RTI (h)(s) Health Solutions

SA112

Machine Learning and Artificial Intelligence for Supporting Systematic Reviews: A Systematic Review of Recent Methodological Developments and Recommendations for Implementation

José S. Marcano-Belisario, PhD; Michaela Lunan-Taylor, PhD; Sathushan Thurairajah, BSc; Emma Hawe, MSc

RTI Health Solutions, Manchester, United Kingdom

BACKGROUND

Literature reviews (systematic, targeted, or narrative) are a cornerstone of health economics and outcomes research and evidence-based decision-making. However, the volume of scientific data is ever increasing, making systematic literature reviews (SLRs) more time and resource intensive. Machine learning (ML) and artificial intelligence (AI) are being hailed as the answer to streamlining literature reviews through reduced timelines and lower consumption of research resources. Yet, these methods need evaluation to ensure continued robustness and appropriate integration into existing workflows.

OBJECTIVE

- To identify recent evidence on the use, performance, and implementation of ML/AI to support systematic reviews
- To summarise key takeaways messages in the existing evidence base

METHODS

An SLR was performed to identify recent studies reporting on ML/AI as a tool for conducting reviews as described in Table 1.

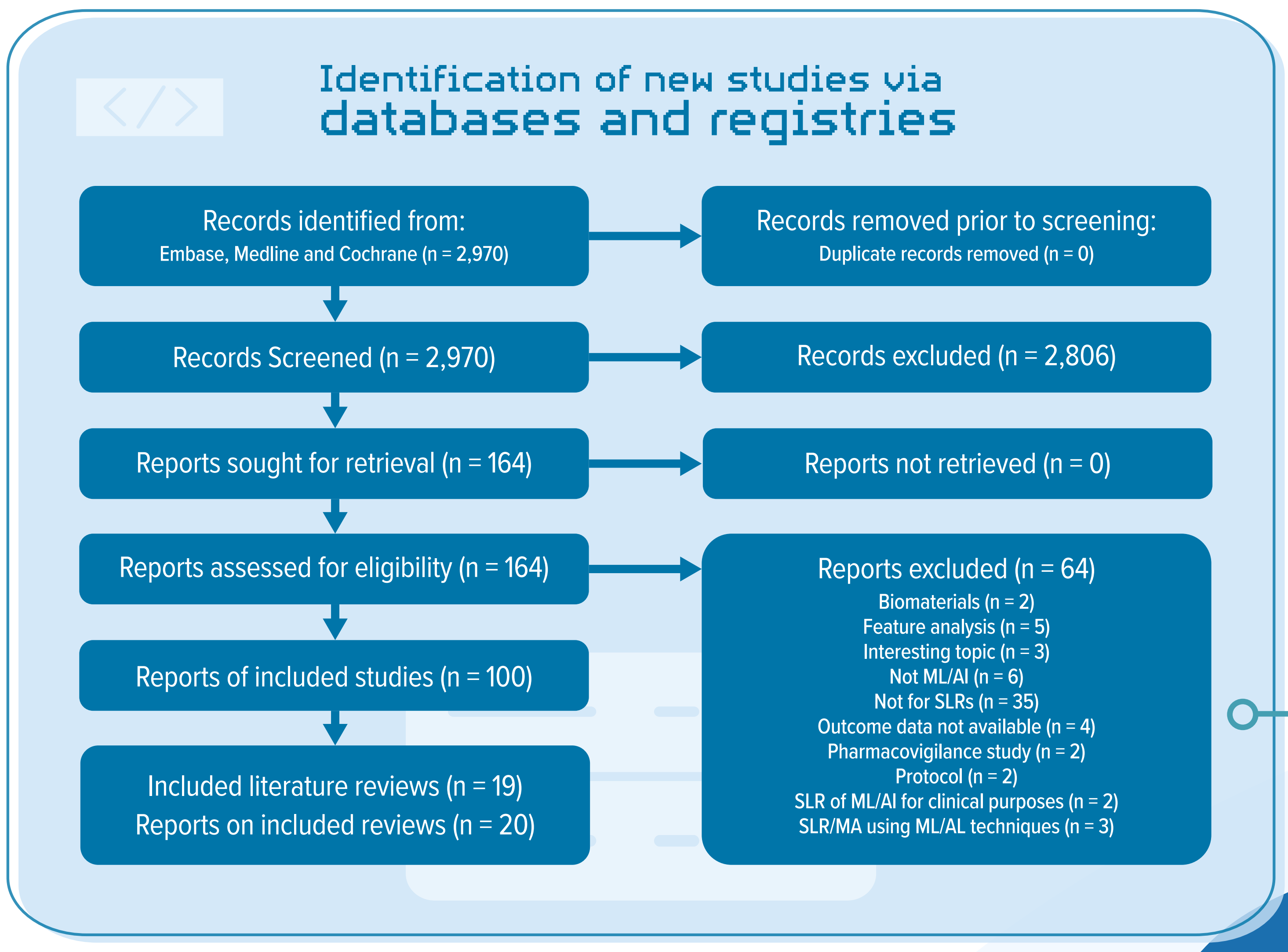
Table 1. SLR Methods

Type of data	• Any SLR process applied to a health-related topic	**Types of study design**	• Any type of study design
Type of methods	• ML/AI methods	**Databases searched (for studies published from 1 January 2022 onwards)**	• Embase • Medline • Cochrane Library • Searched on 23 April 2024
Outcomes	• Type of ML/AI technique • Information about the process used to train the ML/AI tool • Performance of ML/AI tools • Barriers and facilitators to the implementation of ML/AI tools for automating SLRs	**Study selection**	• Double screening by 2 independent researchers • Any discrepancies resolved by third reviewer
		Staged approach to data synthesis	1. Literature reviews 2. Any other study type

RESULTS

Among 2,970 records identified in the SLR, 100 studies met the inclusion criteria. We synthesised results using a staged approach and this poster summarises the findings of the literature reviews identified for inclusion. Of the included studies, 20 records corresponding to 19 literature reviews were extracted (Figure 1).

Figure 1. PRISMA Diagram of Included Studies



Identification of new studies via databases and registries

Records identified from:
Embase, Medline and Cochrane (n = 2,970)

Records removed prior to screening:
Duplicate records removed (n = 0)

Records Screened (n = 2,970)

Records excluded (n = 2,806)

Reports sought for retrieval (n = 164)

Reports not retrieved (n = 0)

Reports assessed for eligibility (n = 164)

Reports excluded (n = 64)
Biomaterials (n = 2)
Feature analysis (n = 5)
Interesting topic (n = 3)
Not ML/AI (n = 6)
Not for SLRs (n = 35)
Outcome data not available (n = 4)
Pharmacovigilance study (n = 2)
Protocol (n = 2)
SLR of ML/AI for clinical purposes (n = 2)
SLR/MA using ML/AL techniques (n = 3)

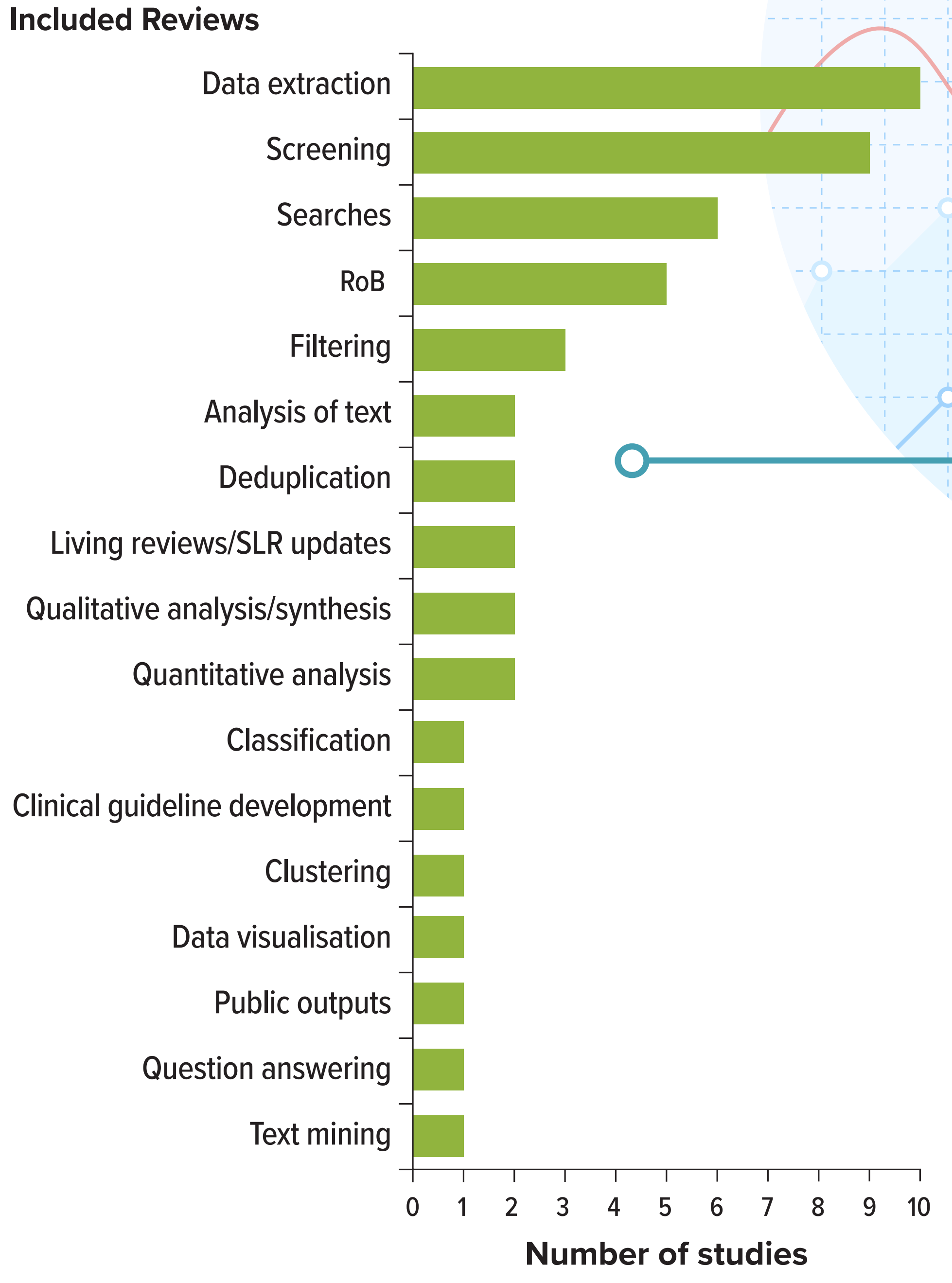
Reports of included studies (n = 100)

Included literature reviews (n = 19)
Reports on included reviews (n = 20)

MA = meta-analysis; PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

The primary objectives of the included reviews were to identify available ML/AI tools (n = 6), describe how ML/AI tools are used (n = 6), evaluate the implementation of ML/AI (n = 1), evaluate the performance of ML/AI tools (n = 3), and identify ML-/AI-assisted methods (n = 3). Figure 2 shows the SLR workflows that were considered in the included reviews (2 reviews did not specify the SLR workflows that were evaluated).

Figure 2. SLR Workflows Considered in the Included Reviews



RoB = risk of bias.

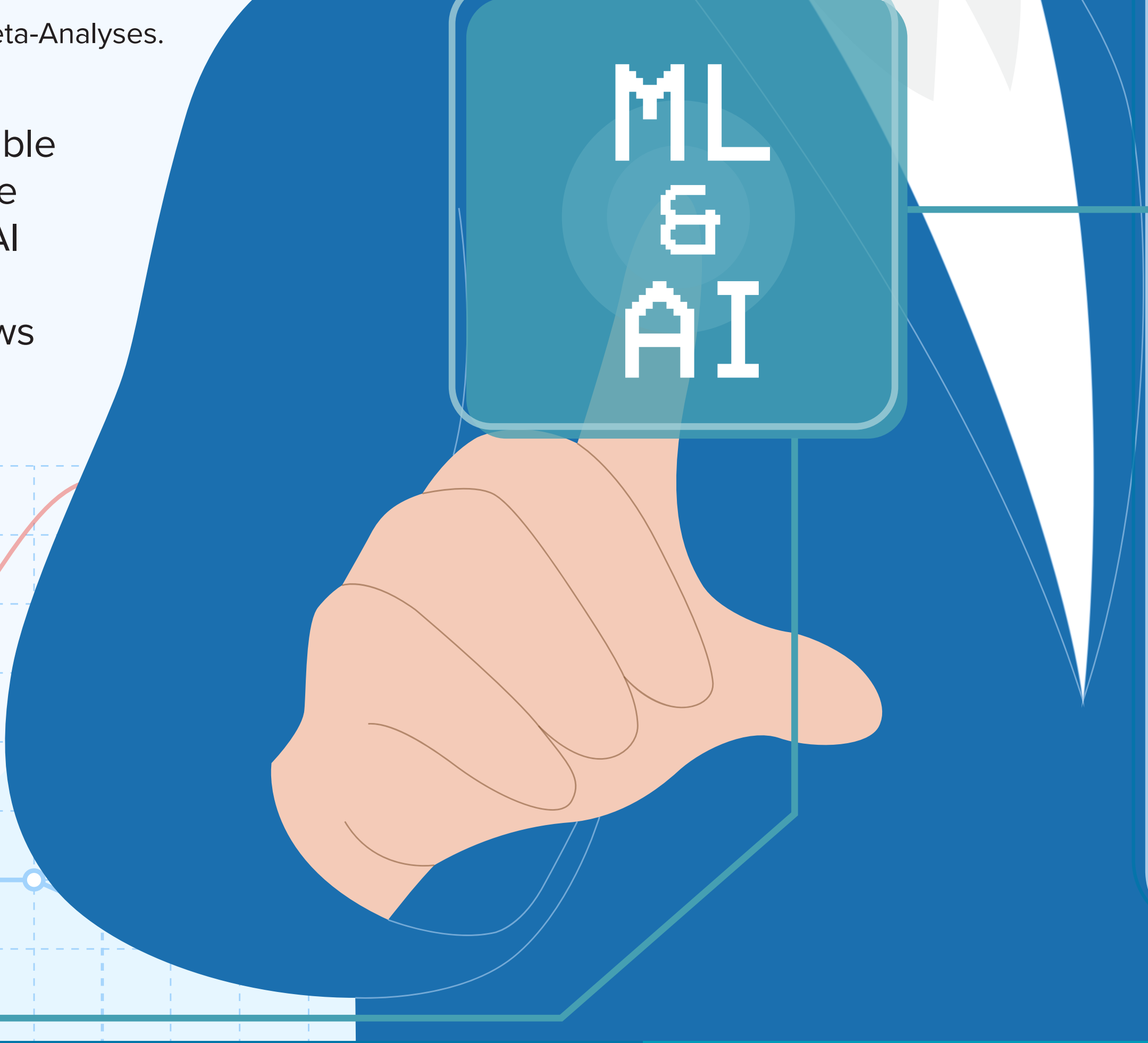
The reviews assessed and/or identified 126 ML/AI tools. (Table 2 lists the most commonly mentioned tools.) The tool purported to have the most features in Kallmes et al.[1] was DistillerSR, with 26/30 features assessed; however, these tools are ever evolving, and this may have now changed.

Table 2. ML/AI Tools Most Utilized in Systematic Reviews

Tool (n)	Tool Capabilities
Abstrackr (8)	Sorting and screening
DistillerSR (7)	Searching, screening, and data extraction, with the capability of RoB assessment
EPPI Reviewer (8)	Sorting, classifying, deduplicating, screening, and clustering
Rayyan (6)	Screening
RobotReviewer (6)	Data extraction and RoB
SWIFT-Review (6)	Searching, screening, and data extraction

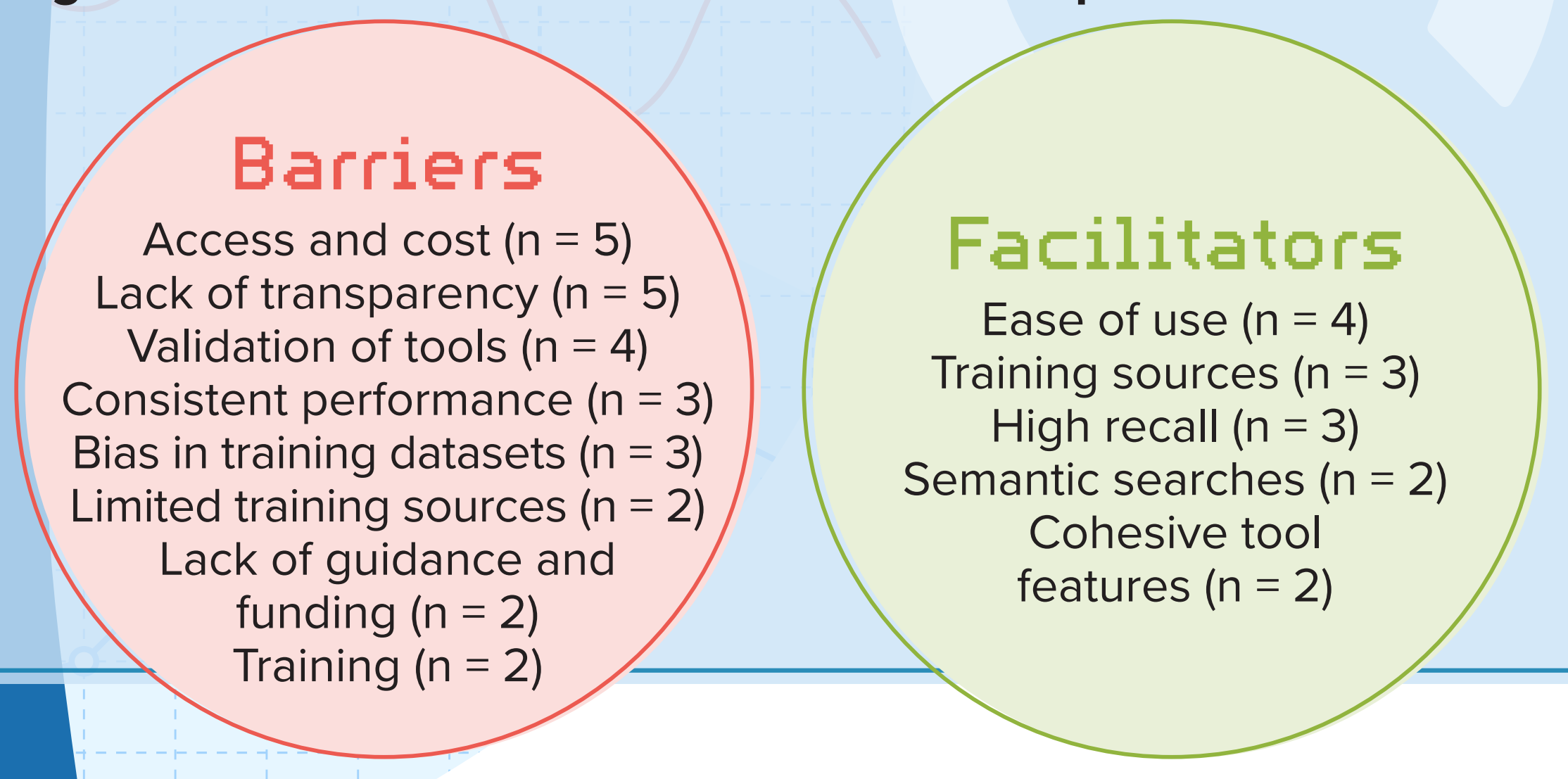
Most ML/AI tools (Table 2) were used to semi-automate workflows within SLRs rather than to achieve full automation. Moreover, the availability of ML/AI tools for conducting SLRs has increased over the years, with an increased focus on discussions around tool development and integration in practice. However, the number of SLRs claiming to have used ML/AI is reportedly still limited.[2] Half of reviews using ML/AI tools were living reviews or rapid reviews.[3] The use of ML/AI tools can lead to workload and time savings in SLRs, and yet, at least for screening, these savings are a function of the threshold used to train the tools[4] – the higher the threshold, the lower the savings.

The tools were largely trained using data from PubMed and abstracts rather than full-text articles. Common performance metrics included recall, precision, and F1 scores; prioritisation of recall optimisation was recommended, as the validity and reliability of most tools is currently not established,[2,5] and there is currently no gold standard for evaluating tools.[6]



Barriers to ML/AI uptake included lack of regulatory agency guidance, costs, training requirements, user-friendliness, and transparency concerns. Facilitators of ML/AI included ease of use, flexibility, high recall (i.e., the proportion of actual included studies identified by the tool), ability to process a large number of citations, and associated workload savings (Figure 3). Suggestions for tool improvements included more diverse training datasets, such as full-text articles and different electronic databases (e.g., Embase and Cochrane); development of more sophisticated natural-language processing techniques, such as semantic searches, rather than keyword searches; and cohesive features to capture all workflows of a review. User, policymaker, and funding agency buy-in are key for the adoption of ML/AI tools for literature reviews, and there is a current expectation that these tools should outperform human researchers.[7]

Figure 3. Barriers and Facilitators to the Adoption of ML/AI Tools



Barriers
Access and cost (n = 5)
Lack of transparency (n = 5)
Validation of tools (n = 4)
Consistent performance (n = 3)
Bias in training datasets (n = 3)
Limited training sources (n = 2)
Lack of guidance and funding (n = 2)
Training (n = 2)

Facilitators
Ease of use (n = 4)
Training sources (n = 3)
High recall (n = 3)
Semantic searches (n = 2)
Cohesive tool features (n = 2)

CONCLUSIONS

This SLR identified recent evidence from published literature reviews on the use, performance, and implementation of ML/AI to support systematic reviews. Unlike many of the reviews identified, this SLR considered the full range of SLR workflows.

The evidence suggests that most efforts in this topic area focus on screening and data extraction. This highlights the need for future research to expand into other SLR workflows, such as searching, risk of bias, and synthesis. A consistent theme across the included reviews was the recommendation to maximise recall of ML/AI tools, particularly for screening and classification, in order to enhance (semi-) automation of these tasks. However, to achieve adequate levels of recall, it is important to consider the complexity of a review question during the training phase of ML/AI tools. This SLR also highlighted the need to expand the data sources used for training ML/AI tools in order to enhance their performance.

To increase their adoption, performance of ML/AI tools should be assessed against current practice, rather than striving for perfection. Moreover, this SLR identified cost and training requirements as one of the key barriers to the uptake of ML/AI tools, highlighting the need to prioritise ease of use during tool development. This SLR also highlighted the tension between business and academic interests in terms of expectations for tool validation, interoperability and transparency, and how this tension may at times interfere with the adoption of ML/AI tools. Lastly, greater guidance from regulatory bodies is needed to determine the best pathway for use of ML/AI in SLRs.

REFERENCES

1. Kallmes K, et al. Int J Technol Assess Health Care. 2022;38:S27.
2. Tercero-Hidalgo J, et al. J Clin Epidemiol. 2022;148:124-34.
3. Yao X, et al. Cancer Epidemiol. 2024;88.
4. Kamra S, et al. Value Health. 2023;26(12):S6.
5. Khalil H, et al. J Clin Epidemiol. 2022;144:22-42.
6. Aletaha A, et al. Med J Islamic Repub Iran. 2023;37:95.
7. Hanegraaf P, et al. BMJ Open. 2024;14(3):e076912.

CONTACT INFORMATION

Jose Marcano-Belisario,
RTI Health Solutions
Manchester, United Kingdom
Phone: 44(0)161.447.6009
Email: jmarcano@rti.org

SCAN QR CODE FOR SUPPLEMENTAL REFERENCE LIST

The power of **knowledge.**
The value of **understanding.**

Presented at: ISPOR Europe 2023; 17-20 November 2024; Barcelona, Spain