861.
18. Wang S, Scells H, Koopman B, Zuccon G. Can ChatGPT write a good Boolean query for systematic review literature search? [Internet]. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '23); 2023 Jul 23-27; Taipei, Taiwan. New York (NY): ACM; 2023 [cited 2026 May 18]. p. 1426-36. Available from: https://dl.acm.org/doi/10.1145/3539618.3591703. https://doi.org/10.1145/3539618.3591703
19. Echer IC. Elaboração de manuais de orientação para o cuidado em saúde. Rev Lat Am Enfermagem. 2005;13(5):754-7. https://doi.org/10.1590/S0104-11692005000500022. PMID: 16308635.
20. How do I change my model improvement privacy settings? [Internet]. San Francisco: Anthropic; 2026 [cited 2026 Aug 23]. Available from: https://privacy.claude.com/en/articles/12109829-how-do-i-change-my-model-improvement-privacy-settings
21. Holst D, Moenck K, Koch J, Schmedemann O, Schüppstuhl T. Transparent reporting of AI in systematic literature reviews: development of the PRISMA-trAIce checklist. JMIR AI. 2025;4:e80247. https://doi.org/10.2196/80247. PMID: 41370833.
22. Khraisha Q, Put S, Kappenberg J, Warraitch A, Had-

field K. Can large language models replace humans in systematic reviews? Evaluating GPT-4's efficacy in screening and extracting data from peer-reviewed and grey literature in multiple languages. Res Synth Methods. 2024;15(4):616-26. https://doi.org/10.1002/jrsm.1715. PMID: 38484744.

23. Ueda D, Kakinuma T, Fujita S, Kamagata K, Fushimi Y, Ito R, et al. Fairness of artificial intelligence in healthcare: review and recommendations. Jpn J Radiol. 2024;42(1):3-15. https://doi.org/10.1007/s11604-023-01474-3. PMID: 37540463.
24. Nazer LH, Zatarah R, Waldrip S, Chen Ke JX, Moukheiber M, Khanna AK, et al. Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digit Health. 2023;2(6):e0000278. https://doi.org/10.1371/journal.pdig.0000278. PMID: 37347721.
25. Chen F, Wang L, Hong J, Jiang J, Zhou L. Unmasking bias in artificial intelligence: a systematic review of bias detection and mitigation strategies in electronic health record-based models. J Am Med Inform Assoc. 2024;31(5):1172-83. https://doi.org/10.1093/jamia/ocae060. PMID: 38520723.
26. Dhuliawala S, Komeili M, Xu J, Raileanu R, Li X, Celikyilmaz A, et al. Chain-of-verification reduces hallucination in large language models [Internet]. In: Findings of the Association for Computational Linguistics: ACL 2024. Stroudsburg: Association for Computational Linguistics; 2024 [cited 2026 Aug 23]. p. 3563-78. Available from: https://aclanthology.org/2024.findings-acl.212/. https://doi.org/10.18653/v1/2024.findings-acl.212
27. Ioannidis JP. Transparency, bias, and reproducibility across science: a meta-research view. J Clin Invest. 2024;134(22):e181923. https://doi.org/10.1172/JCI181923. PMID: 39545412.
28. Siontis GCM, Sweda R, Noseworthy PA, Friedman PA, Siontis KC, Patel CJ. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. BMJ Health Care Inform. 2021;28(1):e100466. https://doi.org/10.1136/bmjhci-2021-100466. PMID: 34969668.
29. Yuan H. Toward real-world deployment of machine learning for health care: external validation, continual monitoring, and randomized clinical trials. Health Care Sci. 2024;3(5):360-64. https://doi.org/10.1002/hcs2.114. PMID: 39479276.

AUTHORSHIP CONTRIBUTIONS

Study conception: Sampaio RS, Fracaroli YR, Franco APMML, Rodrigues MA, Nascimento FC, Lima CS, Costa ICPS.
Data acquisition: Sampaio RS, Fracaroli YR, Franco APMML, Rodrigues MA, Nascimento FC, Lima CS, Costa ICPS.
Data analysis: Sampaio RS, Fracaroli YR, Franco APMML, Rodrigues MA, Nascimento FC, Lima CS, Costa ICPS.
Data interpretation: Sampaio RS, Fracaroli YR, Franco APMML, Rodrigues MA, Nascimento FC, Lima CS, Costa ICPS.
All authors are responsible for the textual writing and critical review of the intellectual content, for the final published version, and for all ethical, legal, and scientific aspects related to the accuracy and integrity of the study.

© 2026 Authors. The authors retain copyright and grant the Online Brazilian Journal of Nursing the right of first publication. This article is licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0).