MetaSLR: AI-Driven Screening and Extraction Platform with Integrated Quality Assurance Checkpoints

MSR122

Rajdeep Kaur[1], Barinder Singh[2], Mrinal Mayank[1], Shubhram Pandey[1],Gagandeep Kaur[1]
[1]Pharmacoevidence, Mohali, India; [2]Pharmacoevidence, London, UK

CONCLUSION

The MetaSLR platform demonstrated efficiency, reliability, and compliance across all phases of the SLR process. With performance metrics comparable to conventional SLRs, it delivered expected accuracy while significantly reducing review timelines. Aligned with NICE and CADTH position papers, the platform ensures transparency, traceability, and human oversight throughout. The addition of automated SLR report generation further streamlines delivery, supporting rapid, HTA-ready evidence synthesis.

Introduction

- Systematic Literature Reviews (SLRs) are essential for healthcare decisions. Conventional review processes are primarily undertaken manually and typically require substantial time and resources[1,2]
- AI methods have the potential to automate various steps in these processes [1,2]
- AI tools like Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) can help speed up and bring consistency to the SLR process
- Additionally, LLM can automate the data extraction from published quantitative and qualitative studies using targeted prompts
- However, AI methods should be based on the principle of augmentation, not replacement, of human involvement (i.e., having a capable and informed human in the loop)

Objective

- To automate the end-to-end SLR process by combining unbiased AI agents with human expertise through a PRISMA-guided workflow. The platform tends to enhance SLR quality through Quality Assurance (QA) checkpoints across SLR steps, while ensuring secured access via role-based access control (RBAC)

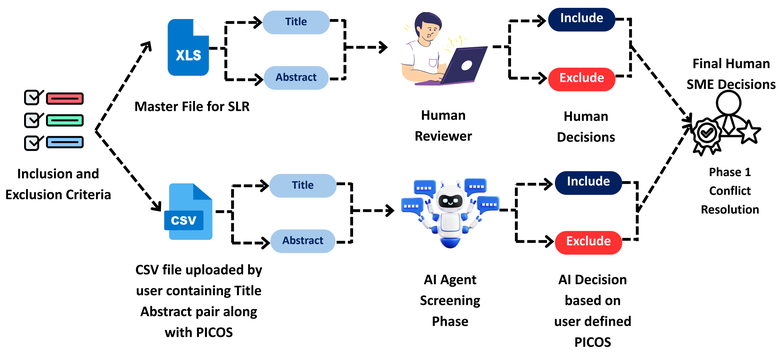
Methodology

The platform was developed using a multi-agent framework built on LangChain and RAG, with AWS Claude as the core language model. This architecture supported distinct phases of the SLR process. The system also incorporated secure authentication and project-specific team management features, enabling multiple SLRs to be executed concurrently within a unified environment. The workflow followed a three-phase approach:

Phase 1: Title and abstract screening

- One or more human reviewers work in parallel with an unbiased AI agent to perform title and abstract screening, followed by QA checkpoints and conflict resolution, as shown in Figure 2 and interface snapshot Figure 5
- Titles and abstracts were screened using the pre-defined inclusion/exclusion criteria by the human reviewers and the AI agent
- Conflicts between the human reviewer and AI decisions were resolved by a human Subject Matter Expert (SME)
- The SME's final decisions were recorded in the database to generate the final inclusion/exclusion list of studies

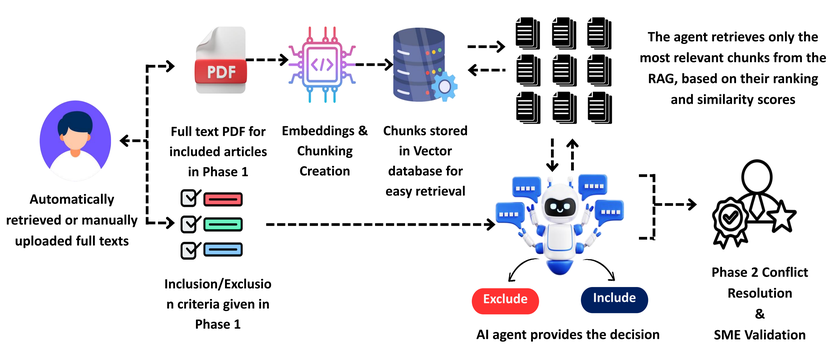
Figure 2. Title and Abstract screening Agent



Phase 2: Full-text identification and screening

- It involved automated and manual full-text search followed by screening of full-text articles (included in the phase-1) by an AI agent and human reviewers in parallel
- Full-text studies were either automatically retrieved or manually uploaded to the platform, then pre-processed through a standardization and RAG pipeline that applied text embedding and chunking to prepare the documents for downstream analysis
- Similar to phase 1, a human SME resolved the conflicts between the human reviewer and AI decisions, and after conflict resolution, the final inclusion list was recorded in the database for data extractions

Figure 3. Full text screening using LLM with RAG



Phase 3: Data extraction and analysis

- The interface included a specialized extraction agent as shown in Figure 4
- The user defined extraction grid column headers and provided a clear description of the extraction parameters (e.g., study design, population size, outcomes)
- The extraction agent then systematically iterates over each included study to identify and extract the relevant information based on the defined columns and data extraction grid
- SMEs reviewed and validated the AI-extracted data
- After approval by the SME (human), the extracted results were saved in the database and could be downloaded as a structured CSV file, as shown in Figure 7

Figure 4. Extraction Agent

- The platform was validated using randomized controlled trials related to schizophrenia, sourcing 980 studies from Embase®, Medline®, and Cochrane databases. SMEs validated the results at each stage of the SLR process performed by the AI agent

Results

- The interface snapshots illustrate the sequential flow of each phase in the workflow, from project setup and title/abstract screening to full-text review and data extractions—all within a unified platform
- By supporting end-to-end functionality and high processing speed, the system streamlines the SLR process while ensuring quality and reliability

Figure 5. Project creation along with PICOS insertion and Data injection

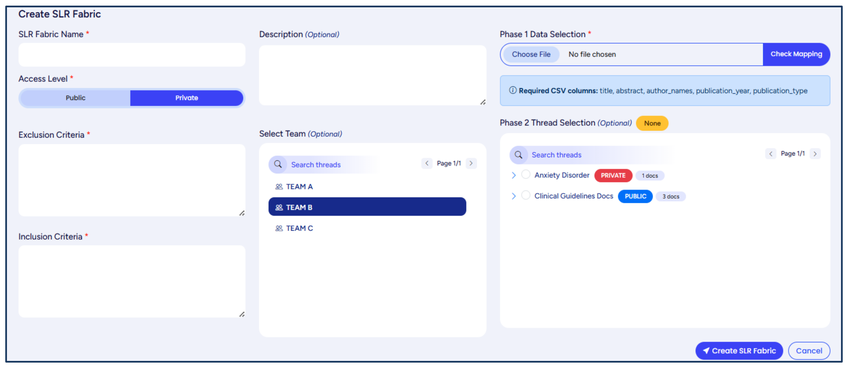
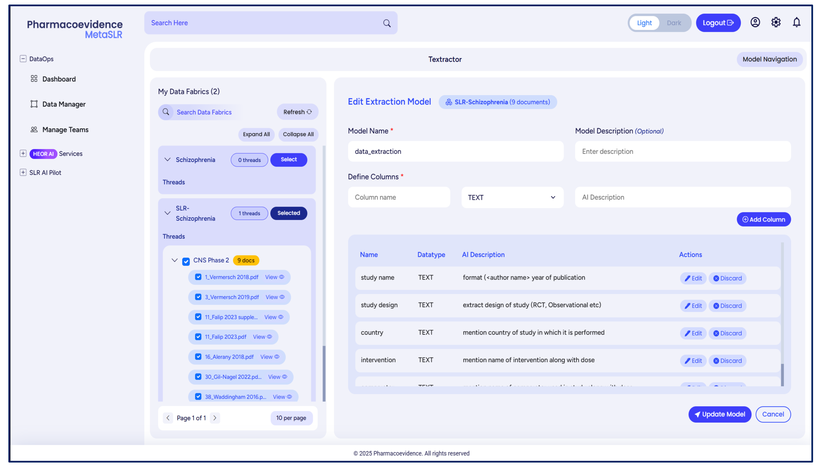


Figure 6. Conflict Resolution phase

Figure 7. Data extraction phase



- ***High screening accuracy: 94.69% accuracy, 88.59% sensitivity, 95.78% specificity vs. human reviewers***
- ***2× faster screening process: Human + AI collaboration reduced screening time from 4 weeks to 2 weeks***
- ***2× faster data extractions: 19 of 20 parameters perfectly extracted; 1 required minor SME edits***
- ***The end-to-end SLR was completed in 1 month, compared to the conventional 2–3 months, while fully adhering to PRISMA guidelines, resulting in substantial time and cost efficiencies***

Compliance with NICE (UK) and CDA-AMC (CANADA)

- The MetaSLR platform was purpose-built to align with key compliance principles outlined in AI position papers by NICE (UK) and CDA (Canada), emphasizing human-in-the-loop oversight, transparency, and responsible AI integration. Leveraging a retrieval-augmented generation (RAG) framework, it ensures traceability of all outputs to source data. With human validation embedded at every stage, the platform enables efficient, compliant execution of SLR workflows in line with HTA expectations [1,2]

References

1. CDA-AMC. New Position Statement Aims to Guide the Use of AI Methods in Health Technology Assessment. 2025
2. NICE. Use of AI in evidence generation: NICE position statement. 2024

For further queries, please contact: barinder.singh@pharmacoevidence.com