MSR107

MetaSLR in Automating the Data Collection in Systematic Literature Reviews

Barinder Singh[1], Gagandeep Kaur[2], Rajdeep Kaur[2], Diaby Karam[3], Sumeet Attri[2]

[1]Pharmacoevidence, London, United Kingdom, [2]Pharmacoevidence, Mohali, India, [3]Department of Pharmaceutical Sciences, University of Florida, Florida, FL, USA

Background

- Systematic literature reviews (SLRs) are critical for evidence synthesis and play a key role in Health Economics and Outcomes Research (HEOR).
- The development of search strategies for SLRs is inherently complex and requires extensive domain expertise and a substantial investment of time.
- The artificial intelligence (AI)-based Large Language Model (LLM) can streamline the labor-intensive process of SLRs, leveraging its remarkable performance in its ability to quickly and accurately analyze large volumes of textual data.
- The growing utilization of AI in the field of research is driven by its widespread potential, offering distinct advantages over traditional methods of conducting SLRs.

Objective

- This study aimed to compare performance of the automated tool, MetaSLR, with the traditional human review process.
- Data from a previously conducted SLR to identify economic evaluations (EEs) of health interventions in Attention-deficit Hyperactivity Disorder (ADHD) was leveraged to validate the accuracy and efficiency of the MetaSLR tool to automate the screening process and improve productivity compared to conventional methodologies.

Methodology

- The EMBASE® database was searched to identify EEs conducted in adults with ADHD and published in the English language.
- A subject matter expert (SME) with a decade of experience in SLRs prepared and optimized a prompt based on the results obtained from a small subset of citations, which was delivered through a Python application programming interface, to identify evidence aligned with the predefined inclusion and exclusion criteria (**Figure 1**).
- The LLM Claude Sonnet 3.5 with an optimized prompt was deployed using a Python script to review the citations and evaluate their inclusion or exclusion.
- An evaluation was conducted to compare the screening results of AI and human reviewers, measuring their agreement levels and assessing the accuracy with which publications were identified for inclusion in the SLR.
- The accuracy, sensitivity, and specificity of the results was calculated using the below mentioned formulae:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

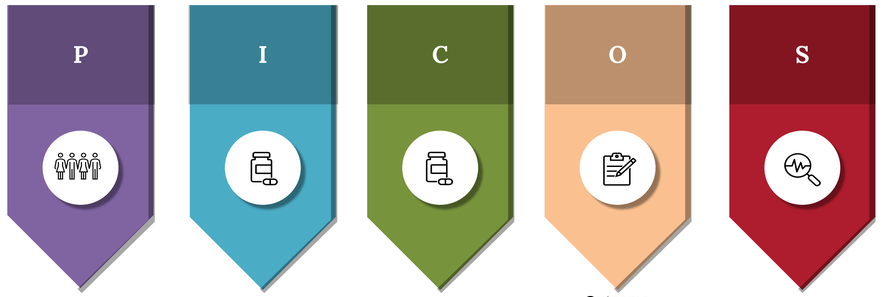
$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TP+FP}$$

FN: False Negative; FP: False Positive; TN: True Negative; TP: True Positive

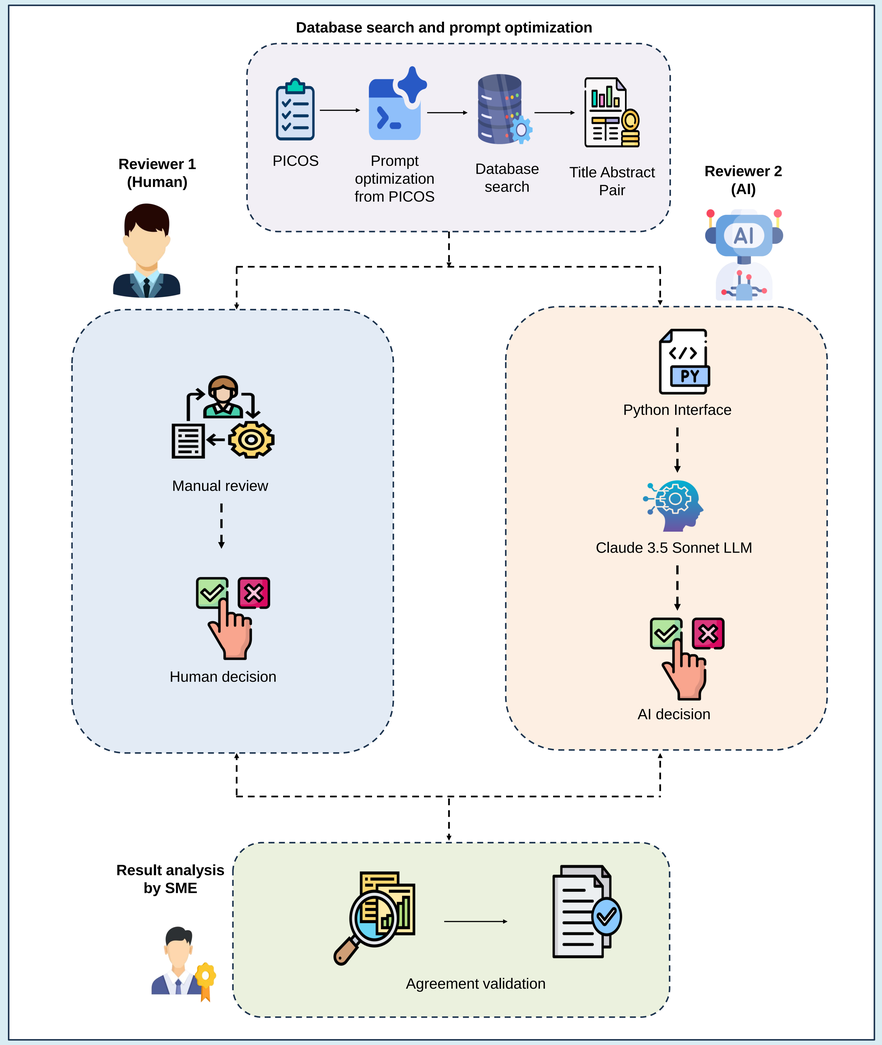
- A comprehensive depiction of the entire process is presented in **Figure 2.**

Figure 1: Eligibility criteria for the SLR



ADHD: Attention-deficit Hyperactivity Disorder; CBA: Cost-benefit Analysis; CEA: Cost-effectiveness Analysis; CMA: Cost-minimization Analysis; CUA: Cost-utility Analysis; EE: Economic Evaluation

Figure 2: Schematic diagram of the two-step screening process using MetaSLR



AI: Artificial Intelligence; SME: Subject Matter Expert

Results

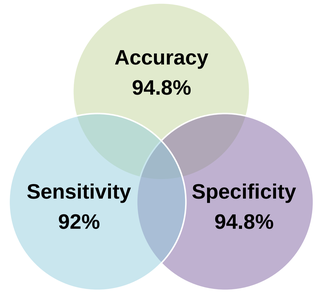
- A total of 4,581 citations retrieved from EMBASE were screened by the AI model Claude Sonnet 3.5.
- During the initial title and abstract screening phase, the inclusion rate of the AI model was higher than that of human reviewers. The AI model included additional articles without missing any relevant publications identified by a human reviewer.
- **Table 1** presents the confusion matrix values to determine the performance metrics such as accuracy, sensitivity, and specificity of the AI model. The model achieved an overall accuracy of 94.8% compared to the human reviewer, with a sensitivity and specificity of 92.0% and 94.8%, respectively (**Figure 3**).
- Both methodologies, i.e., a) two human reviewers, b) replacing one human review with AI, successfully identified all relevant studies, however, the time required for screening differed significantly between the human reviewers and the AI model.

Table 1: Confusion matrix to calculate the performance metrics

		Actual	
		Include	**Exclude**
Predicted	**Include**	**TP 92**	**FP 231**
	Exclude	**FN 8**	**TN 4250**

FN: False Negative; FP: False Positive; TN: True Negative; TP: True Positive

Figure 3: Performance metrics determined using the AI model



AI: Artificial Intelligence

Conclusion

- The findings demonstrate that the MetaSLR tool, using Claude Sonnet 3.5, significantly enhances the efficiency of the SLR process by reducing data collection time by approximately 40 hours.
- This substantial time savings is achieved while maintaining high levels of accuracy and precision, ensuring faster, more consistent outcomes and representing a meaningful advancement in productivity and SLR process optimization.
- The metaSLR platform was designed to support the various steps of the SLR process in alignment with core compliance principles outlined by NICE and CDA, specifically emphasizing human-in-the-loop oversight, transparency, and responsible use of AI.

Disclosures:
- BS, GK, RK, and SA are employees of Pharmacoevidence, Mohali, India.
- DK was an adjunct professor in the Department of Pharmaceutical Sciences, University of Florida, Florida, FL, USA.
- The authors declare no conflict of interest. No funding was received for this research work.