SA32

Developing Best Practices for Using Automation Tools in Literature Reviews

Md Sohail Aman[1], Abhra Roy Choudhury[1], Kopal Dixit[1], Deepti Rai[1], Iain Fotheringham[2]

[1]PharmaQuant Insights Pvt. Ltd., Kolkata, India, [2]PharmaQuant, Dublin, Ireland.

PharmaQuant

Focused Insights, Better outcomes

INTRODUCTION

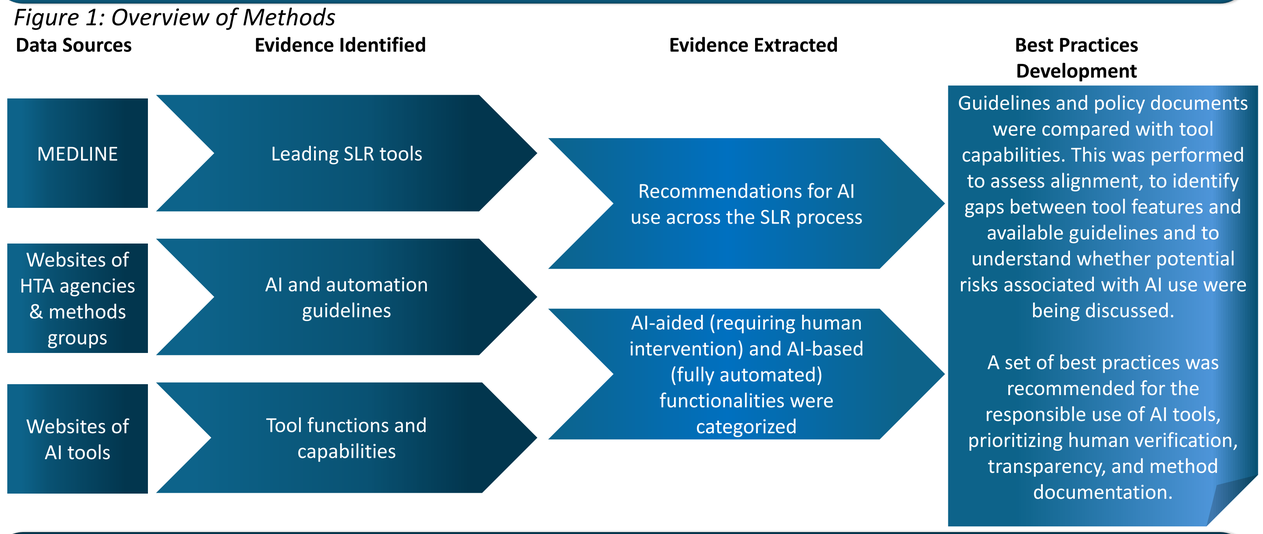
- Systematic literature reviews (SLRs) are a crucial part of health economics and outcomes research (HEOR). Emerging automation and artificial intelligence (AI) technologies can assist in search, screening and data extraction and improve efficiency, but concerns remain about:
  - Incorrect exclusions of citations and incorrect data extraction
  - Reproducibility and algorithm transparency
  - Limited direction concerning human verification
  - Uncertainty around health technology assessment (HTA) agency acceptance[1]
- HTA agencies and methods groups (e.g., National Institute for Health and Care Excellence [NICE], Canada's Drug Agency [CDA], Cochrane) acknowledge AI tools but recommend human verification at all critical steps.[2-4]
- To the authors' knowledge, no research has been performed to inform step-by-step best practices for using automation and AI in SLRs.

OBJECTIVES

- To systematically identify and review methodological guidelines on the use of AI in SLRs
- To assess the AI and automation capabilities of leading SLR tools
- To recommend best practices for responsible, transparent, and unbiased AI use in SLRs

METHODS

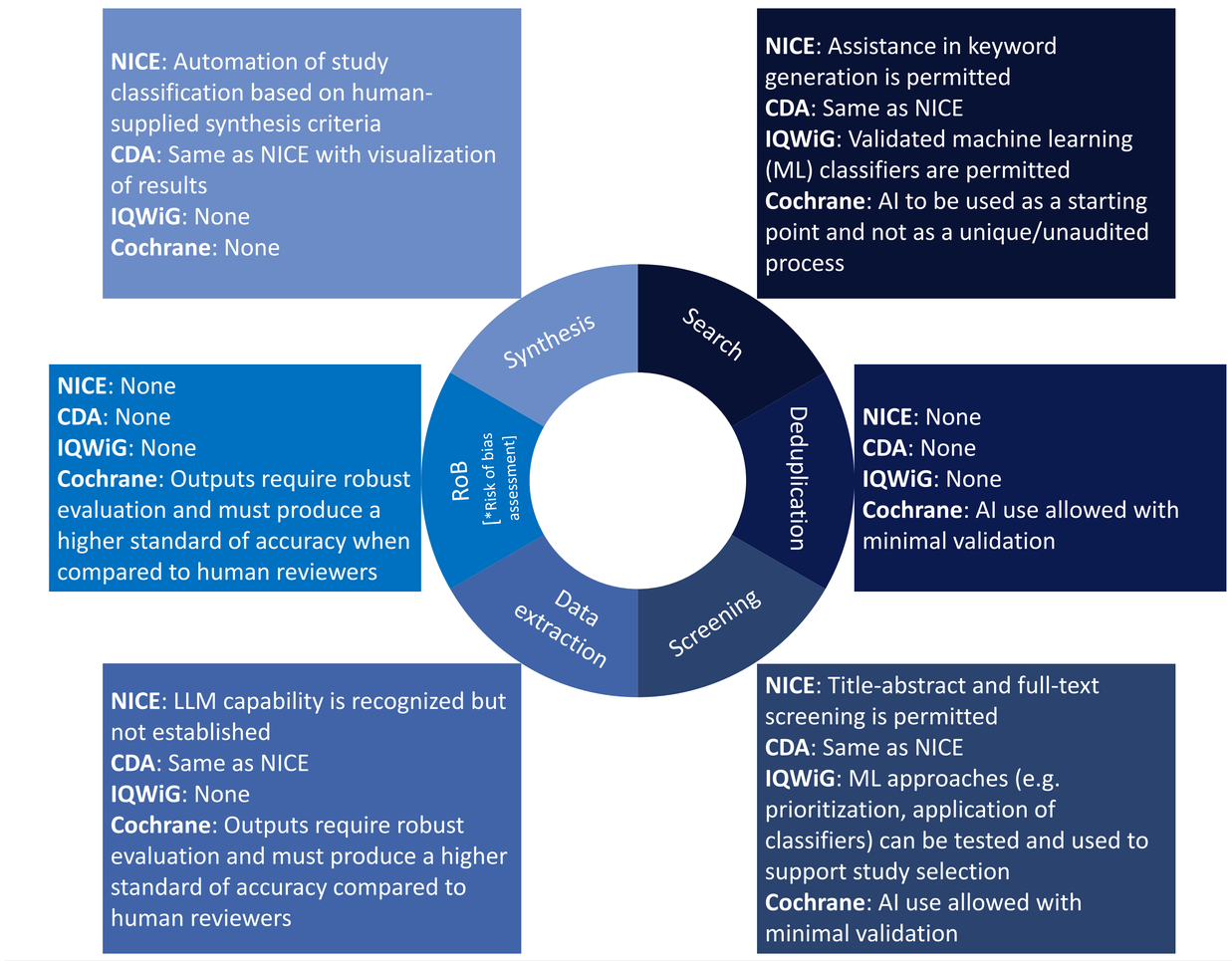
Figure 1: Overview of Methods



RESULTS

- A targeted MEDLINE search (January 2015-June 2025) identified 215 records, of which 5 publications characterized leading SLR tools. No comprehensive best practice frameworks were identified from MEDLINE.
- HTA bodies and SLR methodology working groups urge transparency, methodological rigor, and human oversight when using AI in SLRs. Nine HTA agencies were searched, of which NICE, CDA, Institute for Quality and Efficiency in Health Care (IQWiG), and Cochrane methods group have made recommendations. Clear direction remains largely absent (Figure 2).

Figure 2: Summary of Recommendations from HTA and Methods Groups [2-5]



- While HTA agencies such as The Spanish Agency of Medicines and Medical Devices (AEMPS), The Agenzia Italiana del Farmaco (AIFA), Haute Autorité de Santé (HAS), Dental and Pharmaceutical Benefits Agency (TLV), The National Health Care Institute (ZIN), and Pharmaceutical Benefits Advisory Committee (PBAC) provide guidance on general assessments, they have not yet addressed the use of AI in SLRs.
- ISPOR working groups have explored AI models and their potential applications but have not issued formal recommendations.[6]

- Automated rank-ordering, screening and data extraction are increasingly integrated with leading SLR tools, augmenting human effort rather than replacing it. SLR tools that have at least two review functions (e.g. screening and data extraction, etc.) were reviewed (Table 1).
- The proposed best practice recommendations aim to address gaps in existing guidance, prioritizing rigor and minimizing bias while leveraging possible AI efficiencies (Table 2).

RESULTS continued

Table 1: Leading SLR Tools, Their AI Features, and Other Functionalities

Tool	Free/Paid	AI Features	Other Functionalities
ActiveSLR	Free	AI-based deduplication AI-aided screening (AI-generated PICO) AI-aided linking of studies	Free full-text retrieval PRISMA Data extraction
Colandr	Free	AI-aided screening (rank ordering) AI-aided data-extraction (suggestion-based response)	Not available
Covidence	Paid	AI-aided screening (rank ordering) AI-aided data-extraction (suggestion-based response)	Search Deduplication Free full-text retrieval Risk of bias assessment
DistillerSR	Paid	AI-aided title-abstract screening (rank ordering) AI-aided data extraction (suggestion-based response)	Search Deduplication Free full-text retrieval
EPPI-Reviewer	Paid	AI-aided screening (rank ordering)	Deduplication (possibly) Data extraction (possibly)
EasySLR	Paid	AI-aided search (AI recommended keywords) AI-aided screening (rank ordering) AI-aided data-extraction (suggestion-based response)	Deduplication Free full-text retrieval Data extraction
Nested Knowledge	Paid	AI-aided search (AI recommended keywords) AI-based screening (replaces one human reviewer in double screening) AI-aided data extraction (suggestion-based response) AI-aided synthesis (visual creation)	Deduplication Critical appraisal
PICOPortal	Paid	AI-aided title-abstract screening (rank ordering) AI-aided critical appraisal AI-aided data extraction (suggestion-based response)	Deduplication Free full-text retrieval
Rayyan	Free & paid	AI based deduplication AI-aided screening (AI-generated PICO, rank ordering) AI-aided data extraction (suggestion-based response)	PRISMA
SWIFT-by Sciome	Paid	AI-aided screening (rank ordering)	Search (refinement)
Laser AI	Paid	AI-aided screening (rank ordering) AI-aided data extraction (suggestion-based response)	PRISMA
Lindexer	Paid	AI-aided screening (rank ordering) AI-aided data extraction (suggestion-based response)	Synthesis Critical appraisal PRISMA

Keys: PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; PICO, Population-Intervention-Comparator-Outcome

Table 2: Suggested Best Practices

SLR Steps	Suggested Best Practices
Search	AI assistance (e.g. LLM-suggested keywords) may be useful as a starting point but requires human verification. Suggested keywords should be validated for each database with careful identification of additional and contextually relevant terms to ensure comprehensive coverage.
Screening	AI review of one set of studies and human review of the other followed by human reconciliation may be acceptable in double screening. Exclusive reliance on AI for decision-making in single screening is not advised. However, AI can assist by prioritizing studies by relevance or by accelerating screening via AI-driven PICO extraction or keyword highlighting.
Data extraction	AI assistance may be acceptable in the retrieval of certain study characteristics (e.g. design, enrollment region, therapies administered) with moderate human verification. Considering current AI capabilities, human verification is recommended during automated extraction of quantitative data (e.g., hazard ratios, numerical values). In case of reviews with double reviewers, AI may extract one side, and human reviewer may extract the other followed by reconciliation of any differences by reviewer.
Synthesis	AI assistance in data summarization and visualization is acceptable with final human verification. Caution should be exercised when summarizing quantitative values.
Additional features	AI-based deduplication is acceptable for instances of exact duplicate studies. A human verification is suggested for removing any AI-suggested partially matched duplicates (e.g. author mismatch, capitalization, abstract vs full-text, linked publications). AI can also assist in study linking (e.g. trial name suggestion). AI prompts may be used to generate evidence summaries, tables, graphs, etc.
Documentation	Tools used, their version, settings, and the extent of use must be reported.

DISCUSSION

- Automated rank-ordering, screening, and data extraction are increasingly integrated into leading SLR tools, enhancing efficiency while augmenting human judgment (Table 1). However, their "black box" nature raises concerns around transparency, accuracy, and validation.
- Industry-wide standardization concerning best practices for the use of AI and automation tools in SLRs is therefore crucial.
- The recommended best practices are consistent with Cochrane's RAISE guidelines and the recent guidance from the UK Health Equity Evidence Centre (October 2025), while also account for available HTA recommendations and tool functionalities with the aim of preserving methodological rigor and quality.[4,7] Importantly, these recommendations address existing gaps in current guidance, emphasizing rigor, bias minimization, and responsible leveraging of AI efficiencies.
- Best practices will require ongoing revision as SLR tools, AI, and automation technologies continue to advance.
- To our knowledge, among the tools assessed, only ActiveSLR, Colandr, and Rayyan are freely available, making them suitable candidates for future testing of methodological alignment.

REFERENCES

1.Ge L, Agrawal R, Singer M, Kannapiran P, De Castro Molina JA, Teow KL, Yap CW, Abisheganaden JA. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. Syst Rev. 2024 Oct 25;13(1):269. doi: 10.1186/s13643-024-02682-2. PMID: 39456077; PMCID: PMC11504244
2. NICE. Use of AI in evidence generation: NICE position statement. Version number: 1.0. 2024 August 15. Accessed in June 2025
3. CDA. New Position Statement Aims to Guide the Use of AI Methods in Health Technology Assessment. 2025 April 25. Accessed in June 2025
4. Cochrane. Responsible AI in Evidence Synthesis (RAISE) 3: selecting and using AI evidence synthesis tools. Accessed in June 2025
5. IQWiG. General Methods. Version 7.0. 2023 September 19.
6. Fleurence RL, Bian J, Wang X, Xu H, Dawoud D, Higashi M, Chhatwal J; ISPOR Working Group on Generative AI. Generative Artificial Intelligence for Health Technology Assessment: Opportunities, Challenges, and Policy Considerations: An ISPOR Working Group Report. Value Health. 2025 Feb;28(2):175-183. doi: 10.1016/j.jval.2024.10.3846. Epub 2024 Nov 12. PMID: 39536966; PMCID: PMC11786987.
7. Health equity evidence centre. A practical guide to using AI tools to assist Evidence Synthesis. Accessed in October 2025.

Presented at ISPOR Europe 2025; Glasgow, Scotland, UK; November 9-12, 2025