Empowering Systematic Literature Reviews: Utilizing Generative AI for Comprehensive Literature Screening From Titles and Abstracts to Full-Text

MSR140

Barinder Singh[1], Rajdeep Kaur[2], Pankaj Rai[2]
[1]Pharmacoevidence, London, UK, [2]Pharmacoevidence, Mohali, India

PharmacoEvidence®

CONCLUSION

Large Language Models (LLM) and generative AI can potentially improve the Systematic Literature Review (SLR) process by increasing the efficiency and optimizing the resources. Integrating Claude 3.5 Sonnet in a generative AI interface demonstrated high performance and efficiency in automating the SLR process. Its seamless full-text interaction handling and consistent efficiency across both screening stages significantly accelerated and streamlined SLRs, yielding substantial time savings over traditional manual methods.

INTRODUCTION

- SLRs provide a comprehensive analysis of the research, which is essential for producing robust evidence, directing healthcare decisions, and impacting policymaking-specific issues
- However, the manual procedure of generating a search strategy and selecting articles for inclusion in an SLR is time-consuming and error-prone
- Generative AI has the potential to automate the evidence-generation methodologies used in Health Economics and Outcomes Research (HEOR)
- The LLM models are trained on a large amount of data; these trained models can classify the articles according to the inclusion and exclusion criteria
- Prompt engineering is also important to automate the SLR process, as prompts provide clear and concise instructions to the LLM for the review procedure
- This study aims to address these challenges by developing and evaluating an AI-powered system for literature screening
- By leveraging LLM[1], agentic approach, and advanced techniques like Retrieval-Augmented Generation (RAG), we seek to create a tool that can significantly accelerate the SLR process while maintaining high standards of accuracy and reliability[2]

OBJECTIVE

- Recent advancements in generative AI are transforming literature reviews by automating and expediting literature screening.
- This study aims to develop and evaluate an automated system that utilizes advanced language models and embedding techniques for rapid and accurate literature screening using title, abstract and full text data

METHODOLOGY

- A Python-based interface was developed utilizing the Claude 3.5 Sonnet generative AI model to automate the SLR process in two distinct phases:

Phase 1: Title and abstract-based screening

- A data sheet containing relevant details (titles, abstracts, etc.) extracted from the literature database was uploaded into the interface
- Data sheet along with inclusion and exclusion criteria and optimized prompt was given as input to the LLM model for the review process[3]
- The prompt template provided clear and concise instructions to the LLM models for the initial screening. The LLM model processes the data in the datasheet row by row according to the inclusion and exclusion criteria and provides the decision according to the inclusion and exclusion criteria for the initial screening
- Articles with a different decision between the human expert reviewer and AI reviewer were flagged as conflicts
- A highly experienced subject matter expert (SME) critically analysed the variations in the decisions made by both human experts and AI reviewers and provided the final decision for articles flagged as a conflict
- The articles marked for inclusion are passed to the next phase for the full-text analysis

Phase 2: Full Text-based screening

- In this phase, the full texts of the included articles were either automatically retrieved or manually uploaded for the detailed screening process. Full text was analyzed as per the inclusion and exclusion criteria
- Full-text articles were pre-processed and chunked into smaller, manageable sections to create a Retrieval-Augmented Generation (RAG) pipeline, as shown in Figure 2
- Embeddings were stored in a PostgreSQL vector database that captures the semantic meaning of the text
- Metadata was stored alongside the embeddings for efficient indexing and fast retrieval
- An agent was developed to analyze the full text and retrieve the relevant chunks as per the inclusion and exclusion criteria.
- This agent used advanced natural language processing techniques to find the relevant chunks as per the criteria and to give the final decision as included or excluded
- Similar to Phase-1, SME resolves the conflict between the decision of the human expert reviewer and the AI reviewer.

Figure 1: SLR automation workflow diagram

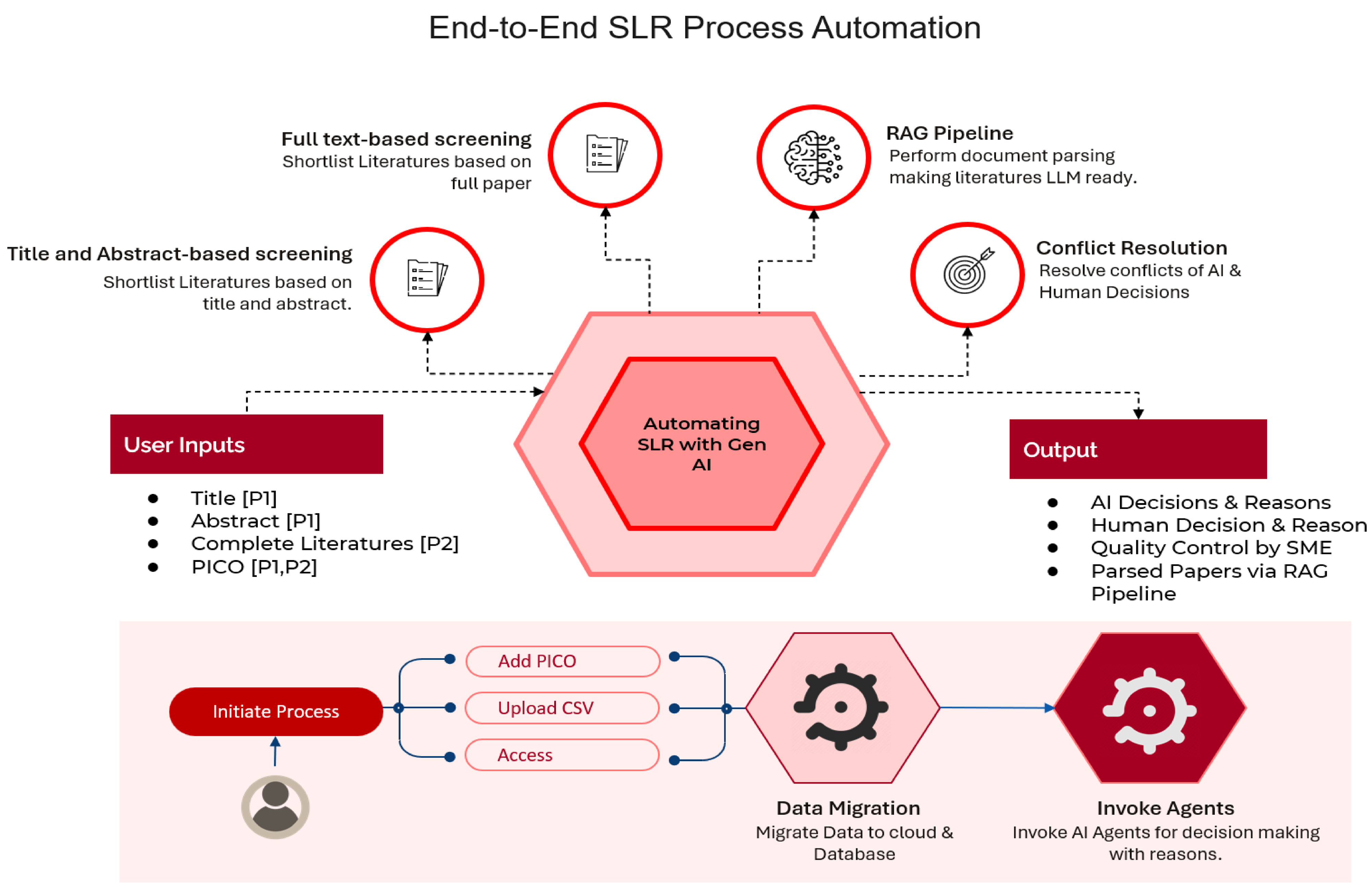
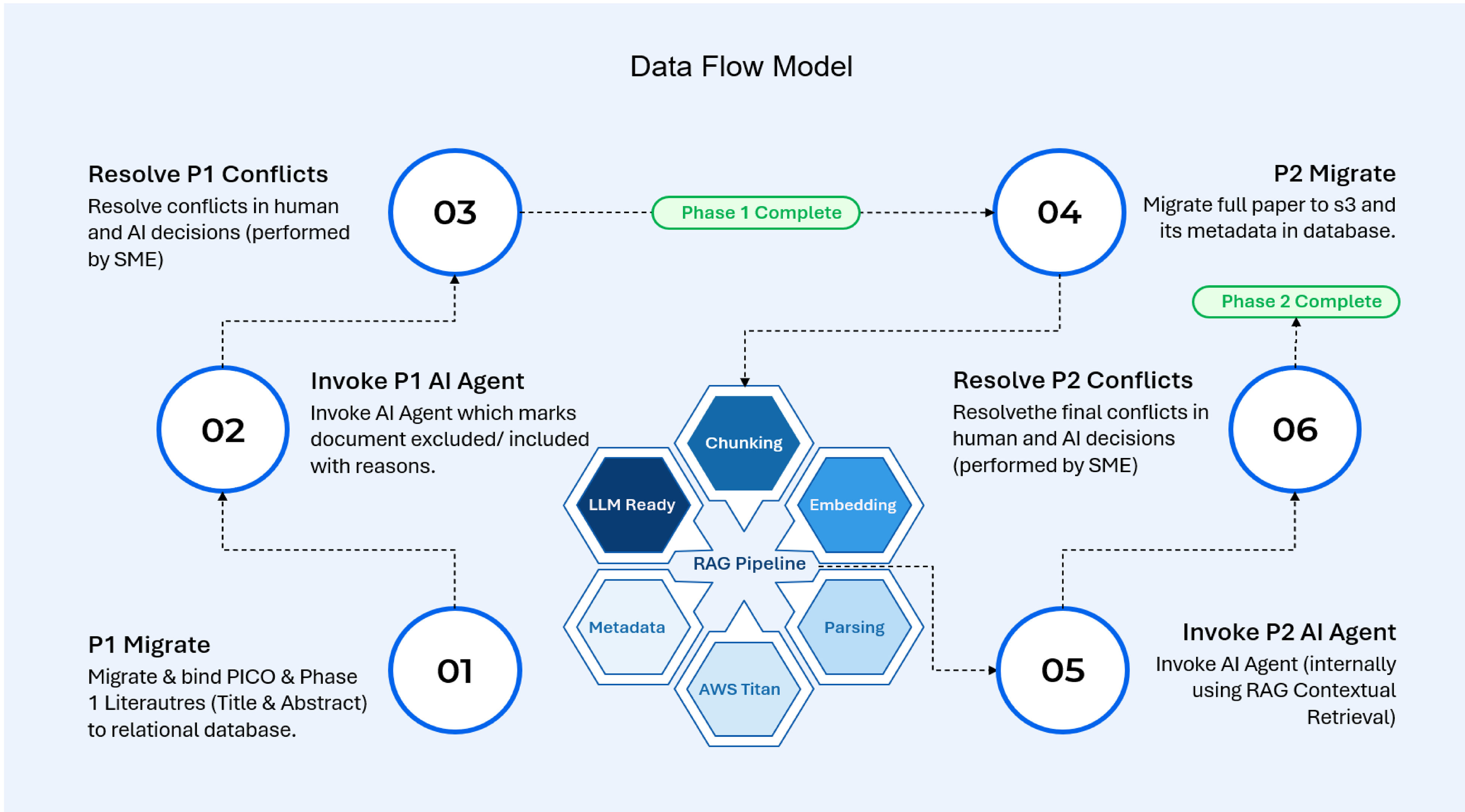


Figure 2: systematic representation of SLR Automation with GenAI



RESULTS

- The various performance metrics (accuracy, sensitivity, specificity, and precision) were used to evaluate the performance of the interface[2]

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \qquad Eq.(1)$$

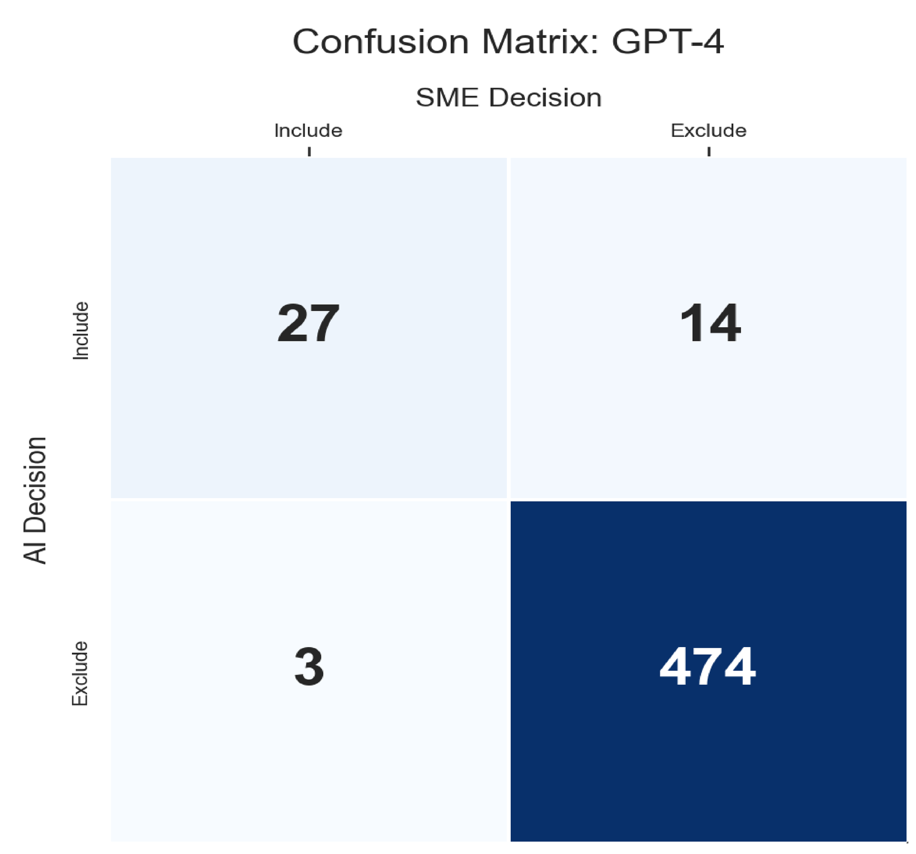
$$\text{Sensitivity} = \frac{TP}{TP+FN} \qquad Eq.(2)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \qquad Eq.(3)$$

$$\text{Precision} = \frac{TP}{TP+FP} \qquad Eq.(4)$$

- The first-stage screening of the publications, using title and abstracts, was conducted by the both human reviewer and Claude 3.5 Sonnet
- The generative AI demonstrated an outstanding performance achieving an accuracy rate of 96.72%, a sensitivity rate of 90.00%, and a specificity of 97.13% as shown in Figure 3.
- In subsequent steps, the AI interface effectively interacted with the full texts
- The model utilized full texts against the eligibility criteria and achieved screening efficiency comparable to that of the first stage
- Together, screenings across both stages resulted in saving approximately 40% of hours compared to traditional human review processes, despite the small sample size (518 publications)

Figure 3: Model Confusion Matrix



References

1. Mahuli et al. *Br Dent J.* 2023;235(2):90-92
2. Issaiy et al. *BMC Med Res Methodol.* 2024;24(1):78
3. Susnjak et al. *arXiv preprint.* 2023:1-20

Disclosures

BS, RK, and PR authors declare that they have no conflict of interest.

Need a poster copy?

Scan here

Hands-on experience with AI tools?

Contact: rshiny@pharmacoevidence.com

ISPOR