Assessing the Recall of Artificial Intelligence in Conducting Systematic Literature Reviews

Aastha Radotra, MPH[1], Geetank Kamboj, M.Pharm.[1], Surabhi Aggarwal, M.Pharm.[1], **Hemant Rathi, MSc[1,2]**

[1]Skyward Analytics, Gurugram, India, [2]EasySLR, Gurugram, India

MSR111

INTRODUCTION

- Systematic literature reviews (SLRs) are **time-consuming** and **resource-intensive**, requiring substantial effort across multiple stages mainly screening and data extraction
- Completing an SLR typically takes over 15 months, placing significant burden on researchers[1]
- To address these challenges, **artificial intelligence** (AI) tools have been developed to accelerate the screening process and reduce the overall workload[1-2]
- The objective of this study was to evaluate the **screening recall** of AI in replicating the study selection of previously published SLRs

METHODS

- We conducted an exploratory analysis to evaluate the recall of AI-assisted SLRs using the **EasySLR™** platform (**Figure 1**)
- Published SLRs were selected based on the following inclusion criteria: (1) presence of a clearly defined search strategy, (2) availability of a complete list of included studies, and (3) use of a single freely accessible database for literature retrieval
- A total of five eligible SLRs were identified for the evaluation: **Kaegi et al., 2022**[3], **Sharifian-Dorche et al., 2021**[4], **Nicholas et al., 2020**[5], **Kaegi et al., 2020**[6], and **Bruurs et al., 2013**[7]
- The original PubMed search strategies from these SLRs were replicated to obtain the respective citation sets
- Retrieved citations were then cross-referenced against the list of studies included in each original SLR
- Inclusion and exclusion criteria were extracted from the selected SLRs to construct study-specific screening rules
- These screening rules were fed into EasySLR™ to facilitate AI-assisted screening for each of the five included SLRs
- Recall was calculated as the proportion of human-included studies correctly included by AI

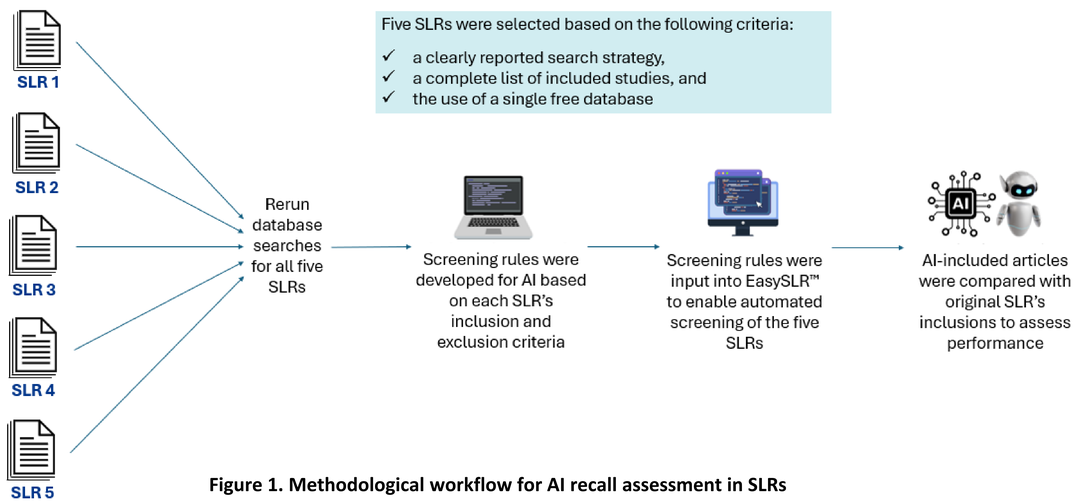

Figure 1. Methodological workflow for AI recall assessment in SLRs

RESULTS

- The characteristics of the five included SLRs, including population, intervention, outcomes, and study design, are summarised in **Table 1**
- AI demonstrated high recall across the five evaluated SLRs, varying from **73% to 100%** (**Figure 2**)
  - **100% recall** was observed for Kaegi et al., 2020 where perfect concordance with the originally included studies was observed
  - **High recall** was observed for Bruurs et al., 2013 (90%) and Nicholas et al., 2020 (84%)
  - **Moderate recall** were recorded for Kaegi et al., 2022 (74%) and Sharifian-Dorche et al., 2021 (73%)
- Replicated search hits and AI-included studies were largely consistent with the original SLRs and human reviewer selections (**Table 2**)
- The observed variation in AI performance may be due to the complexity of research questions and the variation between context-driven decisions made by human reviewers in the original SLRs and the standardized screening rules applied during this analysis

Table 1. Study characteristics of included SLRs presented using the PICOS framework

Systematic review	Population (P)	Intervention (I) vs. Comparators (C)	Outcomes (O)	Study Design (S)
Kaegi et al., 2022[3]	Patients with immune-mediated diseases (e.g., MS, RA, SLE)	CD20-targeting mAbs vs. placebo/standard care	Safety, efficacy	RCTs, case series, and open-label studies
Sharifian-Dorche et al., 2021[4]	Patients with MS or NMOSD on DMDs	Various DMDs vs. no treatment or other therapies	COVID-19 risk, severity, mortality	Observational studies, case series and reports
Nicholas et al., 2020[5]	Patients with MS on oral DMDs	Oral DMDs	Adherence, discontinuation	Observational studies
Kaegi et al., 2020[6]	Patients with immune-mediated diseases (e.g., MS, RA, SLE)	Atacicept vs. placebo, conventional treatment, or other biologics	Safety, efficacy	RCTs
Bruurs et al., 2013[7]	Patients with asthma	Physiotherapy vs. control	QoL, symptom reduction, cardiopulmonary fitness	RCTs

Figure 2. Screening recall of AI across published SLRs

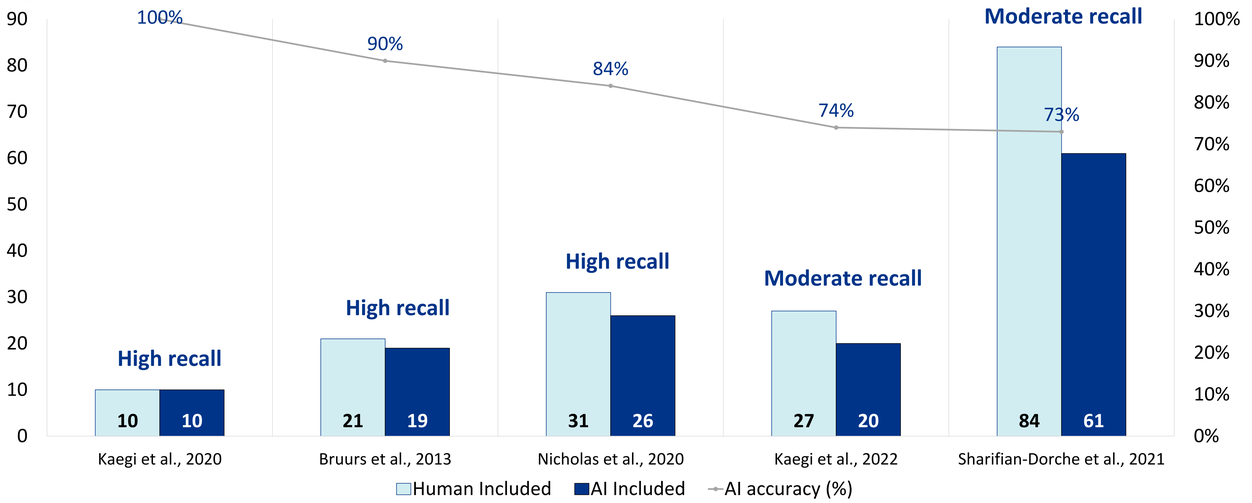

Table 2. Search output and study inclusion: AI versus human screening across published SLRs

Systematic review	Database Searched	Original Search Hits	Human Included Studies, Final Report	Replicated Search Hits*	Studies Included by AI**
Kaegi et al., 2022[3]	PubMed	2,220	27	3,182	20
Sharifian-Dorche et al., 2021[4]	PubMed	262	84	278	61
Nicholas et al., 2020[5]	PubMed	510	31	657	26
Kaegi et al., 2020[6]	PubMed	118	10	166	10
Bruurs et al., 2013[7]	PubMed	237	21	517	19

*Replicated search hits are higher than the original, despite using the same year limits, as the search was re-run post-publication—potentially capturing articles retrospectively indexed, newly reclassified, or added through routine database updates and corrections.
**Studies included by AI represent the number of correctly identified studies from the original human-included set, not the total number of AI inclusions.

CONCLUSION

- **EasySLR™** demonstrated high recall in screening across multiple SLRs, indicating its potential to streamline evidence synthesis workflows
- A key limitation of our approach is that screening decisions often involve **subjective judgment** beyond the protocol, which cannot be fully captured by predefined rules and may therefore limit AI performance. An additional limitation was that we did not attempt to contact the authors of the original SLRs to explore around this aspect
- However, observed variability in recall highlights the importance of maintaining **human oversight**

FUNDING: No funding was received for the conduct of this study.

Poster presented at ISPOR 2025, Montreal, QC, Canada (Wednesday, May 15th, 2025)

REFERENCES: 1. Van Mossel S et al. Artificial Intelligence as a New Research Ally? Performing AI-Assisted Systematic Literature Reviews in Health Economics. PharmacoEconomics. 2025 Apr 10:1-4; 2. Burns JK et al. Using an artificial intelligence tool can be as accurate as human assessors in level one screening for a systematic review. Health Information & Libraries Journal. 2024 Jun;41(2):136-48; 3. Kaegi C et al. Systematic review of safety and efficacy of second-and third-generation CD20-targeting biologics in treating immune-mediated disorders. Frontiers in Immunology. 2022 Feb 2;12:788830; 4. Sharifian-Dorche M et al. COVID-19 and disease-modifying therapies in patients with demyelinating diseases of the central nervous system: a systematic review. Multiple sclerosis and related disorders. 2021 May 1;50:102800; 5. Nicholas JA et al. Real-world adherence to, and persistence with, once-and twice-daily oral disease-modifying drugs in patients with multiple sclerosis: a systematic review and meta-analysis. BMC neurology. 2020 Dec;20:1-5; 6. Kaegi C et al. Systematic review of safety and efficacy of atacicept in treating immune-mediated disorders. Frontiers in immunology. 2020 Mar 24;11:433.; 7. Bruurs ML et al. The effectiveness of physiotherapy in patients with asthma: a systematic review of the literature. Respiratory medicine. 2013 Apr 1;107(4):483-94.

ABBREVIATIONS: AI, artificial intelligence; CD20, cluster of differentiation 20; COVID-19, Coronavirus Disease 2019; DMDs, disease-modifying drugs; mAbs, monoclonal antibodies; MS, multiple sclerosis; NMOSD, neuromyelitis optica spectrum disorder; PICOS, Population, Intervention, Comparator, Outcomes, and Study Design; QoL, quality of life; RA, rheumatoid arthritis; RCTs, randomised controlled trials; SLE, systemic lupus erythematosus; SLRs, systematic literature reviews.