Artificial intelligence is different: Is it time to update systematic literature review workflows?

MRS210

Scan QR code for additional information and e-poster

Indicates additional information is included in the supplement

AI-augmented literature screening tools can be **integrated** easily into our **existing workflows** to help improve the **efficiency** and **integrity** of SLRs

Why did we undertake this research?

- **Objective:** to explore how well two commercially available machine learning-based AI screening tools perform in making inclusion and exclusion decisions during title/abstract screening for an SLR.
- With the increasing demand for comprehensive, up-to-date literature reviews, machine learning-driven AI tools are expected to become integral to delivering high-quality SLRs efficiently.

How did we perform this research?

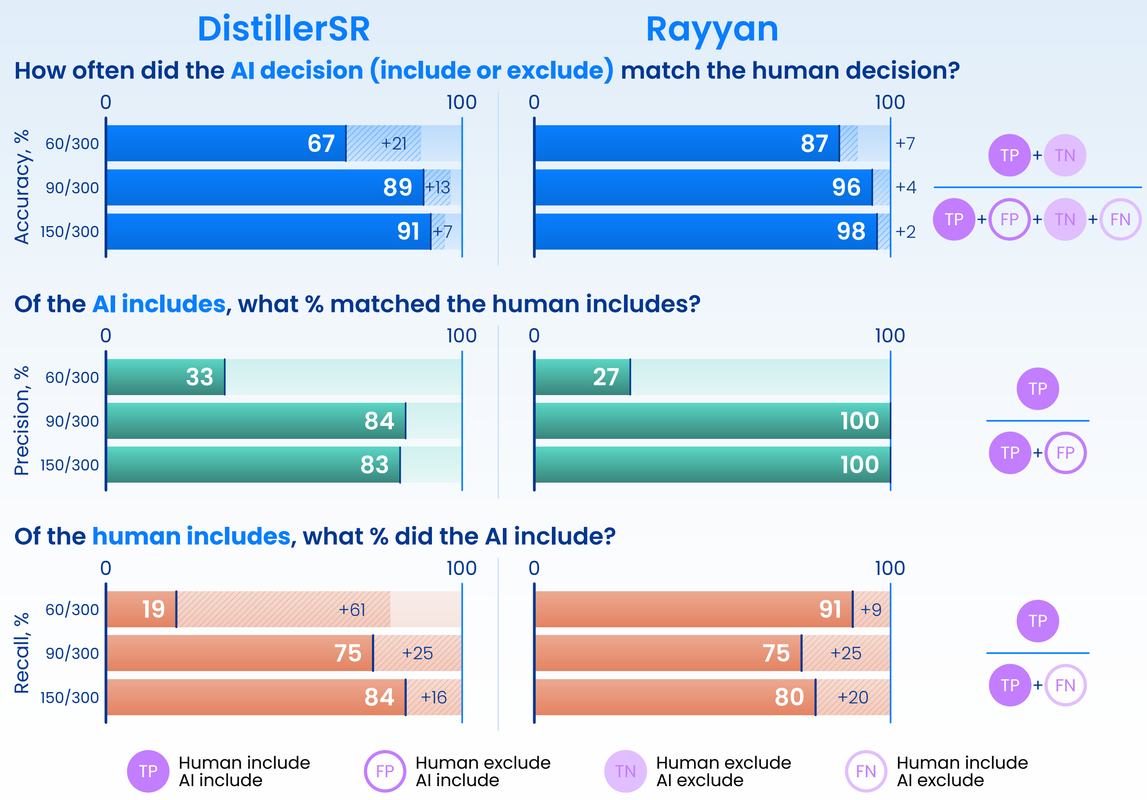
- Accuracy, precision and recall of machine learning-based models from DistillerSR and Rayyan trained on 60, 90, and 150 records were compared with a human review of 300 titles/abstracts on dual vs triple-inhaled therapy in patients with COPD (**Fig 1**).

SLR methodology

Fig 1. Study overview

What did we find?

Fig 2. Accuracy, precision and recall in DistillerSR and Rayyan using different sized training sets



Solid bars represent the base case analysis and shaded bars represent the scenario analysis in which FN were replaced with human FTR decisions. In Rayyan, 15, 55 and 43 records had available decisions in the 60, 90 and 150 training sets, respectively.

AI, artificial intelligence; FN, false negative; FN, false positive; FTR, full text review; TN, true negative; TP, true positive.

- **Three-quarters** of records included by the human were also included by the AI (recall; training sets, ≥90 records) at title/abstract screening when using either platform (**Fig 2**).
- This **increased to 100%** in a scenario where human-included records missed by the AI (FN) were substituted with human FTR decisions.

Did the AI miss anything? Not in this instance, but the SLR sample size was small, and reliability may depend on training set size.

- **Use case:** when applied post-title/abstract screening, AI tools could help identify human-included records that have a high likelihood of exclusion at FTR, thus reducing the need for human FTR of these records.

Fig 3. Exclusion reasons for 6 titles/abstracts with ≥1 AI FP

AI decision by training set (n/N)	Ref#	200	330	140	273	279	289
	Human decision	Ex	Ex	Ex	Ex	Ex	Ex
	60/300	FP	TN	FP	TN	TN	FP
	90/300	TN	FP	TN	FP	FP	TN
	150/300	FP	FP	FP	FP	FP	FP
	Human re-review decision	Ex	Ex	Ex	Ex	Ex	Ex
	Exclusion category	S	I/C	I/C	S	I/C	P

AI, artificial intelligence; Ex, exclude; FP, false positive; I/C, intervention/comparator; N, full sample size; n, training set sample size; P, population; Ref, reference; S, study design; TN, true negative.

Exclusion reasons

- None of the 6 titles/abstracts (all in DistillerSR) categorised as AI FPs across multiple training sets were subsequently included on human re-review (**Fig 3**).

Did the human miss anything? Not in this instance.

- **Use case:** AI tools could be applied to check the validity of the human decision.

Authors: Lucia Giles,[1] George Miles,[2] Gillian Sibbring[2]
Author affiliation: [1]Prime, Oxford, UK; [2]Prime, Knutsford, UK.

Abbreviations: **AI**, artificial intelligence; **COPD**, chronic obstructive pulmonary disease; **FN**, false negative; **FP**, false positive; **FTR**, full text review; **SLR**, systematic literature review.

Acknowledgements: The authors thank Lloyd Gwishiri, MD, for writing assistance, Adam Moriarty and Joey Cusano, for creative input; Sinead Stewart, for editorial support, as well as the Prime Digital and Production teams for their support with development of this poster. All authors are employees of Prime, UK. All opinions are the authors' own and do not necessarily reflect those of our employer.

Presented at ISPOR Europe 2024, Barcelona, Spain, Wednesday, 20 November 2024.