Is Artificial Intelligence Replacing Humans in Systematic Literature Reviews? A Systematic Literature Review

Queiros L[1], Mearns ES[2], Ademisoye E[3], McCarvil M[1], Alarcão J[4], Garcia MJ[1], Abogunrin S[1]

[1]F. Hoffmann-La Roche Ltd., Basel, BS, Switzerland; [2]Genentech, San Francisco, CA, USA; [3]AdvantageTechnoeconomics, London, KEN, UK; [4]Roche Farmacêutica e Química, Lda., Amadora, Portugal

BACKGROUND

- Systematic literature reviews (SLRs) seek to answer research questions and form comprehensive, rigorous evidence-based conclusions.[1]
- As methodologies have advanced, so have the standards, time, and costs to produce rigorous, high-quality, relevant SLRs increased. Despite, SLRs still rely on a heavily manual process and can be out of date by the time of completion.[2,3]
- To address SLR workload challenges, much research has been done to incorporate artificial intelligence (AI) methods in the SLR process[4]. As a result, there are now multiple SLR applications that provide artificial intelligence (AI)-as-a-service capability (AIsAPP).[5]
- However, it is unclear if researchers are utilizing the AI component of these applications.

OBJECTIVE

- We performed an SLR to assess whether AI is being utilized and/or reported as part of the methods of published SLRs or scoping reviews or protocols of such articles.

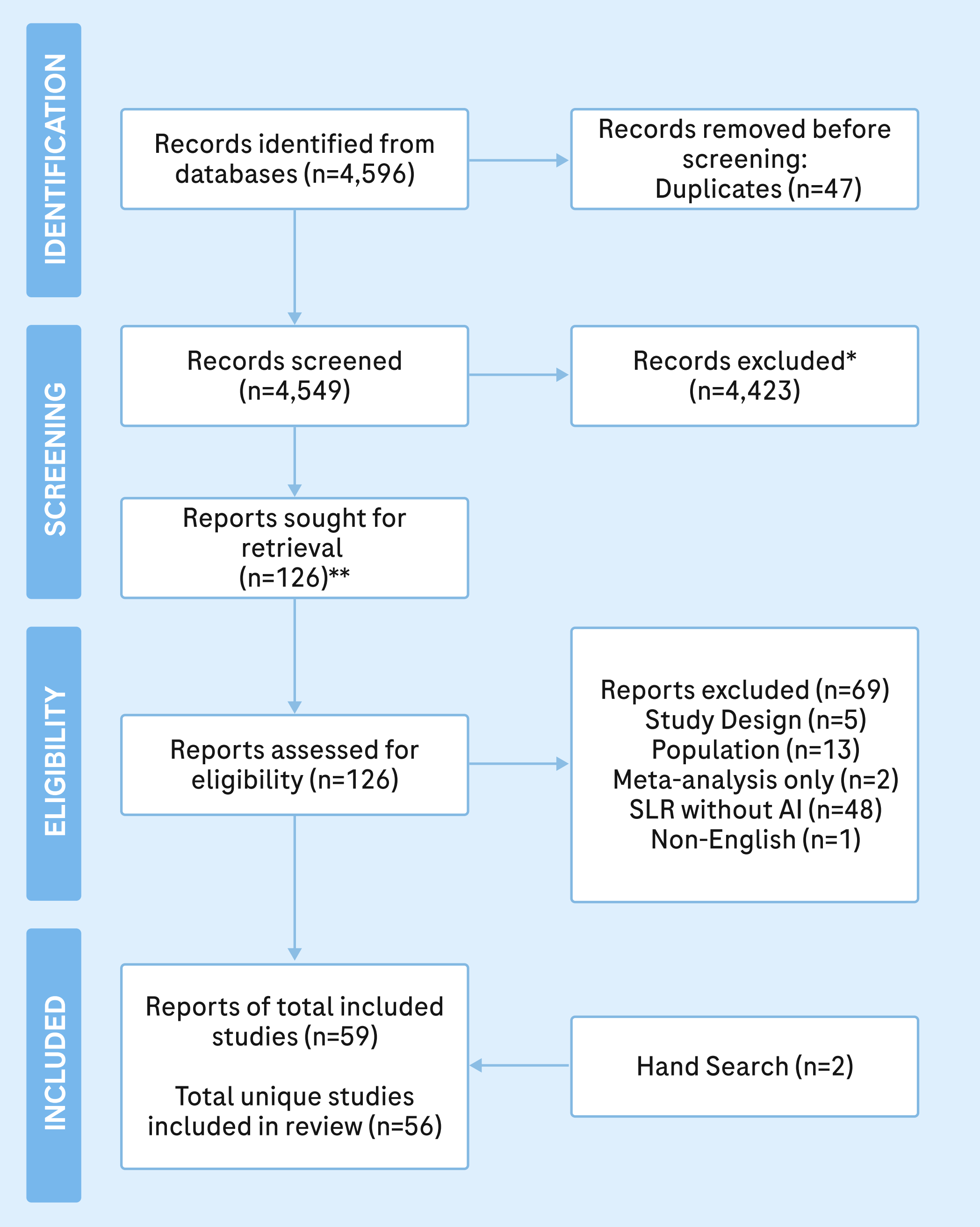
METHODS

- A systematic review was performed following PRISMA and Cochrane Collaboration guidelines.[1,6]
- Relevant articles were systematically searched in Embase, MEDLINE or MEDLINE in Process (via embase.com) on June 21, 2021. Grey literature were included as supplementary information.
- English language articles of SLRs or scoping reviews (including protocols of such articles) with or without meta-analyses, reporting the utilization of, or plan to utilize, AI and/or an AIsAPP in any SLR step were eligible for inclusion provided they were studying a health condition and/or intervention in humans.
- Two investigators independently screened titles and abstracts and full-texts of potentially relevant citations; discrepancies were resolved by a third independent reviewer at both levels. Title and abstract, and full-text screening were performed using Rayyan (without AI).
- Data were extracted from eligible studies by one independent reviewer into a standardized Excel template, and a second investigator verified entries for correctness and accuracy.
- The quality of the SLRs was assessed using six domains related to the methods sections of the PRISMA Expanded Checklist[5] and PRISMA-P Checklist[6]. Each domain was rated as 0=not reported, 1=low quality or 2=high quality); overall quality scores ranged from 0 (lowest methodological rigor) to 12 (highest methodological rigor).

RESULTS

- After screening, de-duplication, and backwards citation tracking, a total of 59 records (corresponding to 56 unique studies) met all eligibility criteria and were included in this SLR (**Figure 1**).
- The included protocols, SLRs and scoping reviews were heterogeneous in terms of patient populations, interventions, country of origin, and AI methodologies.
- The most frequently reported AIsAPPs were Rayyan (N=22), DistillerSR (N=11) and EPPI-reviewer (N=9). Other AIsAPPs utilized were Abstrackr, RobotReviewer, Robot Analyst and LIvE platform (**Figure 2**). Python packages were used in most of the bespoke algorithms (N=7).
- AIsAPPs were mainly used to support the title and abstract screening step (N=31) (**Figure 3**). AIsAPP/softwares were also reported to be used at other SLR steps, such as search, full-text screening, data extraction and qualitative analysis (**Figure 3**).

Figure 1. PRISMA Flow Diagram



Abbreviations: AI, artificial intelligence; SLR, systematic literature review
* automation tools were not included in screening; all were excluded by a human
** non-retrievable articles were considered based on title and abstract information available

Table 1. Summary of the Use of AI (n=16 studies)[7-22]

Integration of AI in the SLR workflow	AIsAPP Utilized	AI Method
Full-Automation (automatic method implemented with minimal human intervention, e.g. training)		
Title and abstract screening conducted with automatic method only	Bespoke algorithm (n=2)†	NLP [9,13], Text mining [9], ML [9]
Semi-Automation (automatic method with significant human intervention)		
Reduce the number of records that move to title and abstract screening	Bespoke algorithm (n=2)†	NLP [10], ML [21]
	Bespoke algorithm (n=1)❖	Text Mining [19]
Rank-order records and reduce the number of records manually reviewed in title and abstract screening	Bespoke algorithm (n=1)*	Text mining [8]
	Abstrackr (n=2)	Abstrackr AI capabilities [7,15]
	EPPI-Reviewer (n=1)	ML [18]
Rank-order records in title and abstract screening	RobotAnalyst (n=1)	Text Mining and ML [22]
Second reviewer in title and abstract screening	Bespoke algorithm (n=1)†	ML [12]
	Rayyan (n=1)	Rayyan AI capabilities [14]
	EPPI-Reviewer (n=1)	ML [17]
Second reviewer in full-text screening	Wordstat and QDA Miner (n=1)	Text Mining [15]
Topic modeling in data extraction	Bespoke algorithm (n=1)❖	Text Mining [19]
Second Reviewer in data extraction	RobotReviewer (n=1)	RobotReviewer AI capabilities [16]
Second Reviewer in quality assessment	RobotReviewer (n=1)	RobotReviewer AI capabilities [16]
AI used for screening but workflow not clear	LIvE platform (Python) (n=1)	ML [20]
Implementation of AI for screening not clear or not reported		
Not Clear/ Not reported	EPPI-Reviewer (n=1)	ML [11]

Abbreviations: AI, artificial intelligence; AIsAPP, applications that provide artificial intelligence (AI)-as-a-service capability; ML, Machine Learning; NLP, Natural Language Process; SLR, systematic literature review
† Custom Python implementation; ❖ R based software; * No details provided on the software used

Figure 2. AIsAPPs Used in SLRs

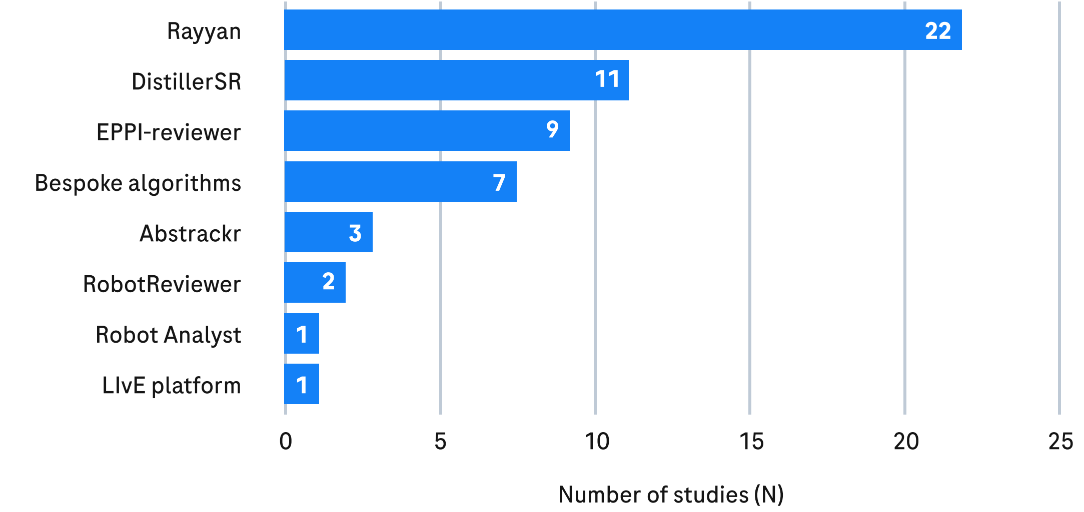
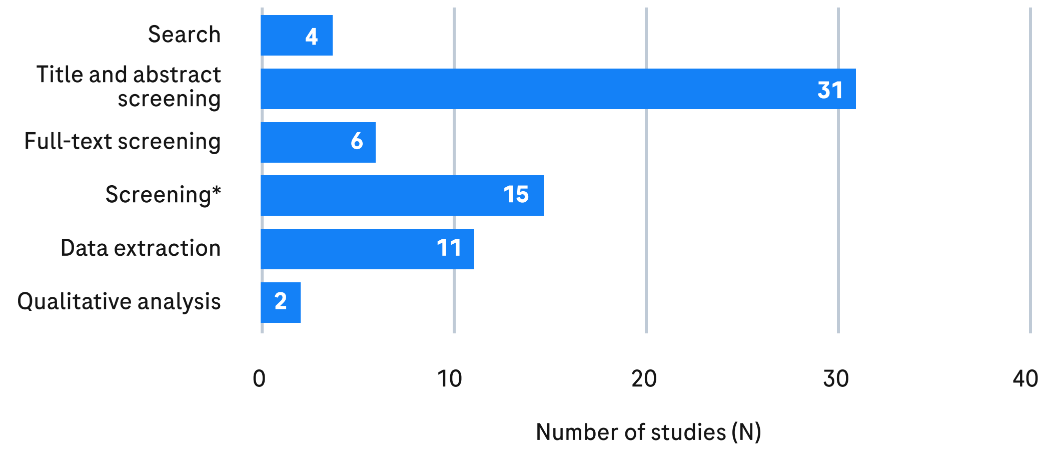
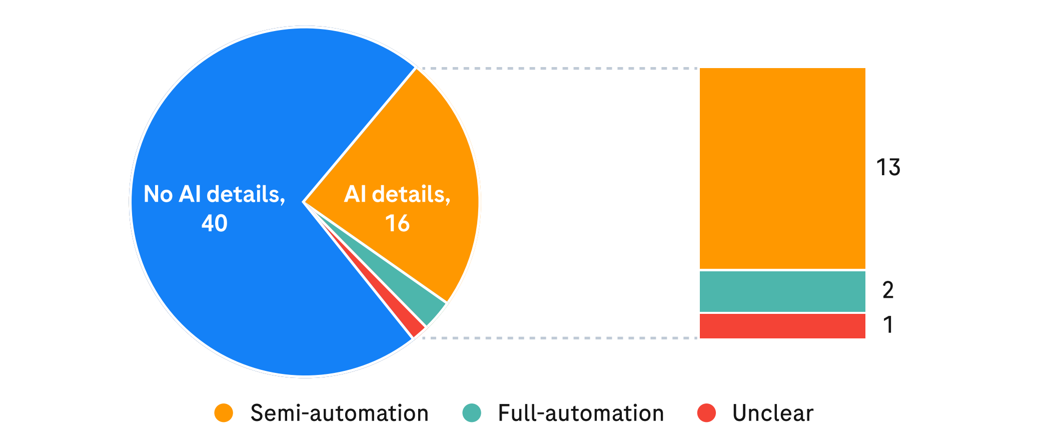


Figure 3. SLR Steps Where AIsAPPs were Utilized



* Screening step in which the AIsAPP was used not specified

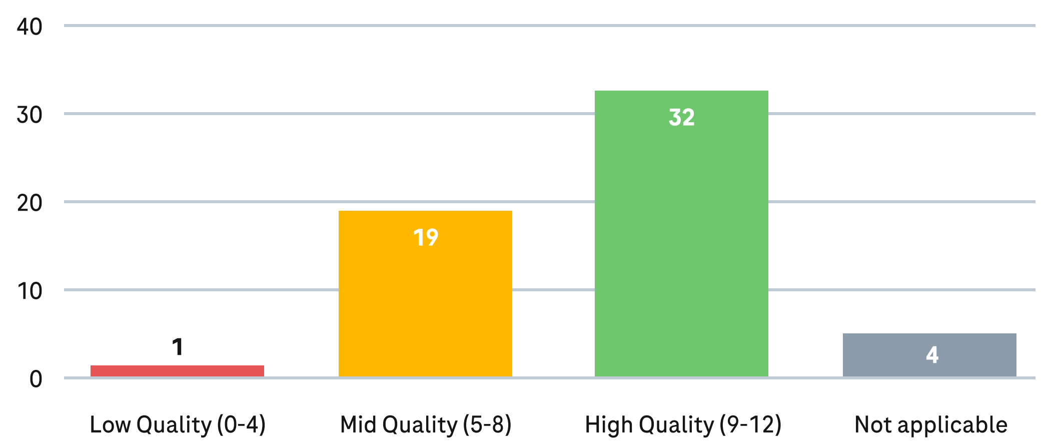
Figure 4. Automation Processes



Abbreviations: AI, artificial intelligence

- Only 16 studies (29% of the 56 included) provided details on how AI was used during the review process (**Table 1**); 13 used AI for semi-automation, two for full automation, and it was unclear how automation was used in one study (**Figure 4**). Semi-automation was mainly used for screening (N=12) and extraction (N=2). Full automation was used for abstract screening in the two studies that reported this information. (**Figure 4**).

Figure 5. Quality Assessment Results



- Where quality assessment was applicable, most studies had a score between 5 and 12 (**Figure 5**), indicating most included studies were of mid- to high-quality for methods reporting.

CONCLUSIONS

- Despite the increasing effort to automate SLRs, it seems AI is not yet a common practice; few studies reported utilizing or planned to use AI in the SLR process.
- In contrast to current PRISMA guidelines, when an AIsAPP was utilized, details on automation steps were often not described.
- Screening is the step of the SLR process where AI was applied more often. This might be related with the fact that this is the most time consuming step of the process and also the step for which more research is available.
- A limitation of our study is that our search was limited to certain controlled vocabulary. Additionally, the screening rules were restricted to only include titles and abstracts with the mention of AI or an AIsAPP. Studies that mentioned AI methods or AIsAPP in the full-text only (and left it out of the abstract methods) were thus not considered for inclusion.
- Further research should evaluate the limitations and barriers of fully incorporating and reporting the utilization of AI as standard practice in SLRs
- Reporting the use of AI methods in the abstract of published SLRs should be encouraged to follow the updated PRISMA guidelines.

REFERENCES

1. Higgins et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. Cochrane, 2022 **2.** Borah R et al. BMJ Open. 2017 Feb 27;7(2):e012545 **3.** Créquit, P. et al. BMC Med 14, 8 (2016) **4.** Jaspers S. et al. EFSA supporting publication 2018: 15(6):EN-1427 **5.** Marshall, I.J. et al. Syst Rev 8, 163 (2019) **6.** Page MJ, et al. 2021;372:n71 **7.** Aucoin M et al. Neuropsychobiology. 2020;79(1):20-42. **8.** Baron JA et al. Drugs R D. 2013;13(1):9-16. **9.** Buchlak QD et al. J Clin Neurosci. 2021;89:177-198 **10.** Buchlak QD et al. Neurosurg Rev. 2020;43(5):1235-1253. **11.** Croker et al. Obes Facts. 2019;12(suppl 1):77 **12.** Eun MY et al. J Stroke Cerebrovasc Dis. 2021;30(6):105742 **13.** Foulquier N et al. Hum Vaccin Immunother. 2018;14(11):2553-2558. **14.** Gaskins NJ et al. Disabil Rehabil. 2021;43(17):2382-2396 **15.** Giummarra MJ et al. Accid Anal Prev. 2020;135:105333 **16.** Goldkuhle M et al. Cochrane Database Syst Rev. 2018;7(7):CD012556 **17.** Hanckel B et al. Syst Rev. 2019;8(1):252 **18.** Karwatowska L et al. BMJ Open. 2020;10(9):e038258 **19.** Pinaire J et al. Health Inf Sci Syst. 2017;5(1):1 **20.** Riaz IB et al. Eur Urol. 2021 Dec;80(6):712-723 **21.** Thiabaud A et al. J Med Internet Res. 2020;22(8):e18747 **22.** Yamamoto R et al. J Nephrol. 2021;34(1): 155-164